



HAL
open science

GrAPHiST: Une approche d'analyse exploratoire pour l'identification des dynamiques des phénomènes spatio-temporels.

Jacques Gautier

► **To cite this version:**

Jacques Gautier. GrAPHiST: Une approche d'analyse exploratoire pour l'identification des dynamiques des phénomènes spatio-temporels.. Géographie. Université Grenoble Alpes (France), 2018. Français. NNT: . tel-02067484

HAL Id: tel-02067484

<https://hal.science/tel-02067484>

Submitted on 14 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Modèles, méthodes et algorithmes en biologie, santé et environnement**

Arrêté ministériel : Arrêté n°MENS0602083A du 7 août 2006

Présentée par

Jacques Gautier

Thèse dirigée par **Mme Paule-Annick Davoine**
et co-encadrée par **Mme Claire Cuntz**

préparée au sein du **Laboratoire d'Informatique de Grenoble**
et de l'**École Doctorale Ingénierie pour la Santé, la Cognition et l'Environnement**

GrAPHiST

Une approche d'analyse exploratoire pour
l'identification des dynamiques des
phénomènes spatio-temporels.

Thèse soutenue publiquement le **2 octobre 2018**,
devant le jury composé de :

M. Didier Josselin

Directeur de Recherche, HDR, CNRS, UMR ESPACE, Président

Mme Anne Ruas

Directrice de Recherche, HDR, IFSTTAR, Rapportrice

M. Christophe Claramunt

Professeur, HDR, École navale, Rapporteur

Mme Luciana Nedel

Professeure, Université Fédérale du Rio Grande do Sul (UFRGS), Examinatrice

Mme Paule-Annick Davoine

Professeure, HDR, Université Grenoble Alpes, Directrice de thèse

Mme Claire Cuntz

Maître de Conférences, Université Lumière-Lyon-II, Co-Encadrante de thèse



*La science nous informe sur les moyens,
mais elle reste muette sur les fins.*

Usul, d'après Platon

Remerciements

Je souhaite exprimer ma plus grande reconnaissance aux membres de mon jury de thèse pour avoir accepté d'évaluer mes travaux. Je remercie tout d'abord M. Didier Josselin, directeur de recherche HDR au CNRS (UMR ESPACE), d'avoir accepté de prendre en charge la présidence du jury. Je remercie ensuite mes rapporteurs, Mme Anne Ruas, directrice de recherche HDR à l'Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR), et M. Christophe Claramunt, professeur HDR à l'École navale, pour leur lecture attentive, leurs retours qui ont permis d'améliorer la qualité de ce manuscrit, et leurs remarques et questions très intéressantes qui ont fait émerger de nombreuses perspectives pour la suite de mes travaux. Un grand merci également à Mme Luciana Nedel, professeure à l'Université Fédérale du Rio Grande do Sul (UFRGS), pour avoir accepté de participer au jury de thèse.

Je tiens à exprimer ma gratitude envers ma directrice de thèse Mme Paule-Annick Davoine, professeure HDR à l'Université Grenoble Alpes, et envers ma co-encadrante de thèse Mme Claire Cuntty, maître de conférences à l'Université Lumière-Lyon-II. Je les remercie pour m'avoir permis d'effectuer cette thèse, pour m'avoir soutenu et conseillé durant ces années, tout en m'accordant leur confiance et une totale liberté pour le développement de *GrAPHiST*.

Je souhaiterais également remercier les membres de l'équipe *STeamer* dans laquelle j'ai eu le plaisir de travailler durant cette thèse. En premier lieu, je tiens à remercier Mme Marlène Villanova-Oliver pour son écoute et ses conseils. Je la remercie également, ainsi que Mme Paule-Annick Davoine, de m'avoir offert la possibilité de poursuivre au sein de l'équipe *STeamer*. Je tiens aussi à remercier M. Philippe Genoud et M. Sylvain Bouveret pour m'avoir aidé à mettre en place mon prototype.

Ma sympathie va également à mes collègues Mme Danielle Ziebelin, M. Jérôme Gensel, Mme Sonia Chardonnel, à mes camarades de bureau Mme Camille Bernard, Mme Yagmur Cinar, M. Adrien Dulac, M. Clément Chagnaud, M. Thibault Thonet, M. Benoit Le Rubrus, Mme Clémence Merzeau, ainsi qu'à M. David Noël, Mme Lina Toro, M. Anthony Hombiat, Mme Aline Menin, Mme Cécile Saint-Marc, Mme Léa Branjonneau, M. Matthew Sreeves, Mme Maeva Sefar, M. Pierre Brunisholz, M. Timothy Claeys, Mme Élodie Morin, M. Tristan Braud, Mme Marie-Jeanne Natete, Mme Camille Cavalière, M. Alexandre Audouard, M. Oton Čopi, M. Tales Paiva Nogueira, M. André Sales-Fontales, M. Iacob Juc, M. Emmanuel Perier, M. Martin Kirchgessner, Mme Lyuba Mancheva, M. Henry-Joseph Audéoud, M. Gabriel Lopes, M. Axel Vincent, M. Irving Hernandez-Gallegos, M. Adel Mehiris, et à Mme Latifa Boudiba, Mme Anne-Laure Bernardin, Mme Samantha Sanchez et Mme Pascale Poulet.

Je tiens particulièrement à remercier Léa Branjonneau, avec qui j'ai eu le grand plaisir de travailler, pour ses remarques et ses observations lors de son utilisation du prototype, qui m'ont énormément aidé à finaliser ce qu'allait devenir *GrAPHiST*. Je veux également exprimer ma gratitude envers Camille Bernard pour tout ce temps passé en sa compagnie au laboratoire, pour avoir toujours été présente pour m'aider et me conseiller durant ma thèse, et pour toutes nos discussions engagées sous caféine. Je tiens également à remercier Cécile Saint-Marc, sans qui j'aurais été perdu lors de chacune des conférences auxquelles nous sommes allés, pour son aide et ses conseils durant ma thèse.

Je tiens à exprimer ma gratitude envers David Noël, Pierre Brunisholz et Anthony Hombiat pour tous les instants passés en leur compagnie. Je les remercie d'avoir transformé ce changement dans ma vie professionnelle en une réelle évolution dans ma façon de voir le monde, de vivre et de penser. Je les remercie également, ainsi qu'Ugo Hévin, de m'avoir fait redécouvrir la musique, et de m'avoir redonné l'envie de jouer d'un instrument. Je tiens à remercier Élodie Mo-

rin, Colin Chargy, Timothy Claeys, Tristan Braud, Cindy Athanase, Fabian Grüber, Hyunyoung Kim, François Rousse, Alexandre Hoffmann, Xavier Bouvard et George Christodoulis pour tous ces moments passés à leur côté, ainsi que Vivien pour m'avoir accompagné dans mes premiers pas à Grenoble. Je remercie également mes amis d'au-delà Grenoble, Frédéric, Pierre, Samuel, Raphaël, Astrid, Raihere, Stridou, Cécile, Anne, Nicolas et tous les autres pour leur affection, et pour les voyages que j'ai effectués à leur côté.

Je souhaite remercier mes parents François et Élisabeth Gautier, et mes soeurs Amandine Stein et Camille Wiplier, pour m'avoir encouragé lorsque j'ai décidé de retourner vers la recherche, pour m'avoir aidé lors de mon installation à Grenoble, et pour avoir été à l'écoute et prêts à me soutenir à chaque instant. Je souhaite également remercier ma famille, et particulièrement Marthe Aeby pour leurs encouragements.

Je souhaiterais encore remercier toutes celles et ceux que je n'ai pas cités, que j'ai croisés durant ces années, et qui ont fait en sorte que mon passage à Grenoble m'a littéralement transformé, bien au-delà de la thèse.

Y finalmente, yo quiero decir muchas gracias a Lina por su apoyo, su ayuda, et por todas las otras cosas que yo vivo desde nos conocimos.

Résumé

Les données permettant de décrire des phénomènes spatio-temporels sont de plus en plus nombreuses. Ces nouvelles données peuvent alors être éloignées de celles habituellement observées pour l'étude de certains phénomènes. Leur analyse, selon une approche hypothético-déductive telle qu'elle est majoritairement effectuée en statistique et dans les *SIG*, peut ainsi passer sous silence certaines informations insoupçonnées, mais pertinentes, sur les dynamiques de ces phénomènes spatio-temporels.

Il peut alors être intéressant de simplement donner à voir les données, pour observer ce qu'elles ont à montrer, avant de les analyser. Ce principe est celui de l'analyse exploratoire : le procédé est de permettre à un utilisateur d'effectuer une exploration libre des données, au moyen de représentations visuelles, afin de mettre en lumière des structures ou des relations insoupçonnées. Aujourd'hui, l'analyse exploratoire est notamment possible au moyen d'environnements de visualisation, intégrant différentes représentations graphiques et cartographiques interactives.

Les environnements de visualisation sont majoritairement développés de manière *ad hoc*, dans le cadre d'une thématique particulière. Or l'émergence constante de nouvelles données incite à promouvoir des méthodes d'analyse applicables à des phénomènes de différentes natures. En fonction de la problématique dans laquelle s'insèrent ces derniers, les dynamiques sur lesquelles va se focaliser l'analyse diffèrent. Analyser un phénomène météorologique dans un but de prévision implique de s'intéresser aux récurrences cycliques du phénomène. Analyser l'évolution d'une population pour la mise en place de politiques publiques implique d'analyser ce phénomène sur le temps long et selon différentes zones de l'espace.

Notre objectif est de proposer une méthode d'analyse exploratoire des phénomènes spatio-temporels et de leurs dynamiques, indépendante du thème traité. Pour cela, nous proposons un environnement de géovisualisation, *GrAPHiST* (*Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels*), permettant l'analyse de différentes dynamiques, selon différentes échelles spatiales et temporelles (linéaires ou cycliques). Développer cet environnement implique de s'interroger sur la modélisation du changement dans l'espace, la nature des dynamiques spatio-temporelles à étudier, et les outils visuels et interactifs permettant de les identifier.

Ainsi, les contributions de notre recherche se situent à plusieurs niveaux :

- une modélisation générique des phénomènes spatio-temporels, sous la forme de séries événementielles ;
- de nouvelles méthodes de représentations graphiques et interactives, autorisant la recherche et l'identification des dynamiques spatio-temporelles, notamment :
 - l'introduction de diagrammes temporels interactifs permettant la recherche visuelle de récurrences cycliques dans les données spatio-temporelles ;
 - l'utilisation de règles de symbologie permettant la visualisation des relations entre les composantes temporelle et spatiale des phénomènes ;
 - de nouvelles méthodes de représentations des agrégats d'événements proches, permettant d'identifier des structures dans leur distribution spatio-temporelle ;
- la formalisation d'une approche d'analyse exploratoire des dynamiques spatio-temporelles, déclinée en plusieurs scénarios selon l'objectif poursuivi.

Nous validons notre approche en l'appliquant à l'analyse de différents jeux de données. L'objectif est de vérifier la possibilité d'identifier des dynamiques, relatives au temps linéaire ou cyclique, au moyen de *GrAPHiST*, et d'illustrer le caractère générique de l'approche, ainsi que les opportunités d'analyse offertes par l'environnement.

Abstract

Datasets allowing the description of spatio-temporal phenomena are becoming ever more numerous. These new data can be very different from those usually observed for studying spatio-temporal phenomena. An analysis through a hypothetic-deductive approach, like is mainly done in statistic and *GIS* domains, can lead to miss some unsuspected, but relevant, information about the dynamics of these spatio-temporal phenomena.

It can be interesting then, to just present the data, to observe what they have to show, before analysing them. This is the principle of the exploratory data analysis : the process is to allow a user to freely explore data, through visual representations, in order to highlight unsuspected structures or relationships. Today, exploratory data analysis is possible through visualization environments, which integrate different graphic or cartographic interactive representations.

Visualization environments are mainly developed in an *ad hoc* manner, in the context of a particular thematic field. However, the constant appearance of new data encourages promoting analysis methods, which could be applied to several types of phenomena. According to the domain related to these phenomena, the analysis will be focused on different dynamics. Analysing a meteorological phenomenon, in a forecasting purpose, implies a focus on the cyclic recurrences of the phenomenon. Analysing the increase of a population, for the purpose of deciding public policies, implies an analysis of the phenomenon on a long-term, through different spatial areas.

Our objective is to propose a method for the exploratory analysis of spatio-temporal phenomena and their dynamics, which would be independent of the topic. In order to achieve this, we propose a geovisualization environment, *GrAPHiST* (*Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels ; Geovisualization for spatio-temporal phenomena analysis*), allowing the analysis of several dynamics, through different spatial and temporal (linear or cyclic) scales. Developing this environment implies to focus on how spatial changes are modelled, on the nature of the spatio-temporal dynamics we have to study, and on the visual and interactive tools, which allow the identification of these dynamics.

So, the contributions of our research can be found at several levels :

- a generic modelling approach of spatio-temporal phenomena, in the form of event series ;
- new graphical and interactive representation methods, which allow the investigation and the identification of spatio-temporal dynamics, including :
 - the introduction of interactive temporal diagrams, which allow the visual research of cyclic recurrences in spatio-temporal data ;
 - the use of symbology rules, which allow the visualization of relationships between the spatial and temporal components of phenomena ;
 - new methods to represent aggregated closed events, which allow to identify structures in their spatio-temporal distribution ;
- the formalization of an exploratory approach for the spatio-temporal dynamics analysis, divided into several scenarios, according to the purpose of the analysis.

We validate our proposition by applying it to the analysis of several datasets. The objective is to verify the possibility to identify dynamics, related to linear or cyclic time, through the use of *GrAPHiST*, and to illustrate the generic aspect of the approach, as well as the analysis opportunities given by the environment.

Table des matières

Introduction générale	19
I L'étude des dynamiques spatio-temporelles par l'analyse exploratoire de la distribution spatio-temporelle des événements	25
Introduction	27
1 La modélisation des phénomènes spatio-temporels	29
1.1 La modélisation des dimensions temporelle et spatiale	30
1.1.1 La modélisation de la dimension spatiale	30
1.1.2 La modélisation de la dimension temporelle	32
1.2 La modélisation de l'information spatio-temporelle	37
1.2.1 L'approche <i>endurante</i> et l'approche <i>perdurante</i>	37
1.2.2 Décrire les phénomènes au travers des entités <i>SNAP</i> et <i>SPAN</i> : les modèles <i>3D+1</i> et <i>4D</i>	39
1.2.3 <i>Triad Model</i> et <i>Pyramid Model</i>	41
1.3 Le phénomène spatio-temporel modélisé selon un point de vue endurant	42
1.3.1 La modélisation implicite du changement	42
1.3.2 La modélisation explicite du changement	45
1.3.3 Le choix de l'approche <i>4D</i>	47
1.4 Le phénomène modélisé selon un point de vue occurrent	49
1.4.1 <i>Événements</i> et <i>processus</i>	49
1.4.2 L'approche de Grenon et Smith	50
1.4.3 L'approche de Reitsma	50
1.4.4 Le choix de l'approche de Reitsma	50
1.5 <i>Événements</i> et <i>Phénomènes</i>	52
1.5.1 L' <i>Événement</i> , entité spatio-temporelle représentant le changement	52
1.5.2 Les <i>Phénomènes</i> , regroupements d' <i>événements</i>	53
1.5.3 Une modélisation générique du changement au travers des <i>Phénomènes</i>	54
2 L'étude des dynamiques des phénomènes spatio-temporels par l'analyse exploratoire	57
2.1 Les dynamiques de phénomènes modélisés par des séries événementielles	58
2.1.1 Les <i>dynamiques spatio-temporelles</i> , structures du <i>phénomène</i> dans le temps et l'espace	58
2.1.2 Les dynamiques spatio-temporelles dans le cadre de notre modélisation événementielle du changement dans l'espace	58
2.1.3 Les différents types de dynamiques sur lesquels nous nous focalisons	59
2.2 Identifier les dynamiques spatio-temporelles des phénomènes au moyen de l'analyse exploratoire	66
2.2.1 L'analyse de données	66

2.2.2	L'analyse exploratoire	68
2.2.3	L'exploration des données au moyen de représentations graphiques	72
	Conclusion	75
II La visualisation des phénomènes spatio-temporels		79
	Introduction	81
3	Représentation graphique d'événements spatio-temporels	83
3.1	L'intégration du temps et de l'espace au sein de la même représentation graphique	84
3.1.1	Au travers d'une représentation non cartographique	84
3.1.2	L'intégration du temps dans les représentations cartographiques	86
3.2	L'intégration du temps et de l'espace à travers la structure de l'environnement de représentation	92
3.2.1	Les dispositifs <i>small-multiples</i>	92
3.2.2	L'animation	95
3.3	Les environnements de géovisualisation à multi-fenêtrage	98
3.3.1	Les caractéristiques de la géovisualisation	98
3.3.2	Le multi-fenêtrage pour représenter les données spatio-temporelles	102
3.3.3	Les environnements de géovisualisation à multi-fenêtrage pour la recherche de patterns spatio-temporels	104
4	L'identification de patterns dans les environnements à multi-fenêtrage	107
4.1	Les représentations graphiques au service de la recherche de patterns temporels	109
4.1.1	Des légendes temporelles aux diagrammes des environnements de géovisualisation	109
4.1.2	<i>Timelines</i> et <i>timewheels</i> pour l'identification de patterns	114
4.1.3	La recherche de récurrences cycliques au travers des diagrammes temporels <i>mixtes</i>	123
4.1.4	Synthèse	133
4.2	La visualisation du lien entre les composantes spatiale et temporelle pour l'identification de patterns	133
4.2.1	L'analyse des données spatio-temporelles au travers du temps linéaire et du temps cyclique : les approches <i>TimeWave</i> et <i>PerSE</i>	134
4.2.2	Les limites de <i>Time Wave</i> et de <i>PerSE</i>	142
4.2.3	L'intégration du temps dans les représentations cartographiques des environnements à multi-fenêtrage	149
4.2.4	Conclusion	158
	Conclusion	161
III GrAPHiST		165
	Introduction	167
5	Structure de GrAPHiST	171
5.1	Modélisation des événements	172
5.1.1	Entités <i>SPAN</i> , et <i>intervalles (spatio-)temporels</i>	172
5.1.2	Notre modèle de données	173
5.2	Structure générale de l'interface de <i>GrAPHiST</i>	175

5.2.1	Un environnement de géovisualisation à multi-fenêtrage	175
5.2.2	La représentation de la dimension thématique	177
5.2.3	Limitation des données à représenter aux événements ponctuels	179
5.3	Représentation de la composante temporelle des événements	180
5.3.1	Le diagramme temporel mixte	180
5.3.2	Représentation des événements dans le diagramme mixte	184
5.3.3	Intérêts et limites des deux approches	188
5.3.4	Diagramme temporel circulaire	189
5.3.5	Relations entre le diagramme mixte et le diagramme circulaire	191
5.4	Représentation de la composante spatiale des événements	192
5.4.1	Principe	192
5.4.2	Représentation des événements dans la carte	193
5.5	Représentation de la composante spatio-temporelle des événements	196
5.5.1	Sélection et <i>brushing</i> des événements selon un critère dimensionnel	197
5.5.2	Représentation des événements selon leur lieu d'apparition	199
5.5.3	Représentation des événements selon leur moment d'apparition relatif à une échelle linéaire	203
5.5.4	Représentation des événements selon leur moment d'apparition relatif à une échelle cyclique	206
5.6	Implémentation de <i>GrAPHiST</i>	209
5.6.1	Implémentation de la base de données	209
5.6.2	Implémentation de l'interface de géovisualisation	210
5.7	Conclusion	211
6	Scénarios d'analyse exploratoire pour l'identification de patterns au moyen de <i>GrAPHiST</i>	215
6.1	Analyse de patterns unidimensionnels	217
6.1.1	Identification de concentrations d'événements dans le temps et l'espace	218
6.1.2	Identification d'une récurrence cyclique d'événements	222
6.1.3	Proposition de pistes d'exploration à l'utilisateur par suggestion de valeurs à affecter à l'échelle temporelle cyclique	228
6.2	Structure des scénarios d'analyse	232
6.3	Analyse de patterns spatio-temporels, au travers de zones spécifiques	234
6.3.1	Procédé général	234
6.3.2	Analyse de clusters relatifs au temps linéaire, selon une entrée spatiale	236
6.3.3	Analyse de clusters relatifs à une échelle temporelle cyclique, selon une entrée spatiale	238
6.3.4	Analyse d'une propagation spatiale d'événements au cours du temps, selon une entrée spatiale	240
6.3.5	Analyse de récurrences cycliques spatialisées, selon une entrée spatiale	242
6.4	Analyse de patterns spatio-temporels, sans hypothèse <i>a priori</i> sur la distribution spatiale des événements	246
6.4.1	Procédé général	246
6.4.2	Analyse de clusters relatifs au temps linéaire, selon une entrée temporelle	248
6.4.3	Analyse de clusters relatifs à une échelle temporelle cyclique, selon une entrée temporelle	250
6.4.4	Analyse d'une propagation spatiale d'événements au cours du temps, selon une entrée temporelle	252
6.4.5	Analyse de récurrences cycliques spatialisées, selon une entrée temporelle	253
6.5	Synthèse	259
	Conclusion	261

IV	Cas d'étude	267
	Introduction	269
7	Analyse de données historiques sur la guerre de Sécession	275
7.1	Objectif	276
7.1.1	Identifier des patterns spatio-temporels relatifs au temps linéaire dans les séries événementielles	276
7.1.2	Recréer le récit d'un phénomène à partir de données historiques	276
7.1.3	Le choix des données du cas d'étude	276
7.2	Données	278
7.2.1	Nature des données	278
7.2.2	Implémentation dans notre modèle de données	279
7.3	Procédé	279
7.3.1	Principe	279
7.3.2	Choix des épisodes à analyser, et objectifs de ces analyses	280
7.3.3	Chronologie des épisodes analysés	280
7.3.4	Scénarios d'analyse	284
7.4	Résultats	285
7.4.1	Analyse des différentes phases du conflit	285
7.4.2	Analyse des engagements du théâtre de l'Est	290
7.5	Synthèse	293
8	Identification de récurrences cycliques spatialisées dans les données événementielles	295
8.1	Objectif	296
8.1.1	Identifier des patterns spatio-temporels relatifs au temps cyclique dans les séries événementielles	296
8.1.2	Le choix des données du cas d'étude	298
8.2	Données	299
8.2.1	Nature des données	299
8.2.2	Construction des données	299
8.2.3	Implémentation dans notre modèle de données	303
8.3	Procédé	304
8.3.1	Principe	304
8.3.2	Les différents jeux de données	305
8.3.3	Scénario d'analyse	306
8.4	Résultats	307
8.4.1	Analyse des jeux de données du groupe A	307
8.4.2	Analyse des jeux de données du groupe B	316
8.5	Limites et synthèse	323
8.5.1	Distribution temporelle des événements cycliques dans les jeux de données	323
8.5.2	Conclusion	324
9	Analyse de clusters spatio-temporels dans des jeux de données réels	329
9.1	Principe et objectifs	330
9.1.1	Application de notre approche sur des données réelles	330
9.1.2	Choix des données	330
9.1.3	Structures recherchées	330
9.2	Données	331
9.2.1	Données d'accidents routiers	331
9.2.2	Données d'inondations	332

9.2.3	Données criminologiques	334
9.3	Procédé général	335
9.3.1	Analyse selon une entrée temporelle	336
9.3.2	Analyse selon une entrée spatiale	337
9.4	Analyse des données d'accidents routiers	337
9.4.1	Scénario d'analyse	337
9.4.2	Résultats	339
9.4.3	Synthèse	342
9.5	Analyse des données d'inondations	342
9.5.1	Scénario d'analyse	342
9.5.2	Résultats	344
9.5.3	Synthèse	349
9.6	Analyse des données criminologiques	350
9.6.1	Scénarios d'analyse	350
9.6.2	Résultats	355
9.6.3	Synthèse	366
9.7	Synthèse générale	367
	Conclusion	369
	Conclusion générale	377
V	Annexes	391
10	Manuel d'utilisation de <i>GrAPHiST</i>	393
10.1	Vue générale	394
10.2	Fenêtre temporelle	394
10.3	Interface de contrôle de la fenêtre temporelle	396
10.4	Interface de contrôle de la fenêtre spatiale	398
10.5	Import d'événements	398
10.6	Transformée de Fourier	400
10.7	Modification de la représentation graphique des événements	402
10.7.1	Diagramme temporel circulaire	402
10.7.2	Représentation des événements dans le diagramme mixte	402
10.7.3	Représentation des événements dans la carte	404
10.8	Sélection des événements	406
10.9	Utilisation de la symbologie	409
10.9.1	Représentation des événements selon leur lieu d'apparition	409
10.9.2	Représentation des événements selon leur moment d'apparition relatif à une échelle linéaire	411
10.9.3	Représentation des événements selon leur moment d'apparition relatif à une échelle cyclique	411
	Bibliographie	413
	Liste des Figures	432

Introduction générale

Contexte et problématique

Aujourd'hui, de plus en plus de données, intégrant les dimensions temporelles et spatiales et relatives à différents domaines, sont disponibles et en constante augmentation. Cette augmentation s'explique par la conjonction de plusieurs facteurs. Tout d'abord, le développement et l'accumulation au cours du temps de bases de données spatiales. Ensuite, la multiplication d'appareils de mesures permettant de créer ces données spatio-temporelles. Enfin, une disponibilité croissante de ces données, au travers de portails de données consultables sur le Web (*Open Data*).

Ces données peuvent permettre une description des changements affectant l'environnement géographique. Ceux-ci sont généralement décrits dans la littérature comme la conséquence de *phénomènes géographiques dynamiques*, aussi appelés *phénomènes spatio-temporels* [Haddad, 2009]. Un *phénomène géographique* peut être considéré comme un ensemble de faits et de caractéristiques observables et localisables dans l'espace. Le terme *dynamique* renvoie à l'évolution de ces faits et caractéristiques au cours du temps.

L'étude des *dynamiques spatio-temporelles* des phénomènes consiste alors à identifier comment ces derniers évoluent dans l'espace au cours du temps. Ces dynamiques peuvent se traduire sous différentes formes : augmentation de l'intensité d'un phénomène au cours du temps, selon différentes zones de l'espace ; propagation spatiale d'un phénomène au cours du temps ; réapparitions cycliques localisées d'un phénomène ; etc. L'étude de ces dynamiques permet alors de répondre à différents besoins et enjeux, en fonction de la problématique dans laquelle s'insère le phénomène. L'étude du changement dans l'espace constitue ainsi un pan des notions fondamentales de la géographie depuis les années 1970 [Morrill, 1970, Lefebvre, 1974].

En parallèle de l'émergence constante de nouvelles données spatio-temporelles, on observe depuis une vingtaine d'années le développement de l'analyse exploratoire (*exploratory data analysis*).

L'étude des données effectuée en statistiques et dans les *SIG* suit majoritairement une démarche hypothético-déductive [Andrienko and Andrienko, 2006] : l'analyse doit valider une hypothèse, énoncée à priori sur les données étudiées. Cette approche convient à l'étude de données connues de l'analyste, dans le cadre d'une problématique de recherche clairement définie. Elle devient cependant plus complexe lorsque les données sont inconnues de l'analyste, que ce dernier ne sait pas quoi y chercher, et souhaite juste découvrir ce que les données pourraient lui donner à voir.

Permettre cette deuxième approche est l'objectif de l'*analyse exploratoire* [Andrienko and Andrienko, 2006]. Le principe est de rendre possible une exploration libre des données, dans le but d'y identifier des structures ou des relations insoupçonnées. Ainsi, l'analyse exploratoire n'est pas basée sur la validation d'hypothèses énoncées à priori, mais sur la génération d'hypothèses au cours de l'analyse [Tukey, 1977, Andrienko and Andrienko, 2006].

Dans le cadre de l'apparition croissante de nouvelles données spatio-temporelles, l'analyse exploratoire apparaît alors comme une solution permettant d'étudier ces nouvelles données, pouvant être éloignées de celles habituellement utilisées pour étudier les phénomènes spatio-temporels, et pour lesquelles il est difficile d'émettre des hypothèses sur ce qu'il faut y chercher.

L'analyse doit permettre à l'utilisateur de répondre à une série de questions, énoncées au cours de l'analyse et de plus en plus spécifiques, en partant d'un point de vue global sur les données, jusqu'à aborder un point de vue de plus en plus détaillé [Klir, 1985, Shneiderman, 2003, Andrienko and Andrienko, 2006]. L'analyse exploratoire s'effectue ainsi au travers d'environnements

interactifs, mettant à disposition de multiples outils devant permettre à l'utilisateur de répondre à l'ensemble des questions qu'il peut être appelé à formuler [Andrienko and Andrienko, 2006]. Ces outils peuvent provenir de différents domaines, tels que la statistique, l'analyse multicritère, l'analyse visuelle, etc.

Les environnements d'analyse exploratoire sont ainsi basés sur trois principes fondamentaux :

- l'utilisation combinée de plusieurs outils de représentation graphiques de la donnée [Tufte, 1983, Andrienko and Andrienko, 2006] ;
- la place centrale de l'analyste (l'utilisateur) dans le procédé de recherche [Tukey, 1977, Andrienko and Andrienko, 2006] ;
- les relations d'interaction entre l'utilisateur et la représentation graphique.

Proposer une analyse exploratoire des phénomènes spatio-temporels passe alors par l'élaboration d'un environnement de visualisation, permettant à un utilisateur d'explorer des données spatio-temporelles et d'y identifier des patterns. Cet environnement de visualisation n'est pas un outil de retranscription, permettant d'afficher les résultats d'une analyse, mais un outil par lequel l'utilisateur effectue son processus de recherche [Card et al., 1999, Spence, 2001, Andrienko and Andrienko, 2006].

Plusieurs travaux portent sur l'élaboration de tels environnements [Andrienko and Andrienko, 2006, Aigner et al., 2007, Aigner et al., 2011, Kaddouri et al., 2014]. Certains proposent une formalisation des différents outils de représentation graphique pouvant être intégrés dans les environnements de visualisation de données spatio-temporelles [Andrienko and Andrienko, 2006]. Ces travaux permettent la mise en place d'une méthodologie générique pour la conception d'environnements d'analyse exploratoire, définissant les approches de représentation graphique à utiliser en fonction des données analysées et des besoins de l'étude.

On observe cependant que les environnements de visualisation des données spatio-temporelles sont le plus souvent implémentés de manière *ad hoc* [Kaddouri et al., 2014]. Ces environnements sont développés de manière à proposer une analyse exploratoire de données spatio-temporelles dans le cadre de projets portant sur une thématique particulière. Les données analysées peuvent alors présenter une structure propre, être analysées au travers d'échelles temporelles et spatiales spécifiques, et être observées à travers une représentation du temps linéaire ou cyclique, selon le sujet d'étude.

Notre problématique est de proposer une approche d'analyse exploratoire de données spatio-temporelles indépendante du thème traité, autorisant une étude des dynamiques spatio-temporelles applicable à différents phénomènes, dans le cadre de différentes problématiques de recherche.

Verrous scientifiques

Notre recherche se positionne dans le domaine de la visualisation pour l'analyse exploratoire de données spatio-temporelles.

Notre objectif est de proposer un ensemble de méthodes, permettant de donner à voir les dynamiques de phénomènes, au sein d'un environnement de géovisualisation intégrant la dimension temporelle. Notre ambition est d'autoriser l'analyse de différents phénomènes, indépendamment du sujet traité, et de permettre l'identification de différents types de dynamiques spatio-temporelles, selon différentes échelles spatiales et différentes échelles temporelles, linéaires

ou cycliques.

Afin de répondre à cet objectif, nous cherchons à proposer notre propre environnement de visualisation de données spatio-temporelles. Pour concevoir ce dernier, nous devons nous intéresser :

- aux approches de modélisation de la donnée spatio-temporelle, dans le but de proposer une modélisation générique permettant l'étude de phénomènes de différentes natures au sein d'un même environnement d'analyse ;
- aux approches de visualisation de la donnée spatio-temporelle, afin de proposer des méthodes de représentation graphique, interactives, permettant l'identification de différentes dynamiques, à travers différentes échelles spatiales et temporelles ;
- à la formalisation d'une stratégie d'analyse, décrivant les actions que doit accomplir un utilisateur à travers cet environnement de visualisation, pour mener l'analyse exploratoire de chaque dynamique spatio-temporelle.

Choix méthodologiques

Les dynamiques décrivant l'évolution des phénomènes spatio-temporels formant un ensemble vaste, difficilement traitable par un seul environnement de visualisation, nous nous sommes focalisés sur quatre types de dynamiques spatio-temporelles dans le cadre de cette thèse. Cette limitation de notre champ de recherche nous permet de réduire le nombre de questions que l'analyste peut être amené à se poser sur les données, rendant plus aisée la création d'un environnement d'analyse exploratoire des phénomènes spatio-temporels. Ainsi, nous décidons de nous focaliser sur¹ :

- l'existence de pics d'intensité d'un phénomène dans l'espace et dans le temps, selon une échelle temporelle linéaire ou cyclique : une augmentation brusque de l'intensité sismique le long d'une faille tectonique, entre le *25 août* et le *10 septembre 2011* ; la variation du nombre d'oiseaux dans une zone spatiale *A* en été, et dans une zone *B* en hiver (*zone de nidification* et *zone d'hivernage*) ;
- l'existence de tendances locales dans la variation d'intensité d'un phénomène durant un intervalle de temps : augmentation, diminution ou stagnation du taux de criminalité, sur une période de dix ans, selon différents quartiers d'une agglomération ;
- l'existence d'une évolution spatiale d'un phénomène au cours du temps : diffusion d'une épidémie au cours du temps ;
- l'existence de récurrences cyclique d'un phénomène, localisées dans certaines zones de l'espace : réapparition d'épisodes de fortes marées, tous les dix ans en un point *A* d'une façade maritime, et tous les vingt ans en un point *B* d'une façade maritime.

Notre approche est basée sur la représentation du changement au travers de séries d'entités spatio-temporelles, limitées dans le temps et l'espace, et possédant un moment et un lieu d'apparition. Ces entités peuvent représenter un accident routier, un cambriolage, un feu de forêt, etc. Nous désignons ces entités spatio-temporelles sous le nom d'*événements spatio-temporels*.

Cette approche permet de modéliser des phénomènes de différentes natures au travers de séries événementielles, et redéfinit également les éléments que nous cherchons à identifier lors d'une analyse des dynamiques spatio-temporelles. Celle-ci consiste alors à étudier la distribution des événements dans le temps et l'espace, afin d'en dégager des structures, que nous nommons *patterns spatio-temporels*. Ces derniers peuvent être :

1. Les exemples cités ici ont uniquement pour but d'illustrer notre propos. Ils ne correspondent pas nécessairement à des cas réellement observés.

- des *clusters spatio-temporels*, relatifs à une échelle linéaire ou cyclique, révélant l'existence de pics, ou de tendances localisés, dans la variation d'intensité du phénomène ;
- des *propagations d'événements* dans l'espace au cours du temps, traduisant l'évolution spatiale d'un phénomène au cours du temps ;
- des *réurrences cycliques spatialisées* dans la distribution des événements, reflétant les composantes cycliques d'un phénomène.

Structure du mémoire

Notre recherche s'articule autour de quatre axes :

- la spécification d'une approche générique pour la modélisation des données spatio-temporelles, et la formalisation des éléments à rechercher pour permettre l'identification de dynamiques spatio-temporelles lors d'une analyse exploratoire ;
- l'identification, à travers l'établissement d'un état de l'art des approches de visualisation des données spatio-temporelles, des méthodes de représentation graphique susceptibles d'aider à la mise en place de notre approche d'analyse exploratoire ;
- la conception d'un environnement de visualisation autorisant une analyse exploratoire de données spatio-temporelles, en mesure d'identifier les dynamiques citées précédemment, et la formalisation de scénarios d'analyse pouvant être effectuées au moyen de cet environnement ;
- la validation de notre proposition par l'application de notre approche à différentes études de cas.

La modélisation des phénomènes en séries d'événements, et la recherche de patterns

La première partie de notre recherche (Partie I) s'est consacrée à la spécification de nos choix méthodologiques, concernant la modélisation des données spatio-temporelles et la formalisation des éléments à identifier lors d'une analyse exploratoire.

Un premier chapitre (Chapitre 1) présente les étapes du raisonnement qui nous a conduits à opter pour une représentation du changement sous la forme de séries d'événements. Ces *événements* sont regroupés en *phénomènes*, rassemblant des *événements* similaires. Le *phénomène* du volcanisme est constitué de plusieurs *événements* liés entre eux comme des *événements* de coulées de lave et des *événements* sismiques. On peut ainsi considérer le *phénomène* comme représentant l'ensemble d'une série d'*événements* traitant d'une problématique particulière.

Dans un second chapitre (Chapitre 2), nous avons spécifié les dynamiques spatio-temporelles auxquelles nous nous intéressons, puis nous formalisons les structures devant être identifiées dans la distribution spatio-temporelle des événements pour permettre l'étude de ces dynamiques. Dans ce chapitre, nous présentons également les raisons de ce positionnement dans le champ de l'analyse exploratoire. Nous concluons en formalisant les objectifs de notre recherche, concernant l'élaboration d'un environnement de géovisualisation permettant une analyse exploratoire des phénomènes spatio-temporels.

État de l'art des approches de représentation de la donnée spatio-temporelle

Une seconde partie de notre recherche (Partie II) s'est consacrée à l'établissement d'un état de l'art des approches de représentation graphique de la donnée spatiale et temporelle, dans le but d'identifier :

-
- comment ces méthodes intègrent simultanément le temps et l'espace, et comment y sont représentées les différentes composantes de la donnée spatio-temporelle ;
 - quelles approches de représentation graphique et interactives permettent de rechercher et d'identifier des structures dans les distributions spatiale, temporelle, et spatio-temporelle des séries d'événements ;
 - quelles sont les limites des environnements de visualisation actuels pour l'analyse des phénomènes spatio-temporels.

Un troisième chapitre (Chapitre 3) présente les étapes du raisonnement qui nous a conduits à opter pour l'utilisation du principe du *multi-fenêtrage*, formalisé par Davoine [Davoine et al., 2001, Davoine et al., 2004] et Kraak [Kraak and Ormeling, 2011].

Dans un quatrième chapitre (Chapitre 4), nous présentons les limites des environnements de géovisualisation actuels pour l'analyse des phénomènes spatio-temporels. Nous montrons comment les problèmes posés par ces environnements sont susceptibles d'être résolus au moyen d'approches proposées dans d'autres travaux. Cet état de l'art nous a orientés vers l'utilisation de diagrammes temporels de type *timeline* et *timewheel* et *mixte* (représentant le temps de manière linéaire et cyclique) [Edsall and Peuquet, 1997, Li, 2010], combinée à une carte d'inventaire [Swedberg and Peuquet, 2017], ainsi que vers l'utilisation de règles de symbologie pour effectuer le lien entre les composantes temporelle et spatiale des événements représentés [Li, 2010, Bak et al., 2009, Andrienko et al., 2011].

GrAPHiST

Une troisième partie de notre recherche (Partie III) s'est consacrée à l'élaboration d'une approche d'analyse exploratoire des phénomènes.

En parallèle de notre travail bibliographique, nous avons implémenté notre propre environnement de géovisualisation, *GrAPHiST*, pour (*environnement de*) *Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels*.

Le développement de ce dernier s'est effectué selon une approche *bottom up*. L'implémentation de *GrAPHiST* a été effectuée d'entrée, avant que soient formalisés l'ensemble de ses fonctions et objectifs. Le développement de *GrAPHiST* a ainsi suivi une trajectoire empirique : la structure de l'environnement, et les différents outils qu'il intègre, ont évolué au cours de l'établissement de notre état de l'art, de l'avancement du développement, et jusqu'à l'achèvement de notre phase d'expérimentation. La formalisation des différentes composantes de l'environnement, a été effectuée une fois ce dernier implémenté.

Notre proposition s'inscrivant dans une approche exploratoire, l'utilisateur conserve un rôle central dans l'analyse des dynamiques spatio-temporelles. La conception de *GrAPHiST* s'est accompagnée de la formalisation de différents scénarios d'analyse, décrivant les différentes actions que doit mener l'utilisateur au moyen de l'outil, pour identifier chacune des dynamiques spatio-temporelles citées précédemment lors d'une analyse exploratoire. La conception de *GrAPHiST* ayant suivi une approche *bottom up*, la manière dont l'analyste interagit avec l'environnement, pour trouver des patterns spatio-temporels, a évolué au cours de l'avancée du développement.

Un cinquième chapitre (Chapitre 5) est consacré à la structure et les différentes composantes de *GrAPHiST*. Ce chapitre présente :

- le modèle de données développé pour l'environnement ;
- la structure générale de l'interface de *GrAPHiST* ;

- comment les approches de visualisation, identifiées lors de l'établissement de l'état de l'art, sont intégrées à l'interface ;
- comment *GrAPHiST* intègre l'utilisation de règles de symbologie, afin de permettre l'identification de patterns spatio-temporels.

Un sixième chapitre (Chapitre 6) présente les différents scénarios d'analyse exploratoire pouvant être effectués au moyen de *GrAPHiST*.

Phase d'expérimentation

Dans le but de valider notre proposition, nous avons appliqué notre approche d'analyse exploratoire à différents cas d'étude. Cette phase d'expérimentation a en outre permis de parfaire l'implémentation de *GrAPHiST* et la formalisation de son utilisation :

- L'analyse des différents cas d'étude a permis de révéler de nouveaux besoins, pour lesquels de nouvelles fonctionnalités ont été intégrées à *GrAPHiST*.
- L'utilisation concrète de l'interface de *GrAPHiST* a permis de préciser les actions à effectuer pour accomplir une recherche de patterns.

La quatrième partie de ce mémoire (Partie IV) est consacrée aux différentes expérimentations effectuées afin de valider notre proposition.

Un septième chapitre (Chapitre 7) présente une analyse de données historiques représentant les lieux et dates des batailles de la Guerre de Sécession. L'objectif est de prouver la possibilité d'identifier des patterns relatifs au temps linéaire, en recréant un récit du conflit à partir de l'analyse des données historiques, avant de comparer ce récit à celui produit par les historiens.

Un huitième chapitre (Chapitre 8) présente l'analyse de jeux de données simulées, présentant chacun une part d'événements cycliques, localisés dans l'espace, et une part d'événements aléatoires. La période du cycle, et ses étendues temporelle et spatiale, sont définies aléatoirement et gardées cachées jusqu'à l'achèvement de l'expérimentation. L'objectif est de prouver la possibilité d'identifier des patterns relatifs à une échelle cyclique, en identifiant ces récurrences cycliques spatialisées, avant de comparer les résultats trouvés aux caractéristiques des jeux de données créés.

Un neuvième et dernier chapitre (Chapitre 9) présente une application de notre proposition à l'analyse de différents jeux de données réels, liées à des problématiques de différentes natures. Cette expérimentation a pour objectif de montrer le caractère générique de notre approche, et de présenter les opportunités offertes par l'environnement dans l'analyse de données spatio-temporelles.

Première partie

L'étude des dynamiques spatio-temporelles par l'analyse exploratoire de la distribution spatio-temporelle des événements

Introduction

Le changement dans l'espace est décrit dans la littérature comme la conséquence de *phénomènes spatio-temporels* [Haddad, 2009]. Ces derniers peuvent être vus comme des objets limités dans le temps et l'espace, pouvant décrire différents types de changements. Une éruption volcanique, un séisme, ou l'augmentation d'une population peuvent être décrits comme des *phénomènes spatio-temporels*.

Le mécanisme de ces phénomènes, leurs variations, leur rythmes dans le temps et l'espace sont décrits par les *dynamiques spatio-temporelles* de ces phénomènes. L'analyse de ces dernières permet d'améliorer les connaissances existantes sur les phénomènes, d'effectuer une prévision de l'apparition ou de l'évolution de phénomènes semblables, et d'aider à la prise de décisions les concernant. En fonction des phénomènes étudiés, et des problématiques qu'ils font intervenir, l'analyse peut se concentrer sur certaines dynamiques spatio-temporelles. Ainsi :

- l'analyse de phénomènes volcaniques peut aider à la mise en place de politiques de prévention du risque, et implique de s'intéresser aux récurrences cycliques du phénomène ;
- l'analyse de l'évolution d'une population peut aider à la mise en place de politiques publiques, et implique d'analyser les variations du phénomène sur le temps long selon différentes zones de l'espace.

Les données disponibles, permettant de décrire le changement dans l'espace, sont aujourd'hui en constante augmentation. Ces données recouvrent des problématiques diverses, permettent de représenter différents types de phénomènes, et peuvent être éloignées des données habituellement observées pour l'étude des phénomènes spatio-temporels. L'analyse de ces données peut cependant être génératrice de connaissances dans différents domaines d'étude.

La vaste étendue des phénomènes pouvant être représentés par ces données incite l'utilisation d'approches d'analyse génériques, applicables à différents phénomènes dans le cadre de différentes problématiques. Proposer ces approches implique de s'interroger sur :

- la modélisation du changement dans l'espace, de manière à proposer une modélisation générique permettant de traiter différents types de phénomènes spatio-temporels ;
- la nature des différentes dynamiques spatio-temporelles à analyser, et la forme sous laquelle elles se présentent selon le format de données adopté pour l'analyse ;
- quelles approches d'analyse permettent l'étude de données pouvant être éloignées de celles habituellement observées pour l'étude des phénomènes spatio-temporels ;
- comment ces approches peuvent-elles offrir la possibilité d'analyser différents types de dynamiques spatio-temporelles, afin de répondre à différents objectifs, ces derniers étant liés au phénomène observé et à la problématique de l'étude.

Dans un premier chapitre (1), nous passons en revue les différentes approches développées au cours du temps pour modéliser les phénomènes et leurs conséquences dans l'environnement spatial. À partir de cet état de l'art, nous proposons une approche générique de modélisation des

phénomènes spatio-temporels. Ceux-ci sont modélisés au travers de séries d'événements *spatio-temporels* limités dans le temps et l'espace.

Le *phénomène* est représenté comme une série d'événements traitant d'une même problématique. Cette modélisation a pour but de permettre la mise en forme des données spatio-temporelles relatives à différents cas d'étude sous un même format de modélisation, permettant leur analyse dans un seul environnement d'analyse.

Dans une première partie d'un second chapitre (2), nous définissons les différentes *dynamiques spatio-temporelles* sur lesquelles nous nous focalisons. Puis, nous définissons la forme prise par ces dynamiques dans le cadre de notre proposition de modélisation des données spatio-temporelles.

Dans une seconde partie du chapitre 2, nous montrons comment l'approche exploratoire peut répondre à nos besoins, en permettant une étude des phénomènes ne nécessitant pas l'énonciation d'hypothèses *a priori* sur la structure des données analysées, permettant ainsi l'analyse de nouvelles données spatio-temporelles, inconnues de l'utilisateur. Nous montrons enfin quels sont les travaux à mener pour permettre une analyse exploratoire des phénomènes spatio-temporels, applicables à différents cas d'étude.

Chapitre 1

La modélisation des *phénomènes spatio-temporels*

Sommaire

1.1	La modélisation des dimensions temporelle et spatiale	30
1.1.1	La modélisation de la dimension spatiale	30
1.1.2	La modélisation de la dimension temporelle	32
1.2	La modélisation de l'information spatio-temporelle	37
1.2.1	L'approche <i>endurante</i> et l'approche <i>perdurante</i>	37
1.2.2	Décrire les phénomènes au travers des entités <i>SNAP</i> et <i>SPAN</i> : les modèles <i>3D+1</i> et <i>4D</i>	39
1.2.3	<i>Triad Model</i> et <i>Pyramid Model</i>	41
1.3	Le phénomène spatio-temporel modélisé selon un point de vue endurant	42
1.3.1	La modélisation implicite du changement	42
1.3.2	La modélisation explicite du changement	45
1.3.3	Le choix de l'approche <i>4D</i>	47
1.4	Le phénomène modélisé selon un point de vue occurrent	49
1.4.1	<i>Événements</i> et <i>processus</i>	49
1.4.2	L'approche de Grenon et Smith	50
1.4.3	L'approche de Reitsma	50
1.4.4	Le choix de l'approche de Reitsma	50
1.5	<i>Événements</i> et <i>Phénomènes</i>	52
1.5.1	L' <i>Événement</i> , entité spatio-temporelle représentant le changement	52
1.5.2	Les <i>Phénomènes</i> , regroupements d' <i>événements</i>	53
1.5.3	Une modélisation générique du changement au travers des <i>Phénomènes</i>	54

Selon certains auteurs, les notions fondamentales de la géographie sont l'espace, les relations dans l'espace et les changements dans l'espace [Morrill, 1970]. Le terme *changement* sous-entend la dimension temporelle. L'étude du changement dans l'espace consiste à décrire l'évolution de l'environnement spatial d'un état initial à un autre, au travers d'une succession temporelle de modifications [Haddad, 2009]. Dans la littérature, on considère le changement comme étant la conséquence de *phénomènes géographiques dynamiques*.

En philosophie, un *phénomène* est un fait du monde physique (objet, action, etc.), psychique (émotion, pensée, etc.) ou social (produit d'interactions sociales, etc.) qui se manifeste par lui-même [Mathieu, 2015]. Un *phénomène géographique* est un ensemble de faits et de caractéristiques observables, localisables sur la surface terrestre [Haddad, 2009]. Ces *phénomènes géographiques* peuvent être *statiques*, et représenter un ensemble de faits stables, fixes dans le temps, ou *dynamiques* et représenter le passage d'un état à un autre au cours du temps [Janet, 1885, Haddad, 2009]. Les *phénomènes géographiques dynamiques* sont également nommés *phénomènes spatio-temporels* dans la littérature, pour souligner que le changement s'effectue dans le temps et dans l'espace [Haddad, 2009].

La modélisation (ou représentation) des *phénomènes spatio-temporels* est un domaine de recherche relativement récent. Les modèles proposés pour représenter le changement ont souvent été construits de manière à convenir à des problématiques particulières plutôt que selon une approche théorique [Peuquet, 2002, Yuan, 2007, Haddad, 2009]. L'objet de ce chapitre est de définir quelles approches théoriques permettent d'offrir une modélisation générique du changement dans l'espace, dans le but d'analyser les *dynamiques spatio-temporelles* de différents *phénomènes*.

Comme le sous-entend leur nom, l'étude des *phénomènes spatio-temporels* implique l'étude de ces deux dimensions. Dans un premier temps, nous nous intéressons à la modélisation de ces deux composantes dimensionnelles, avant de nous intéresser à la modélisation de l'espace-temps.

1.1 La modélisation des dimensions temporelle et spatiale

La modélisation du changement dans l'espace pose la question de la modélisation de l'espace et du temps, domaines initialement étudiés séparément dans la littérature. Nous donnons ici un aperçu général sur les travaux portant sur ces deux dimensions.

1.1.1 La modélisation de la dimension spatiale

Les deux modèles conceptuels les plus utilisés pour la modélisation de l'espace [Couclelis, 1992, Goodchild, 1992] sont :

- l'approche *objet-centrée*, qui s'inscrit dans une perspective discrète de l'espace ;
- l'approche *champ-centrée*, qui s'inscrit dans une perspective continue.

L'adoption de l'un ou l'autre de ces deux points de vue dépend soit d'une posture a priori du chercheur, soit de son choix face à un objectif précis [Mathian and Sanders, 2014], une même problématique pouvant être approchée selon ces deux points de vue.

1.1.1.1 L'approche *objet-centrée*

L'approche *objet-centrée* consiste à représenter l'espace spatial en une composition d'objets géographiques. Ces derniers sont des entités spatiales discrètes, localisées et limitées dans l'espace. Chacun de ces objets possède une identité propre, une localisation, une géométrie (ils

peuvent être ponctuels, linéaires ou surfacique) et des attributs thématiques.

Ces objets peuvent être simples, ou composites. Différents objets géographiques peuvent aussi être liés selon une relation hiérarchique. Un objet géographique composite de rang n peut être formé d'objets, simples ou composites, de rang $n-1$, eux-mêmes formés d'objets géographiques, simples ou composites, de rang $n-2$. Par exemple, l'Union européenne est formée de plusieurs pays, formés eux-mêmes par plusieurs régions. On modélise ainsi plusieurs catégories d'objets qui sont généralement organisées hiérarchiquement [Haddad, 2009]. La notion de parties est souvent liée à la notion de l'échelle spatiale considérée. Plus une échelle est grande, plus une même entité géographique peut être modélisée comme composite, et formée de sous-parties.

Les objets géographiques peuvent correspondent à des entités concrètes, comme des rivières ou des routes, ou à des entités abstraites, comme des communes ou des pays. La distinction entre entités concrètes ou abstraites renvoie à l'aspect tangible ou abstrait des frontières de ces objets [Smith et al., 1994]. Une route possède des frontières physiques, tangibles. Une commune possède des frontières abstraites, que des personnes ont déterminées. On parle, dans ces cas d'entités concrètes, d'objets *bona fide*¹, et dans le second cas, d'objets *fiat*² [Smith et al., 1994, Smith and Varzi, 2000].

Chaque objet géographique est relié aux autres par des relations topologiques [Haddad, 2009] (Figure 1.1).

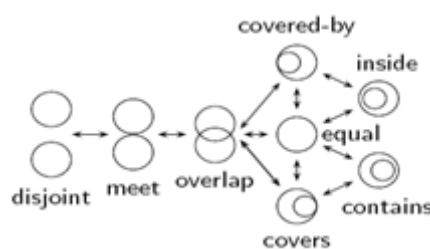


FIGURE 1.1 – Relations topologiques entre objets spatiaux [Haddad, 2009]

L'approche *objet-centrée* est la plus utilisée [Fonseca et al., 2002] et la plus facilement conceptualisable, car plus ancrée dans le langage commun [Galton, 2003, Mathian and Sanders, 2014]. Cette construction permet de définir des objets géographiques correspondant aux entités existantes dans notre représentation du monde (ville, route, etc.).

1.1.1.2 L'approche *champ-centrée*

Dans l'approche *champ-centrée*, appelé aussi *location-based* [Peuquet, 1988], l'espace est divisé en une grille de localisations réparties régulièrement, correspondant à une primitive spatiale qui peut être un point ou une aire. Chacune de ces localisations est rattachée à des attributs prenant une valeur spécifique à la localisation considérée [Galton, 2004, Mathian and Sanders, 2014].

La densité de localisations permettant de couvrir l'espace géographique peut être définie comme la *granularité spatiale* de la modélisation de l'espace. On appelle alors *granules spatiaux* les localisations insécables réparties régulièrement sur l'espace. La granularité spatiale définit ainsi la finesse de la modélisation de l'espace effectuée.

1. du latin *bona fide* : *de bonne foi*, utilisé en langue anglaise pour signifier *authentique*

2. du latin *fiat* : *qu'il soit fait*, utilisé en langue anglaise pour faire référence à un ordre, un décret, une autorisation

L'espace géographique est ainsi conceptualisé comme une grille régulière permettant la représentation de la distribution spatiale d'une grandeur. L'approche *champ-centrée* est très utilisée, par exemple, par les spécialistes en géosciences et en environnement [Mathian and Sanders, 2014], car elle est appropriée pour représenter une information géographique se présentant sous la forme d'une distribution spatiale continue d'une grandeur qualitative ou quantitative (occupation du sol, précipitation, température).

1.1.1.3 L'implémentation des approches *objet-centrée* et *champ-centrée* dans les systèmes d'information géographique

Un système d'information géographique *SIG* est un *système informatique de matériels, de logiciels, et de processus conçus pour permettre la collecte, la gestion, la manipulation, l'analyse, la modélisation, et l'affichage de données à référence spatiale afin de résoudre des problèmes complexes d'aménagement et de gestion* [Moellering, 1988].

La dichotomie *objet-centrée* / *champ-centrée* se retrouve dans les *SIG*, au travers de la dichotomie de la représentation de la donnée géographique en formats *vecteur* et *raster* [Haddad, 2009] (Figure 1.2).

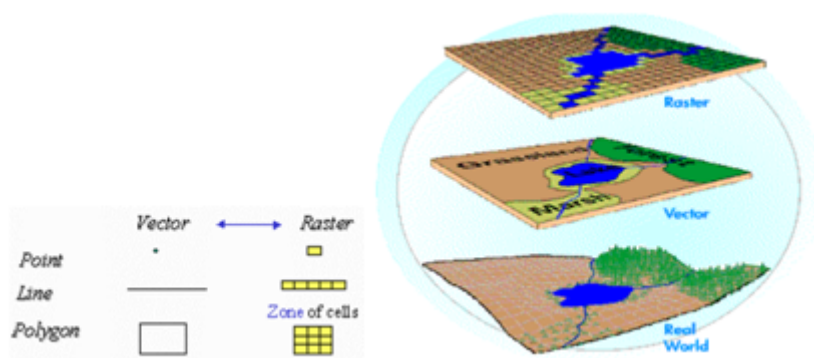


FIGURE 1.2 – Le modèle vecteur vs le modèle raster [Haddad, 2009]

Le format *vecteur* est une implémentation du modèle *objet-centrée*, où l'espace est occupé par des objets discrets, possédant un identifiant, une géométrie (ponctuelle, linéaire ou polygonale), et une suite d'attributs.

Le format *raster* est une implémentation du modèle *champ-centrée*, où l'espace est divisé selon une grille régulière et continue d'entités surfaciques élémentaires, dont l'aire, caractérisant la granularité spatiale de la représentation, est définie par l'utilisateur. Chaque entité élémentaire, ou granule, est rattachée à une valeur prise par une grandeur qualitative ou quantitative.

1.1.2 La modélisation de la dimension temporelle

Parmi les travaux portant sur la modélisation de l'écoulement du temps, on peut discerner quatre distinctions philosophiques, complémentaires et non concurrentes, permettant d'appréhender la dimension temporelle [Haddad, 2009]. Ces distinctions aboutissent sur différentes propositions de représentations temporelles. Ces quatre distinctions sont :

- la distinction entre temps *linéaire* et temps *cyclique* ;
- la distinction entre temps *absolu* et temps *relatif* [Schmidt et al., 2003] ;
- la distinction entre temps *ordinal*, temps *discret*, et temps *continu* [Pelekis et al., 2004, Solana, 2014], accompagnée de la notion de *granularité temporelle* [Kakoudakis and Theodoulidis, 1996] ;

- la distinction entre primitive temporelle *ponctuelle*, et primitive temporelle *durative* (possédant une durée) [Haddad, 2009].

1.1.2.1 Temps *linéaire* et temps *cyclique*

Dans une approche linéaire, l'écoulement du temps est conceptualisé comme la succession chronologique de primitives temporelles (moments, ou intervalles temporels). On parle de représentation du temps au sein d'une *échelle linéaire* (Figure 1.3).

Dans une approche cyclique, l'écoulement du temps est conceptualisé comme la répétition d'une même période temporelle, dont la durée correspond à un nombre fixe de primitives temporelles. On parle de représentation du temps au sein d'une *échelle cyclique* (Figure 1.3). On appelle *durée de l'échelle cyclique* la durée de la période sur laquelle on projette l'écoulement du temps.

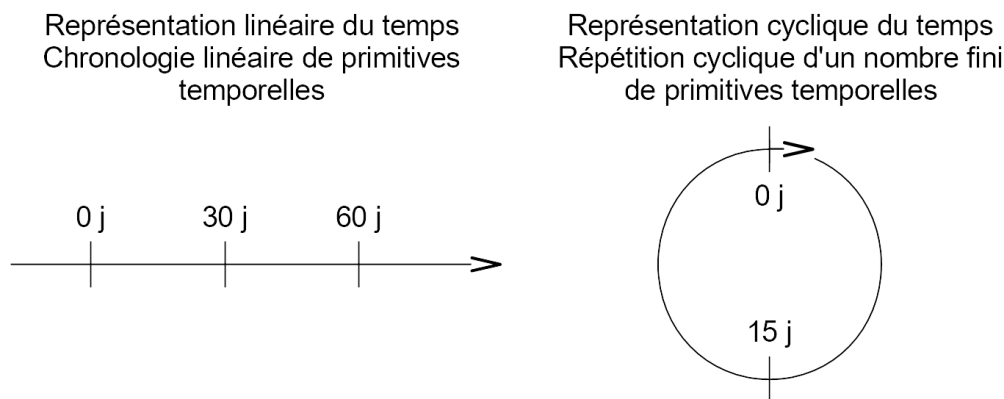


FIGURE 1.3 – Représentations linéaires et cycliques du temps

La représentation selon une échelle temporelle linéaire est prépondérante dans les modélisations de données temporelles. Les bases de données temporelles, par exemple, utilisent un étiquetage des données (une datation des données) selon leur moment d'apparition dans une échelle temporelle linéaire. La représentation selon une échelle cyclique reste cependant intéressante pour mettre en lumière la répétition de certains phénomènes, au travers du cycle des jours ou des saisons par exemple.

Ces deux modélisations sont parfois utilisées de manières combinées, au travers de la représentation du temps comme une succession chronologique d'instances d'une échelle cyclique : l'écoulement du temps peut être représenté comme la succession chronologique de périodes cyclique d'une durée d'un an.

1.1.2.2 Temps *absolu* et temps *relatif*

La représentation du temps comme *absolu* consiste à représenter un moment temporel selon sa "position" au sein d'une échelle temporelle, linéaire ou cyclique. Par exemple :

- le *17 novembre 2002*, pour une échelle temporelle linéaire occidentale ;
- *9h00*, pour une échelle temporelle cyclique d'une journée.

La représentation du temps comme *relatif* consiste à représenter un moment temporel par rapport à d'autres moments temporels. Cette représentation permet d'introduire les notions de *passé* (*deux jours avant un événement*), du *présent* (*au même moment qu'un événement*), et du

futur (deux jours après un événement).

Le temps relatif permet donc d'introduire des relations chronologiques entre différents instants ou durées, semblables aux relations topologiques entre objets spatiaux. Celles-ci sont formalisées dans les *relations d'Allen* [Allen, 1983] (Figure 1.4).

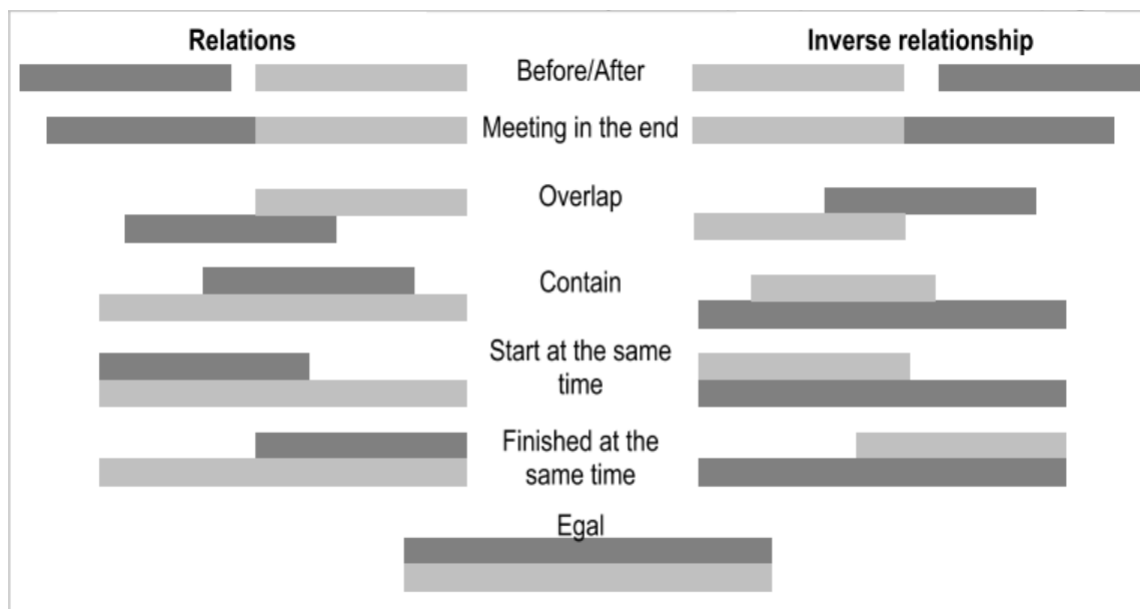


FIGURE 1.4 – Les relations d'Allen [Allen, 1983]

1.1.2.3 Temps *ordinal*, temps *discret* et temps *continu*

La distinction entre temps *ordinal* et temps *discret* d'une part, et temps *continu* d'autre part, est à rapprocher de la distinction entre les modélisations spatiales *objet-centrée* et *champ-centrée*. Ce parallèle est notamment effectué par Galton [Galton, 2004], lorsqu'il étend au temps la dichotomie entre les approches *objet-centrée* et *champ-centrée* utilisées pour l'espace [Mathian and Sanders, 2014].

La représentation temporelle *ordinale* consiste à considérer l'écoulement du temps comme une chronologie de phases temporelles successives, dont on connaît l'ordre de succession, mais ni la durée, ni la position temporelle dans une échelle temporelle absolue [Solana, 2014]. Chaque phase temporelle possède une identité propre, et peut être rattachée à des attributs. Un exemple de représentation temporelle *ordinale* peut être la représentation d'une vie humaine au travers de la succession des phases de l'enfance, de l'âge adulte, et de la vieillesse.

La représentation temporelle *discrète* consiste également à représenter l'écoulement du temps comme la chronologie de phases temporelles successives, mais dont la durée et la position temporelle sont connues. L'espace temporel peut ainsi être décrit comme occupé par des objets temporels, limités dans le temps, pouvant être ponctuels ou étendus dans le temps, possédant une identité propre, et pouvant être rattachés à des attributs.

La représentation *discrète* du temps est à rapprocher de la représentation *objet-centrée* de l'espace. L'objet temporel distinct, limité dans le temps, et aux frontières temporelles définies peut être comparé à l'objet spatial distinct, limité dans l'espace, et aux frontières spatiales définies. De la même manière, les objets temporels peuvent être simples ou composites, et être organisés

hiérarchiquement. Une conférence peut être représentée par un objet temporel occupant une certaine portion du temps. Cette conférence peut être composée de plusieurs sous-objets temporels représentant chaque jour de la conférence.

Le temps *ordinal* et le temps *discret* s'inscrivent tout deux dans une vision discrète, *objet-centrée* du temps, où l'espace temporel est peuplé d'objets temporels. La représentation *ordinaire* décrit ces objets au travers d'une représentation relative du temps. La représentation *discrète* décrit ces objets au travers d'une représentation absolue du temps.

Hewitt [Hewitt, 1983], Allen [Allen, 1983], DiBiase [DiBiase et al., 1992] et MacEachren [MacEachren, 1994] décrivent six *temporalités* permettant de caractériser ces entités temporelles discrètes 1.5.

Trois temporalités caractérisent la composante temporelle d'une entité temporelle en particulier (Figure 1.5) :

- le *moment d'apparition*, l'instant temporel d'occurrence de l'entité ;
- la *durée* de l'entité ;
- la *période de retour*, représentant l'étendue temporelle entre deux instances du même type d'entité temporelle. La notion de *période de retour* est ambiguë. Dans le domaine des risques, la *période de retour* est considérée comme l'écart de temps probable entre deux événements [Arnaud and Davoine, 2009]. La *période de retour* est une notion probabiliste. Elle n'est pas à rapprocher de la *période* des phénomènes périodiques (comme la période de la révolution de la Terre autour du Soleil).

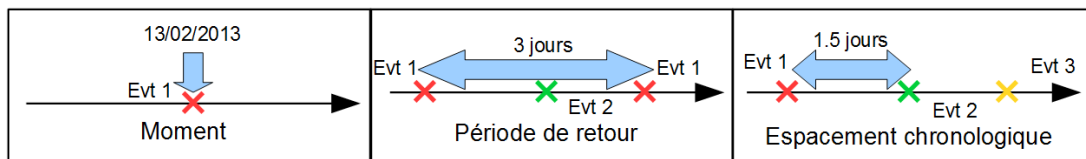


FIGURE 1.5 – Les temporalités caractérisant une entité temporelle [Hewitt, 1983][Allen, 1983, DiBiase et al., 1992, MacEachren, 1994]

Trois temporalités caractérisent les relations temporelles entre plusieurs entités (Figure 1.6) :

- l'*ordre*, décrivant l'enchaînement chronologique de plusieurs entités temporelles ;
- l'*espacement chronologique*, l'étendue temporelle entre deux entités temporelles ;
- la *synchronisation* entre l'apparition de plusieurs entités temporelles.

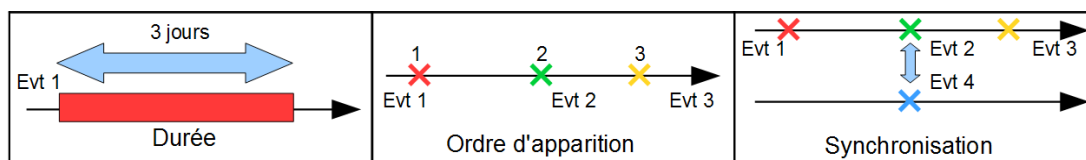


FIGURE 1.6 – Les temporalités caractérisent les relations temporelles entre entités [Hewitt, 1983][Allen, 1983, DiBiase et al., 1992, MacEachren, 1994]

Les trois dernières temporalités se rapprochent des *relations d'Allen* [Allen, 1983] (Figure 1.4), équivalents temporels des relations topologiques reliant entre eux les objets géographiques. Certains travaux proposent d'autres temporalités telle la *fréquence*, représentant le nombre d'occurrences d'un objet temporel durant une période donnée [Arnaud and Davoine, 2009]. Cette *fréquence* n'est pas à rapprocher de la *période de retour*, ou de la fréquence des phénomènes périodiques, inverse de la période de réapparition de ces phénomènes.

La représentation temporelle *continue* est à rapprocher de la modélisation *champ-centrée* de l'espace. Dans cette représentation, il existe toujours, en théorie, un intervalle temporel divisible entre deux points du temps. En pratique, le temps est divisé selon une grille régulière de primitives temporelles, représentées par des instants ou des intervalles temporels de durée égale. Chaque primitive est rattachée à une ou à des grandeurs prenant une valeur propre à l'instant ou à l'intervalle de temps considéré.

À l'instar de la modélisation *champ-centrée* de l'espace, une *granularité temporelle* définit la densité de primitives temporelles insécables, ou *granule temporel*, occupant l'espace temporel. La granularité temporelle définit ainsi la finesse de la modélisation du temps effectuée. Cette approche permet ainsi la représentation d'une distribution temporelle continue d'une grandeur qualitative ou quantitative.

Cette représentation du temps peut apparaître comme discrète, car il n'y a pas d'intervalle de temps divisible entre deux primitives temporelles. Mais le caractère continu de la représentation fait référence à la division régulière (continue) du temps selon une granularité temporelle, qui s'oppose à la division du temps en différentes phases temporelles distinctes, aux frontières non régulières, possédant une identité propre.

Cette distinction entre représentation *continue* et *discrète* du temps se retrouve dans la construction de la donnée temporelle en séries *temporelles* et séries *événementielles*.

Une série *temporelle* consiste à étiqueter l'évolution de la valeur prise par une grandeur G , observée selon un pas de temps régulier appelé *pas d'échantillonnage* (Figure 1.7). Une série *temporelle* s'inscrit dans une représentation *continue*, le *pas d'échantillonnage* correspondant à la *granularité temporelle* utilisée.

Date	Grandeur G
21/01/17	10
22/01/17	12
23/01/17	11
24/01/17	13
25/01/17	14
26/01/17	15

FIGURE 1.7 – Exemple d'une série temporelle

Dans une série *événementielle*, chaque entrée d'information correspond à un moment temporel où quelque chose de remarquable est observable (Figure 1.8). Ce peut être, par exemple, lorsqu'une grandeur G atteint une certaine valeur, ou lorsque quelque chose de succinct est observable.

Événement	Date	Grandeur G, rattachée à l'événement
Evt ₁	21/01/17	10
Evt ₂	21/01/17	13
Evt ₃	23/01/17	15
Evt ₄	24/01/2017 – 26/01/2017	16
Evt ₅	27/01/17	17
Evt ₆	27/01/17	15

FIGURE 1.8 – Exemple d'une série événementielle

Les moments temporels des différentes entrées d'information se succèdent, mais sont séparés

entre eux par des intervalles de temps irréguliers. Ce stockage d'information est à rapprocher d'une représentation *discrète* du temps, chaque entrée d'information représentant un objet temporel.

1.1.2.4 Primitive temporelle *ponctuelle* et primitive temporelle *durative*

La primitive temporelle constitue le plus petit élément, insécable, de représentation du temps.

Dans certains modèles [McCarthy and Hayes, 1969, McDermott, 1982, Haddad, 2009], l'unité temporelle élémentaire est un instant ponctuel. Dans ce modèle, si l'on considère un instant ponctuel T en particulier, il existe un état *anté- T* , et *post- T* . L'instant T ouvre ainsi un ou plusieurs intervalles temporels pouvant se terminer à T' , T'' , etc.

Entre deux instants successifs, on reste cependant dans un état figé. Si un changement a lieu à l'instant T , ce changement a lieu à uniquement à l'instant T . On passe d'un état figé *anté- T* , à un autre état figé *post- T* . Ainsi, s'il est possible de représenter des changements concurrents (démarrant au même instant temporel), ces changements sont ponctuels. Les changements duratifs, étendus dans le temps, ne peuvent pas être représentés en suivant cette approche [Worboys, 2005].

Dans d'autres modèles, comme celui d'Allen [Allen, 1984, Haddad, 2009], l'unité temporelle élémentaire est un intervalle temporel. Cette approche permet ainsi de représenter un changement s'étalant sur une durée, entre un état *anté-changement* et un état *post-changement*, séparés par un état de transition lors duquel s'effectue un changement. Cette approche ne permet cependant pas de représenter des instants ponctuels [Galton, 1990], et donc des changements succincts.

1.2 La modélisation de l'information spatio-temporelle

1.2.1 L'approche *endurante* et l'approche *perdurante*

Les *phénomènes spatio-temporels* sont définis en utilisant deux concepts [Haddad, 2009] :

- d'une part, le changement résultant.
- d'autre part, l'entité spatio-temporelle qui est à l'origine de ce changement, c'est-à-dire un *événement* ou un *processus*.

Ces deux concepts se retrouvent dans la vision du monde proposée par Grenon et Smith [Grenon and Smith, 2004], qui proposent de classer les différentes entités du monde en entités *SNAP* (Repère *A* de la Figure 1.9), et en entités *SPAN* (Repère *B* de la Figure 1.9).

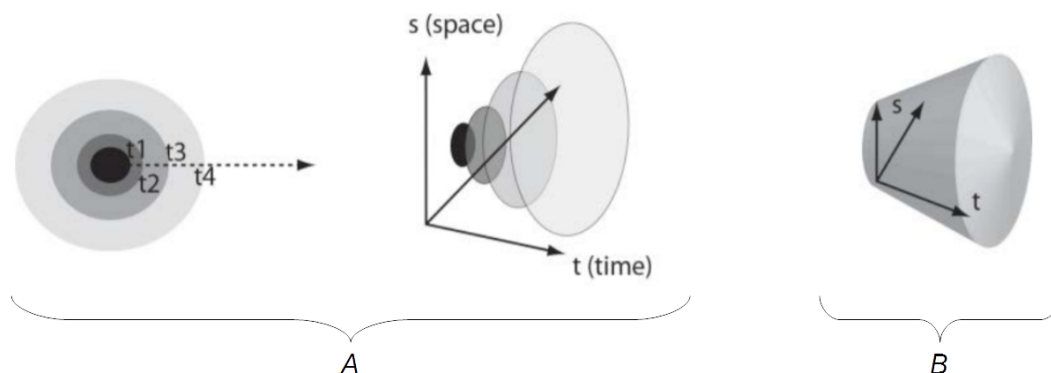


FIGURE 1.9 – Approches *SNAP* et *SPAN* de la modélisation de l'information spatio-temporelle [Mathian and Sanders, 2014]

1.2.1.1 Les entités *SNAP*

Les entités *SNAP*, appelées *continuanes* ou *endurantes* [Mathian and Sanders, 2014], sont des entités spatiales capables de persister dans le temps (d'où le nom *endurantes* : ce sont des entités qui *endurent* des changements).

Ces entités sont pérennes dans le temps, et par là même atemporelles : elles ne possèdent pas de composante temporelle [Hawley, 2004]. Elles possèdent cependant une composante spatiale et une composante thématique qui peut évoluer au cours du temps. Ces entités spatiales gardent leur identité au cours du temps, et peuvent être observées dans leur entièreté à chaque instant de leur existence, d'où le nom de *SNAP* qui se réfère au terme *snapshot*, exprimant l'observation d'une de ces entités à un moment t .

Un cours d'eau, observé sur une étendue temporelle d'un siècle, peut être modélisé comme une entité *endurante* :

- Sur une étendue temporelle d'un siècle, un cours d'eau peut être considéré comme pérenne, atemporel.
- Ce cours d'eau possède une identité propre, qui reste inchangée sur toute l'étendue temporelle d'observation.
- Il possède une dimension spatiale qui peut évoluer au cours du temps. Son étendue spatiale peut ainsi évoluer à la suite d'une crue.
- Il possède une dimension thématique qui peut également évoluer au cours du temps. Au cours d'une crue, son débit peut évoluer.
- Enfin ce cours d'eau, et ses composantes spatiales et thématiques, peuvent être décrits dans leur intégralité, que l'on observe ce cours d'eau le *2 janvier 1920* ou le *20 mars 1993*.

Dans une approche de l'espace *objet-centrée*, les entités *SNAP* peuvent modéliser les objets spatiaux recouvrant l'environnement géographique : forêts, route, ville, individus. Les propriétés thématiques rattachées à ces objets (taille d'une forêt ou d'une ville, longueur d'une route, âge d'un individu) peuvent également être considérées comme des entités *SNAP* [Mathian and Sanders, 2014].

Dans une approche de l'espace *champ-centrée*, les entités *SNAP* peuvent modéliser des regroupements de primitives spatiales, partageant une même propriété [Mathian and Sanders, 2014]. Un tel regroupement définit une étendue spatiale associée à la propriété partagée par les primitives spatiales (ensemble de points présentant une même occupation des sols). Cette étendue spatiale peut alors être considérée comme une entité *endurante* dont on observe l'évolution au cours du temps (évolution de la surface présentant une même occupation des sols).

1.2.1.2 Les entités *SPAN*

Les entités *SPAN*, appelées *occurentes* ou *perdurantes*, sont des entités spatio-temporelles limitées dans l'espace et le temps. Elles possèdent, à l'inverse des entités *SNAP*, une composante temporelle (*SPAN* étant un anagramme de *SNAP*).

Ces entités possèdent une date de début et une date de fin, entre lesquelles se succèdent des parties temporelles [Grenon and Smith, 2004, Livet and Nef, 2009, Mathian and Sanders, 2014]. Chaque partie temporelle correspond à un état de l'entité, dans le temps et l'espace, cet état n'évoluant pas durant l'écart de temps défini par la partie temporelle. La succession des parties temporelles correspond ainsi à l'enchaînement des différentes phases d'existence de l'entité

spatio-temporelle.

Pour cette raison, ces entités ne sont pas observables dans leur entièreté à chaque instant t , à la manière d'une entité *SNAP*. Afin de décrire une entité *SPAN* dans son entièreté, il est nécessaire de l'observer sur l'ensemble de son existence, et de décrire l'ensemble formé par les parties temporelles de l'entité.

Une inondation peut être modélisée comme une entité *occurrente* :

- Elle est limitée dans le temps, et possède une date de début et une date de fin.
- Elle est limitée dans l'espace, et possède une emprise spatiale.
- Entre les dates de début et de fin, on peut diviser la durée d'existence de l'inondation en différentes phases temporelles, chacune représentant un état différent de l'inondation. Cette division peut refléter les différentes extensions spatiales de l'inondation au cours du temps, les différentes variations du débit d'un cours d'eau au cours du temps, ou les différentes séquences de l'inondation (crue, stabilisation, décrue).
- Enfin, il est nécessaire, pour définir l'entièreté de l'inondation, de l'observer sur toute la durée de son existence. L'inondation ne peut être décrite dans son intégralité à partir d'une observation à un instant t .

1.2.1.3 Complémentarité des deux approches

Souvent présentées comme opposées en philosophie, ces deux approches peuvent être vues comme complémentaires [Mathian and Sanders, 2014]. Bittner et al. [Bittner et al., 2004, Mathian and Sanders, 2014] soulignent que ces deux types d'entités sont liés. Des entités *endurantes* existent et occupent l'environnement spatial de façon pérenne. Leurs composantes spatiale ou thématique sont des entités *endurantes* qui subissent des modifications au cours du temps. Ces modifications, et la trajectoire de vie formée par l'ensemble de ces changements, sont des entités *occurrentes*.

Une rivière, entité *endurante*, est impactée par une inondation, entité *occurrente*. Le phénomène de déforestation, entité *occurrente*, se manifeste par une diminution de l'étendue spatiale des forêts, entités *endurantes*. Un navire est une entité *endurante*, son voyage d'un port à un autre une entité *occurrente*. Dans le cadre d'une approche *champ-centrée* de l'espace, un agrégat de primitives spatiales partageant une propriété est une entité *endurante*. L'évolution géométrique de l'agrégat est une entité *occurrente*.

Selon Grenon et Smith [Grenon and Smith, 2004], on distingue ainsi deux relations possibles entre ces entités, la *participation* et l'*implication* :

- les entités *SNAP* participent à des entités *SPAN* : un navire participe à une traversée de la mer.
- les entités *SPAN* impliquent des entités *SNAP* : une inondation implique des parcelles inondées.

1.2.2 Décrire les phénomènes au travers des entités *SNAP* et *SPAN* : les modèles $3D+1$ et $4D$

Galton [Galton, 2004] fait un lien entre la dichotomie entre entités *SNAP* et *SPAN* et la manière de conceptualiser l'espace et le temps [Mathian and Sanders, 2014].

Ce parallèle est à rapprocher de celui que nous faisons précédemment entre l'approche spatiale *champ-centrée* et l'approche temporelle *temps-continu* d'une part, et entre l'approche spatiale *objet-centrée* et l'approche temporelle *temps-discret* d'autre part :

- l'approche *champ-centrée* consiste à représenter l'espace comme continu, au travers d'une division de l'espace en granules spatiaux ; l'approche *temps-continu* consiste à représenter le temps au travers d'une division régulière en granules temporels.
- l'approche *objet-centrée* consiste à représenter l'espace comme discret, et peuplé d'objets limités dans l'espace ; l'approche *temps-discret* consiste à représenter le temps comme discret, et peuplé d'objets limités dans le temps, que nous nommons *événements*.

On peut alors étendre ce parallèle à la conceptualisation de l'espace-temps, en le considérant :

- soit comme étant continu et divisé en granules spatio-temporels.
- soit discret, et peuplé d'objets limités dans le temps et l'espace.

Cette réflexion de Galton aboutit à deux manières de représenter le changement. La première, appelée $3D+1$, s'inscrit dans une vision *endurantiste*, la seconde, appelée $4D$, dans une vision *perdurantiste*.

1.2.2.1 Le modèle $3D+1$

L'approche $3D+1$ consiste en une modélisation continue de l'espace-temps, au travers de la définition d'entités *SNAP* occupant l'espace, dont on suit l'évolution temporelle dans une représentation continue du temps.

Ces entités *SNAP*, sont caractérisées par leur position dans l'espace (x,y,z) , qui forment le terme $3D$. La valeur de ces trois dimensions spatiales décrivent l'état de l'entité à un instant t . Ces entités spatiales peuvent être rattachées à d'autres entités *SNAP* représentant leurs propriétés thématiques. La dimension temporelle est représentée par le terme $+1$.

Les entités *SNAP* peuvent se référer à des objets spatiaux, à des granules spatiaux ou à des regroupements de granules spatiaux. Chacune de ces entités possède, pour chaque granule temporel, un état différent.

Le changement dans l'espace est représenté au travers de l'évolution temporelle des entités *SNAP* pérennes, d'où le terme $3D+1$, la dimension temporelle étant considérée comme orthogonale à la dimension spatiale.

1.2.2.2 Le modèle $4D$

L'approche $4D$ consiste en une modélisation discrète de l'espace-temps, peuplé d'entités *SPAN* limitées dans le temps et l'espace.

Ces entités forment des *hyperobjets*, discrets, occupant l'espace-temps. Les dimensions temporelles et spatiales sont mises sur un pied d'égalité, d'où le terme $4D$. Chaque hyperobjet forme un intervalle spatio-temporel dans lequel il n'y a pas de variation de l'état de l'espace au cours du temps, cette variation étant inscrite dans la structure de l'hyperobjet.

Le modèle n'est ainsi pas basé sur la description du changement au travers de l'évolution temporelle d'entités spatiales, mais sur la description du changement par une entité spatio-temporelle distincte. Le changement dans l'espace est représenté directement au travers des entités *SPAN*, que nous nommons *événements spatio-temporels*.

1.2.3 *Triad Model* et *Pyramid Model*

Ces approches $3D+1$ et $4D$ peuvent être illustrées au travers de deux modèles souvent cités dans la littérature, le *Triad Model* de Peuquet [Peuquet, 1994] et le *Pyramid Model* de Mennis [Mennis et al., 2000].

1.2.3.1 Le *Triad Model* de Peuquet

Le modèle de la *triade*, ou *Triad Model* [Peuquet, 1994], définit l'information spatio-temporelle comme étant formée de trois dimensions : *spatiale*, *temporelle* et *thématique* (Figure 1.10).

La *triade* décrit les différentes façons d'aborder l'information spatio-temporelle, depuis trois points de vue exprimés sous la forme de trois questions :

- *Where ? (Où ?)*
- *When ? (Quand ?)*
- *What ? (Quoi ?)*

Le *Triad Model* peut s'inscrire dans une représentation continue, enduringe, de l'espace-temps. Ce modèle permet d'illustrer comment est caractérisé le changement dans l'espace dans une approche $3D+1$, en décrivant les différentes entrées permettant d'observer l'évolution d'entités endurantes :

Where = When + What Considérant un point dans l'espace, on peut observer l'évolution temporelle de la valeur de différentes grandeurs thématiques observables en ce point.

When = Where + What Considérant un moment dans le temps, on peut observer la distribution spatiale de différentes grandeurs thématiques observables à ce moment.

What = Where + When Considérant une valeur prise par une grandeur thématique, on peut observer à quels points dans l'espace on observe cette valeur, et à quels moments dans le temps.

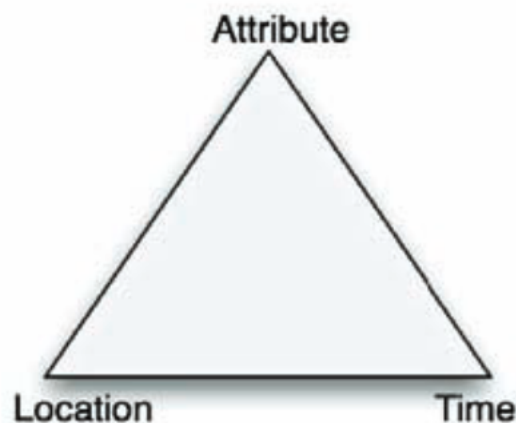


FIGURE 1.10 – Le Triad Model de Peuquet [Peuquet, 1994, Li, 2010]

1.2.3.2 Le *Pyramid Model* de Mennis

Le modèle de la *triade* fut par la suite étendu vers celui de la *pyramide*, ou *Pyramid Model* [Mennis et al., 2000], dans lequel a été ajoutée aux trois composantes de l'information spatio-temporelle, une composante de niveau supérieur : l'*objet (spatio-temporel)* (Figure 1.11).

Le *Pyramid Model* peut s'inscrire dans une représentation discrète, occurrente, de l'espace-temps. Ce modèle permet de décrire les trois composantes dimensionnelles (*spatiale*, *temporelle* et *thématique*) des entités *SPAN* peuplant l'espace-temps dans le cadre d'une approche $4D$. La (méta-)composante *objet* du *Pyramid Model* introduit ainsi un quatrième point de vue, *Whether?* (*Si?*), interrogeant sur l'existence d'un objet spatio-temporel :

Where = When + What + Whether Considérant un point dans l'espace, on peut observer quels objets spatio-temporels occupent cette position spatiale, et quelles sont ses composantes temporelles et thématiques.

When = Where + What + Whether Considérant un moment dans le temps, on peut observer quels objets spatio-temporels existent à ce moment temporel, et quelles sont ses composantes spatiales et thématiques.

What = Where + When + Whether Considérant une valeur prise par une grandeur thématique, on peut observer quels objets spatio-temporels possédant cette valeur existent, et quelles sont ses composantes spatiales et temporelles.

Whether = Where + When + What Considérant un objet en particulier, on peut observer quelles sont ses composantes spatiales, temporelles et thématiques.

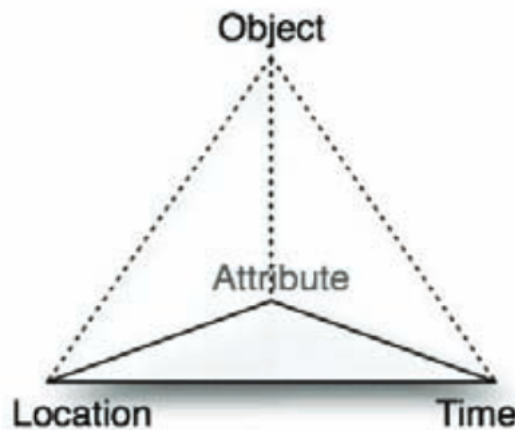


FIGURE 1.11 – Le Pyramid Model de Mennis [Mennis et al., 2000, Li, 2010]

1.3 Le phénomène spatio-temporel modélisé selon un point de vue endurant

La représentation des phénomènes spatio-temporels selon une approche $3D+1$ peut s'effectuer de deux manières, en modélisant implicitement ou explicitement les modifications apportées aux entités *SNAP*, entraînant leur évolution dans le temps.

1.3.1 La modélisation implicite du changement

Dans une modélisation implicite du changement, la modification apportée aux entités pérennes n'est pas explicitement représentée. Le principe consiste à représenter différents états temporels des entités *SNAP*, et d'en déduire le changement par la comparaison des différents états. Nous présentons ici quelques exemples de modélisation basés sur cette approche.

1.3.1.1 Le modèle *snapshot*

Cette approche [Haddad, 2009] s'inspire des bases de données temporelles [Tansel et al., 1993]. Les premières solutions pour introduire l'information temporelle dans les bases de données furent

d'étiqueter temporellement les tables [Gadia and Vaishnav, 1985], les tuples [Snodgrass and Ahn, 1985], ou les attributs d'objets [Gadia and Yeung, 1988].

L'extension de cet étiquetage temporel au spatio-temporel consiste à étiqueter les couches spatiales (Modèles en snapshots [Armstrong, 1988]), les attributs (Composites spatio-temporels [Langran and Chrisman, 1988]) et les objets spatiaux (Objets spatio-temporels [Worboys, 1992]). Cet étiquetage permet des représentations de différents états de l'espace, représenté selon une approche *champ-centrée* ou *objet-centrée*, pour différents instants temporels, qui sont appelés *snapshots temporels* [Haddad, 2009].

L'évolution dynamique du territoire est alors représentée par une succession de *snapshots* de la même région spatiale. Il n'existe pas une relation temporelle explicite entre les différents *snapshots*. De ce fait, le changement n'est pas explicitement représenté, mais il doit être interprété au travers de la comparaison de différents *snapshots*. Ce modèle demeure cependant redondant et simpliste, l'ensemble des données relatives aux entités *SNAP* étant représentée dans chaque *snapshots*, même dans le cas d'une absence de changement.

1.3.1.2 Les modèles des *composites spatio-temporels* et des *objets spatio-temporels*

Ce modèle, également appelé *Space-Time Composite Model* [Langran and Chrisman, 1988], consiste à projeter différents états temporels de l'espace sur une seule couche spatiale, de manière à définir de nouvelles entités *SNAP* appelées *composites*. Ces entités *composites* forment alors un découpage de l'espace, où chaque région regroupe un ensemble de points présentant le même historique (Figure 1.12). Chaque *composite* possède alors un historique différent de ses voisins.

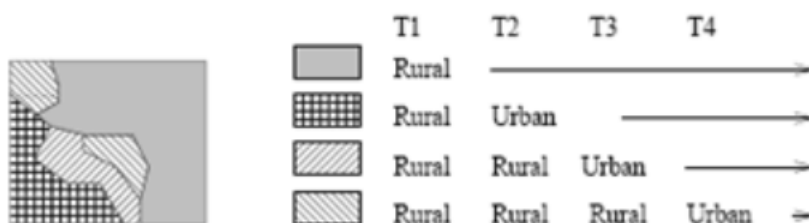


FIGURE 1.12 – *Space-Time Composite Model* [Langran and Chrisman, 1988, Renolen, 1995]

Le principe est donc de ne stocker que le *vecteur de modification* de chaque entité *SNAP* subissant une trajectoire de vie distincte. La position spatiale et l'historique de chaque *composite* permettent de retranscrire l'état de l'espace pour n'importe quel instant temporel, en évitant le stockage redondant de données de l'approche *snapshot*.

La modélisation du changement impactant la zone d'étude s'effectue au travers de la modélisation des différents états temporels de chaque *composite* (Figure 1.13).

Chaque composite est à considérer comme un objet *SNAP*, atemporel, mais pouvant passer d'un état d'existence à un état de non-existence. Ce procédé reste ancré dans une modélisation implicite du changement, car les changements en eux-mêmes (comment s'effectue la modification portant sur les composites) ne sont pas représentés. Seuls sont représentés des états *anté-* et *post-*changement des entités spatiales.

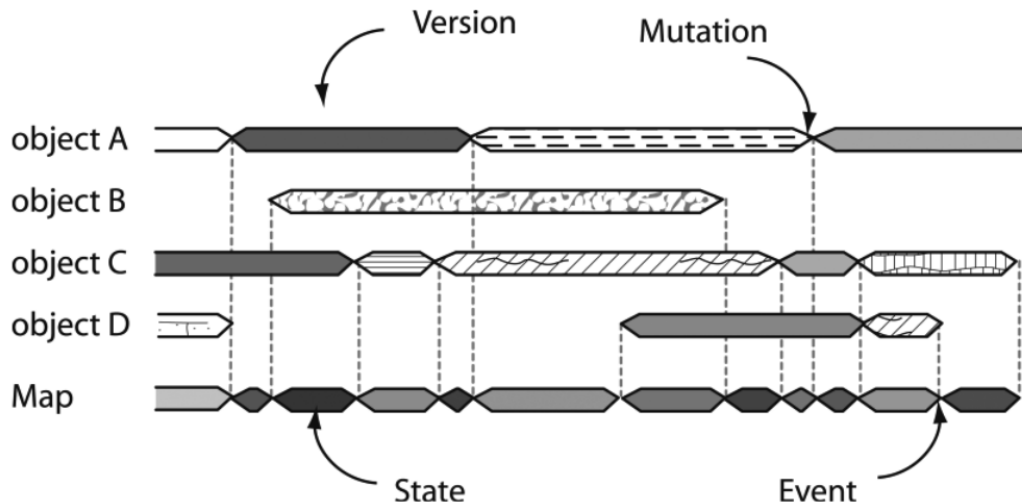


FIGURE 1.13 – Modélisation du changement dans le *Space-Time Composite Model* [Langran and Chrisman, 1988, Mathian and Sanders, 2014]

Le modèle d’objets spatio-temporels de Worboys [Worboys, 1992] (Figure 1.14) peut être considéré comme une version en trois dimensions du modèle des *composites spatio-temporels*, où seul le vecteur de modification d’entités à la trajectoire de vie distincte est stocké.

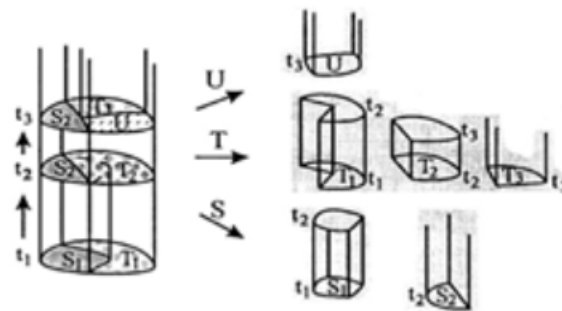


FIGURE 1.14 – Objets spatio-temporels [Worboys, 1992, Yuan, 1999]

Dans la Figure 1.14, l’espace est divisé en trois zones, dont l’étendue temporelle évolue au cours du temps :

- une zone agricole S .
- une zone urbaine T .
- une zone industrielle U .

Ces trois zones peuvent être considérées comme des entités *SNAP*. Les différents états temporels de chacune des entités, entre les dates t_1 , t_2 et t_3 sont représentés par une version de l’entité *SNAP* correspondante.

Les différents états temporels de chacune des zones sont cependant désignés sous le terme d’*atomes spatio-temporels* [Haddad, 2009], occupant une zone de l’espace et du temps. Ce modèle peut alors s’apparenter à une représentation occurrente du changement, où chaque *atome* est une entité *SPAN* représentant une tranche de vie du territoire.

Cependant, ces *atomes* sont la représentation d’un état temporel d’un objet spatial de référence. L’entité spatio-temporelle à l’origine du changement entre deux états n’est jamais ré-

présentée dans ce modèle. Le changement est toujours représenté au travers d'une évolution des objets de l'espace. Aussi, si ce modèle peut être considéré comme étant inscrit dans une approche occurrente de la modélisation du changement, nous préférons le considérer comme inscrit dans une approche endurente. Nous considérons alors ces *atomes* comme les états temporels d'un objet spatial atemporel.

Cet exemple montre cependant que la frontière entre modèles occurrents et perdurants peut être difficile à définir.

1.3.2 La modélisation explicite du changement

D'autres modèles, toujours inscrits dans la description de l'évolution temporelle d'entités *SNAP*, et donc au travers de l'approche $3D+1$, s'appliquent à représenter la nature du changement affectant les entités pérennes. Les modèles de cette famille partagent certaines caractéristiques [Haddad, 2009] :

- Ces modèles sont basés sur une représentation de l'espace comme occupé par des objets géographiques [Kuhn, 2001, Haddad, 2009]. La représentation peut porter sur l'évolution temporelle d'un objet *rivière*.
- Au contraire de la modélisation implicite du changement, ces modèles ne mettent pas l'emphase sur différents états des objets à différents instants temporels, mais sur les changements intervenant entre ces différents états. Une crue d'un objet *rivière*, entraînant un élargissement de l'étendue du cours d'eau, n'est pas représentée par deux états temporels de l'objet *rivière*, mais par une la datation d'un élargissement de l'objet *rivière*.
- Les changements affectant les entités *SNAP* sont explicitement représentés, mais seulement en tant que modifications portant sur ces entités. Les entités spatio-temporelles responsables de ces changements ne sont pas représentées [Haddad, 2009]. La modification affectant l'objet *rivière* est représentée, mais la crue à l'origine de cette modification ne l'est pas.

Les modèles s'inscrivent toujours dans une approche $3D+1$ et non dans une approche $4D$. L'emphase est mise sur la modélisation des changements affectant des objets persistants [Hornsby and Cole, 2007], ce qui inscrit ces modèles dans une vision endurente du changement. Ces modèles restent basés sur le principe que ces changements sont le résultat d'entités occurrentes, mais ceux-ci ne sont pas explicitement modélisés.

Certains modèles tentent d'exprimer le changement en composants élémentaires, agissant sur les composantes de l'information spatio-temporelle telles qu'elles sont décrites dans le modèle de la *triade* de Peuquet [Peuquet, 1994, Mathian and Sanders, 2014, Haddad, 2009]. Ces modèles étant basés sur l'étude de l'évolution d'entités atemporelles, le changement peut porter sur leur composante spatiale (localisation / forme), et sur leur composante thématique (état d'un attribut qualitatif / état d'un attribut quantitatif) [Mathian and Sanders, 2014].

À ces deux composantes, certains modèles ajoutent une catégorie de changement pouvant porter sur l'identité de l'entité spatiale dont on suit l'évolution. Cette catégorie permet de répertorier des modifications portant sur l'apparition ou la disparition de ces objets spatiaux, ou sur des relations de succession entre différents objets [Mathian and Sanders, 2014]. On peut faire le parallèle entre cette catégorie de modifications et les modèles d'objets *étiquetés* de Hunter et Williamson [Hunter and Williamson, 1990] et du *Space-Time Cube* de Langran [Langran, 1992], modèles basés sur une représentation implicite du changement, et qui consistaient à étiqueter les dates de création et de suppression des objets géographiques [Haddad, 2009].

Il convient de noter que ces objets spatiaux ne sont pas considérés, au sein de ces modèles, comme des objets spatio-temporels limités dans le temps. Ces modifications portant sur l'apparition et la disparition d'objets spatiaux sont à interpréter comme des transitions d'entités *SNAP* d'un état d'*existence* à un état de *non-existence*.

Dans le modèle proposé par Cheylan [Cheylan et al., 1994, Mathian and Sanders, 2014], présenté dans la Figure 1.15, le suivi de l'évolution de l'objet est ainsi relatif à ces trois aspects : au travers du mouvement de l'objet pour sa dimension spatiale, au travers de la vie de l'objet pour sa dimension thématique, au travers de la généalogie de l'objet pour son identité.

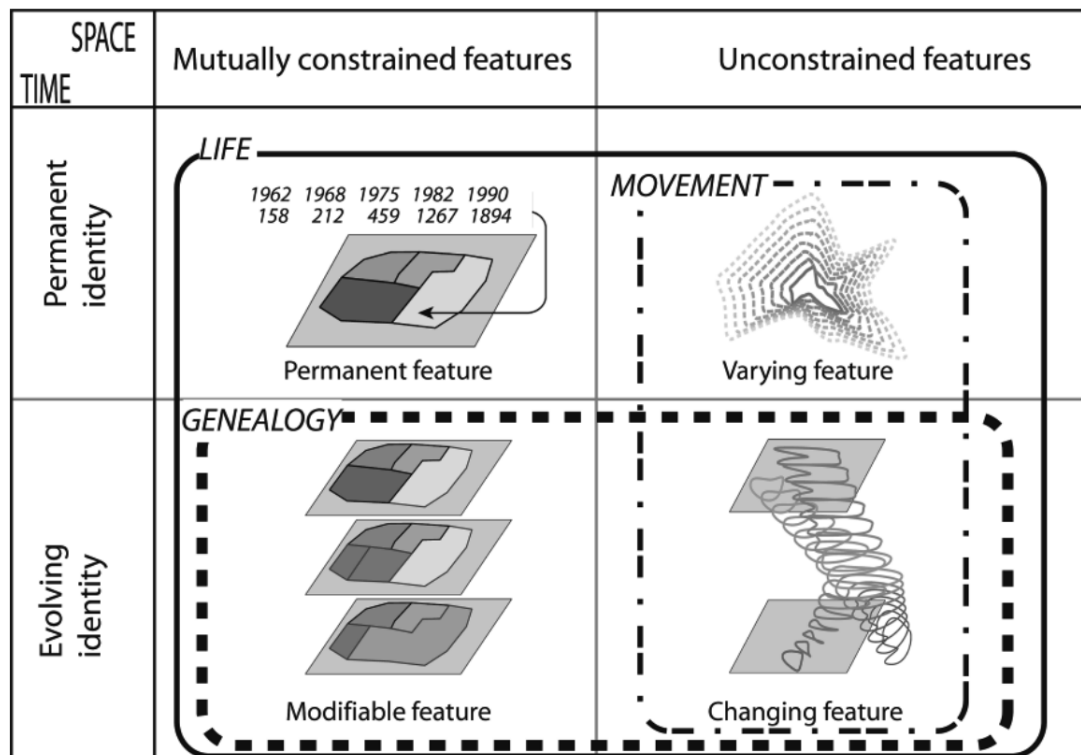


FIGURE 1.15 – Types de changements et catégorie des objets [Cheylan and Lardon, 1993]

D'autres modèles s'attachent à représenter comment s'effectuent les modifications sur les différentes composantes des entités *SNAP*, et les relations entre différentes entités *SNAP* impliquées dans la même modification [Mathian and Sanders, 2014]. C'est le cas du modèle des *processus spatio-temporels* de Claramunt et Thériault [Claramunt and Thériault, 1995, Claramunt et al., 1997, Claramunt et al., 1998], qui propose une typologie des modifications spatio-temporelles des objets spatiaux (Figure 1.16).

Cette typologie est structurée en trois groupes [Mathian and Sanders, 2014] :

- les changements portant sur l'évolution d'un objet, que ce soit sur sa composante spatiale ou thématique ;
- les changements portant sur la généalogie des objets et leur diffusion spatiale, faisant intervenir des relations entre différents objets spatiaux (succession, duplication, etc.) ;
- les changements portant sur la structure spatiale d'un ensemble d'objets (modification de frontières entre différents objets contigus, etc.).

D'autres travaux s'intéressant à la modélisation des changements d'objets spatiaux ont abouti à des typologies de changements similaires [Haddad, 2009]. Parmi eux se trouvent les travaux

sur le modèle de *graphe historique* de Renolen [Renolen, 1996] (création d'un objet, cessation d'un objet, etc.), ou le modèle de *changement de régions spatiales* de [Kauppinen et al., 2008] (scission de régions, fusion de région, etc.).

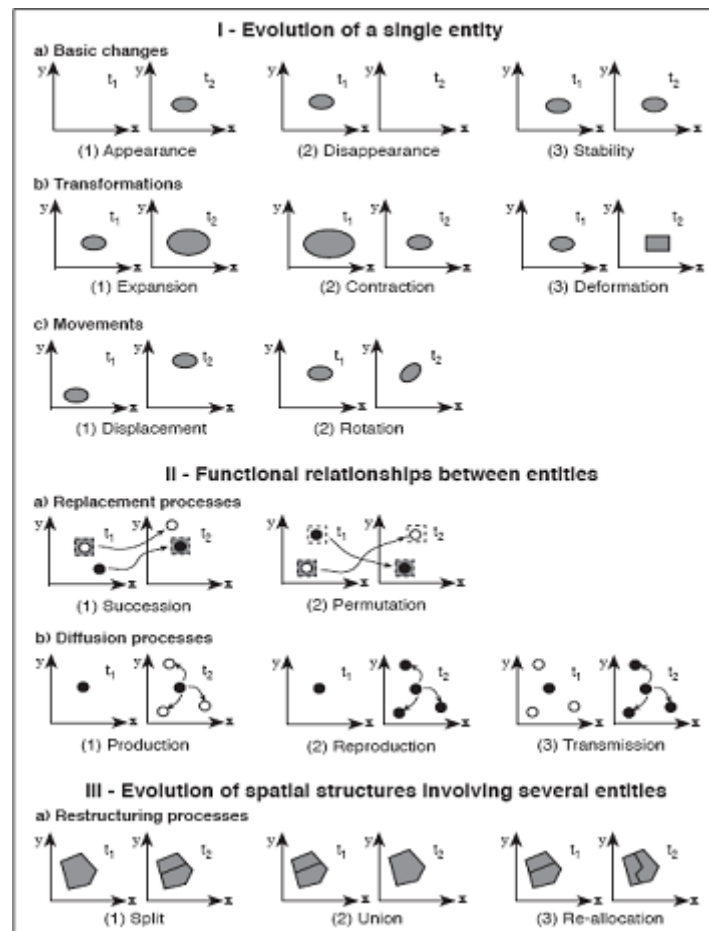


FIGURE 1.16 – Les processus spatio-temporels de Claramunt et Thériault [Claramunt and Thériault, 1995]

1.3.3 Le choix de l'approche $4D$

Certains auteurs affirment que la représentation et le raisonnement sur les dynamiques géographiques nécessitent plus que l'introduction du temps dans une approche enduringiste [Mathian and Sanders, 2014] et qu'il est nécessaire d'introduire des éléments occurrents dans les représentations des phénomènes spatio-temporels [Yuan, 2001, Renolen, 2000, Galton and Worboys, 2005, Worboys, 2005, Mathian and Sanders, 2014], et de représenter le changement au sein de modèles $4D$.

1.3.3.1 Limites de l'approche $3D+1$

Dans le cadre de notre objectif d'analyse des dynamiques des phénomènes spatio-temporels, la représentation des phénomènes au travers de l'évolution temporelle d'entités spatiales nous pose principalement deux problèmes.

D'une part, l'emphasis mise, dans les modèles $3D+1$, sur la description des entités spatiales pérennes et des modifications les impactant au cours du temps, occulte la modélisation des *événements* et *processus* à l'origine de ces modifications. Ces entités sont alors décrites de manière

ad hoc pour des besoins spécifiques selon le cas d'étude, et non en se basant sur une approche théorique [Peuquet, 2002, Yuan, 2007, Haddad, 2009]. En découlent des définitions diverses des notions de *processus* et d'*événements* [Haddad, 2009].

Cette absence d'approches théoriques rend difficile une modélisation générique des *événements* et *processus* à l'origine du changement, qui puisse être utilisée de manière indifférenciée sur différentes problématiques. Cette approche générique nous est pourtant nécessaire pour proposer une méthode d'analyse des dynamiques des phénomènes spatio-temporels, pris au sens large.

D'autre part, la représentation étant centrée sur l'évolution temporelle individuelle d'entités spatiales, le changement est décrit au travers de l'ensemble de ces évolutions individuelles, plutôt qu'au travers de la description d'un phénomène par une entité globale interagissant l'ensemble des entités spatiales. Un phénomène se déroulant dans une région spatiale n'est observable qu'au travers des évolutions de chaque entité *SNAP* peuplant la région considérée.

La modélisation du phénomène par une entité distincte des entités spatiales qu'il impacte offrirait une vision d'ensemble du phénomène, rendue difficile dans une approche $3D+1$ qui tend au cloisonnement de la représentation par entité spatiale. La représentation utilisée serait centrée sur les évolutions spatio-temporelles d'un phénomène dans son ensemble, et non sur l'évolution temporelle de chaque entité spatiale traitée de manière distincte. Cette vision d'ensemble du phénomène serait plus adaptée pour analyser les dynamiques spatio-temporelles de celui-ci.

Pour ces deux raisons, nous nous positionnons dans une approche occurrente de la représentation de l'espace-temps, nous permettant de représenter le changement au travers d'entités spatio-temporelles distinctes. Ce choix présente également d'autres atouts.

1.3.3.2 Avantages du format de données événementiel

Un autre limite des approches $3D+1$ provient de la nature des données à disposition, souvent sous la forme de séries temporelles relatives à des objets spatiaux fixes, comme des stations de mesure où l'évolution d'une variable a été observée, ou des entités territoriales pour lesquelles des données brutes événementielles ont été agrégées.

Baser notre représentation sur des entités spatio-temporelles, étendues dans le temps et l'espace, permettrait de nous affranchir du caractère "fixe" des entités *SNAP* de référence, et permettrait analyse des phénomènes où les dimensions temporelle et spatiale sont simultanément prises en compte

L'approche $4D$ nous permet également d'étudier, directement au travers de données brutes, certains phénomènes ne pouvant être enregistrés que sous un format événementiel (accidents, feux de forêt, homicides, etc.). Les séries temporelles représentant ce type de phénomène sont obtenues par agrégation de données événementielles brutes, selon une zone et une fréquence d'échantillonnage.

Enfin l'approche $4D$ présente une certaine polyvalence, les données sous format événementiel pouvant :

- soit provenir directement sous la forme de données brutes,
- soit être recrées à partir de données temporelles, en définissant comme *événement* remarquable le moment où une variable relative à une position spatiale de référence atteint un certain seuil.

1.4 Le phénomène modélisé selon un point de vue occurrent

Galton oppose à l'approche $3D+1$, la représentation $4D$ [Galton, 2004]. L'emphase n'est plus mise sur la représentation de l'évolution temporelle d'entités durantes, mais sur la représentation d'entités occurrentes à l'origine du changement, appelées *événements* et *processus*. On peut distinguer deux approches de cette représentation, reposant sur deux visions du monde géographique [Haddad, 2009].

La première approche, introduite par Grenon et Smith [Grenon and Smith, 2004], représente de manière complémentaire des entités *SPAN* agissant sur la trajectoire de vie d'entités *SNAP*. La seconde approche, introduite par Reitsma [Reitsma, 2005], se concentre sur la représentation d'entités occurrentes, sans représenter les entités durantes impactées.

1.4.1 *Événements et processus*

La construction d'approches $4D$ a engagé des débats sur la définition des termes *événements* et *processus*, et en quoi ils diffèrent l'un de l'autre [Haddad, 2009].

Dans la littérature [Haddad, 2009], le *processus* est en général défini comme l'entité responsable du changement [Forbus, 1984, Claramunt et al., 1997, Thériault et al., 1999, Galton, 2001], tandis que l'*événement* est souvent lié à l'apparition de quelque chose de remarquable pour le domaine étudié [Peuquet, 1994, Galton, 2001, Worboys, 2001, Yuan, 2001]. Pierre Livet [Livet et al., 2010, Mathian and Sanders, 2014] considère également les *événements* comme *caractérisant les apparitions, disparitions, et changements brusques* des objets géographiques, et les *processus* comme *référant à ce qui fait changer* les objets, leurs propriétés, et les relations entre objets au cours du temps.

Le critère temporel a également été avancé pour effectuer une distinction entre les deux entités. Pour certains, les *processus* sont des intervalles temporels ouverts (ils ont un début, mais n'ont pas de fin), tandis que les *événements* sont des intervalles fermés [Yuan, 2001, Galton, 2006, Bennett and Fellbaum, 2006]. Pour d'autres, les *processus* sont étendus dans le temps, alors que les *événements* sont des transitions instantanées [Grenon and Smith, 2004, Frank, 2007].

D'autres critères ont été proposés, tels que *l'homogénéité* [Mourelatos, 1978, Galton, 2006, Bennett and Fellbaum, 2006, Galton, 2008] : les *processus* seraient homogènes, dans le sens que les subdivisions temporelles d'un processus sont de la même nature que le processus lui-même : le processus "courir" consiste à courir dans toutes ses subdivisions temporelles. À l'inverse, les événements ne seraient pas homogènes. L'*événement* "excursion" est composé d'activités de natures différentes (courir, nager, marche, etc.), représentées par différentes subdivisions temporelles.

Comme prolongement de ces distinctions, des relations entre ces deux entités ont été proposées [Haddad, 2009]. Certains considèrent l'*événement* comme composé de *processus* [Claramunt et al., 1997, Yuan, 2001, McIntosh and Yuan, 2005, Galton, 2001, Galton, 2004, Galton, 2008], d'autres considèrent que les *processus* sont composés d'*événements* [Worboys, 2001]. D'autres relations, concernant les rapports entre plusieurs *événements*, ont également été proposées [Grenon and Smith, 2004, Worboys and Hornsby, 2004, Galton and Worboys, 2005]. Un *événement* peut, par exemple, causer (initier ou faciliter), bloquer ou terminer l'existence d'un autre *événement*.

La définition de *processus* et d'*événement* n'est pas tranchée et reste un sujet de débat. Cependant tout comme la représentation du changement, la définition de ces termes dépend du problème étudié [Galton, 2006, Bennett and Fellbaum, 2006, Galton, 2008, Haddad, 2009].

1.4.2 L'approche de Grenon et Smith

L'approche de Grenon et Smith de la représentation $4D$ repose sur le principe que les approches endurantes et occurrentes de représentation des dynamiques spatio-temporelles sont complémentaires [Mathian and Sanders, 2014]. Elle peut être associée à la modélisation explicite du changement selon une approche $3D+1$ [Haddad, 2009].

Dans ce modèle, le monde est composé d'entités *SNAP*, pérennes, et d'entités *SPAN* qui existent et disparaissent. Les entités *SNAP* représentées sont des objets géographiques et les propriétés thématiques de ces derniers. Les entités *SPAN* sont des *événements* et des *processus*. Comme le soulignent Grenon et Smith [Grenon and Smith, 2004], des objets géographiques peuvent participer à des *événements* et des *processus*, des *événements* et des *processus* peuvent impliquer des objets géographiques.

L'implication fait référence à l'effet produit par une entité *SPAN* sur les entités *SNAP*. La participation définit le rôle que joue une entité *SNAP* dans une entité *SPAN* [Haddad, 2009]. Ce critère permet de distinguer un type particulier d'entité *SPAN* : l'action. L'action est un *événement* ou un *processus* initié ou terminé par un agent (un objet géographique) [Worboys and Hornsby, 2004, Haddad, 2009]. Les *événements* ou *processus*, non initiés ou terminés par un agent, ne font pas participer d'objets géographiques, mais en impliquent. Un individu (un objet géographique) peut participer au processus *courir*, mais peut être impliqué par un événement *maladie*.

Les objets géographiques sont donc représentés comme subissant les modifications apportées par les *événements* et les *processus*, les deux types d'entités étant explicitement modélisés. Ce modèle permet d'associer le suivi de l'évolution des objets géographiques, et une vision d'ensemble des *événements* ou *processus* à l'origine de ces évolutions.

1.4.3 L'approche de Reitsma

La deuxième approche, introduite par Reitsma [Reitsma, 2005], se base sur la philosophie des *processus* de Bohm [Bohm, 1980] qui stipule que le monde est composé uniquement de changements appelés *flux*.

Dans cette approche, seules les entités occurrentes sont représentées [Reitsma, 2005, Reitsma and Albrecht, 2005], les objets géographiques pérennes constituant le monde ne sont pas modélisés. Les entités occurrentes possèdent leurs propres attributs, une dimension temporelle (moment d'apparition et/ou étendue temporelle), et une étendue spatiale (localisation et/ou étendue spatiale).

Ce modèle présente la limite d'être adapté uniquement à la modélisation des processus physiques (comme dans le cas de phénomènes météorologiques par exemple) [Reitsma and Dubayah, 2007, Haddad, 2009]. La vision de Grenon et Smith [Grenon and Smith, 2004] reste la plus répandue dans la littérature.

La vision de Reitsma peut cependant être satisfaisante dans le cas où l'on dispose d'informations sur des *événements*, mais non sur les objets impactés ou sur les objets agissant lors de ces *événements*.

1.4.4 Le choix de l'approche de Reitsma

Nous choisissons de nous positionner sur une approche occurrente de la modélisation du changement, et de baser notre proposition sur une modélisation des données spatio-temporelles sous

un format événementiel. La modélisation des *événements* suit la décomposition de l'objet spatio-temporel en trois composantes, proposée par le *Pyramid Model* de Mennis [Mennis et al., 2000] (Repère *b* de la Figure 1.17), lui-même issu du *Triad Model* de Peuquet [Peuquet, 1994] (Repère *a* de la Figure 1.17).

Les composantes de l'objet spatio-temporel peuvent également être décomposées (Repère *c* de la Figure 1.17) :

- la composante spatiale peut être décomposée en coordonnées x , y et z , représentant les trois axes de l'espace ;
- la composante temporelle peut être décomposée en temps d'*observation* (quand l'*événement* est observé), temps de l'*événement* (quand l'*événement* a lieu), et temps de *transaction* (quand l'*événement* est enregistré) ;
- la composante thématique peut être décomposée selon une liste d'attributs.

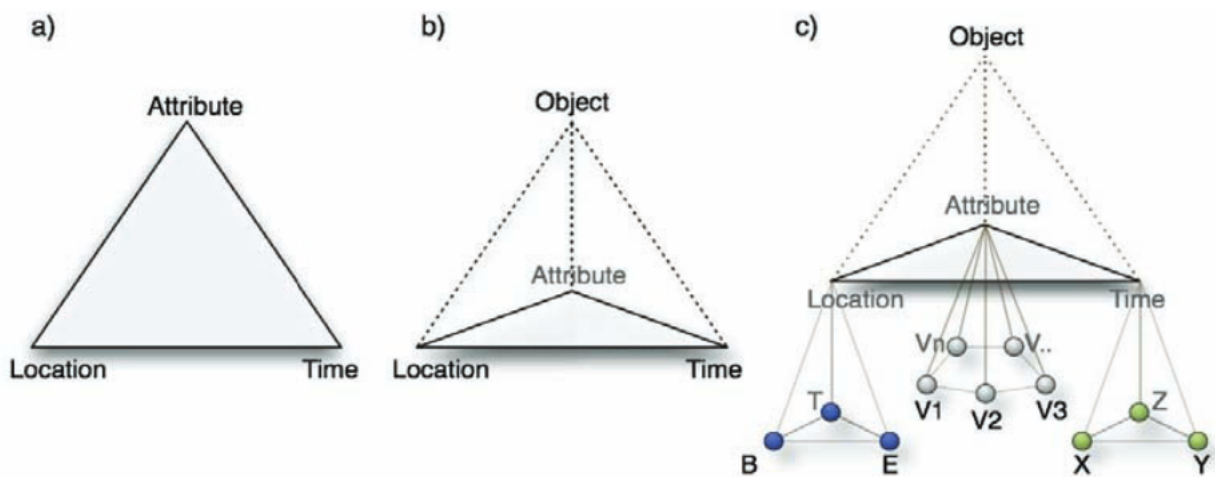


FIGURE 1.17 – Du Triad Model au Pyramid Model [Peuquet, 1994, Mennis et al., 2000, Li, 2010]

Pour illustrer notre proposition, et afin de déterminer une modélisation de la donnée spatio-temporelle la plus générique possible, nous nous sommes intéressés à différents cas d'études, et au format des données à disposition.

Nous nous sommes, entre autres, intéressés à des phénomènes d'accidentologie routière, de criminologie, et à des phénomènes liés aux risques naturels, notamment des phénomènes liés au volcanisme (coulées de lave, séismes), et des phénomènes d'inondation.

Pour chacune de ces problématiques, nous avons accès à des données sous la forme de séries événementielles :

- Concernant les données relatives aux événements d'accidents de la route, nous disposons pour chaque accident de la date et du lieu où avait eu lieu l'accident, ainsi que diverses informations sur les circonstances de l'accident (voirie, visibilité, conditions météorologiques, etc.), ainsi que sur les personnes et les véhicules impliqués.
- Concernant les données criminologiques, nous disposons d'une série de données événementielles comprenant, pour chaque entrée, une date et un lieu d'apparition, ainsi que des informations sur les circonstances et la nature du délit ou du crime.
- Pour le cas des coulées des laves, nous disposons pour chaque coulée des dates de début et de fin de l'éruption, de l'empreinte géométrique de la coulée en fin d'éruption, et d'informations sur la coulée de lave (débit, nature de la coulée, etc.).

- Pour le cas des séismes, nous disposons de la date du séisme, de la position du foyer, et d'informations sur le séisme (magnitude, etc.).
- Enfin pour le cas des inondations, nous disposons de la date et de la durée de l'inondation, du barycentre de la (ou des) commune(s) touchée(s), et d'informations thématiques sur l'événement.

Ces données événementielles concernent des problématiques métiers différentes, mais présentent un format proche, puisque nous disposons pour chaque entrée d'information :

- d'une date d'apparition, éventuellement d'une durée ;
- d'un lieu d'apparition, éventuellement d'une empreinte géographique ;
- d'une liste de données attributaires.

Les données à disposition ne font pas référence à des objets spatiaux pérennes, mais à des entités limitées dans le temps et l'espace, sauf pour le cas des inondations, où la composante spatiale portée par la donnée fait référence à des communes touchées.

Ce format de données s'inscrit dans une conceptualisation occurrentiste du temps. Le changement est représenté de façon explicite par une série d'entités spatio-temporelles, et non de façon implicite par la représentation de deux états de l'espace anté-changement et post-changement ou par une série de modifications affectant des objets spatiaux. Cependant, aucune, ou peu d'information est donnée sur la façon dont les entités représentées dans ces séries événementielles affectent les entités pérennes peuplant l'espace géographique.

Si l'on replace ces phénomènes dans le monde réel, il est évident qu'ils impactent des objets persistants : une inondation impacte un territoire, peut impacter des zones urbaines ou des axes de communication. Cependant, les données à disposition ne nous permettent pas de représenter ces objets durables et leur évolution dans le temps lors de l'apparition d'une entité modificatrice. Les données à disposition s'inscrivent ainsi dans l'approche de la modélisation de la donnée spatio-temporelle de Reitsma [Reitsma, 2005], où seuls sont représentées les entités *SPAN*, *événements* et *processus*, et non les entités *SNAP* impliquées ou participant aux *événements* et *processus*.

L'étude des dynamiques de phénomènes, représentés au travers de séries d'entités spatio-temporelles, ne nécessitant pas de représenter le suivi de l'évolution d'entités spatiales pérennes, nous nous sommes positionnés sur une modélisation occurrente de la donnée spatio-temporelle selon le modèle de Reitsma.

1.5 *Événements et Phénomènes*

1.5.1 *L'Événement, entité spatio-temporelle représentant le changement*

La définition des termes *événements* et *processus* n'étant pas tranchée dans la littérature, il nous est nécessaire de préciser notre propre terminologie.

Notre approche est basée sur la représentation du changement au travers de séries d'entités spatio-temporelles, limitées dans le temps et l'espace, et possédant une composante thématique, conformément à la formalisation de l'objet spatio-temporel représentée dans le *Pyramid Model* [Mennis et al., 2000]. Comme évoqué précédemment, cette entité spatio-temporelle peut être simple, ou composite et formée de différentes sous-entités spatio-temporelles, correspondant à différentes emprises spatiales de l'entité au cours du temps, ou à différentes valeurs portées par

la composante thématique de l'entité au cours du temps.

Nous utilisons le terme d'*événement* (*spatio-temporel*) pour désigner ces entités spatio-temporelles occurrentes. Les sous-entités des *événements* composites sont désignées par le terme de *sous-parties spatio-temporelles* (des *événements*).

Le terme *événement* est, dans notre proposition, décorrélé du critère d'étendue temporelle : un *événement* peut être ponctuel dans le temps, ou posséder une durée.

Il n'est également pas défini selon un critère spatial. Un *événement* peut être ponctuel dans l'espace ou posséder une emprise spatiale. Un *événement* peut également posséder une composante spatiale linéaire. L'*événement* "trajet de l'individu A" possède comme composante spatiale la trajectoire linéaire de l'individu A durant la durée de l'événement. Différentes *sous-parties spatio-temporelles* de cet *événement* peuvent décrire les différentes phases sur trajet de l'individu A.

L'*événement* est également indépendant d'un critère "déclencheur", comme chez Pierre Livet [Livet et al., 2010, Mathian and Sanders, 2014], où l'*événement* désigne les changements brusques marquant les étapes d'une modification des objets géographiques apportée par des *processus*. Le terme *processus* est absent de notre terminologie, le terme *événement* étant utilisé pour désigner, dans son ensemble, l'objet spatio-temporel responsable du changement.

L'*événement* désigne ainsi un objet spatio-temporel limité dans le temps et l'espace. Les *sous-parties spatio-temporelles* de l'*événement* désignent les différents états spatio-temporels de l'*événement* au cours de son existence. La représentation d'une entité spatio-temporelle en *événement* composé de *sous-parties* dépend des données à disposition, cette représentation n'étant possible que si l'on connaît les différents états d'une même entité spatio-temporelle au cours de son existence.

Considérons comme *événement* la crue d'une rivière, d'une durée de trois jours. Si l'on connaît l'extension de la rivière pour chaque journée, l'*événement* peut être représenté comme composite, et composé de trois *sous-parties*. Si l'on connaît uniquement l'extension maximale de la rivière durant ces trois jours, l'*événement* ne peut être représenté que comme simple.

1.5.2 Les *Phénomènes*, regroupements d'*événements*

Notre représentation de la donnée spatio-temporelle s'inscrit dans une approche générique, où nous pourrions appliquer notre format de modélisation à différents cas d'études. Différents types de changement pourraient être représentés de manière analogue par des *événements spatio-temporels*. Les *événements* se rapportant au même type de changement sont alors regroupés dans un ensemble que nous désignons par le terme *phénomène*.

Ce que nous appelons ici *phénomène* se rapporte plus à la définition d'un ensemble d'*événements spatio-temporels* semblables, à un type d'*événements*, plutôt qu'à un *événement* en particulier. Un séisme sur l'île de la Réunion le 19 août 1994 est un phénomène sismique, mais nous préférons réserver le terme de *phénomène* à l'ensemble des séismes similaires plutôt que pour ce séisme en particulier. Un séisme en particulier sera décrit par le terme *événement*.

La notion de *phénomène* est renvoyée à une vision globale du changement spatio-temporel. L'entité *phénomène* et l'entité *événement* sont de même nature, mais il faut voir le *phénomène* comme un *méta-événement* rassemblant plusieurs *événements* similaires ou liés entre eux : le phénomène du volcanisme est constitué de plusieurs événements liés entre eux, comme plusieurs

événements de coulées de lave combinés à plusieurs événements sismiques. On peut ainsi considérer le *phénomène* comme représentant l'ensemble d'une série événementielle traitant d'une problématique particulière.

Cette similarité entre les *événements* d'un même *phénomène* s'établit avant tout sur un critère thématique. On peut cependant considérer que des *événements* semblables sur le plan thématique, mais apparaissant dans deux régions spatiales ou à deux époques distinctes, constituent deux *phénomènes*.

La définition d'un *phénomène* est ainsi relative, et dépend du point de vue adopté par l'observateur, ainsi que des échelles spatiales et temporelles utilisées pour la représentation du changement.

La frontière entre ce qui relève du *phénomène* et de l'*événement* est également relative :

- À une échelle locale, on peut considérer comme *phénomène* un feu de forêt touchant une région, et comme *événements* les différents départs de feu.
- À une échelle plus globale, on peut considérer comme *phénomène* l'ensemble des feux de forêt, et comme *événements* chaque feu de forêt en particulier.

La distinction que nous effectuons entre *phénomène* et *événement* n'est cependant pas la même que celle que nous effectuons entre *événement* et *sous-partie* d'un *événement* :

- Un *phénomène* regroupe des *événements*. C'est un ensemble de plusieurs entités spatio-temporelles semblables, mais dont chaque membre est un objet distinct.
- Un *événement* est composé de *sous-parties*. C'est un seul et même objet, composite. Les *sous-parties* représentent les différents états spatio-temporels d'un même objet spatio-temporel.

Nous reviendrons sur les relations entre *phénomènes*, *événement* et *sous-parties* dans le chapitre 5.

1.5.3 Une modélisation générique du changement au travers des *Phénomènes*

Le format de modélisation choisi permet ainsi de représenter plusieurs types de changement, au travers de séries d'entités spatio-temporelles, les *phénomènes*, dont les membres, les *événements*, suivent la structure représentée par le *Pyramid Model* de Mennis [Mennis et al., 2000] (Figure 1.18). Chaque *événement* possède son identité propre, et trois composantes dimensionnelles : spatiale, temporelle, et thématique.

On peut émettre l'hypothèse qu'au-delà des données auxquelles nous nous sommes intéressés dans le cadre de notre proposition, des phénomènes spatio-temporels de différentes natures peuvent être représentés au travers de ce format. Nous avons décidé, dans le cadre de notre proposition, de nous limiter aux phénomènes n'impliquant pas d'objets mobiles, dont nous devrions analyser les trajectoires.

Ce caractère générique de la représentation des phénomènes pourrait ensuite être utilisé afin d'offrir une analyse des dynamiques spatio-temporelles applicable à différents phénomènes et à différentes problématiques métiers. L'analyse des phénomènes revient à analyser des séries d'événements dont la composante thématique a plus ou moins été fixée.

L'analyse des dynamiques spatio-temporelles nécessite, non pas d'étudier l'évolution des événements formant un phénomène, mais d'étudier la distribution spatio-temporelle de cet ensemble

afin d'en dégager des structures. La définition de ces structures, et la description des procédés utilisés pour les identifier, font l'objet du chapitre 2.

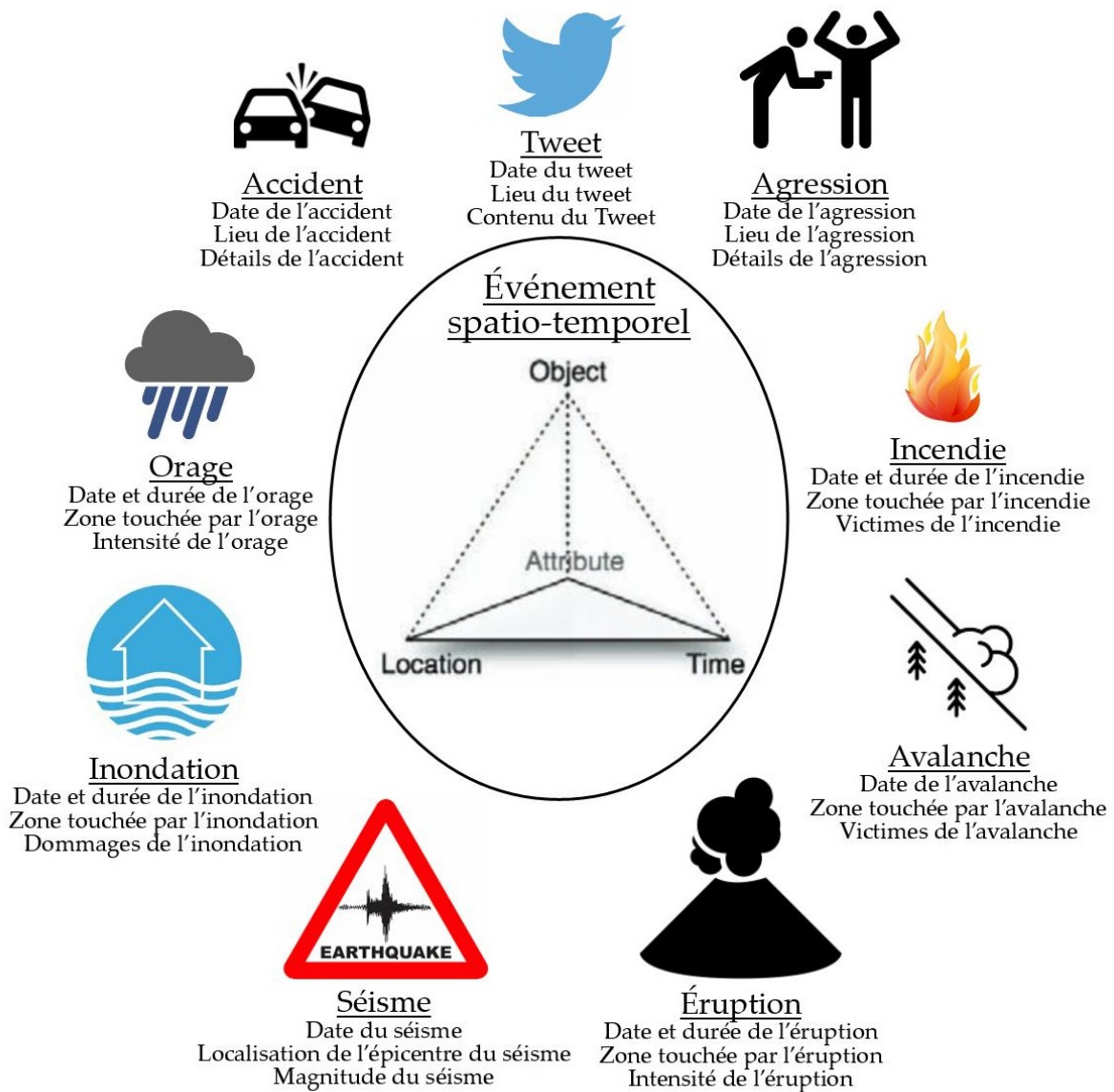


FIGURE 1.18 – Généricité de l'approche. Au centre, le Pyramid Model [Mennis et al., 2000].

Chapitre 2

L'étude des dynamiques des phénomènes spatio-temporels par l'analyse exploratoire

Sommaire

2.1	Les dynamiques de phénomènes modélisés par des séries événementielles	58
2.1.1	Les <i>dynamiques spatio-temporelles</i> , structures du <i>phénomène</i> dans le temps et l'espace	58
2.1.2	Les dynamiques spatio-temporelles dans le cadre de notre modélisation événementielle du changement dans l'espace	58
2.1.3	Les différents types de dynamiques sur lesquels nous nous focalisons . . .	59
2.2	Identifier les dynamiques spatio-temporelles des phénomènes au moyen de l'analyse exploratoire	66
2.2.1	L'analyse de données	66
2.2.2	L'analyse exploratoire	68
2.2.3	L'exploration des données au moyen de représentations graphiques . . .	72

2.1 Les dynamiques de phénomènes modélisés par des séries événementielles

2.1.1 Les *dynamiques spatio-temporelles*, structures du *phénomène* dans le temps et l'espace

Comme nous le précisons dans le chapitre 1, les *phénomènes spatio-temporels*, à l'origine du changement dans l'espace, sont une autre dénomination des *phénomènes géographiques dynamiques* [Haddad, 2009].

Un *phénomène géographique* étant un ensemble de faits et de caractéristiques observables et localisables dans l'espace géographique [Haddad, 2009], le terme *dynamique* fait référence au changement de ces faits et caractéristiques dans l'espace au cours du temps. L'analyse des *dynamiques spatio-temporelles* des *phénomènes* consiste alors à identifier comment s'effectuent ces changements, à identifier la structure et les variations du *phénomène* dans le temps et l'espace.

L'identification de ces dynamiques peut ainsi offrir une meilleure compréhension du mécanisme du phénomène observé, dans le but d'améliorer les connaissances existantes sur des phénomènes semblables, d'effectuer une prévision de leur apparition ou de leur évolution, et d'aider à la prise de décisions les concernant.

Un phénomène pouvant être relié à différentes problématiques, l'analyse des dynamiques spatio-temporelles peut répondre à plusieurs besoins et enjeux. Par exemple :

- L'analyse des dynamiques d'un phénomène sociétal peut permettre de suspecter les facteurs sociaux impactant son mécanisme, et aider à la prise de décisions dans la politique publique.
- L'analyse des dynamiques d'un phénomène naturel permet d'en comprendre les rythmes, et d'effectuer une prévision de l'apparition et de l'évolution de phénomènes de même nature, afin d'aider à la mise en place de politiques de prévention des risques naturels.
- L'analyse de données historiques liées à épisode particulier de l'*Histoire*, représenté sous la forme d'un phénomène spatio-temporel dont on étudie les dynamiques, permet de générer des connaissances sur cet épisode historique, et de récréer un récit de celui-ci.

2.1.2 Les dynamiques spatio-temporelles dans le cadre de notre modélisation événementielle du changement dans l'espace

Notre approche de modélisation du changement dans l'espace se base sur la construction de séries d'*événements*, entités limitées dans le temps et l'espace, et possédant une entité propre ainsi qu'une composante thématique. La structure de ces *événements* se conforme à la formalisation de l'objet spatio-temporel représentée dans le *Pyramid Model* [Mennis et al., 2000].

Ces *événements* sont regroupés en *phénomènes*, rassemblant des *événements* similaires. Cette similarité entre les *événements* d'un même *phénomène* s'établissant avant tout sur un critère thématique, les *événements* composant un *phénomène* sont tous plus ou moins proches selon cette composante. L'analyse d'un phénomène s'effectue ainsi dans un cadre où la dimension thématique est plus ou moins fixée. L'analyse des dynamiques spatio-temporelles porte alors sur les dimensions temporelle et spatiale du phénomène.

L'analyse des dynamiques spatio-temporelles nécessite, non pas d'étudier l'évolution individuelle de chaque événement composant le phénomène, mais d'analyser la distribution spatio-temporelle de l'ensemble d'événements formé par le phénomène afin d'en dégager des structures, d'identifier des patterns, des tendances, ou des relations entre la distribution temporelle et la

distribution spatiale de cette série d'objets.

Les événements composant un phénomène sont des entités spatio-temporelles pouvant être ponctuelles ou étendues dans le temps ou l'espace. Certaines dynamiques spatio-temporelles peuvent ainsi être relatives à l'évolution de la durée des événements en fonction de leur localisation dans l'espace, ou relatives à l'évolution de la région impactée par des événements en fonction de leur localisation dans le temps.

Nous effectuons le choix de nous concentrer sur les dynamiques décrivant l'évolution de la structure globale du phénomène. Nous ne cherchons pas à identifier l'évolution moyenne des caractéristiques temporelles et spatiales individuelles des événements composant un phénomène, mais à identifier l'évolution spatio-temporelle et les rythmes du phénomène pris dans son ensemble.

L'analyse de ces dynamiques spatio-temporelles demande alors d'étudier les possibles relations entre la localisation des événements et les temporalités définissant les relations entre ces événements (*l'ordre d'apparition*, *l'espacement temporel*, et la *synchronisation*) à travers l'analyse de leur distribution spatio-temporelle.

2.1.3 Les différents types de dynamiques sur lesquels nous nous focalisons

Les dynamiques spatio-temporelles que nous cherchons à analyser peuvent être relatives au temps linéaire, ou relatives à une échelle temporelle cyclique, et représentent différentes évolutions pouvant impacter la structure du phénomène. Nous cherchons ainsi à identifier des variations de l'intensité du phénomène dans le temps et l'espace, et à identifier la possible propagation spatiale d'un phénomène au cours du temps.

2.1.3.1 Analyse de variations de l'intensité du phénomène dans le temps et l'espace

Une variation d'intensité d'un phénomène dans le temps et l'espace correspond, dans le cadre de notre modélisation, à une variation du nombre d'apparitions d'événements dans le temps et l'espace. Nous cherchons à identifier trois formes de variations de l'intensité d'un phénomène.

Analyse de pics d'intensité localisés dans le temps et l'espace. Nous cherchons dans un premier temps à identifier des variations brusques de l'intensité du phénomène, apparaissant dans des zones restreintes de l'espace et durant un intervalle de temps limité.

L'observation peut s'effectuer sur un intervalle de temps linéaire afin de localiser, dans le temps et l'espace, les pics d'intensité d'un phénomène au cours de son existence. L'analyse d'accidents routiers peut, par exemple, permettre d'identifier un pic du nombre d'accidents, localisé dans le temps à l'instant T_1 , et dans l'espace au lieu-dit *La Grande Motte* (Figure 2.1).

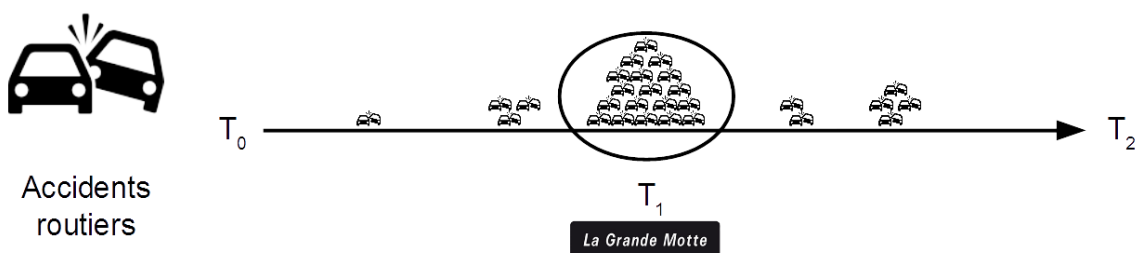


FIGURE 2.1 – Pics d'intensité du phénomène au travers d'une échelle temporelle linéaire

Les variations de l'intensité d'un phénomène peuvent également être observées au travers d'échelles temporelles cycliques. L'analyse d'accidents routiers peut ainsi être effectuée au travers d'une échelle temporelle d'une année, afin d'identifier si l'arrivée de l'hiver coïncide avec l'augmentation du nombre d'accidents survenant en zone de montagne, par rapport au nombre d'accidents survenant en pleine vallée (Figure 2.2).

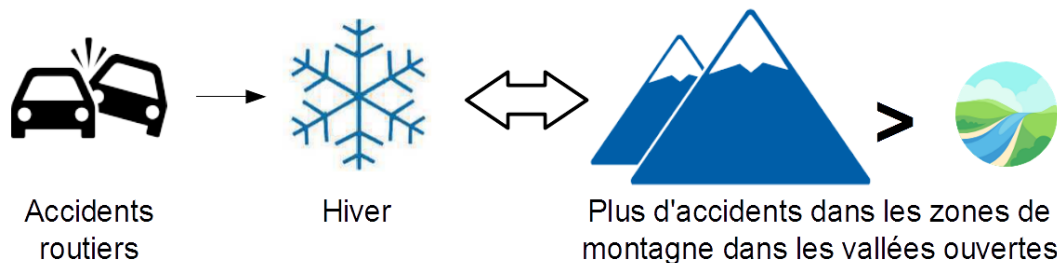


FIGURE 2.2 – Pics d'intensité du phénomène relatifs à une échelle temporelle cyclique d'un an

L'utilisation d'une telle échelle n'a pas pour objectif d'identifier le caractère cyclique d'un phénomène, ou la fréquence de ce cycle. Le but est d'identifier à quels moments de l'échelle représentée ces phénomènes sont les plus actifs. L'analyse peut alors porter sur l'identification des pics d'intensité de phénomènes connus d'emblée comme étant cycliques.

Le phénomène des accidents routiers peut être considéré comme observant un cycle journalier, car il est lié aux activités humaines. L'analyse de ce phénomène, au travers d'une échelle temporelle d'une journée, peut permettre d'identifier à quelles heures surviennent le plus d'accidents routiers, et où se situent ces pics d'intensité dans l'espace (Figure 2.3).



FIGURE 2.3 – Pics d'intensité du phénomène relatifs à une échelle temporelle cyclique d'une journée

Les échelles temporelles utilisées peuvent ainsi être des échelles calendaires, reflétant les rythmes de la nature et des sociétés humaines (heure, jour, semaine, mois, année). L'analyse permet alors de mettre en comparaison les rythmes d'apparition de pics d'intensité du phénomène avec les rythmes naturels et sociétaux représentés par ces échelles temporelles.

Analyse de tendances localisées au cours du temps. Dans un second temps, nous cherchons à identifier, non pas des variations brusques, mais des tendances dans la variation de l'intensité d'un phénomène durant un intervalle de temps linéaire : l'intensité du phénomène augmente-t-elle ? diminue-t-elle ? est-elle en stagnation ?

Ces tendances, globales sur le plan temporel, peuvent présenter des disparités selon l'espace : l'intensité du phénomène augmente au cours du temps dans une zone spatiale A , et diminue dans une zone spatiale B . L'analyse de tendances localisées dans des données d'accidents routiers (Figure 2.4) doit ainsi permettre d'identifier quelles sont les zones présentant une augmentation, une diminution, ou une stagnation du nombre d'accidents.

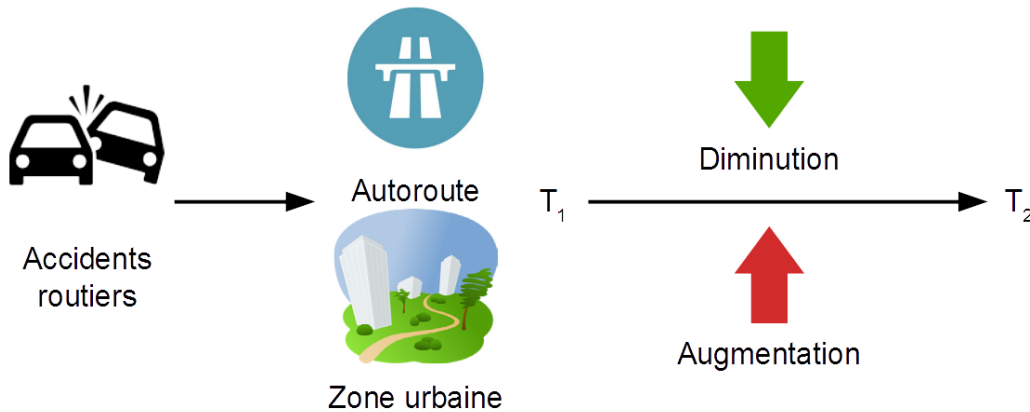


FIGURE 2.4 – Variations localisées du phénomène au cours du temps

Analyse de récurrences cycliques localisées. Dans un troisième temps, nous cherchons à identifier les possibles récurrences cycliques dans la variation d'intensité du phénomène. À l'inverse de l'analyse de la localisation de pics d'intensité d'un phénomène, au travers d'une échelle temporelle cyclique définie, le but de cette analyse est d'identifier si la variation d'un phénomène est cyclique et quelle est la fréquence de ce cycle.

De la même manière que le sont les pics d'intensité du phénomène, ces variations cycliques peuvent être localisées dans l'espace. Ceci peut se traduire par la circonscription du caractère cyclique de l'intensité du phénomène à une zone de l'espace, ou par l'existence de différentes zones où cette évolution est cyclique, mais pour lesquelles la fréquence du cycle observé est différente.

L'analyse de récurrences cycliques spatialisées dans des données liées aux phénomènes naturels, comme des épisodes de fortes pluies (Figure 2.5), doit ainsi permettre d'identifier les fréquences de réapparition du phénomène selon différentes zones de l'espace.

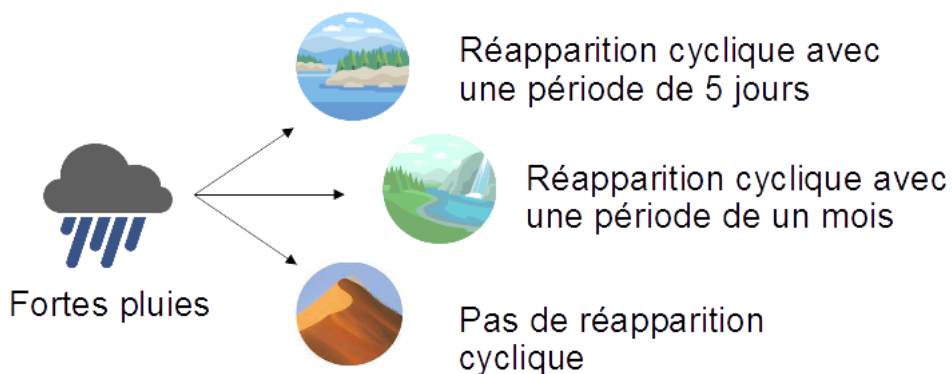


FIGURE 2.5 – Réapparition cyclique localisée d'un phénomène

2.1.3.2 Analyse de l'évolution spatiale d'un phénomène au cours du temps

Nous cherchons enfin à identifier des tendances globales dans l'évolution de l'extension spatiale du phénomène, ou dans le déplacement de sa zone d'activité au cours du temps. Nous cherchons notamment à identifier de possibles axes et foyers de diffusion du phénomène.

Un possible cas d'application peut être l'analyse de la diffusion, au cours du temps, d'une épidémie (Figure 2.10). L'analyse doit alors permettre d'identifier les foyers de l'épidémie et ses différents axes de diffusion.

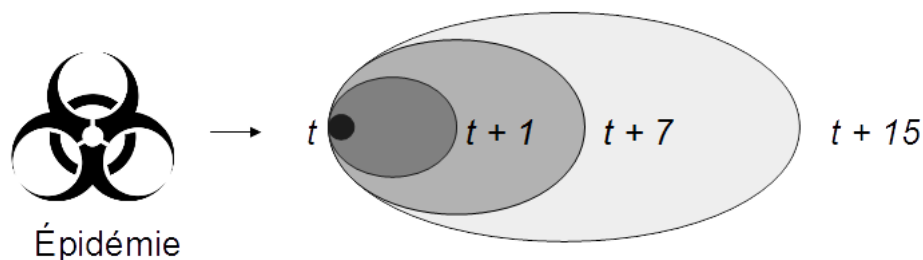


FIGURE 2.6 – Propagation spatiale d'un phénomène au cours du temps

2.1.3.3 Les patterns traduisant les dynamiques dans la distribution des événements

L'analyse des dynamiques spatio-temporelles des phénomènes consiste à mettre en lumière certains *motifs d'intérêts*, que nous appelons *patterns spatio-temporels*, dans la distribution des séries d'événements.

Clusters spatio-temporels. Un pic d'intensité d'un phénomène dans le temps et l'espace se traduit, dans la série événementielle correspondante, par une concentration d'événements dans une zone spatiale et durant un intervalle temporel. Nous nommons ce motif d'intérêt *cluster spatio-temporel*. Selon que le temps soit représenté comme linéaire ou cyclique, ces patterns prennent une forme différente.

Clusters relatifs au temps linéaire. Dans une représentation linéaire du temps, un cluster se traduit par la concentration spatiale d'événements apparaissant durant les mêmes intervalles temporels linéaires.

Des événements apparaissant sur le territoire de Grenoble peuvent être concentrés temporellement dans la période allant du *24 mai 2012* au *30 mai 2012*. Des événements apparaissant sur le territoire de Lyon peuvent être concentrés temporellement dans la période allant du *10 août 2012* au *16 août 2012*.

Ces deux groupes d'événements peuvent être considérés comme des clusters relatifs au temps linéaire (Figure 2.7).

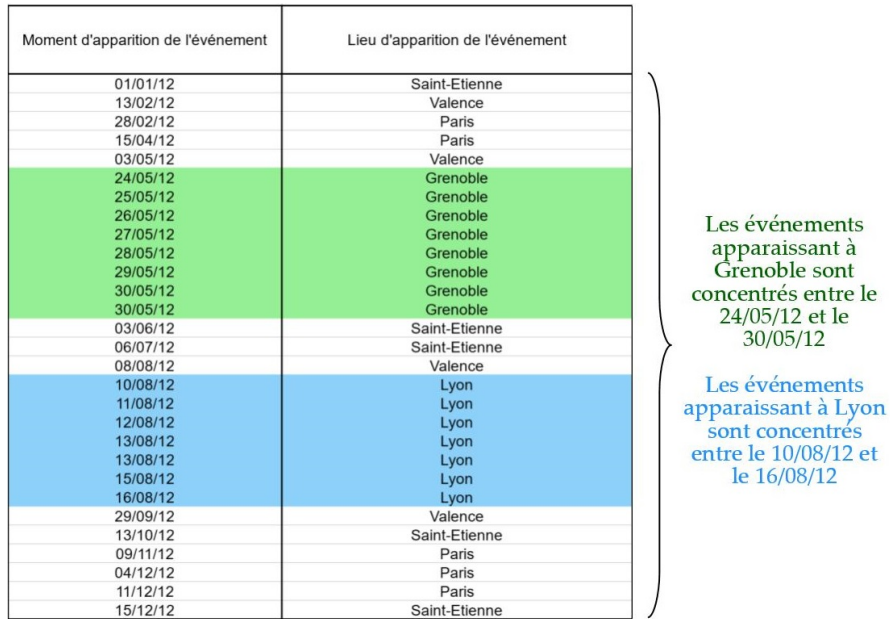


FIGURE 2.7 – Clusters spatio-temporels relatifs au temps linéaire

L'identification de clusters spatio-temporels sur un intervalle temporel linéaire permet également d'identifier des tendances dans la variation d'intensité d'un phénomène selon différentes zones de l'espace.

Dans ce cas précis, l'étude consiste à analyser la distribution temporelle des événements selon différentes zones de l'espace, et à identifier la proportion d'événements anciens ou récents (Figure 2.8) :

- Une zone spatiale présentant une majorité d'événements anciens et une minorité d'événements récents subit une augmentation de l'intensité du phénomène.
- Une zone spatiale présentant une minorité d'événements anciens et une majorité d'événements récents subit une augmentation de l'intensité du phénomène.
- Une zone spatiale présentant un nombre égal d'événements anciens et récents subit une stagnation de l'intensité du phénomène.



FIGURE 2.8 – Identification de clusters relatifs au temps linéaire pour l'analyse de tendances

Clusters relatifs au temps cyclique. Dans une représentation cyclique du temps, un cluster se traduit par la concentration spatiale d'événements apparaissant durant les mêmes intervalles temporels cycliques.

Moment d'apparition de l'événement	Lieu d'apparition de l'événement
02:33:00	Lyon
03:18:00	Valence
07:46:00	Grenoble
07:55:00	Saint-Etienne
08:00:00	Saint-Etienne
08:05:00	Saint-Etienne
08:10:00	Saint-Etienne
08:20:00	Saint-Etienne
09:30:00	Grenoble
10:14:00	Valence
10:16:00	Valence
10:22:00	Grenoble
13:26:00	Lyon
13:46:00	Grenoble
14:25:00	Paris
14:28:00	Paris
14:30:00	Paris
14:32:00	Paris
14:35:00	Paris
15:15:00	Grenoble
16:02:00	Valence
16:34:00	Lyon
17:06:00	Lyon
17:29:00	Valence
20:10:00	Lyon
21:12:00	Valence
22:14:00	Grenoble
22:25:00	Grenoble
23:32:00	Valence

Les événements apparaissant à Saint-Etienne sont concentrés entre 07:55 et 8:20

Les événements apparaissant à Paris sont concentrés entre 14:25 et 14:35

FIGURE 2.9 – Clusters spatio-temporels relatifs au temps cyclique

Des événements apparaissant à Saint-Étienne peuvent être concentrés temporellement sur une plage horaire journalière allant de 07 :55 à 8 :20. Des événements apparaissant à Paris peuvent être concentrés temporellement sur une plage horaire journalière allant de 14 :25 à 14 :35. Ces deux sous-groupes d'événements peuvent être considérés comme des clusters relatifs à une échelle cyclique de 24 heures (Figure 2.9).

L'existence de ces clusters, à l'inverse des clusters relatifs au temps linéaire, dépend de l'échelle temporelle utilisée. Ces clusters sont relatifs à une échelle cyclique donnée :

- Soit l'exemple d'événements d'accidents routiers. Selon une échelle temporelle cyclique d'une semaine, il peut y avoir plus d'accidents vers 18h dans les zones d'activité (à forte concentration d'emplois) que dans les zones résidentielles. Sur une échelle temporelle cyclique d'une année, il peut y avoir plus d'accidents en hiver dans les zones de montagne que dans les zones de plaine ou urbaines.
- Soit l'exemple d'événements de cambriolages. Sur une échelle temporelle d'une journée, il peut y avoir plus de cambriolages en journée dans les zones résidentielles, plus de cambriolages durant la nuit dans les zones d'activité.

Propagation spatiale d'événements. Afin d'identifier une évolution spatiale d'un phénomène au cours du temps, nous cherchons à repérer des patterns dans l'évolution de l'étendue de la distribution spatiale des événements au cours du temps. Afin notamment d'analyser les possibles foyers et axes de diffusion du phénomène, nous cherchons à identifier une propagation spatiale des événements au cours du temps.

Dans le cas de l'analyse de données d'épidémie, le phénomène peut être modélisé par des événements représentant chaque cas de maladie. La diffusion de l'épidémie peut être identifiée par la propagation des cas de maladie dans l'espace au cours du temps, depuis le foyer de l'épidémie le long des axes de diffusion (Figure 2.10).



FIGURE 2.10 – Propagation spatiale d'un phénomène au cours du temps

Contrairement à la recherche de clusters spatio-temporels, la recherche d'une propagation spatiale d'événements au cours du temps ne passe pas par l'identification de concentration d'événements dans le temps et l'espace. Celle-ci passe par l'identification d'une corrélation entre la position spatiale des événements, les uns par rapport aux autres sur un axe directionnel, et leur ordre d'apparition dans le temps.

Récurrences cycliques spatialisées d'événements. Afin d'identifier des récurrences cycliques localisées dans la variation de l'intensité du phénomène, nous cherchons à repérer des récurrences cycliques spatialisées dans la distribution des événements.

Celles-ci se caractérisent par la réapparition de clusters d'événements, relatifs à une échelle cyclique dont la durée correspond à la période du cycle, présentant les caractéristiques suivantes :

- l'apparition des événements des différents clusters dans la même zone spatiale : *les événements correspondants apparaissent tous dans la région de Grenoble* ;
- l'apparition des événements des différents clusters durant le même intervalle temporel cyclique : *les événements correspondants apparaissent tous durant le mois de juin* ;
- l'apparition d'un cluster à chaque instance de l'échelle cyclique : *les événements correspondants apparaissent chaque année*.

Si l'existence de récurrences cycliques fait l'objet d'une analyse temporelle des événements, nous nous intéressons ici à la spatialité de ces récurrences cycliques. Ainsi, dans une série d'événements, ne réapparaissant pas de manière cyclique de manière globale, des événements circonscrits à une zone spatiale peuvent réapparaître de manière cyclique (Figure 2.11).

Cette spatialité peut se traduire de deux façons :

- les événements réapparaissant de manière cyclique sont circonscrits à une même zone spatiale ;
- des événements réapparaissant de manière cyclique apparaissent dans différentes zones spatiales, mais la période du cycle de réapparition diffère selon la zone d'apparition des événements.

L'analyse diffère ici de la recherche de clusters spatio-temporels, où l'analyse a pour but d'analyser des concentrations d'événements pour une échelle cyclique donnée, fixée à l'avance. Cette approche peut également conduire à l'identification de cycles ayant pour période l'échelle cyclique choisie pour l'analyse. L'analyse de récurrence spatialisée a au contraire pour objectif l'identification de l'échelle cyclique pour laquelle on observe une réapparition cyclique de concentration d'événements dans le temps et l'espace.

Moment d'apparition de l'événement	Lieu d'apparition de l'événement
12/12/12	Grenoble
14/12/12	Valence
31/12/12	Lyon
04/01/13	Lyon
11/01/13	Saint-Etienne
11/01/13	Grenoble
29/01/13	Valence
13/02/13	Grenoble
18/02/13	Lyon
19/02/13	Lyon
28/02/13	Valence
06/03/13	Paris
12/03/13	Grenoble
15/04/13	Grenoble
17/04/13	Paris
21/04/13	Lyon
24/04/13	Saint-Etienne
01/05/13	Paris
08/05/13	Valence
10/05/13	Grenoble
11/05/13	Saint-Etienne
11/05/13	Lyon
14/06/13	Saint-Etienne
16/06/13	Grenoble
30/06/13	Paris
13/07/13	Grenoble
21/07/13	Valence
01/08/13	Lyon
14/08/13	Grenoble

Les événements localisés sur Grenoble réapparaissent de manière cyclique avec une période de réapparition oscillant autour d'un mois

FIGURE 2.11 – Récurrences cycliques spatialisées d'événements

2.2 Identifier les dynamiques spatio-temporelles des phénomènes au moyen de l'analyse exploratoire

2.2.1 L'analyse de données

2.2.1.1 L'analyse de données en statistique

Il existe différentes définitions de l'analyse de données dans la littérature. En statistique, l'analyse de données est définie comme *le procédé de calculer une série de synthèses et de valeurs dérivées depuis une collection de données définie* (d'après [Berthold and Hand, 2007, Andrienko and Andrienko, 2006]).

L'analyse de données peut alors être séquentielle [Andrienko and Andrienko, 2006]. Une première analyse peut être effectuée sur les données au moyen d'une première technique analytique. En fonction des résultats obtenus, une seconde analyse peut être effectuée afin d'examiner les données selon un second point de vue, nécessitant parfois de modifier les données. Une troisième analyse peut être effectuée en revenant au format de données d'origine en utilisant un outil d'analyse différent (d'après [Berthold and Hand, 2007, Andrienko and Andrienko, 2006]).

Ce procédé peut alors être répété de nombreuses fois, chaque outil d'analyse donnant à voir un aspect radicalement différent des mêmes données, dans le but de répondre à différentes questions sur les données ([Berthold and Hand, 2007, Andrienko and Andrienko, 2006]).

2.2.1.2 L'analyse de données dans les *Systèmes d'Information Géographiques*

Dans le domaine des Systèmes d'Information Géographiques (*SIG*), l'analyse de données est définie comme *un procédé permettant de rechercher des patterns géographiques dans les données spatiales, ainsi que des relations entre des entités spatiales* (d'après [Mitchell, 1999, Andrienko and Andrienko, 2006]).

Le procédé d'analyse a pour point de départ la formulation d'une question à laquelle l'analyse

doit répondre, suivie du choix d'une méthode analytique correspondant à la question posée, au type de données disponibles [Andrienko and Andrienko, 2006]. Les données sont ensuite mises en forme pour permettre l'utilisation de la méthode analytique choisie.

Une fois l'analyse effectuée, les résultats sont représentés dans l'interface visuelle du *SIG* [Andrienko and Andrienko, 2006]. Cette représentation visuelle permet à l'analyste de décider quelles sont les informations valides ou utiles parmi les résultats obtenus, et quelles analyses doivent au contraire être réeffectuées, en modifiant les paramètres de l'analyse ou la méthode d'analyse utilisée.

2.2.1.3 Point commun de ces deux approches : l'énonciation d'une hypothèse *a priori*

Dans les deux domaines des statistiques et des *SIG*, le procédé d'analyse suit une structure commune [Andrienko and Andrienko, 2006] :

- la formalisation d'une question à laquelle l'analyse doit répondre ;
- le choix d'une méthode d'analyse des données ;
- la mise en forme des données pour effectuer l'analyse selon la méthode choisie ;
- l'application de la méthode d'analyse choisie ;
- l'interprétation et évaluation des résultats obtenus.

À partir des résultats obtenus, ce procédé peut être réitéré afin d'approfondir l'analyse.

La différence entre l'analyse statistique et l'analyse de données effectuée au moyen de *SIG* semble alors liée aux données traitées et aux outils d'analyse proposés. Dans les deux cas, l'analyse est orientée par une hypothèse formulée *a priori*, avant que l'analyse en elle-même ne renvoie de l'information sur les données étudiées [Andrienko and Andrienko, 2006]. Les questions soulevées par cette hypothèse sont la cause de l'analyse, mais orientent également le choix des données et des outils d'analyses utilisés, ainsi que l'interprétation des résultats [Andrienko and Andrienko, 2006].

2.2.1.4 Le choix de l'analyse exploratoire

Ce procédé d'analyse, basé sur la validation d'une hypothèse énoncée *a priori* sur les données, se justifie dans le cadre de problématiques de recherche où la nature des données analysées est connue, et où les éléments à identifier sont précisément définis [Andrienko and Andrienko, 2006].

En revanche, cette approche pose problème dans le cas où les jeux de données à analyser, ainsi que les éléments à identifier, sont inconnus. Énoncer *a priori* une hypothèse à valider sur les données, afin d'opter pour la méthode d'analyse adéquate, nécessite d'être *familier* avec les données étudiées [Andrienko and Andrienko, 2006].

Or, l'émergence constante de nouveaux éléments permettant de décrire les phénomènes peut amener à observer des données pour lesquelles l'utilisateur ne sait pas dans quelle direction focaliser son analyse. Cette étude peut s'effectuer dans le cadre d'un domaine de recherche connu de l'utilisateur, mais pour lequel il n'a pas eu l'occasion d'analyser des jeux de données similaires aux nouveaux éléments à disposition. L'utilisateur peut alors ne pas savoir quels éléments y chercher. L'utilisateur doit d'abord devenir *familier* avec ces données, avant de définir sur quelle dynamique spatio-temporelle il doit approfondir sa recherche.

Dans le cadre de notre proposition, la modélisation des phénomènes en séries d'entités spatio-temporelles nous permet de définir, grossièrement, le format des données étudiées, et les structures

recherchées dans la distribution spatio-temporelle des événements. Si les dynamiques à identifier, et les patterns correspondants, sont fixés, leur définition reste relativement vague et ne portent que sur la structure globale de ces patterns.

Or, le développement d'une analyse générique implique que les données représentées puissent se référer à des phénomènes de différentes natures, s'inscrivant chacun dans des problématiques particulières. La nature de ces données et les problématiques associées influant sur l'objectif de l'analyse, les dynamiques à identifier dans chaque cas d'étude différent. Les patterns recherchés peuvent être de nature différente, mais également varier dans leur étendue temporelle, spatiale, ou dans le taux de variation d'intensité représenté.

Ce manque de précision dans la définition des patterns à identifier, ainsi que la possibilité, pour l'utilisateur, de faire face à des données dont il n'a pas l'expérience, rend difficile l'établissement de méthodes d'analyse basées sur une approche hypothético-déductive. Ainsi, pour offrir une analyse des phénomènes, applicable à différents cas d'étude et permettant d'exploiter l'apparition constante de nouveaux éléments décrivant les phénomènes, il peut être intéressant de simplement montrer les données, de permettre de les explorer, et de montrer ce qu'elles peuvent donner à voir.

Cette analyse, effectuée sans *a priori* sur la structure des données analysées, permet, en outre, à l'analyste de découvrir de nouvelles informations insoupçonnées [Tukey, 1977, Andrienko and Andrienko, 2006], son raisonnement n'étant pas focalisé sur l'identification d'éléments précisément définis en amont de l'analyse.

Dans le cadre de notre proposition, il faut donc que nous utilisions une approche d'analyse autorisant l'étude de données inconnues, et permettant à l'analyste :

- dans un premier temps, de devenir *familier* avec les données observées ;
- dans un second temps, de définir les dynamiques spatio-temporelles sur lesquelles il doit approfondir sa recherche.

Permettre ceci est l'objectif suivi par l'analyse exploratoire (*exploratory data analysis*) [Andrienko and Andrienko, 2006].

2.2.2 L'analyse exploratoire

2.2.2.1 Principe originel : l'exploration des données pour la génération d'hypothèses

Le principe de l'analyse exploratoire n'est pas de répondre à une question spécifique, comme en analyse statistique, mais *d'explorer un jeu de données et de voir ce qu'il aurait à nous dire* [Andrienko and Andrienko, 2006].

L'analyse exploratoire ne représente pas un regroupement spécifique d'outils d'analyse, mais plutôt une philosophie différente sur la manière d'effectuer l'analyse de données [Andrienko and Andrienko, 2006].

John Tukey [Tukey, 1977, Andrienko and Andrienko, 2006] définit cette philosophie comme une manière de contrebalancer un biais durablement installé dans le domaine de l'analyse statistique, qu'il attribue au recours quasi systématique au développement d'outils mathématiques, dans le cadre de la validation d'hypothèses énoncées *a priori*. John Tukey considère l'analyse exploratoire comme un retour au but originel de la statistique : la détection et la description de patterns, de tendances et de relations dans les données, permettant de générer par la suite une hypothèse sur ces données. Le principe de l'analyse exploratoire est ainsi basé, selon John Tukey,

sur la génération plutôt que sur la validation d'hypothèses. Si l'analyse de données effectuée dans les domaines de la statistique et des systèmes d'information géographique cherche à valider une hypothèse énoncée *a priori*, l'analyse exploratoire cherche à permettre une exploration des données dans le but de générer ces hypothèses *a posteriori* [Andrienko and Andrienko, 2006].

2.2.2.2 D'anciennes méthodes analogues

L'analyse des cas de choléra effectuée par John Snow au moment de l'épidémie de Londres en 1854 [Bonfim and Medeiros, 2008] peut être considérée comme un cas d'analyse exploratoire.

John Snow, considéré comme le créateur de l'épidémiologie moderne, est un médecin britannique du XIXe siècle. En 1854, Londres est confrontée à une épidémie de choléra. Les autorités locales, ne connaissant pas le mécanisme de propagation de la maladie, s'interrogent sur la manière de contenir l'épidémie.

Snow décide de représenter sur une carte de Londres l'ensemble des cas de choléra répertoriés (entités ponctuelles noires dans la carte de la Figure 2.12). En analysant visuellement la distribution spatiale des cas de choléra, Snow identifie une concentration des cas de choléra dans l'espace. En analysant de plus près la zone spatiale où se situe ce cluster, Snow découvre que les cas de choléra sont concentrés autour d'une pompe à eau publique installée sur Broad Street. Snow émet alors l'hypothèse que le choléra se transmet par l'eau, et que l'eau contaminée fournie par la pompe de Broad Street est à l'origine de la surreprésentation de cas de choléra dans cette partie de la ville de Londres.



FIGURE 2.12 – Carte de cas de choléra pendant l'épidémie de 1854 à Londres

Suite à cette analyse, la pompe de Broad Street est mise hors service. Cette action est communément admise comme à l'origine de la fin de l'épidémie, mais Snow observa que l'épidémie pouvait déjà être en rapide déclin lors de la mise hors service de la pompe.

Le principal reproche fait aux thèses de Snow sur la propagation du choléra est qu'il considère l'eau comme seule responsable de la propagation de la maladie. Cependant, son analyse de l'épidémie de 1854 est un bon exemple de ce que l'analyse exploratoire permet lorsqu'un analyste fait face à une série de données sans savoir quoi y chercher. Conformément à la philosophie de l'analyse exploratoire, l'analyse de Snow permet de générer une hypothèse à partir de l'exploration libre des données.

2.2.2.3 Le besoin d'un objet d'étude et d'un objectif de recherche

D'après Klir [Klir, 1985, Andrienko and Andrienko, 2006], une recherche empirique implique un objet d'investigation, un objectif d'investigation, et des contraintes imposées sur cette investigation.

Dans le cas de l'analyse effectuée par Snow :

- L'objet d'investigation correspond aux cas de choléra sur la ville de Londres.
- L'objectif d'investigation est la découverte des vecteurs de propagation du choléra.
- Les contraintes imposées correspondent au format des données disponibles.

Si Klir n'utilise pas le terme d'analyse exploratoire dans son ouvrage, cette dernière, en tant que recherche empirique de structures dans les données, nécessite également un objet d'étude et un objectif de recherche, aussi global soit-il [Andrienko and Andrienko, 2006].

Les données collectées en vue d'une analyse exploratoire sont sélectionnées en regard d'un certain objet d'étude, et l'analyste effectue son exploration afin de répondre à un certain objectif de recherche.

Ben Shneiderman, dans son *Information Seeking Mantra* [Shneiderman, 2003, Andrienko and Andrienko, 2006], définit le procédé général de l'analyse exploratoire selon un procédé en trois étapes, *Overview first, zoom and filter, and then details on demand* :

- La première étape consiste en une observation globale de l'ensemble d'un jeu de données.
- La seconde étape consiste en une focalisation de l'analyse sur des objets d'intérêt.
- La dernière étape consiste en une exploration plus poussée, focalisée sur les caractéristiques des objets d'intérêts identifiés dans la précédente étape.

Ce procédé peut être répété plusieurs fois de manière itérative. Shneiderman ne précise pas quels critères définissent les objets d'intérêts, que l'analyste doit identifier, et sur lesquels doit être effectuée une analyse approfondie. Ces derniers sont choisis par l'analyste en fonction de son objectif de recherche [Andrienko and Andrienko, 2006].

Ainsi, l'analyse exploratoire semble elle aussi être orientée dès le départ par une certaine hypothèse, formulée par l'analyste. Cette hypothèse est à l'origine de la collecte du jeu de données, et oriente l'analyste dans son exploration.

Cependant, l'analyse exploratoire se distingue de l'analyse de données effectuée en statistiques et au moyen des *SIG* [Andrienko and Andrienko, 2006] :

- l'analyse exploratoire a pour objectif de répondre, non pas à une question spécifique, mais à une série de questions dans le cadre d'un objectif de recherche global.
- les questions posées par l'analyse exploratoire varient en termes de niveaux de généralité. Certaines sont très générales et portent sur l'ensemble du jeu de données, d'autres sont plus spécifiques et portent sur une partie du jeu de données, identifiée lors de l'analyse.

- la plupart des questions posées par l’analyse exploratoire sont formulées au cours de celle-ci et non à l’avance. L’hypothèse générale permet de guider l’analyste dans les premiers instants de son exploration. D’autres hypothèses, plus précises et plus spécifiques, sont générées en cours d’analyse, à partir de l’interprétation par l’analyste de ce qu’il voit.

Ces particularités sont liées au caractère *générique* de l’analyse exploratoire. Elles rendent cependant difficile la formalisation d’une méthode universelle d’analyse exploratoire [Andrienko and Andrienko, 2006], l’exploration effectuée par l’analyste pouvant se révéler très différente selon l’objet d’étude et l’objectif de recherche. De plus, la multiplicité des questions, pouvant être posées par l’analyste au cours de sa recherche, implique de mettre à sa disposition de multiples outils d’analyse permettant de répondre à chacune de ces questions [Andrienko and Andrienko, 2006]. L’approche idéale d’analyse exploratoire doit alors proposer différents outils permettant à l’utilisateur de répondre à chaque hypothèse qu’il pourrait émettre au cours de sa recherche.

Les environnements d’analyse exploratoire sont ainsi souvent élaborés dans le cadre de domaines thématiques particuliers [Kaddouri et al., 2014], ce qui permet de réduire le nombre de questions que l’utilisateur peut formuler au cours de l’analyse.

2.2.2.4 L’analyse exploratoire au service de la recherche de dynamiques spatio-temporelles

Le caractère *générique* de l’analyse exploratoire apporte la possibilité d’une approche d’analyse des dynamiques applicable à différents phénomènes spatio-temporels. Chaque phénomène constitue ainsi un objet d’étude, pour lequel des données spatio-temporelles peuvent être collectées, en vue d’être analysées.

Ces données sont d’abord mises sous la forme de séries événementielles, conformément à l’approche de modélisation générique que nous présentions dans le chapitre 1. Cette modélisation générique du changement dans l’espace, permettant de mettre en forme les données relatives à différents types de phénomènes selon un même format, permet leur utilisation dans un même outil d’analyse exploratoire.

L’objectif de recherche de chaque analyse exploratoire portant sur un phénomène est l’identification de ses dynamiques. En fonction du phénomène analysé, des données utilisées, et de la problématique métier dans laquelle elles s’inscrivent, les dynamiques recherchées peuvent être différentes :

- Analyser des phénomènes naturels dans un but de prévention implique de s’intéresser aux récurrences cycliques du phénomène.
- Analyser des phénomènes sociaux afin d’aider à la mise en place de politiques publiques implique d’analyser les variations du phénomène sur le temps long selon différentes zones de l’espace.

Nous avons souligné qu’une approche d’analyse exploratoire idéale doit proposer les outils permettant de répondre à n’importe quelles questions auxquelles l’utilisateur pourrait être confronté. Le nombre de ces questions pouvant être infini, un environnement d’analyse exploratoire permettant de répondre à la totalité des objectifs d’analyse possibles ne pourra sans doute jamais être proposé [Andrienko and Andrienko, 2006].

Cependant, en limitant le champ de recherche aux dynamiques spatio-temporelles que nous présentons au début de ce chapitre, il nous est possible de proposer une méthode d’analyse ex-

ploratoire permettant d'analyser ce type de structures.

Les différents patterns que nous cherchons à analyser au sein des données étant définis, il nous est possible d'identifier quels outils sont en mesure de permettre une analyse exploratoire de chacune des dynamiques spatio-temporelles citées.

Combinée à notre modélisation générique des données spatio-temporelles, une interface d'analyse exploratoire doit alors permettre une analyse des dynamiques spatio-temporelles des phénomènes, applicable à différentes données dans le cadre de différentes problématiques métier.

2.2.3 L'exploration des données au moyen de représentations graphiques

L'analyse exploratoire est fortement associée à l'utilisation de représentations graphiques des données [Andrienko and Andrienko, 2006]. La plupart des techniques d'analyse exploratoire sont ainsi de nature graphique.

Ces dernières sont parfois associées à des techniques d'analyse quantitatives. Des outils mathématiques existent permettant l'analyse de gradients, de clusters, ou de récurrences cycliques dans des jeux de données. Cependant, l'exploration visuelle des données peut mettre en lumière certaines caractéristiques qui ne seraient pas directement décelables par une simple analyse mathématique [Tufte, 1983].

La raison pour laquelle ces techniques reposent essentiellement sur des outils graphiques est que le principe de base de ces analyses est l'exploration libre des données de la part de l'utilisateur, sans que cette recherche soit orientée par un parti pris sur les données ou une problématique de recherche particulière [Andrienko and Andrienko, 2006].

La représentation graphique, associée à la capacité naturelle que possède chaque utilisateur dans la reconnaissance de patterns graphiques, se présente alors comme le meilleur moyen d'effectuer cette exploration libre, et de révéler des structures insoupçonnées dans les données [Heckert and Filliben, 2003, Andrienko and Andrienko, 2006].

Il est ainsi établi que les outils d'analyse, offerts par les environnements de visualisation, contribuent à l'extraction d'informations utiles sur les structures et les relations présentes dans les données spatiales [Kaddouri et al., 2014].

Cette possibilité d'analyse visuelle n'est cependant pas réservée à la représentation spatiale. D'autres domaines se sont appliqués à fournir des méthodes d'analyses visuelles, permettant de mettre en valeur des motifs d'intérêt dans des données de différentes natures [Kaddouri et al., 2014]. Des domaines comme l'*Info Viz* ou la *Dataviz* ont ainsi produit des techniques de représentation graphique permettant une analyse visuelle de données statistiques.

Si la représentation visuelle de la donnée spatiale peut ainsi permettre une identification de patterns spatiaux, une représentation visuelle de la donnée spatio-temporelle pourrait permettre une identification de patterns spatio-temporels. Une telle représentation nécessite cependant d'intégrer la dimension temporelle à la représentation cartographique.

2.2.3.1 Le manque de représentations graphiques applicables à l'analyse de phénomènes liés à différentes thématiques

Plusieurs travaux ont porté sur l'élaboration d'environnements de visualisations pouvant permettre d'effectuer une analyse exploratoire des données spatio-temporelles [Andrienko and

Andrienko, 2006, Aigner et al., 2007, Aigner et al., 2011, Kaddouri et al., 2014].

Certains travaux adoptent une approche généraliste et essayent de mettre en place une méthode d'élaboration générale des environnements d'analyse exploratoire [Andrienko and Andrienko, 2006], en recensant différentes méthodes de représentations graphiques pouvant être intégrées dans un environnement de visualisation, et en classant celles-ci en fonction du format des données étudiées, des patterns recherchés, etc.

Ces travaux permettent, en fonction des données étudiées et des besoins de l'analyse, de sélectionner les méthodes de représentation les plus adéquates, et d'intégrer celles-ci au sein d'un environnement de visualisation répondant à un cas particulier. Cependant, ces travaux ne portent pas sur l'implémentation d'un environnement de visualisation offrant la possibilité d'une analyse exploratoire générale, applicable à différentes problématiques de recherche. La majorité des environnements de visualisation existants sont ainsi développés de manière *ad hoc*, dans le cadre de projets portant sur une thématique particulière [Kaddouri et al., 2014].

Peu d'environnements de visualisation offrent ainsi la possibilité d'une analyse exploratoire indépendante du thème traité. Or, l'émergence constante de nouvelles données, incite à l'utilisation d'environnement de visualisation permettant :

- d'une part la représentation de phénomènes de différentes natures ;
- d'autre part, l'analyse de différentes dynamiques spatio-temporelles, à travers la recherche et l'identification de différents patterns.

Si la représentation de phénomènes de différentes natures est permise par la modélisation générique que nous présentons dans le chapitre 1, permettre une analyse exploratoire, applicable à différents phénomènes, implique de définir quelles méthodes de représentation graphiques autorisent l'analyse visuelle des différentes dynamiques spatio-temporelles, et peuvent être intégrées au sein d'un même environnement de visualisation.

Un état de l'art des outils de visualisation des données temporelles et spatiales est ainsi nécessaire, afin d'identifier ces méthodes et outils. Cet état de l'art doit notamment porter sur les techniques mises en place pour visualiser les relations entre les composantes temporelle et spatiale des données. Il faut s'interroger sur les représentations graphiques utilisées, mais également sur les fonctions d'interaction mises en place et sur la place de l'utilisateur dans l'utilisation de ces outils visuels interactifs.

Enfin, cet état de l'art doit permettre d'identifier quelles sont les limites rencontrées par les environnements actuels dans l'analyse des phénomènes spatio-temporels, et comment notre proposition peut dépasser ces limites.

Conclusion

Nous adoptons une modélisation générique du changement dans l'espace, de manière à pouvoir traiter des données représentant différents *phénomènes spatio-temporels*.

Cette modélisation est basée sur une vision occurrente du changement [Grenon and Smith, 2004, Galton, 2004], ce dernier étant représenté au travers d'entités spatio-temporelles limitées dans le temps et l'espace, que nous nommons *événements spatio-temporels*. Nous choisissons, conformément au modèle de Reitsma [Reitsma, 2005], à limiter notre modélisation aux événements spatio-temporels, sans représenter les objets spatiaux pérennes potentiellement impactés par ces événements.

La structure de chaque *événement* découle du *Pyramid Model* de Mennis [Mennis et al., 2000], et possède son identité propre, et trois composantes dimensionnelles : spatiale, temporelle, et thématique.

Un *phénomène spatio-temporel* est considéré comme un regroupement d'*événements* traitant du même objet d'étude. Ce regroupement s'effectue principalement sur le plan thématique, mais peut également s'effectuer sur un plan spatial ou temporel. Les données relatives à un même objet d'étude peuvent alors être collectées, et mises sous la forme d'une série d'événements composant un même phénomène, en vue d'être analysées. Cette modélisation générique permet de mettre en forme les données relatives à différents types de phénomènes sous un même format, permettant leur utilisation dans un même environnement d'analyse.

Un phénomène peut être décrit comme un ensemble d'objets spatio-temporels semblables. L'étude des *dynamiques* d'un phénomène revient alors à analyser la distribution spatio-temporelle des événements le composant afin d'y identifier des structures. Ces dynamiques peuvent être relatives au temps linéaire, ou relatives à une échelle temporelle cyclique. Nous cherchons à identifier :

- des pics d'intensité d'un phénomène, localisés dans le temps et l'espace ;
- des tendances dans la variation d'intensité du phénomène sur un intervalle de temps linéaire, localisées dans différentes zones de l'espace ;
- l'évolution spatiale d'un phénomène au cours du temps ;
- des récurrences cycliques, localisées dans l'espace, dans la variation d'intensité du phénomène.

Ces dynamiques sont observables par des motifs d'intérêt, que nous nommons *patterns spatio-temporels*, présents dans la distribution spatio-temporelle des événements. Ceux-ci peuvent être :

- des *clusters* d'événements spatio-temporels, relatifs au temps linéaire ;
Ces clusters permettent d'identifier des pics d'intensité localisés dans l'espace ou dans le temps linéaire. L'analyse de la proportion d'événements récents et d'événements anciens selon l'espace au moyen de ces clusters permet également d'identifier des tendances localisées dans la variation d'intensité du *phénomène*.

- des *clusters* d'événements spatio-temporels, relatifs à une échelle temporelle cyclique ;
Ces clusters permettent d'identifier des pics d'intensité localisés dans l'espace, et dans certains intervalles temporels relatifs à une échelle temporelle cyclique définie.
- des *propagations spatiales* d'événements au cours du temps, permettant d'identifier l'évolution spatiale d'un phénomène au cours du temps, notamment sa diffusion ;
- des *réapparitions cycliques localisées* d'événements ;
Ces structures peuvent retranscrire une évolution périodique de l'intensité d'un phénomène, circonscrite à une zone spatiale, ou présentant une fréquence différente selon la localisation du phénomène.

En fonction du phénomène étudié et de la problématique dans laquelle il s'insère, les dynamiques recherchées et les patterns à identifier diffèrent. De plus, l'émergence constante de nouveaux jeux de données permettant de décrire les phénomènes spatio-temporels peut amener un utilisateur à faire face à des données pour lesquelles il ne sait pas sur quelles dynamiques focaliser son analyse. Ces deux critères rendent difficile l'élaboration d'une analyse générique des phénomènes spatio-temporels selon une approche hypothético-déductive. Proposer une approche d'analyse applicable à différents cas d'étude nécessite d'offrir à l'utilisateur la possibilité d'analyser un phénomène, sans être "familier" ce dernier. Pour cela, nous adoptons une approche exploratoire.

Le principe est d'effectuer l'exploration d'un jeu de données inconnu, de manière visuelle et interactive, sans formuler d'hypothèse a priori sur la structure des données. L'analyste est amené à identifier des patterns, des tendances, ou des relations insoupçonnées dans les données, permettant de générer des hypothèses sur les dynamiques du phénomène étudié [Tukey, 1977, Andrienko and Andrienko, 2006].

L'analyse exploratoire est basée sur l'utilisation d'environnement de visualisation, permettant d'explorer visuellement un jeu de données au moyen de représentations graphiques et interactives. Or, si des travaux existent proposant de définir quelles sont les méthodes de visualisation les plus adéquates en fonction des données étudiées et de l'objectif de recherche, la plupart des environnements de visualisation proposant une analyse exploratoire des données spatio-temporelles sont développés de manière *ad hoc*, pour répondre à une problématique particulière [Kaddouri et al., 2014].

Une solution permettant d'offrir une analyse exploratoire des dynamiques spatio-temporelles, indépendante du thème traité, est alors l'élaboration d'un environnement de visualisation :

- utilisant notre proposition de modélisation générique du changement dans l'espace, permettant d'analyser des phénomènes de différentes natures au sein d'un même environnement ;
- permettant une recherche visuelle des différents patterns spatio-temporels cités précédemment, au sein de séries d'événements.

Les différents patterns que nous cherchons à analyser au sein des données étant définis, il nous est possible d'identifier quels outils graphiques et interactifs sont en mesure de permettre une analyse exploratoire de chacune des dynamiques spatio-temporelles citées, au sein de notre environnement.

Un état de l'art des approches de visualisation des données temporelles et spatiales est ainsi nécessaire. Cet état de l'art a pour objectif d'identifier les outils et techniques les plus adéquats pour révéler des relations entre les composantes temporelle et spatiale des séries d'événements spatio-temporelles, et ainsi permettre une analyse exploratoire des dynamiques des phénomènes.

Cet état de l'art a également pour objectif d'identifier les limites rencontrées par les environnements de visualisation actuels, dans l'analyse des dynamiques spatio-temporelles citées précédemment. L'état de l'art doit alors permettre d'identifier des méthodes de représentation graphique, pouvant être des sources d'inspiration pour dépasser ces limites.

Plusieurs domaines de recherche sont susceptibles de proposer de tels outils. On peut retrouver parmi eux :

- la géovisualisation (geovisualization en anglais, contraction de *geographic visualization*), qui renvoie à un ensemble d'outils et de techniques de représentation cartographique interactive permettant la visualisation de données spatiales ;
- la *DataViz* (contraction de *data visualization*), qui renvoie à un ensemble d'outils et de techniques permettant de représenter des données ou information sous la forme d'objets visuels, dans le but d'améliorer leur compréhension par un utilisateur ;
- l'*InfoViz* (contraction d'*info visualization*), qui renvoie à un ensemble d'outils et de techniques de représentation visuelle interactive permettant la visualisation de données abstraites, telles que des données numériques, textuelles ou géographiques.

Cependant si l'arrivée de l'informatique a permis d'introduire de nombreuses formes de représentation des données temporelles et spatiales, les représentations graphiques portant sur ces deux dimensions sont anciennes. Certaines de ces représentations permettent, comme le montre l'exemple de l'analyse des cas de choléra par John Snow, une analyse exploratoire de données spatiales.

Aussi, si les outils informatiques sont prépondérants dans les environnements d'analyse exploratoire et permettent une plus grande liberté dans l'exploration des données, les représentations graphiques et cartographiques classiques peuvent être une source d'inspiration permettant de proposer de nouveaux outils, informatiques, de représentation et d'exploration des données spatio-temporelles.

Deuxième partie

La visualisation des phénomènes
spatio-temporels

Introduction

L'analyse exploratoire de patterns dans les séries d'événements spatio-temporels nécessite l'utilisation d'environnements de visualisation permettant une représentation visuelle de la distribution spatio-temporelle de séries d'événements [Andrienko and Andrienko, 2006]. Ces environnements doivent également offrir des procédés de représentation graphique et d'interaction permettant d'explorer les séries événementielles, selon différents points de vue, afin d'identifier les patterns.

Il est ainsi nécessaire de s'interroger :

- d'une part, sur les approches de représentation graphique permettant de représenter les séries d'événements, en intégrant leurs différentes composantes dimensionnelles ;
- d'autre part, sur les catégories de représentation les plus adéquates pour permettre une identification visuelle de structures dans la distribution spatio-temporelle d'une série d'événements, et sur les différents outils d'interaction nécessaire à cette identification.

Nous effectuons dans un premier chapitre 3 un état de l'art des différentes méthodes de représentations graphiques de l'information spatio-temporelles, pouvant être appliquées à la représentation de séries d'événements. Notre définition du phénomène comme un regroupement d'événements, traitant de la même problématique et souvent proches sur le plan thématique, implique que nous devons particulièrement nous intéresser à la représentation graphique des composantes temporelles et spatiales des événements. Nous nous intéressons dans ce chapitre à des approches originaires de différents domaines de recherche, principalement axées sur l'intégration du temps dans les représentations cartographiques.

Nous présentons des approches issues de la geovisualisation, l'émergence de nouvelles technologies en informatique ayant permis l'intégration de la dimension temporelle, de différentes manières, dans les représentations cartographiques [Kaddouri et al., 2014]. Cependant, le développement de méthodes visant à représenter visuellement les phénomènes spatio-temporels est un ancien sujet d'étude [Kaddouri et al., 2014]. Ainsi, nous nous intéressons également à des méthodes plus anciennes de représentation du temps dans les cartes. Ce legs méthodologique peut alors servir d'inspiration pour l'élaboration d'environnements de visualisation intégrant les dimensions temporelle et spatiale.

Cet état de l'art doit nous permettre d'établir une liste de méthodes et de procédés nous permettant de représenter visuellement des événements spatio-temporels. Cependant, effectuer une *visualisation* des données, dans le cadre d'une analyse exploratoire, ne se limite pas à représenter graphiquement une série de données, ou à renvoyer visuellement le résultat d'une analyse. Faire de la *visualisation* ne consiste pas à transmettre au sein d'une représentation graphique, fût-elle interactive, le résultat d'un processus cognitif, mais à *assister* ce processus par la représentation graphique pour la production d'un résultat. La *visualisation* des données n'est pas un but, mais un moyen mis au service de l'analyse exploratoire [Spence, 2001], permettant le raisonnement de l'analyste au travers de la représentation graphique [Card et al., 1999].

Dans un second chapitre 4, nous effectuons ainsi un second état de l'art, centré sur les approches graphiques et les outils d'interaction permettant, au sein d'environnements de visualisation intégrant les dimensions temporelle et spatiale, d'explorer visuellement les données dans le but d'identifier des patterns spatio-temporels. Nous nous intéressons dans cet état de l'art à des approches traitant de données géographiques, mais également à d'autres procédés graphiques interactifs traitant de données numériques ou abstraites. Nous présentons ainsi des approches issues de la géovisualisation, mais aussi issues de la *Data Visualization* et de l'*Info Visualization*. Nous nous intéressons également à des méthodes plus anciennes de représentation graphique de la donnée, pouvant être des sources d'inspiration pour le développement d'outils d'analyse exploratoire des données spatio-temporelles. Ce second état de l'art doit nous permettre d'effectuer une synthèse des différents procédés pouvant être intégrés dans un environnement de visualisation, afin de rendre possible une analyse exploratoire des patterns spatio-temporels dans les séries d'événements.

Chapitre 3

Représentation graphique d'événements spatio-temporels

Sommaire

3.1	L'intégration du temps et de l'espace au sein de la même représentation graphique	84
3.1.1	Au travers d'une représentation non cartographique	84
3.1.2	L'intégration du temps dans les représentations cartographiques	86
3.2	L'intégration du temps et de l'espace à travers la structure de l'environnement de représentation	92
3.2.1	Les dispositifs <i>small-multiples</i>	92
3.2.2	L'animation	95
3.3	Les environnements de géovisualisation à multi-fenêtrage	98
3.3.1	Les caractéristiques de la géovisualisation	98
3.3.2	Le multi-fenêtrage pour représenter les données spatio-temporelles	102
3.3.3	Les environnements de géovisualisation à multi-fenêtrage pour la recherche de patterns spatio-temporels	104

L'analyse exploratoire de données spatio-temporelles nécessite de mettre en place des représentations visuelles et interactives de ces données. Avant de s'intéresser aux outils d'interaction permettant d'explorer les jeux de données spatio-temporelles dans le but d'y déceler des structures, des relations entre leurs composantes temporelle et spatiale, il faut s'interroger sur les moyens permettant de représenter visuellement ces deux composantes.

Il faut donc dans un premier temps s'intéresser aux méthodes permettant de représenter graphiquement la dimension spatiale et la dimension temporelle, et définir comment ces approches intègrent les différentes composantes des séries d'événements. Dans ce chapitre, nous présentons un état de l'art de ces différentes approches, et nous montrons lesquelles sont les plus adaptées à l'analyse des données événementielles.

3.1 L'intégration du temps et de l'espace au sein de la même représentation graphique

La représentation des données spatio-temporelles peut s'effectuer au travers d'approches reliant dans une même représentation graphique les dimensions temporelle et spatiale. Ces approches sont majoritairement basées sur une représentation cartographique intégrant le temps, mais d'autres disciplines tentent de représenter les interactions espace-temps par d'autres approches [Kaddouri et al., 2014].

3.1.1 Au travers d'une représentation non cartographique

De nombreux travaux en *visual analytics* ont pour thématique la représentation des dynamiques spatio-temporelle [Scharl and Tochtermann, 2009, Elmqvist et al., 2010], mais le caractère spatial des informations à analyser n'est pas toujours considéré comme le point d'entrée privilégié pour visualiser l'information. Ainsi, à la carte peuvent se substituer d'autres dispositifs graphiques [Kaddouri et al., 2014].

On retrouve également des exemples de représentations non cartographiques de données possédant une composante temporelle et une composante spatiale. *L'horaire de train* de Marey [Marey, 1885] est considéré comme un des exemples les plus représentatifs des méthodes de visualisation non cartographiques des phénomènes spatio-temporels [Tufte, 1991]. Marey représente dans ce graphique des temporalités (durée, vitesse) par de l'espace, tout en représentant la composante spatiale, réduite à un espace à une dimension, par un autre axe spatial (Figure 3.1).

Le graphique se lit de la façon suivante :

- Verticalement, l'espace entre Paris et Lyon, entrecoupée de plusieurs sections séparées par des gares intermédiaires.
- Horizontalement, le temps selon une échelle temporelle cyclique de 24 heures.
- Obliquement, les trains, de Paris vers Lyon (oblique vers la droite et vers le bas) et de Lyon vers Paris (oblique vers la droite, de bas en haut).
- Chaque ligne tracée sur le diagramme correspond à un voyage en train.
- Chaque section oblique d'un tracé correspond à un voyage sur l'axe Paris-Lyon, vers Lyon si la section est oblique à droite vers le bas, vers Paris si la section est oblique à droite vers le haut.
- Les points de début et fin d'une section oblique correspondent à la gare et à l'heure, respectivement du début et de la fin du voyage.
- L'angle qu'effectue chaque section oblique avec l'horizontale indique la vitesse de chaque train.

- Chaque section horizontale d'un tracé représente une escale en gare, dont la durée correspond à la longueur de la section.

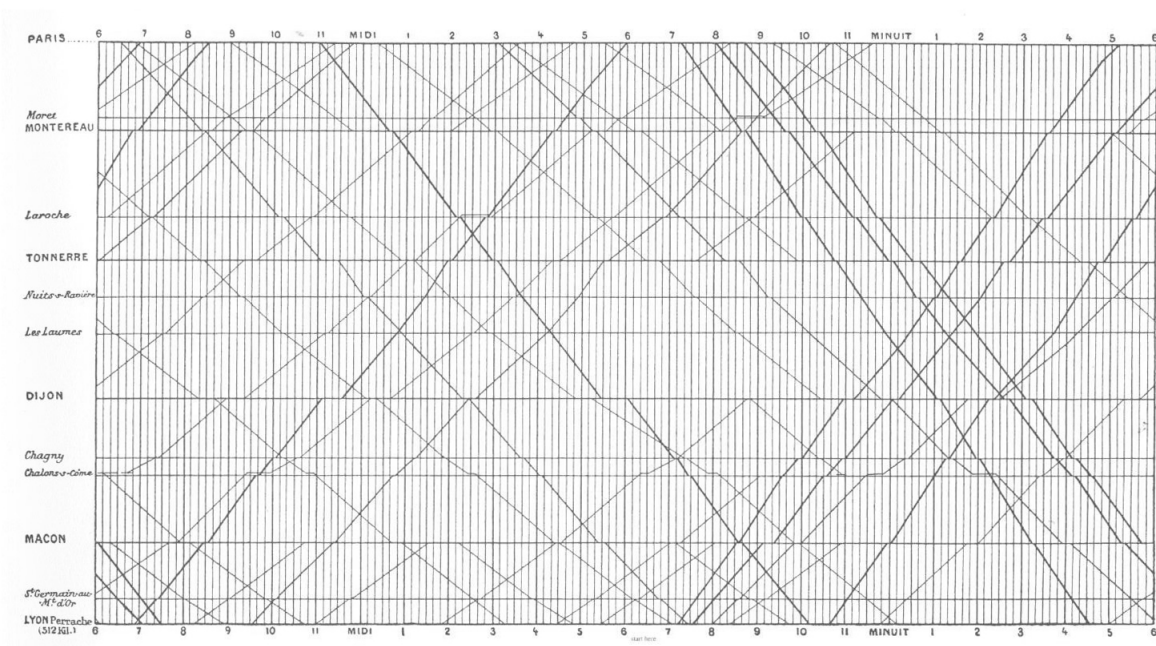


FIGURE 3.1 – L'horaire de train de Marey [Marey, 1885] [Tufté, 1991]

Chaque voyage en train peut être considéré comme un événement, limité dans l'espace et le temps :

- Sa composante spatiale, consistant en un trajet entre deux villes, est représentée par l'axe des ordonnées.
- Sa composante temporelle (les horaires de départ, d'arrivée, ainsi que la durée du trajet) est représentée par l'axe des abscisses.

La représentation de Marey illustre un concept que Tufté appellera la règle du $1+1=3$ [Kadouri et al., 2014]. En représentant, au travers des deux axes, deux données brutes de chaque événement (sa composante et sa composante temporelle), ce diagramme permet la visualisation d'informations supplémentaires. En représentant la distance et le temps sur son graphique, il permet, entre autres, d'identifier une troisième variable, la vitesse.

Ce diagramme permet ainsi de visualiser un nombre important d'informations :

- la durée des escales ;
- le nombre d'escales ;
- les horaires de chaque train ;
- la durée de chaque voyage ;
- la vitesse de chaque train ;
- le nombre de gares sur l'axe Paris-Lyon ;
- l'influence du nombre d'escales sur la durée totale d'un voyage ;
- la densité des départs et arrivées sur une durée de 24h.

3.1.2 L'intégration du temps dans les représentations cartographiques

La composante spatiale des données peut également être visualisée au travers d'une représentation cartographique. La composante temporelle des données peut alors être traitée comme une variable à cartographier.

3.1.2.1 La représentation du temps par la symbologie

L'utilisation des règles de sémiologie graphique de Bertin. Les travaux de Jacques Bertin sur la sémiologie graphique [Bertin, 1973] ont permis de définir un ensemble de règles d'utilisation des variables visuelles pour la représentation de variables thématiques, en fonction de la nature des variables, et de l'implantation géométrique des objets géographiques représentés (Figure 3.2).

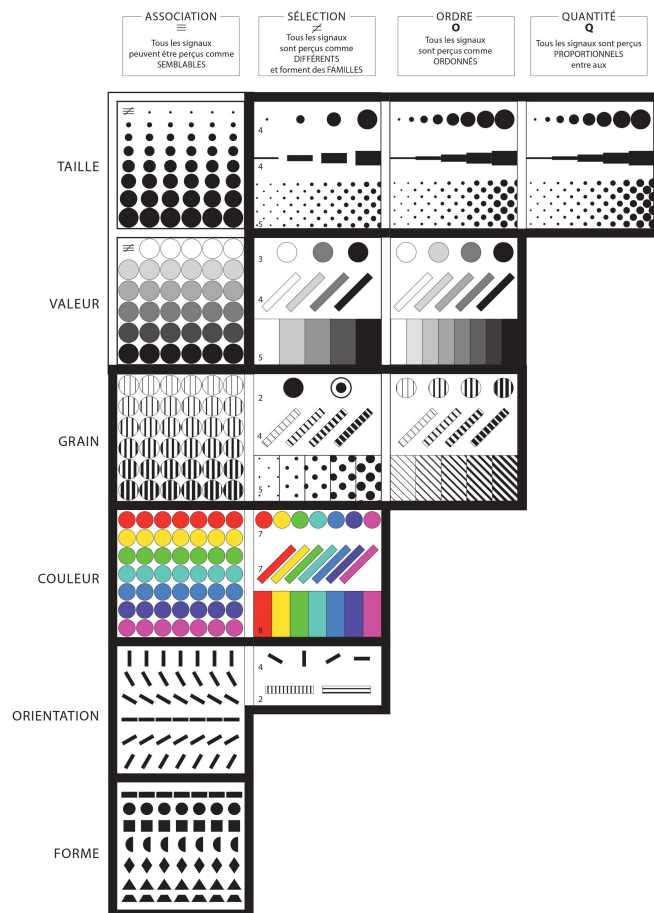


FIGURE 3.2 – Sémiologie graphique et variables visuelles proposées par Bertin [Bertin, 1973]

L'objectif de Bertin est de proposer des procédés visuels permettant de faciliter la lecture et la compréhension de l'information au travers des cartes en faisant apparaître des relations d'association, de ressemblance, de différenciation, d'ordre, de proportionnalité entre les données rattachées aux objets géographiques [Bertin, 1973, Kaddouri et al., 2014].

Les propositions de Bertin portent également sur la représentation des données temporelles [Bertin, 1973]. Les variables visuelles sont alors utilisées pour représenter la composante temporelle des objets représentés.

La caractéristique temporelle cartographiée peut alors être différente selon les données utilisées et l'objectif de la carte. Afin de visualiser l'évolution d'objets pérennes, Bertin propose

l'utilisation de la variable visuelle de la valeur pour représenter le taux de changement d'un attribut entre deux dates, ou le moment rattaché aux différentes positions spatiales d'un objet, considérés comme des variables ordonnées (Figure 3.3).

L'utilisation de la valeur permet de visualiser l'ordre chronologique des différentes positions spatiales de l'objet. L'utilisation de la transparence permet de pallier à une superposition des différentes traces. Cette approche présente cependant des limites de lecture lorsque le nombre d'éléments devient important [Davoine et al., 2012b].

Afin de mieux visualiser le mouvement d'un objet ou son changement de position, Bertin [Bertin, 1973] préconise également l'utilisation de cartes de flux, utilisant des symboles *flèches* pour le mouvement.

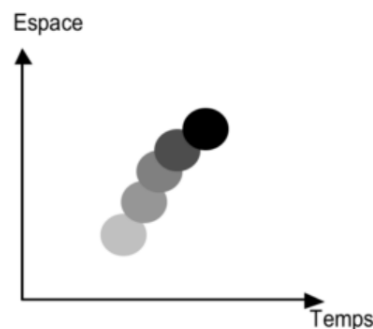


FIGURE 3.3 – Représentation de la trace d'un objet selon les règles préconisées par Bertin [Bertin, 1973]

Dans le cas d'événements limités dans le temps, les variables visuelles peuvent être utilisées pour représenter les temporalités de l'événement. Ainsi la valeur peut être utilisée pour représenter la durée ou le moment d'apparition d'un événement, considérée comme des variables ordonnées.

L'utilisation de plusieurs variables visuelles pour représenter le temps. L'application des règles de la sémiologie graphique à la représentation du temps reste ambiguë, car elle dépend de la manière de qualifier les composantes temporelles en tant que variables à cartographier [Kaddouri et al., 2014]. La cartographie du moment d'apparition des événements ne présentera pas les mêmes caractéristiques selon que l'on considère le moment d'apparition comme caractérisant l'objet en lui-même (variable qualitative utilisant la variable visuelle de la couleur), ou comme une variable établissant une relation d'ordre entre différents événements (variable quantitative utilisant la variable visuelle de la valeur).

L'utilisation d'échelles temporelles cycliques empêche également d'établir une relation d'ordre entre les moments d'apparition de différents événements. D'autres variables visuelles peuvent alors être utilisées pour représenter un moment d'apparition relatif à une échelle cyclique. Par exemple :

- la couleur, pour représenter durant quelle saison apparaît un événement (variation de couleurs *bleu, vert, jaune, rouge* pour représenter les saisons *hiver, printemps, été, automne*) ;
- des symboles, pour représenter si un événement apparaît en été (*soleil*) ou en hiver (*flocon*) ;
- l'orientation, pour représenter à quelle heure apparaît un événement au cours d'une journée (une analogie peut être effectuée entre l'orientation de l'objet graphique sur la carte et l'orientation d'une aiguille sur une horloge).

Certains auteurs analysent ainsi l'adaptation des variables visuelles de Bertin [Bertin, 1973] dans la représentation des temporalités dans les représentations cartographiques [Köbben and Yaman, 1996, Dukaczewski, 2005, Kaddouri et al., 2014] (Figure 3.4).

Temporalités Variables cartographiques	Moment	Durée	Fréquence	Chronologie	Changement
Taille		+++	++	++	+++
Valeur		+++	++	++	+++
Grain / Trame	+++	+++	++	++	+++
Couleur	+		(+)	(+)	
Orientation	+++		++	++	
Forme	+++		++	++	

+++ Très adapté ++ Adapté + Peu adapté

FIGURE 3.4 – Pertinence des variables visuelles dans des cartes dynamiques [Köbben and Yaman, 1996, Dukaczewski, 2005, Kaddouri et al., 2014]

La Figure 3.5 [Davoine et al., 2012b] illustre une approche utilisant deux variables visuelles pour représenter la composante temporelle des données. La carte présente la chronologie d'une série de coulées de lave survenue sur l'île de la Réunion entre 1972 et 2010. Afin de visualiser l'ordre d'apparition des différentes coulées de lave dans le temps, les auteurs ont choisi de considérer le moment d'apparition de chaque coulée comme une variable quantitative et d'utiliser la variable visuelle de la valeur.

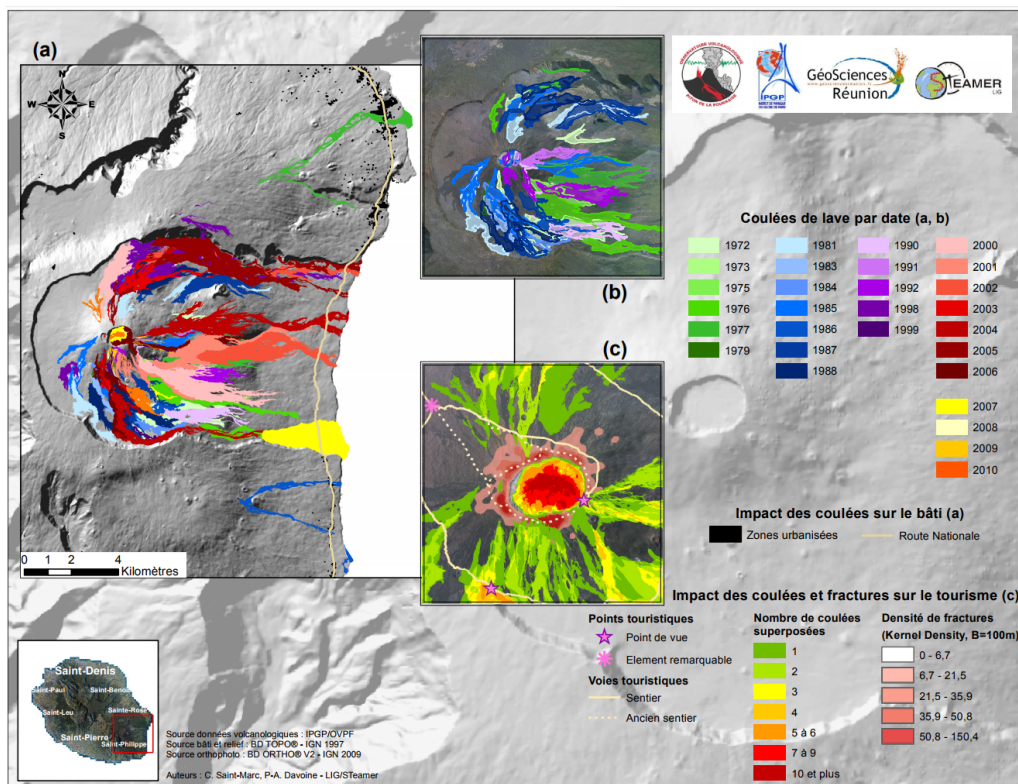


FIGURE 3.5 – Évaluation de l'aléa volcanique du Piton de la Fournaise. L'impact des coulées de lave [Cécile Saint-Marc, 2012].

Cependant, l'apparition de ces coulées de lave s'effectue sur une période longue (38 ans), et les coulées de lave sont représentées par année. Le nombre de modalités de la variable visuelle valeur à représenter devient alors trop important pour proposer une carte lisible. Les auteurs choisissent de regrouper les coulées de lave en cinq séquences de quatre à sept ans, et de considérer l'époque de chaque séquence comme une variable qualitative. Les auteurs associent alors une couleur à chaque séquence de coulées de lave, puis utilisent la valeur pour représenter l'ordre d'apparition des coulées à l'intérieur de chaque séquence [Davoine et al., 2012b].

L'extension de la sémiologie graphique de Vasilev. Vasilev (cité par Cauvin [Cauvin et al., 2008]) propose une extension de la sémiologie graphique (Figure 3.6), en tenant compte de la géométrie de l'objet et de différentes caractéristiques temporelles à cartographier.

Celles-ci comprennent :

- le moment d'apparition d'un événement ;
- la durée (soit la durée d'un événement, soit l'espacement chronologique entre deux événements successifs) ;
- les composantes de ce que Vasilev nomme le *temps structurant*, comme la fréquence, ou le temps comme distance, qui consiste en la quantification d'une distance par un temps de parcours.

Vasilev préconise ainsi, pour la représentation du moment d'apparition des événements, l'utilisation d'informations textuelles. Ces propositions, apportant notamment une utilisation massive des labels, impliquent un niveau de lecture plus détaillé que celui suggéré par Bertin [Kaddouri et al., 2014]. Ce niveau détaillé de lecture rend plus difficile l'identification de patterns spatiaux ou spatio-temporels [Kaddouri et al., 2014].

Catégorie	Dimensions géométriques		
	Point	Ligne	Surface
Moment Dates des événements	● juin 2007 Ouverture de la gare TGV Champagne-Ardennes	— juin 2007 Mise en service de la ligne TGV Est	▲ avril 2006 Crues du Danube
Durée Continuité des événements	● Gare non desservie du 25 au 29 décembre 2006 9h45 ● stations du tram ligne A, avec 9h55 ● heures de passage 10h05 ●	Trajet du tram (arrêts et liaisons) 9h55 9h45 — 10h05	■ surface d'un parking : occupation du jeudi au samedi jeudi vendredi samedi
Temps structurant Fréquence Temps standard	Jours d'ouverture d'un hyper-marché ● lundi ● mardi ● mercredi Fuseau horaire central UTC	fréquences de lignes ferroviaires — une fois par jour — deux fois par jour — toutes les heures	
Temps comme distance Intervalle de temps Direction et/ou distance temporelles	Position temporelle des lieux à partir de la place centrale 12 min 15 min	Distance-temps en voiture Place centrale 12 min 15 min Gare	Espace géographique de référence Espace fonctionnel des transports ferroviaires
Espace comme horloge	Est : lever du soleil Ouest : coucher du soleil		Fuseaux horaires

FIGURE 3.6 – Sémiologie graphique pour la représentation du temps dans les cartes selon Vasilev [Cauvin et al., 2008]

La sémiologie dynamique. L'apparition de l'animation dans les représentations cartographiques a permis de reconsidérer la sémiologie graphique, l'animation permettant de définir de nouvelles variables visuelles pour la représentation des temporalités [Kaddouri et al., 2014], comme le mouvement ou le clignotement [Green, 1998], ou la date et la durée d'apparition d'objets graphiques [MacEachren, 2004].

L'analyse de l'adaptation des variables visuelles de Bertin [Bertin, 1973] dans la représentation des temporalités dans les représentations cartographiques [Kaddouri et al., 2014, Köbben and Yaman, 1996, Dukaczewski, 2005] (Figure 3.4), a montré que ces variables étaient peu adaptées à la représentation de certaines temporalités comme la synchronisation de l'apparition d'événements, la période de retour, et la fréquence [Kaddouri et al., 2014]. Cette analyse conforte les propositions de Dibiase [DiBiase et al., 1992], MacEachren [MacEachren, 1994, MacEachren, 2004] et de Green [Green, 1998] sur l'utilisation de variables visuelles dynamiques pour représenter les temporalités [Kaddouri et al., 2014].

Cependant, dans le cadre d'une analyse exploratoire où nous cherchons à représenter, et à explorer visuellement la distribution spatio-temporelle de données brutes, des temporalités comme la synchronisation, la période de retour ou la fréquence ne sont pas celles que la représentation graphique doit permettre de montrer. Représenter visuellement que deux événements sont synchronisés implique d'avoir connaissance de cette synchronisation *a priori*, ce qui va à l'inverse de l'approche exploratoire.

L'accent doit être placé sur la représentation de la localisation des événements dans le temps, soit leur moment d'apparition, voire leur durée, temporalités pouvant être représentées par les variables préconisées par Bertin. À partir de la représentation de ces temporalités, d'autres caractéristiques temporelles des données peuvent être analysées, illustrant la règle du $1+1 = 3$ énoncée par Tufte [Kaddouri et al., 2014].

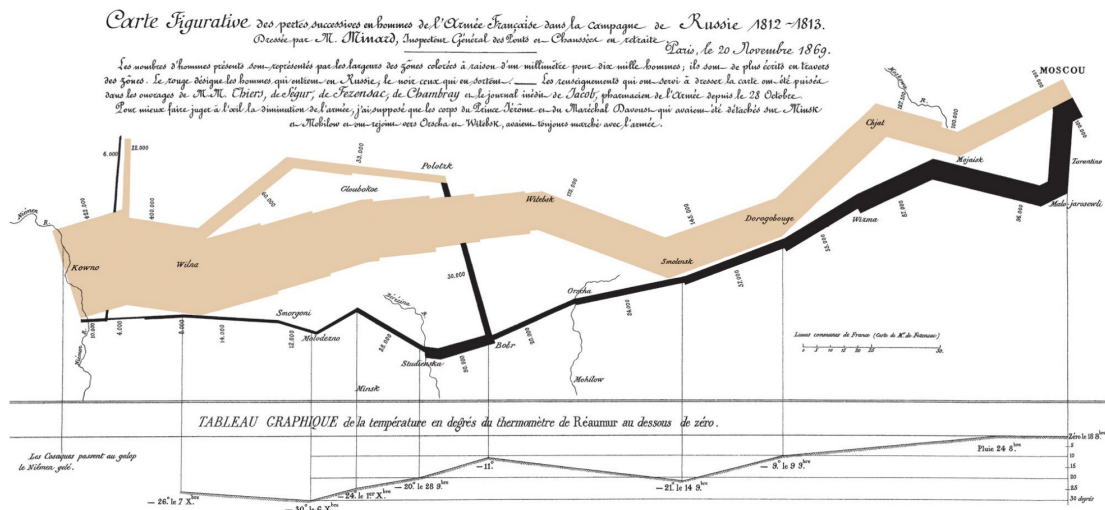
3.1.2.2 La représentation du temps par une variable spatiale : la *Time-Geography*

Pour parer à l'ambiguïté dans la représentation de la dimension temporelle dans un environnement cartographique, une solution consiste à utiliser une dimension spatiale pour représenter le temps. Cette approche prend donc la forme d'une représentation en trois dimensions, représentant l'espace sur un plan horizontal, et le temps sur un axe vertical.

Cette proposition s'inscrit dans le courant de la *Time-Geography* d'Hägerstrand [Hägerstrand, 1970], et a été initialement utilisée pour traiter des problématiques de déplacements et de mobilité. Les trajectoires d'individus peuvent alors être représentées dans l'espace sur un plan horizontal, et dans le temps sur l'axe vertical. Cette solution a été implémentée sous le nom de *cube spatio-temporel*.

La Figure 3.7 présente un exemple de la représentation d'un flux de personne au travers d'un cube spatio-temporel : la carte de la retraite de Russie de Charles Joseph Minard [Minard, 1980, Marey, 1878] (Repère *A* de la Figure 3.7), représentant la progression des troupes françaises lors de la campagne de Russie en 1812 est projetée dans un cube spatio-temporel (Repère *B* de la Figure 3.7).

Les règles de la sémiologie graphique peuvent être utilisées de concert avec le cube spatio-temporel, pour renforcer la représentation des caractéristiques temporelles des objets représentés. Elles peuvent également être utilisées pour représenter l'évolution de variables thématiques associées à des objets spatiaux pérennes évoluant dans le temps.



A: Carte de Minard

B: Projection de la carte de Minard dans un cube spatio-temporel

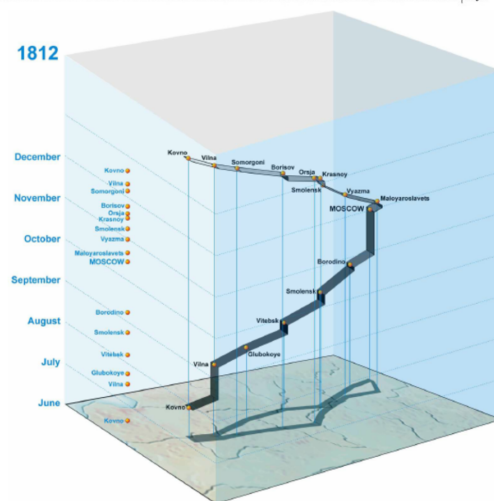


FIGURE 3.7 – Le temps représenté par une variable spatiale. Projection de la carte de Minard dans le cube spatio-temporel [Minard, 1980, Marey, 1878, Kraak, 2003b].

S’inspirant des travaux de la *Time-Geography*, Andrienko [Andrienko et al., 2003b] et Gatalsky [Gatalsky et al., 2004] proposent une adaptation du cube spatio-temporel à des problématiques basées sur des données événementielles. Dans la Figure 3.8 [Gatalsky et al., 2004], une série d’événements sismiques sont représentés par des points disposés dans un cube spatio-temporel. Le lieu d’apparition des événements est ainsi représenté par la position des objets graphiques sur le plan horizontal, tandis que le moment d’apparition est représenté par la position des objets sur l’axe vertical.

Les variables visuelles proposées par la sémiologie graphique peuvent également être appliquées à la représentation de la composante thématique des événements. Dans la Figure 3.8, la taille des objets graphiques représente la magnitude de chaque séisme.

Le cube spatio-temporel apparaît ainsi comme une approche permettant de modéliser les *hyperobjets* proposés par Galton [Galton, 2004] dans son approche de modélisation 4D du changement dans l’espace. L’axe vertical représentant le temps, un point sur cet axe représente un moment, une distance y représente un intervalle de temps. Un événement étendu dans l’espace et dans le temps peut alors être représenté dans le cube spatio-temporel par un objet en trois dimensions. Sa projection sur le plan horizontal représente sa composante spatiale, et sa projection sur l’axe vertical correspond à sa composante temporelle.

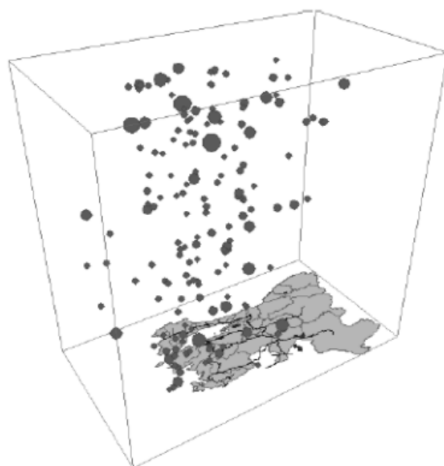


FIGURE 3.8 – Cube spatio-temporel pour la représentation des événements [Gatalsky et al., 2004]

Cette proposition ouvre ainsi de nouvelles opportunités pour la représentation des dynamiques spatio-temporelles, mais sa difficulté de lecture réserve son usage à des spécialistes [Kaddouri et al., 2014].

3.2 L'intégration du temps et de l'espace à travers la structure de l'environnement de représentation

Le temps peut ainsi être intégré aux représentations cartographiques en le considérant, soit comme variable à cartographier, soit comme variable à représenter par une variable spatiale supplémentaire aux deux dimensions de la carte, au sein d'un environnement en trois dimensions.

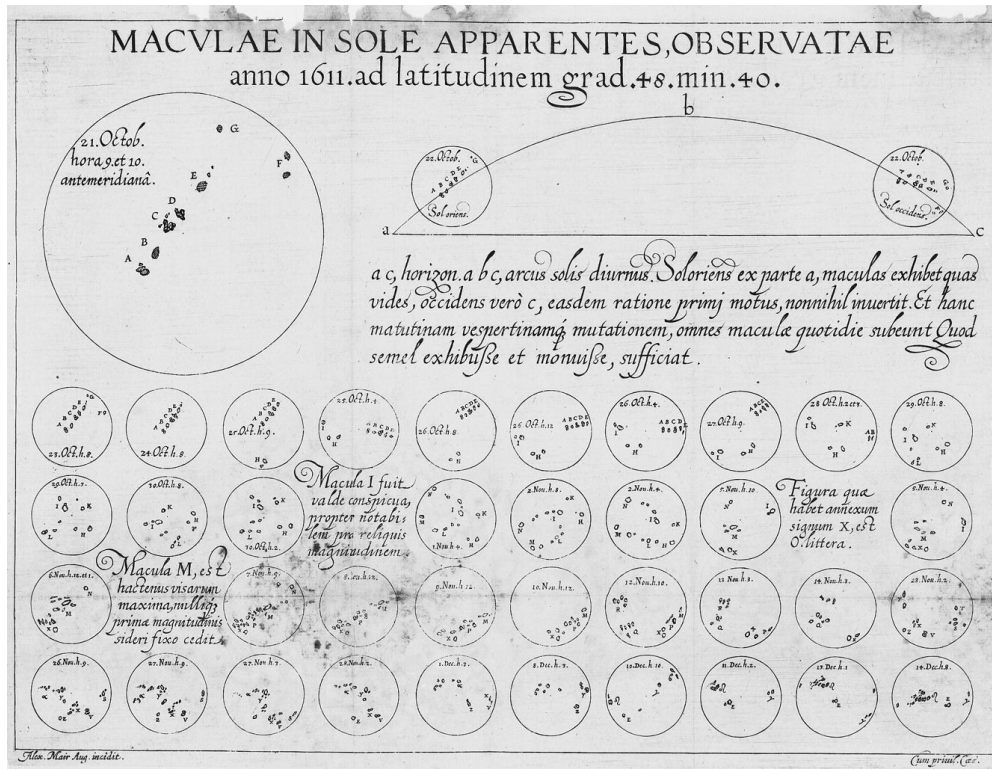
Le temps peut cependant être traité comme une variable orthogonale à la carte, qu'il convient de représenter à travers la structure de l'environnement de représentations. Soit par la multiplication de supports, soit par l'ajout de variables autres que graphiques. Ces environnements peuvent être construits autour d'une représentation cartographique, ou bien se baser sur d'autres supports graphiques, combinés ou non à une représentation cartographique.

3.2.1 Les dispositifs *small-multiples*

3.2.1.1 Le temps comme composante projetée sur la carte : la collection de cartes

Une seconde approche proposée par Bertin dans la sémiologie graphique [Bertin, 1973, Kaddouri et al., 2014] consiste à considérer le temps comme une composante orthogonale à la carte. L'état de l'environnement géographique est représenté dans une carte, pour une primitive temporelle (instant ou intervalle de temps), au sein d'une collection de cartes. Le changement est ainsi représenté en produisant plusieurs cartes décrivant l'état de l'espace pour différentes primitives temporelles, que celles-ci soient des instants ou des intervalles de temps.

Cette approche s'inspire des travaux d'observation des taches solaires de Christoph Scheiner effectués dans le premier quart du *XVIIe* siècle [Kaddouri et al., 2014] (Figure 3.9). Scheiner présente, côte à côte, une série de représentations de la surface du Soleil, à différents instants. La succession de représentation de la surface du Soleil permet de visualiser l'évolution de la distribution des tâches solaires au cours du temps. Ce format visuel sera plus tard baptisé *small-multiples* par Tufte [Kaddouri et al., 2014].



Source gallica.bnf.fr / Observatoire de Paris

FIGURE 3.9 – Observation des taches solaires par Christoph Scheiner

Dans le cas, majoritaire, où la primitive temporelle est un instant, la collection de cartes est à rapprocher d’une collection de *snapshots* décrivant l’espace. La Figure 3.10 montre une série de cartes, parue en 2012 dans le *New York Times* [The New York Times, 2012] utilisant le principe des *small-multiples*. Ces cartes sont extraites d’une série de cartes représentant l’évolution des États américains touchés par la sécheresse au XXe siècle.

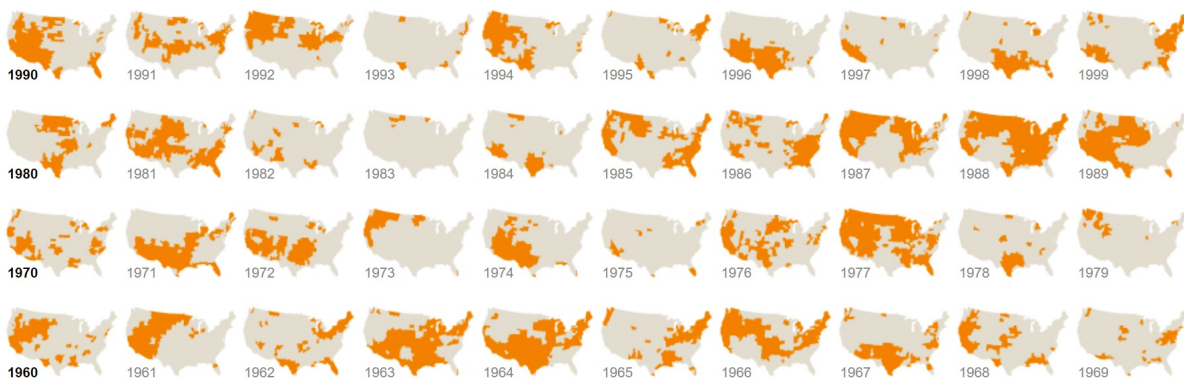


FIGURE 3.10 – États américains touchés par la sécheresse de 1960 à 1999 [The New York Times, 2012]

La représentation de données au travers de *small-multiples* relatifs à une primitive temporelle ponctuelle est adaptée à la représentation de l’évolution d’entités spatiales pérennes, mais ne permet pas la représentation de données événementielles.

En effet, conformément à la définition des entités spatio-temporelles occurrentes par Grenon et Smith [Grenon and Smith, 2004], celles-ci doivent être représentées dans leur intégralité. Hors les représentations spatiales de *small-multiples* relatifs à une primitive ponctuelle représente des

états instantanés de l'espace.

Une représentation utilisant des *small-multiples* relatifs à des intervalles temporels peut cependant permettre une représentation de la distribution spatiale d'événements apparaissant durant différents intervalles de temps.

Cette représentation permet une vue d'ensemble des impacts du changement, mais présente également une difficulté de lecture dans le cas d'un trop grand nombre de cartes [Kaddouri et al., 2014].

3.2.1.2 L'espace comme composante projetée : la collection de diagrammes temporels

L'approche *small-multiples* peut être décrite comme une multiplication de supports de même nature, représentant l'état d'une même dimension, pour plusieurs modalités, fixées, d'une dimension particulière. La collection de cartes consiste en une multiplication de représentations cartographiques, représentant l'état de l'espace, à différents instants ou intervalles temporels, fixés pour chaque carte.

L'approche *small-multiples* peut alors être mise en oeuvre, non plus en représentant l'espace pour différentes modalités du temps, mais en représentant le temps pour différentes modalités de l'espace. Le principe est alors de représenter par une série de représentations graphiques du temps, comme des diagrammes temporels, l'historique de différentes régions de l'espace, correspondant à des localisations ponctuelles ou à des zones étendues.

La Figure 3.11 présente un exemple d'une telle approche. *EventViewer* [Beard et al., 2008] est un environnement développé dans le but de visualiser et d'explorer les dimensions spatiale, temporelle et thématique dans les données de capteurs. Dans l'exemple de la Figure 3.11, chaque capteur est associé à une localisation spatiale, et enregistre l'apparition d'épisodes de basse pression atmosphérique ou de fort vent. L'historique des événements, enregistré pour chaque localisation spatiale, est représenté par un diagramme temporel.

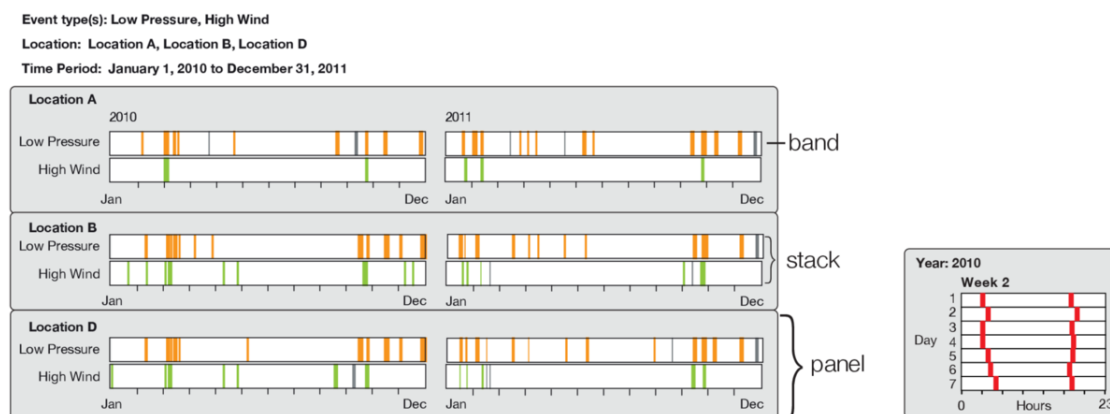


FIGURE 3.11 – Copie d'écran d'*EventViewer* [Beard et al., 2008]

3.2.1.3 L'introduction de collections de diagrammes temporels dans la carte

Une alternative à la collection de diagrammes temporels, mais qui peut s'inscrire dans la logique des *small-multiples* consiste à intégrer la représentation du temps dans différentes régions de l'espace à l'intérieur de la représentation cartographique.

L'approche consiste toujours à représenter l'évolution temporelle d'une variable ou l'apparition d'événements dans le temps, pour différentes régions de l'espace, au travers d'une collection de diagrammes temporels. Cependant, les diagrammes sont ici intégrés directement dans la carte, à la manière de symboles rattachés à des objets spatiaux. La carte devient alors un cartogramme [Bertin, 1973].

La Figure 3.12 [Andrienko and Andrienko, 2004] montre l'évolution temporelle du nombre de vols d'automobiles dans les différents États des *États-Unis d'Amérique*. Pour chaque État, un diagramme temporel représente l'évolution du nombre de vols. Cependant, à l'inverse de l'exemple de la Figure 3.11, chaque diagramme est placé sur la carte, superposé à la région spatiale correspondant aux données représentées dans le diagramme. Cette représentation permet ainsi une meilleure prise en compte de la composante spatiale des données.

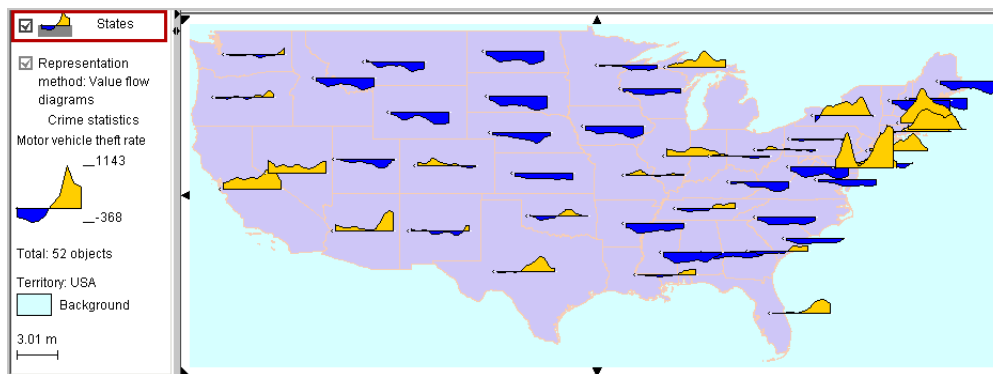


FIGURE 3.12 – Copie d'écran de *Value Flow Map* [Andrienko and Andrienko, 2004]

Cet exemple montre le cas d'une représentation de l'évolution temporelle d'objets spatiaux pérennes. Cependant, une représentation graphique similaire peut être reproduite afin d'être appliquée aux données événementielles. Pour une série de régions spatiales, la distribution temporelle des événements apparaissant dans chaque région peut être représentée au travers de diagrammes temporels projetés sur la carte.

3.2.2 L'animation

L'animation est considérée comme une des solutions des plus récentes pour représenter les évolutions d'objets dans l'espace en fonction du temps [Antoni et al., 2004, Kaddouri et al., 2014].

Le procédé consiste à afficher, de façon successive et rapide, une série de cartes représentant l'état de l'espace pour différentes primitives temporelles, en représentant ou non le passage entre les différents instants de manière brute ou fluide. Dans le cas de primitives temporelles ponctuelles, l'animation peut être vue comme un affichage successif et rapide de *snapshots* de l'espace. Le temps est ainsi intégré aux représentations cartographiques par l'ajout d'une variable autre que graphique : le temps de l'animation.

Cette succession permet de donner l'illusion d'un mouvement, d'un déplacement ou d'un changement [Cauvin et al., 2008, Kaddouri et al., 2014], et l'impression d'une dynamique de l'espace. Kraak [Kraak, 2003a] souligne que *les animations sont très utiles pour clarifier des tendances et des processus, aussi bien dans l'explication que pour donner un aperçu des relations spatiales*.

3.2.2.1 La représentation continue du temps par le temps de l'animation

L'animation peut être utilisée afin de représenter l'écoulement du temps de manière continue au travers de l'écoulement de l'animation, en représentant le temps réel par le temps de l'animation. On parle alors d'animation de série chronologique ou d'*animation par image* [Kaddouri et al., 2014].

Le temps réel est représenté selon des pas de temps relatifs à la granularité temporelle utilisée pour modéliser l'information. Le pas de temps de l'animation est utilisé pour représenter le pas de temps réel proportionnellement à la granularité utilisée. On peut par exemple représenter un pas de temps réel d'une année par une seconde d'animation.

La série chronologique permet de visualiser les changements de localisation, de géométrie, d'attributs d'objets spatiaux au cours du temps [Kaddouri et al., 2014]. L'animation d'une série chronologique permet également de représenter l'apparition d'événements limités dans l'espace et le temps. La Figure 3.13 suivante montre un arrêt sur image de l'animation d'une série chronologique présentant l'apparition des principaux séismes apparus de 2001 à 2015.

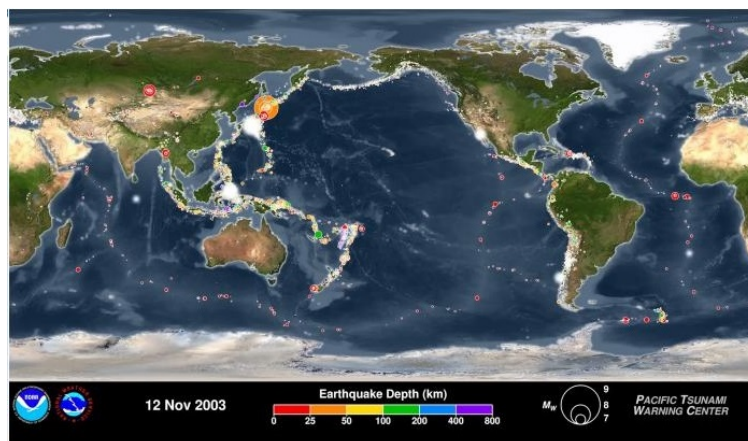


FIGURE 3.13 – Time Lapse des principaux tremblements de Terre sur les cent dernières années, et sur les quinze dernières années [NOAA, 2016]

La représentation du temps par un pas de temps d'animation régulier et constant permet de visualiser le moment d'apparition, la durée, et le rythme des phénomènes spatio-temporels. Il est en revanche plus difficile de voir un ordre de successions de différents événements, surtout si beaucoup d'événements ont lieu durant un temps très court.

3.2.2.2 La représentation séquentielle des dynamiques spatio-temporelles par l'animation

L'animation peut également être utilisée afin de représenter les phénomènes spatio-temporels de manière séquentielle. Le principe n'est plus de représenter le temps réel par le pas de temps de l'animation, mais de représenter différentes séquences des phénomènes spatio-temporels par différentes séquences de l'animation.

Le temps de l'animation n'est donc plus proportionnel au temps réel. Il ne représente plus le temps réel, mais l'enchaînement temporel des séquences des phénomènes. L'animation doit ici attirer l'attention de l'utilisateur sur chaque changement, déplacement, ou apparition d'événements induite par le phénomène, et sur l'enchaînement de ces différentes séquences. Ceci peut être effectué de deux manières :

- La première approche consiste à attirer l'attention de l'utilisateur en animant sur la carte les entités spatio-temporelles impliquées par la séquence en question, par le clignotement d'un objet ou son déplacement sur la carte. On nomme ainsi ce type d'animation *animation par objet* [Kaddouri et al., 2014].
- La seconde approche consiste à animer la carte elle-même. Ce type d'animation inclut les glissements de vue et des effets de zoom pour attirer l'attention de l'utilisateur sur la zone où quelque chose a lieu [Kaddouri et al., 2014].

Cette représentation séquentielle permet une visualisation des différentes phases des phénomènes spatio-temporels et de leur enchaînement dans le temps, mais nécessite de connaître les différentes phases des phénomènes représentés. Ainsi, elle ne peut être utilisée dans le cadre d'une analyse exploratoire.

3.2.2.3 Animation et *small-multiples*

L'animation peut être considérée comme une extension du dispositif *small-multiples* puisqu'il s'agit d'une représentation successive de différents états temporels de l'espace. L'animation permet une meilleure perception du passage d'un état temporel à un autre, que dans une représentation en *small-multiples* [Kaddouri et al., 2014]. Il devient cependant difficile d'effectuer une comparaison entre deux états temporels dans le cadre d'une animation, surtout si ces deux états sont temporellement éloignés [Kaddouri et al., 2014]. Il peut être intéressant de considérer ces deux solutions comme complémentaires, mais peu d'environnements proposent cette approche mixte [Kaddouri et al., 2014].

Une étude de l'Université de Lausanne [Boyandin et al., 2012] cherchant à comparer l'impact de ces deux représentations lors d'une analyse de données montre que la majorité du public favorise les animations, mais obtient de meilleurs résultats au moyen des dispositifs *small-multiples*. L'animation se révèle plus efficace pour souligner et localiser des rythmes, des densités de changements, mais peu efficace pour comparer des états distants dans le temps [Kaddouri et al., 2014].

3.2.2.4 Animation et interaction

Si l'apparition de l'animation dans les représentations cartographiques et la formalisation de nouvelles variables visuelles ont ouvert de nouvelles opportunités pour la prise en compte de la dimension temporelle dans la carte, l'utilisation de l'animation ne peut être dissociée du principe de l'interactivité *utilisateur-interface* pour être efficace [Kaddouri et al., 2014].

L'animation à elle seule fournit à l'utilisateur une impression du mouvement, de la densité de changement apporté par un phénomène, de l'enchaînement temporel des différentes modifications, mais sans retenir les différentes phases des phénomènes. L'utilisateur ne garde pas en mémoire les différentes images, uniquement le mouvement [Ormeling, 1995, Kaddouri et al., 2014]. L'animation seule reste peu efficace pour mémoriser les différentes phases des phénomènes spatio-temporels. Pour permettre une réelle prise en compte des différentes séquences des phénomènes au cours du temps, il convient d'offrir à l'utilisateur un moyen d'interagir avec l'animation.

La carte animée doit permettre à l'utilisateur d'interagir avec elle, et de modifier les critères de l'animation, par exemple modifier son rythme de déroulement, revenir sur un passage, etc. Pour cela, l'animation doit s'inscrire dans un environnement de visualisation, permettant à l'utilisateur d'interagir avec les données représentées [Kaddouri et al., 2014]. Au-delà cette possibilité d'interaction, il faut que l'utilisateur puisse contextualiser dans le temps ce qu'il voit sur

la carte animée, afin d'avoir une vision d'ensemble de l'agencement des modifications affectant l'environnement spatial dans le temps.

3.3 Les environnements de géovisualisation à multi-fenêtrage

L'émergence de nouvelles techniques de représentation de la donnée, combinées à de nouveaux outils et méthodes d'interactivité, a permis le développement d'environnements de visualisation dynamiques adaptés aux données géographiques. Ces outils se définissent comme des applications de géovisualisation, et regroupent des outils plus ou moins complexes, allant de cartes animées et interactives, à des environnements proposant une interface multidimensionnelle à plusieurs fenêtres, utilisant différentes formes de représentation de l'information combinées à un niveau élevé de possibilité d'interaction.

Ces environnements sont basés sur l'utilisation de nouvelles formes de cartographie, telles que les cartes interactives, multimédias, voire animées, et sur l'association de différents supports de visualisation au sein d'une même interface. Ces supports pouvant être graphiques, cartographiques, ou basés sur d'autres médias (son, etc.). Associées à la possibilité donnée à l'utilisateur d'effectuer des requêtes, visuelles ou non, sur les données représentées, ces applications permettent à l'utilisateur de pousser l'analyse et l'exploration des données géographiques sur le plan spatial, temporel, ou informationnel, et ce à différentes échelles [Kaddouri et al., 2014]. Les principaux outils utilisés par les applications de géovisualisation sont donc l'interactivité, le multimédia, la multiplication de support de visualisation, aussi appelée multifenêtrage, et la synchronisation.

3.3.1 Les caractéristiques de la géovisualisation

L'interactivité et le multimédia. La première caractéristique des environnements de visualisation est l'intégration d'outils d'interaction permettant à l'utilisateur de modifier les paramètres de la représentation graphique [Kaddouri et al., 2014]. L'ajout de l'interactivité dans les applications de géovisualisation engendre la transformation de ces interfaces en interfaces multimédias, permettant de relier les données représentées à des bases de données, des liens hypertextes, des documents, permettant à l'utilisateur d'interroger les données par des requêtes visuelles [Kaddouri et al., 2014]. Ces outils d'interaction permettent ainsi de modifier la représentation des données à l'intérieur de l'interface de visualisation, et font ainsi partie intégrante du processus d'exploration des données au travers des environnements de géovisualisation.

Trois types d'outils d'interaction se distinguent parmi les possibilités offertes par ces environnements [Kaddouri et al., 2014] :

- les outils de navigation, permettant de modifier les paramètres de la représentation graphique, de changer l'emprise et l'échelle de représentation des données ;
- les outils de positionnement, permettant à l'utilisateur de se situer sur une carte ou sur un diagramme ;
- les outils de sélection, permettant à l'utilisateur de sélectionner et d'interroger les données.

On peut rajouter à ces trois types d'interaction les outils permettant de modifier la structure même de l'interface de représentation des données. On peut citer par exemple les outils permettant de modifier le rythme de lecture d'une carte animée, des outils permettant de passer d'une représentation en deux dimensions de l'espace à une représentation en trois dimensions, ou des outils permettant de modifier la représentation de l'espace d'une représentation cartographique à une vue satellitaire.

On peut enfin rajouter à cette classification les outils permettant d'effectuer des analyses quantitatives sur les données et de retourner les résultats sous une forme visuelle. On peut citer

les fonctions de calcul d'itinéraires dans les outils de navigation, ou les fonctions d'interpolation dans les SIG d'analyse de données spatiales.

L'utilisation de l'interactivité couplée à l'utilisation du multimédia présente donc l'intérêt de permettre une exploration d'une masse importante de données sur différents critères et selon différentes approches exploratoires. Elles offrent la possibilité de contextualiser les données en accédant à de l'information complémentaire sur les données représentées.

La multiplication de supports de représentation et la synchronisation. Une des innovations majeures des environnements de géovisualisation concerne la structure même de l'interface. L'application n'est plus pensée comme une version améliorée de la carte, mais comme une fusion de différents supports de visualisation.

À la différence de l'approche *small-multiples*, la multiplication de supports ne concerne pas des représentations de même nature, présentant l'état d'une dimension pour plusieurs modalités, fixées, d'une dimension particulière. Les différents supports concernent ici des représentations graphiques de différentes natures, consacrées à la représentation de différentes dimensions de la donnée. Le principe n'est plus de représenter une série de cartes représentant l'état de l'espace pour différents intervalles temporels, mais d'associer une représentation graphique de l'espace avec une représentation graphique du temps. Chaque support peut être mis en place au sein d'une fenêtre dédiée de l'interface. On parle alors d'environnements à multi-fenêtrage [Kaddouri et al., 2014].

Les environnements de géovisualisation utilisant l'animation intègrent ainsi des représentations interactives du temps permettant de contrôler l'animation, mais aussi de contextualiser les différents instants de l'animation dans le temps (Figure 3.14) [Kraak et al., 1997, Peterson, 1999, Kaddouri et al., 2014]. Ces représentations du temps sont considérées comme des légendes temporelles [Kaddouri et al., 2014].

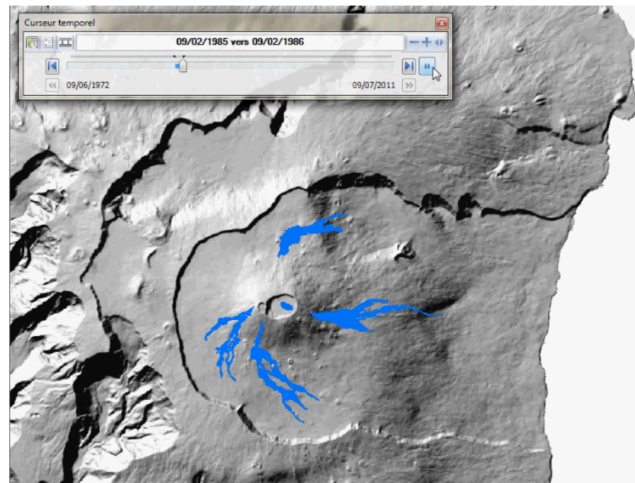


FIGURE 3.14 – Historique des coulées de lave du Piton de la Fournaise de 1972 à 2011 [Davoine, 2014]

Le lien entre la carte et ces légendes s'effectue par la synchronisation entre ces représentations. La légende temporelle sert d'outil de représentation du temps, mais également d'outil d'interaction permettant de contrôler ce qui est représenté dans la carte [Kaddouri et al., 2014]. Cette synchronisation permet à l'utilisateur de faire le lien entre ce qui est représenté dans la carte, et le temps.

Chaque fenêtre est dédiée à la représentation d'une même dimension, mais chaque dimension peut également être représentée au travers de différents supports de représentations, graphiques ou non. Un outil de géovisualisation à multifenêtrage peut ainsi combiner plusieurs outils de représentation dédiés à la dimension spatiale de l'information, et d'autres outils de représentation dédiés à la dimension temporelle de l'information. La dimension spatiale peut être représentée au travers d'une seule représentation cartographique, ou bien être représentée de manière combinée par une carte et un cartogramme. La dimension temporelle peut être représentée au travers d'une simple horloge, ou au moyen d'un ou de plusieurs diagrammes temporels.

La multiplication de supports de visualisation au sein d'une même interface, s'accompagne de la synchronisation des différents supports entre eux, chaque fenêtre étant reliée aux autres par des liens dynamiques. Chaque fenêtre offre des possibilités d'interaction, et chaque action, effectuée par l'auteur dans une des fenêtres de l'interface, a une incidence dans les autres fenêtres.

La donnée représentée dans l'interface à multi-fenêtrage est représentée sous différentes formes dans chacune des fenêtres de l'interface, et ces représentations, qu'elles soient graphiques, textuelles, cartographiques ou tabulaires sont liées entre elles. Ainsi la sélection d'une entité, au travers de la sélection de l'élément le représentant dans une des fenêtres de l'interface, entraîne la sélection des éléments représentant l'entité dans les autres fenêtres de l'interface [Kaddouri et al., 2014].

Ainsi, en fonction des objectifs d'analyse de l'environnement de géovisualisation, le format d'application à multi-fenêtrage permet de combiner au sein d'un même environnement différents supports cartographiques, graphiques ou tabulaires pour la visualisation de l'information. Ces différents supports permettent de visualiser les différentes composantes dimensionnelles de l'information, selon différents procédés, et d'effectuer une exploration visuelle des données selon plusieurs axes d'approches.

Le choix du multi-fenêtrage. Comme nous l'évoquons précédemment, si la carte reste au cœur des environnements de géovisualisation, ces derniers forment un ensemble vaste, allant d'applications basiques de représentation cartographique de la donnée, à des environnements beaucoup plus complexes, à haut degré d'interaction, permettant une analyse visuelle des données géographiques.

Cette diversité dans le nombre d'outils disponibles répond à une diversité dans les usages des applications de géovisualisation, pouvant aller du besoin de localiser, de présenter ou de synthétiser la donnée, jusqu'à des besoins d'analyse et d'exploration des données spatiales. MacEachren [MacEachren, 1994] établit quatre usages des environnements de géovisualisation [Kaddouri et al., 2014] :

- présenter la donnée et transmettre les résultats d'une étude à un public ;
- analyser une série de données ;
- synthétiser un ensemble d'observations ;
- explorer des données.

MacEachren définit dans le modèle *User-Map* [MacEachren, 1994] (Figure 3.15) la place de ces différents usages selon trois dimensions :

- le niveau d'expérience de l'utilisateur, entre expert et novice (grand public) ;
- la connaissance des données représentées, définissant si l'objectif de l'environnement est de présenter le connu ou de révéler l'inconnu ;
- le niveau d'interaction proposé à l'utilisateur, de faible à élevé.

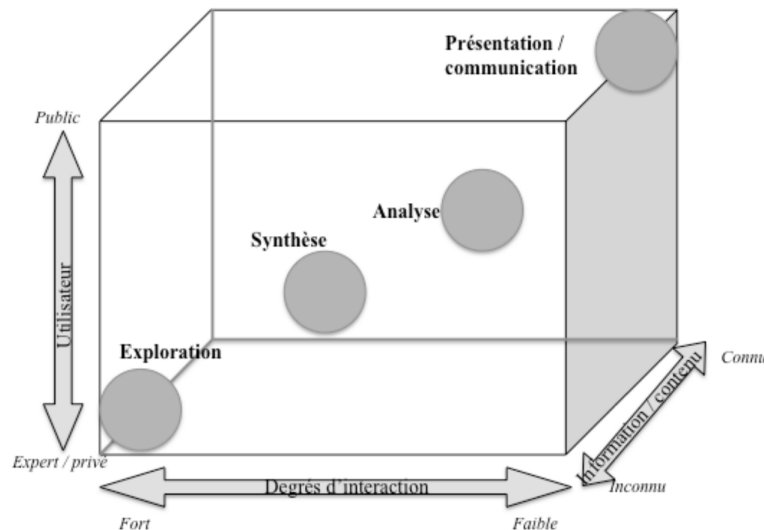


FIGURE 3.15 – Géovisualisation au prisme de 3 dimensions [MacEachren, 1994]

On peut opposer, selon leur position dans les extrêmes du cube de MacEachren :

- d’une part les environnements de géovisualisation dont l’objectif est de transmettre une information connue, peu interactifs et s’adressant à un large public ;
- d’autre part les environnements d’exploration, à haut niveau d’interactivité, s’adressant à un public spécialisé, dont l’objectif est de révéler des structures dans des données inconnues.

Un environnement de géovisualisation permettant une analyse exploratoire de données spatio-temporelles inconnues doit ainsi proposer une représentation de la donnée intégrant un haut niveau d’interactivité.

Davoine [Davoine, 2014] effectue une classification des différentes options de représentation et d’interaction offertes par les environnements de géovisualisation, en fonction de leur objectif (Figure 3.16).

Outils / actions	Services rendus	Objectifs
Zoom / pan/ survol/ Infobulles / vue d’ensemble / modification de l’angle de vue	Navigation	Synthèse
Curseur de position / carton de situation	Situation	
Clic sur la carte / Activation des éléments de la légende via des boutons radios ou menus déroulants	Sélection	
Requêtes multicritères	Interrogation	Analyse
Changement de l’apparence des cartes, de la symbologie, du fond de cartes	Modification	
Agencement multi-vues synchronisées / Requêtes visuelles / Filtrage / Switch multionglets / transition vers d’autres vues	Contextualisation / Ré-évaluation (ou re-calcul) / Ré-expression / consultation multiniveaux	Exploration

FIGURE 3.16 – Outils de l’interactivité et services rendus [Davoine, 2014]

Selon cette classification, l’utilisation de différents supports de représentation synchronisés s’impose pour permettre une exploration des données. Le multi-fenêtrage s’impose ainsi comme le principe général dans le développement d’un environnement permettant une analyse exploratoire de données spatio-temporelles.

3.3.2 Le multi-fenêtrage pour représenter les données spatio-temporelles

Le multi-fenêtrage peut être utilisé pour représenter les différentes composantes de la donnée spatio-temporelle telles qu'elles sont décrites dans le *Triad Model* [Peuquet, 1994] et le *Pyramid Model* [Mennis et al., 2000].

Davoine [Davoine et al., 2001, Davoine et al., 2004] (Repère *A* de la Figure 3.17) et Kraak [Kraak and Ormeling, 2011] (Repère *B* de la Figure 3.17) ont proposé une formalisation de la structure des environnements de géovisualisation à multi-fenêtrage, permettant de représenter au sein de fenêtres dédiées les trois composantes spatiale, temporelle et thématique de la donnée spatio-temporelle.

L'environnement de géovisualisation est divisé en trois fenêtres. La composante spatiale de l'information est retranscrite au travers d'une représentation cartographique occupant une des fenêtres de l'environnement, tandis que les composantes temporelle et thématique sont représentées dans deux autres fenêtres, occupées par des représentations graphiques ou tabulaires.

Ces différentes fenêtres sont synchronisées entre elles, et offrent chacune des possibilités d'interaction. Les actions effectuées par l'utilisateur dans une des fenêtres de l'interface, donc selon une de ses dimensions, se répercutent dans les autres fenêtres de l'interface. Ainsi, les résultats d'une requête visuelle effectuée dans une des fenêtres sont représentés sur les autres fenêtres de façon dynamique. Si une sélection d'entités est effectuée dans une des fenêtres, elle entraîne la sélection des entités correspondantes dans les autres fenêtres.

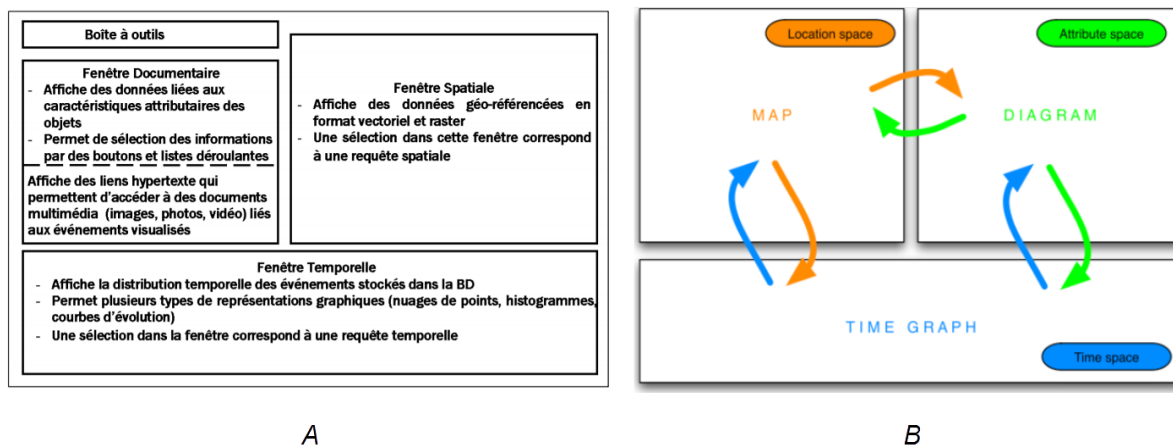


FIGURE 3.17 – Formalisme des environnements de géovisualisation [Davoine et al., 2001, Davoine et al., 2004, Kraak and Ormeling, 2011]

L'environnement de géovisualisation à multi-fenêtrage apparaît donc, non seulement comme un outil de visualisation de la donnée selon les trois dimensions de l'information spatio-temporelle, mais également comme un outil permettant d'interroger la donnée selon ces trois dimensions, au moyen de requêtes visuelles dans chacune des trois fenêtres de l'interface. Ces environnements de géovisualisation consacrés à la représentation de l'information spatio-temporelle sont parfois appelés *Systèmes d'Information Spatio-Temporels*, ou *SIST*.

Ce formalisme d'environnements de visualisation des données spatio-temporelles a été implémenté sous la forme de différents *SIST*, dont un des représentants est *GenGHIS* [Davoine et al., 2012a, Kaddouri et al., 2014], développé par l'équipe *Steamer* du *Laboratoire d'Informatique de Grenoble* (Figure 3.18).

GenGHIS est un environnement générique, permettant la création de *SIST* spécifiques à certaines problématiques métier. L'interface générée permet une visualisation des trois composantes dimensionnelles de la donnée, la production de cartes interactives et multimédias, et l'interrogation des données via des requêtes spatiales, temporelles ou informationnelles.

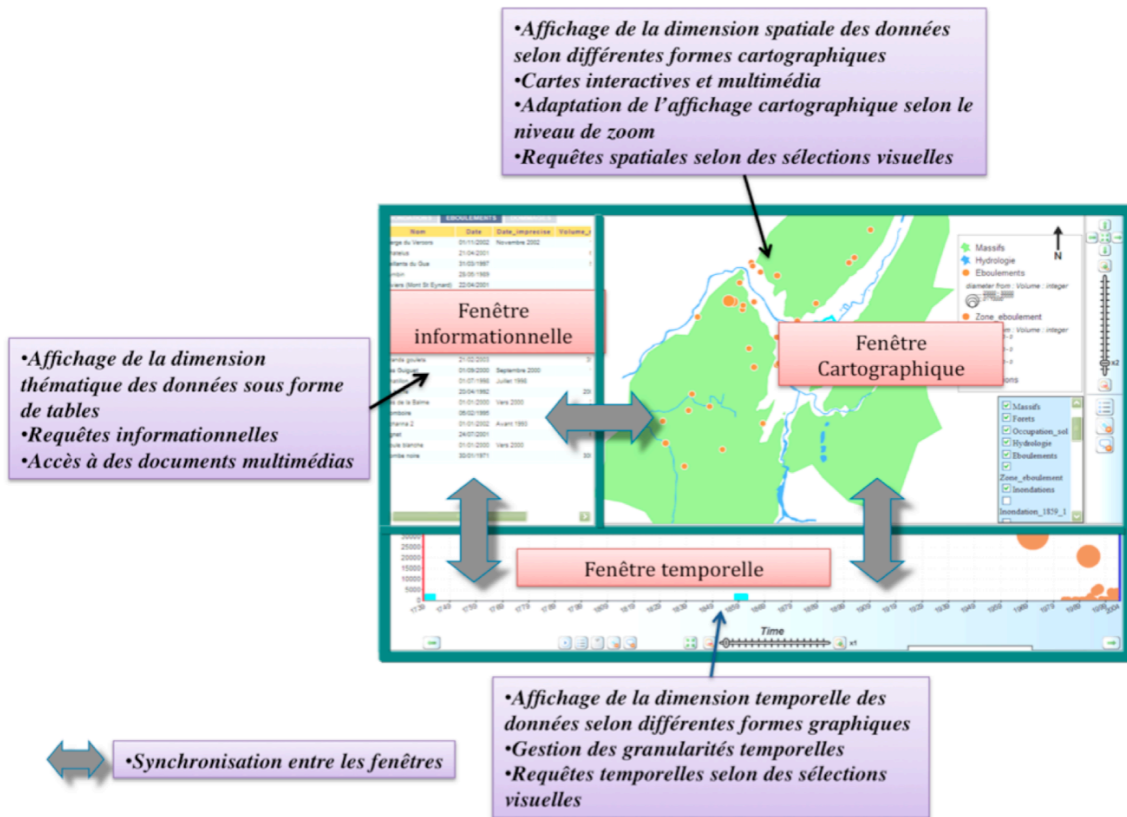


FIGURE 3.18 – Fonctionnement d'une interface de géovisualisation, l'exemple de GenGHIS [Davoine et al., 2012a, Kaddouri et al., 2014]

L'utilisation du multi-fenêtrage, par une représentation décomposée des dimensions temporelle, spatiale, et thématique, permet de représenter les différentes composantes de l'information géographique telles qu'elles sont décrites dans le *Triad Model* [Peuquet, 1994, Kaddouri et al., 2014]. Chaque fenêtre peut alors être occupée par le support de représentation le plus adapté à la composante de l'information à représenter (Figure 3.19).

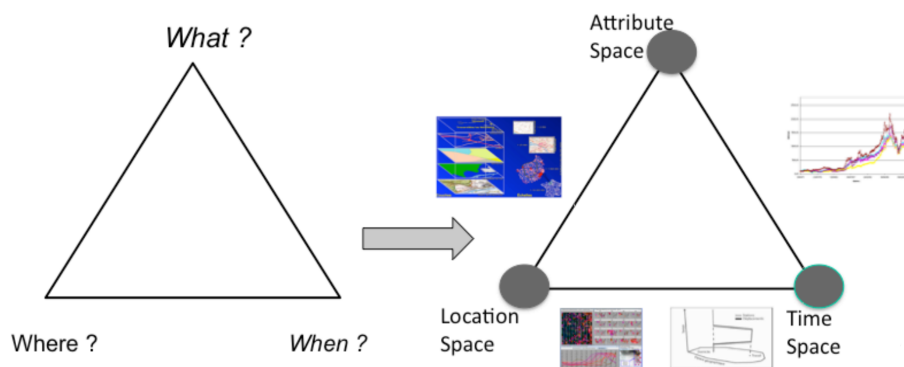


FIGURE 3.19 – De la Triade de Peuquet à la géovisualisation [Kraak and Ormeling, 2011, Kaddouri et al., 2014]

3.3.3 Les environnements de géovisualisation à multi-fenêtrage pour la recherche de patterns spatio-temporels

L'approche la plus adaptée pour l'analyse exploratoire de données spatio-temporelles consiste donc en un environnement de géovisualisation à multi-fenêtrage, à haut degré d'interaction. Cet environnement doit ainsi permettre la visualisation et l'exploration des données selon différentes composantes dimensionnelles.

Les événements que nous analysons étant rattachés à un même phénomène pour chaque cas d'analyse, les modalités de leur dimension thématique sont plus au moins fixées, aussi faut-il surtout s'intéresser aux représentations des composantes spatiale et temporelle permettant d'identifier des structures dans les données spatio-temporelles.

Il nous faut donc, dans un premier temps, définir quelles sont les représentations graphiques ou cartographiques les plus adéquates pour identifier ces patterns dans les différentes fenêtres d'un environnement à multi-fenêtrage :

- quelles sont les représentations graphiques ou cartographiques permettant de mettre en évidence une concentration d'événements dans le temps ou dans l'espace ?
- quelles sont les représentations permettant d'identifier une modification de l'intensité d'un phénomène dans le temps ou dans l'espace ?
- quelles sont les représentations graphiques permettant de repérer des récurrences cycliques dans les données ?

Cependant, si l'utilisation du multi-fenêtrage permet une visualisation combinée des composantes spatiales et temporelles, nous cherchons ici à repérer des structures dans la distribution spatio-temporelle des événements représentés, faisant intervenir des relations entre les deux composantes de la donnée. Il nous faut donc nous interroger sur les différentes techniques de représentation graphique et outils d'interaction permettant de mettre en évidence des relations entre ces deux composantes.

À l'utilisation du multi-fenêtrage, permettant de représenter les composantes spatiale et temporelle de la donnée en différentes fenêtres visuelles, peuvent alors s'ajouter différentes approches présentées dans ce chapitre. Ces approches, utilisées de concert avec le multi-fenêtrage, permettent de faire le lien entre ce qui est représenté dans les différentes fenêtres.

L'intégration d'un cube spatio-temporel, dans la fenêtre spatiale, permet de représenter la composante temporelle des événements, également représentée dans la fenêtre temporelle. L'utilisation de règles de symbologie peut permettre de représenter les éléments de la fenêtre spatiale en fonction du temps pour les relier aux entités correspondantes dans la fenêtre temporelle. Une approche *small-multiples* peut être utilisée, afin de représenter l'état de l'environnement géographique pour différents moments temporels, au moyen de plusieurs supports de représentation intégrés dans la fenêtre spatiale.

Chacune des fenêtres, temporelle ou spatiale, peut alors également servir d'outil d'interaction permettant de contrôler la représentation des événements dans les autres fenêtres de l'environnement, afin de permettre de faire le lien entre ce qui est représenté dans chaque fenêtre.

Le chapitre suivant se concentre sur les différents outils de représentation et d'interaction pouvant permettre une analyse exploratoire des patterns spatio-temporels au sein des environnements à multi-fenêtrage. Nous nous intéressons :

- d'une part aux différentes méthodes de représentation du temps et de l'espace pouvant être utilisées dans les fenêtres d'un environnement à multi-fenêtrage ;

- d'autre part aux différents agencements des techniques représentées dans ce chapitre au sein des environnements à multi-fenêtrage.

Chapitre 4

L'identification de patterns dans les environnements à multi-fenêtrage

Sommaire

4.1	Les représentations graphiques au service de la recherche de patterns temporels	109
4.1.1	Des légendes temporelles aux diagrammes des environnements de géovisualisation	109
4.1.2	<i>Timelines</i> et <i>timewheels</i> pour l'identification de patterns	114
4.1.3	La recherche de récurrences cycliques au travers des diagrammes temporels <i>mixtes</i>	123
4.1.4	Synthèse	133
4.2	La visualisation du lien entre les composantes spatiale et temporelle pour l'identification de patterns	133
4.2.1	L'analyse des données spatio-temporelles au travers du temps linéaire et du temps cyclique : les approches <i>TimeWave</i> et <i>PerSE</i>	134
4.2.2	Les limites de <i>Time Wave</i> et de <i>PerSE</i>	142
4.2.3	L'intégration du temps dans les représentations cartographiques des environnements à multi-fenêtrage	149
4.2.4	Conclusion	158

L'utilisation du multi-fenêtrage permet une visualisation des différentes composantes d'une série d'événements spatio-temporels. Pour permettre une identification de structures dans leur distribution spatio-temporelle, il est nécessaire de s'interroger :

- d'une part sur les représentations graphiques permettant d'identifier des structures dans ces composantes, au travers des différentes fenêtres d'un environnement à multi-fenêtrage ;
- d'autre part sur les approches de représentation et sur les outils d'interaction permettant d'identifier des relations entre ces différentes composantes.

L'identification de patterns spatio-temporels dans les environnements à multi-fenêtrage peut y être facilitée, si les différentes fenêtres permettent d'identifier des motifs d'intérêts issus de la projection des patterns sur un plan strictement temporel, ou spatial (tableau suivant).

L'identification d'une propagation spatiale des événements au cours du temps, selon un axe directionnel ou depuis un point dans l'espace, est impossible si l'analyse n'est pas effectuée conjointement sur les composantes temporelles et spatiales d'une série d'événements. À l'inverse, l'existence de clusters dans la distribution spatio-temporelle des événements, et la réapparition de clusters selon une période cyclique, peuvent se traduire par l'apparition de structures observables sur un plan strictement temporel ou spatial :

- Un cluster spatio-temporel se traduit par l'apparition d'une concentration d'événements dans l'espace (cluster spatial), et dans le temps (cluster temporel).
- Une récurrence cyclique spatialisée se traduit par une concentration spatiale d'événements (cluster spatial), réapparaissant de manière cyclique dans le temps (récurrence cyclique).

Pattern spatio-temporel	Pattern temporel	Pattern spatial
cluster S.T. relatif au temps linéaire	cluster temporel relatif au temps linéaire	cluster spatial
cluster S.T. relatif au temps cyclique	cluster temporel relatif au temps cyclique	cluster spatial
récurrence cyclique spatialisée	récurrence cyclique	cluster spatial

L'utilisation de représentations cartographiques en deux dimensions ou de cubes spatio-temporels, représentant l'ensemble d'une série événementielle, permet d'identifier des concentrations d'événements dans l'espace. Par contre, il faut définir quelles représentations des données temporelles sont les plus adaptées pour fournir une analyse exploratoire de concentrations temporelles ou de récurrences cycliques d'événements.

Dans ce chapitre, nous définissons dans un premier temps quelles sont les représentations du temps les plus adéquates pour identifier les structures temporelles correspondant aux différents patterns spatio-temporels que nous souhaitons analyser.

Dans un second temps, nous nous intéressons à l'insertion de ces représentations temporelles dans les environnements à multi-fenêtrage, et comment différentes méthodes de représentation des données spatio-temporelles peuvent être conjuguées au multi-fenêtrage, pour permettre une analyse exploratoire des structures spatio-temporelles dans les séries d'événements. Nous essayons ensuite de définir quelles sont les limites des environnements existants pouvant permettre une analyse exploratoire des dynamiques spatio-temporelles. Enfin, nous présentons quelques

pistes d'amélioration, susceptibles de permettre le développement d'un environnement d'analyse exploratoire palliant les défauts des environnements existants.

4.1 Les représentations graphiques au service de la recherche de patterns temporels

4.1.1 Des légendes temporelles aux diagrammes des environnements de géo-visualisation

4.1.1.1 La légende temporelle comme outil d'interaction et de contextualisation des données temporelles

Comme évoqué dans le chapitre précédent (3), les représentations du temps associées aux représentations cartographiques sont souvent considérées comme des légendes temporelles [Kadouri et al., 2014] permettant de contextualiser temporellement les éléments représentés dans la carte. Ces légendes sont parfois interactives et permettent à l'utilisateur d'interagir avec la composante temporelle projetée dans la carte.

Si l'utilisation d'une légende temporelle permet d'intégrer à une carte animée un outil d'interaction et de contextualisation temporelle, les légendes temporelles ne sont pas limitées au cadre de l'animation. Des représentations cartographiques statiques peuvent également proposer des légendes temporelles permettant de contextualiser temporellement des événements représentés sur une carte.

La Figure 4.1 représente les différentes coulées de lave dues aux éruptions du Piton de la Fournaise sur l'île de la Réunion, de 1972 à 2010. Le moment d'apparition de chaque coulée est représenté dans la carte par l'utilisation de la symbologie. Un des objectifs de cette carte est de représenter les différentes périodes d'activité du volcan représenté par une couleur différente dans la carte. Cette dernière est associée à une ligne de temps, représentant les différentes séquences temporelles où des coulées de lave ont lieu, et à un diagramme temporel représentant la valeur de deux attributs associés à chaque coulée de lave. Ces deux légendes temporelles permettent de contextualiser temporellement les couleurs représentées dans la carte.

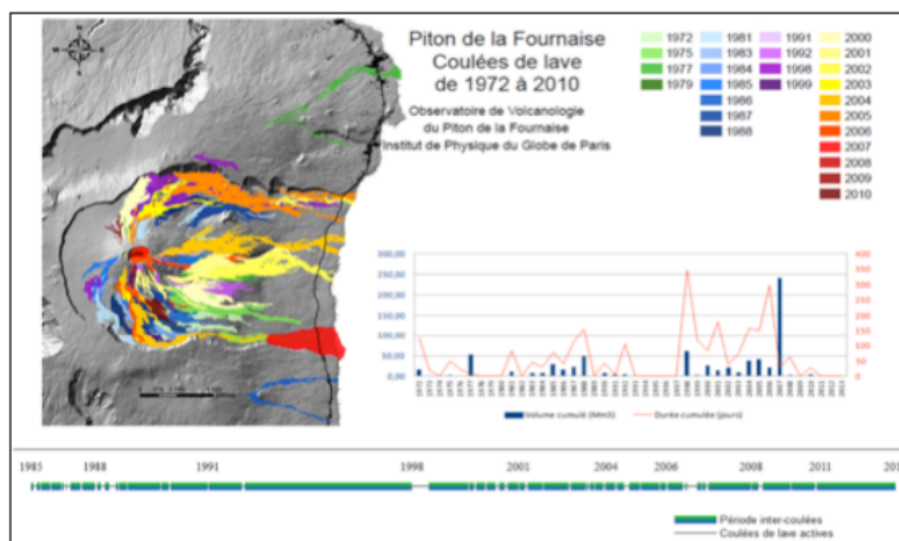


FIGURE 4.1 – Éruptions du Piton de la Fournaise [Saint-Marc et al., 2014]

Les légendes temporelles peuvent ainsi être utilisées dans des environnements utilisant une

représentation cartographique interactive, dynamique, multimédia, et par extension dans les environnements de géovisualisation à multi-fenêtrage.

4.1.1.2 Les différents types de légendes temporelles

Les légendes temporelles peuvent se présenter sous différentes formes dans les environnements de géovisualisation [Kaddouri et al., 2014]. Selon plusieurs classifications des formes de représentation du temps [Chi et al., 1998, Silva and Catarci, 2000, Andrienko et al., 2003a, Müller and Schumann, 2003, Aigner et al., 2007, Li, 2010], cette multiplicité est en partie due aux différentes manières d’appréhender la dimension temporelle, aux différentes manières de modéliser les données, et à la multiplicité des objectifs des environnements de géovisualisation.

Selon que le temps soit appréhendé de manière linéaire ou cyclique, que les données à représenter soient sous la forme d’événements ou sous la forme de séries temporelles, que l’objectif soit de communiquer une information ou d’explorer la donnée, la légende temporelle ne présentera pas les mêmes caractéristiques.

Indépendamment de la représentation du temps et du format de modélisation des données, les légendes temporelles peuvent présenter des différences sur leur structure même. Ces légendes peuvent être [Cauvin et al., 2008] :

- numériques (compteur numérique, affichage d’une date numérique) ;
- textuelles (affichage d’une date textuelle) ;
- iconiques (horloge, calendrier, affichage des saisons au moyen de symboles) ;
- graphiques (ligne de temps, diagramme temporel, barre de défilement).

Différentes approches de représentation du temps peuvent être illustrées au travers de chacune de ces formes de légende temporelle. Prenons l’exemple de la représentation du temps comme linéaire ou cyclique :

- Dans le cas d’une représentation graphique, le temps linéaire peut être représenté par des diagrammes de type *timeline* ou *ligne de temps* (Figure 4.5), le temps cyclique peut être représenté par des diagrammes de type *timewheel* ou *roue du temps* (Figure 4.7) ;
- Dans le cas d’une représentation du temps dans un format textuel, une légende indiquant le jour, le mois et l’année, soit sous la forme *3 mars 2017* représente le temps de manière linéaire. À l’inverse, une légende indiquant le jour de la semaine représente le temps selon une échelle cyclique d’une semaine.

Certaines légendes temporelles offrent des possibilités d’interaction très limitées. D’autres légendes temporelles présentent au contraire des niveaux très élevés d’interaction, permettant de modifier la forme de la représentation du temps, la granularité utilisée, l’échelle et l’emprise temporelle de la représentation, et d’effectuer, de manière graphique, des requêtes temporelles sur les données.

Les légendes temporelles peuvent être simples et limitées à une forme de représentation du temps, ou complexes et associer plusieurs supports de représentation [Edsall and Peuquet, 1997]. Certaines légendes offrent ainsi la possibilité de combiner au sein de la même interface plusieurs formes de représentation du temps. Certaines légendes temporelles proposent de représenter le temps selon plusieurs granularités temporelles, selon plusieurs échelles, ou de représenter simultanément le temps de manière cyclique et linéaire.

La multiplicité des échelles et celle des granularités temporelles de représentation sont d'ailleurs souvent liées, chaque échelle étant parfois rattachée à une certaine granularité. La modification dynamique de l'échelle de représentation dans un diagramme temporel peut alors s'accompagner d'une modification de la granularité utilisée. C'est le cas, par exemple, des diagrammes temporels linéaires offrant des possibilités de *zoom temporel* : lorsque l'emprise temporelle de la représentation diminue, la granularité temporelle utilisée devient plus fine.

Cette représentation de plusieurs échelles et de plusieurs granularités temporelles n'est pas circonscrite à l'utilisation de légendes temporelles sous forme de diagrammes. Le temps peut également être représenté grâce au croisement de représentations graphiques avec des représentations numériques, textuelles, ou iconiques : une ligne de temps couvrant une étendue temporelle d'une journée avec une granularité horaire peut être combinée avec une représentation numérique du temps utilisant la minute comme granularité temporelle.

Ces procédés permettent de combiner des représentations du temps linéaire et du temps cyclique (selon plusieurs échelles). L'association de représentations linéaires et cycliques du temps peut également s'effectuer selon des granularités temporelles différentes. Un affichage textuel peut représenter un moment temporel selon une échelle linéaire ayant pour granule (unité élémentaire) la journée, tandis qu'un affichage numérique peut représenter ce moment temporel à la minute près selon une échelle cyclique d'une journée.

La multiplication de supports de représentation, ou la possibilité de modifier ceux-ci de manière interactive permet d'associer plusieurs formes de représentation de la donnée temporelle, permettant d'étudier celle-ci selon différents points de vue (Figure 4.2).

La représentation d'une série événementielle peut ainsi donner à voir plusieurs informations selon le diagramme utilisé :

- Une représentation sous la forme d'un nuage de points permet de représenter le moment d'apparition de chaque événement, combinée à la représentation d'une valeur attributaire portée par chaque événement (Repère *A* de la Figure 4.2).
- Une représentation sous la forme d'un histogramme permet de représenter le nombre d'événements apparaissant durant des intervalles temporels réguliers (Repère *B* de la Figure 4.2).
- Une représentation sous la forme d'un calendrier permet d'observer quelles sont les séquences d'apparition et de non-apparition d'événements (Repère *C* de la Figure 4.2).

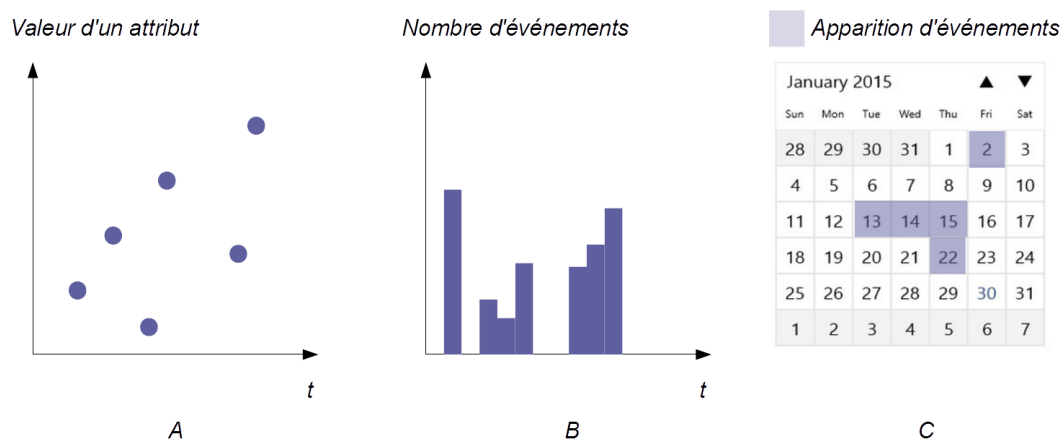


FIGURE 4.2 – Différentes formes de légendes temporelles

4.1.1.3 Le choix d'une représentation du temps par des diagrammes temporels

L'utilisation de diagrammes temporels semble être une solution largement utilisée dans les outils de géovisualisation à multi-fenêtrage pour contextualiser les données dans le temps, effectuer des requêtes de manière visuelle sur la composante temporelle des données, et explorer celle-ci afin d'en dégager des patterns. Ces diagrammes peuvent également présenter différentes formes, selon les données utilisées, la représentation du temps adoptée, et les objectifs de la visualisation :

- Certains diagrammes s'appliquent à représenter le temps de manière linéaire, d'autres à représenter le temps de manière cyclique.
- Certains diagrammes représentent des données modélisées sous la forme de séries temporelles, d'autres des données prenant la forme de séries événementielles.
- Certains diagrammes représentent les données au travers des valeurs d'une ou de plusieurs variables thématiques rattachées aux données, d'autres représentent le nombre d'événements apparaissant au cours du temps.

Plusieurs travaux ont porté sur la classification des différentes méthodes de représentation graphiques du temps, et se basent sur différents critères pour les différencier.

L'approche de Silva et Catarci [Silva and Catarci, 2000], basée sur la distinction entre temps linéaire et temps cyclique, distingue trois classes de représentations :

- les représentations linéaires basées sur les *timelines* [Harrison et al., 1994, Karam, 1994, Allen, 1995, Plaisant et al., 1996, Kumar et al., 1998] ;
- les représentations périodiques basées sur les représentations calendaires [Kincaid et al., 1985, Beard et al., 1990, Mackinlay et al., 1991, Tessler, 1993, Mackinlay et al., 1994, Palen, 1999] ;
- les représentations périodiques basées sur les *timewheels* [Keim, 1996, Keim, 1997, Carlis and Konstan, 1998, Elmasri and Lee, 1998].

L'approche de Muller [Müller and Schumann, 2003] propose une classification basée sur deux critères :

- la modélisation utilisée pour les données, qui peut être sous la forme de séries temporelles, ou sous la forme de séries événementielles ;
- le caractère statique ou dynamique de la représentation graphique.

L'approche *Geo-exploration* d'Adrienko et d'Adrienko [Adrienko et al., 2003a] classent les représentations selon deux critères :

- la forme des données représentées ;
- le type de tâches à effectuer.

On peut enfin citer l'approche d'Aigner [Aigner et al., 2007] (Figure 4.3) qui propose une classification des méthodes de représentation des données temporelles selon trois critères :

- la manière dont le temps est représenté : temps linéaire ou temps cyclique ; primitive temporelle ponctuelle ou étendue dans le temps ;
- les données représentées ;
- la forme générale prise par le diagramme.

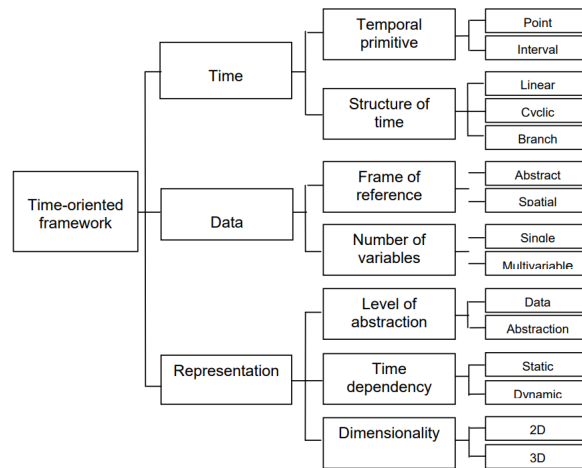


FIGURE 4.3 – Modèle détaillé de classification des représentations temporelles [Aigner et al., 2007, Li, 2010]

Dans *Visualisation of Time-Oriented Data*, Aigner [Aigner et al., 2011] introduit trois critères à prendre en compte pour la construction d’une représentation graphique des données temporelles :

- Le type de donnée représenté, ou *Data Level* : *What is presented?*
- Comment la donnée est représentée, ou *Presentation Level* : *How is it presented?*
- Les tâches à accomplir, ou *Task Level* (Figure 4.4) : *Why is it presented?*

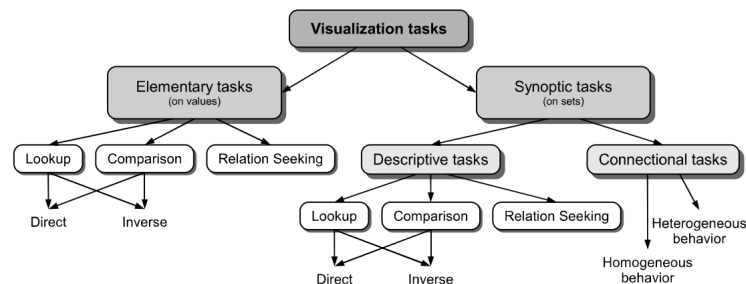


FIGURE 4.4 – Taxonomie des tâches de visualisation [Aigner et al., 2011]

Dans notre cas, les données à représenter sont sous la forme de séries événementielles. Notre objectif est d’identifier visuellement dans ces représentations graphiques les patterns temporels suivants :

- une concentration d’événements dans le temps linéaire ;
- une concentration d’événements relative à une échelle temporelle cyclique ;
- une réapparition cyclique d’événements.

Nous devons donc nous interroger sur les diagrammes temporels permettant une représentation des séries événementielles selon le temps linéaire et selon le temps cyclique, et permettant l’identification des différentes structures temporelles citées précédemment. L’usage des diagrammes temporels n’étant pas limité au domaine de la géovisualisation, d’autres formes développées en *Data Visualization* ou en *Info Visualization* sont également étudiées, et peuvent être des sources d’inspiration pour proposer des représentations graphiques permettant d’identifier des patterns temporels.

4.1.2 *Timelines* et *timewheels* pour l'identification de patterns

4.1.2.1 *Timelines* et *timewheels* dans les environnements de géovisualisation

Les *timelines*. Dans la plupart des méthodes de visualisation graphique du temps, ce dernier est considéré comme linéaire et unidimensionnel [Li, 2010].

Selon Silva et Catarci [Silva and Catarci, 2000], la plupart des représentations visuelles du temps linéaire sont basées sur la *timeline*. Dans ce diagramme, le temps est représenté selon un axe linéaire, une position graphique X_1 correspond à un instant temporel T_1 , et une longueur graphique X correspond à une durée T (Figure 4.5).

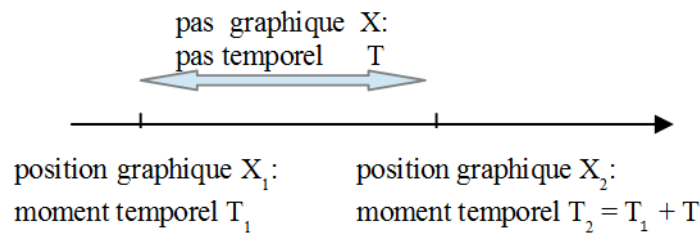


FIGURE 4.5 – Principe de la ligne de temps, ou *timeline*

La composante temporelle d'un événement est représentée dans ce diagramme par un objet graphique. La position de l'objet graphique sur l'axe du diagramme représente son moment d'apparition, et la longueur de l'objet selon ce même axe représente sa durée. Un événement ponctuel dans le temps est ainsi représenté par une entité graphique ponctuelle.

Les *timelines* peuvent être réduites à un simple diagramme unidimensionnel où des événements sont représentés le long de l'axe. Les diagrammes *timelines* peuvent également se présenter sous d'autres formes. La Figure 4.6 [Davoine, 2014] montre un exemple de *timeline* en nuage de points, permettant de représenter sur un second axe la valeur d'un attribut porté par chaque événement.

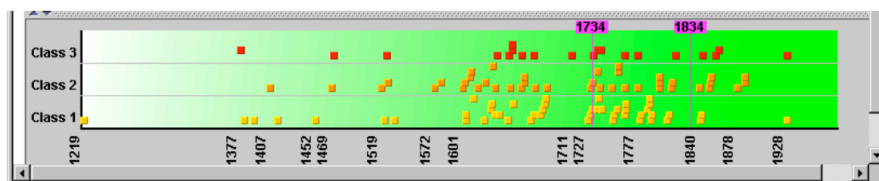


FIGURE 4.6 – Capture d'écran de l'environnement *SPHERE* [Davoine, 2014]

Les *timewheels*. Il existe également plusieurs représentations cycliques du temps, développées afin de refléter les rythmes de la nature ou de l'activité humaine [Hornsby and Egenhofer, 2002].

La *timewheel* est une forme de représentation couramment utilisée pour visualiser le temps cyclique [Edsall et al., 1997, Edsall and Peuquet, 1997]. Dans ce diagramme, le temps est représenté à travers une étendue temporelle, représentée par un cercle (Figure 4.7), que l'on considère comme récurrente dans le temps linéaire (une heure, un jour, un mois).

Le temps est considéré comme cyclique, une *timewheel* représentant le temps sur une journée ne représente pas, par défaut, une journée en particulier. Les différentes instances d'une échelle cyclique d'une journée sont projetées sur la *timewheel*. Une position graphique X_1 sur le cercle

correspond au même instant temporel T_1 dans chacune des instances de l'échelle cyclique représentée. Un angle α formé par une section du cercle correspond à un intervalle de temps T existant dans chacune des instances de l'échelle cyclique représentée.

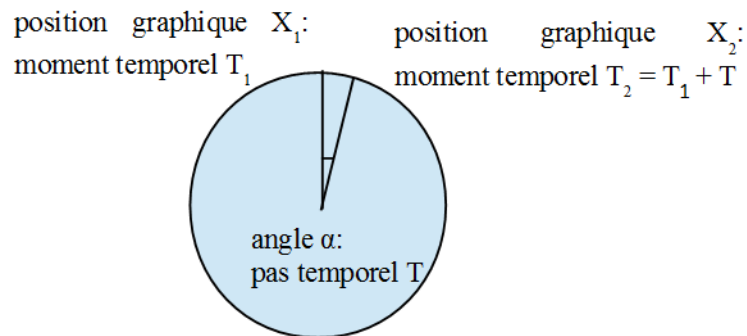


FIGURE 4.7 – Principe de la roue du temps, ou *timewheel*

La composante temporelle d'un événement est représentée dans ce diagramme par un objet graphique. La position de l'objet graphique sur le cercle du diagramme représente son moment d'apparition selon l'échelle cyclique représentée par le cercle, et l'angle formé par la section du cercle occupée par l'objet graphique représente la durée de l'événement. Un événement ponctuel dans le temps est ainsi représenté par une entité graphique ponctuelle.

La *timewheel* représentant une échelle cyclique, les événements représentés sur une *timewheel* peuvent apparaître durant différentes instances de la période cyclique représentée, mais être projetés sur un même diagramme. Une *timewheel* peut également être utilisée pour représenter par un même objet graphique un événement récurrent, et apparaissant à chaque instance de l'échelle cyclique représentée.

De la même manière que pour la *timeline*, une *timewheel* peut représenter l'existence d'événements sous la forme d'entités graphiques positionnées sur le cercle du diagramme, ou représenter une valeur attributaire rattachée à l'événement, à la manière par exemple d'un diagramme en nuage de point. L'axe radial est alors utilisé pour représenter la valeur de l'attribut correspondant.

La Figure 4.8 montre une capture d'écran de l'environnement *Silhouette Graph* [Harris, 2000].



FIGURE 4.8 – Capture d'écran de l'environnement *Silhouette Graph* [Harris, 2000]

Les données représentées sont constituées d'une série temporelle représentant l'évolution de trois variables thématiques. Celles-ci sont représentées sur trois sections concentriques d'une *timewheel*, transformant celle-ci en *triple timewheel*. L'axe radial est utilisé dans chaque section concentrique pour représenter la valeur de la variable correspondante au cours du temps.

L'association des *timelines* et des *timewheels*. Dans une approche à multi-fenêtrage, le temps est représenté par un graphique temporel, majoritairement de type *timeline* ou *timewheel* [Edsall et al., 1997, Edsall and Peuquet, 1997], selon que le temps soit appréhendé de façon linéaire ou cyclique. Certaines analyses nécessitent cependant une approche du temps à la fois linéaire et cyclique.

Plusieurs environnements de géovisualisation ont cherché à intégrer ces deux types de diagrammes au sein d'une même interface. L'application *TEMPEST* développée par Edsall et Peuquet [Edsall and Peuquet, 1997] combine une *timeline* à une *timewheel* dans le même environnement de géovisualisation. La combinaison d'une *timeline* et d'une *timewheel* peut ainsi servir à localiser un événement, à la fois de manière absolue dans le temps linéaire, et selon une ou plusieurs échelles cycliques.

La Figure 4.9 montre une capture d'écran de *TEMPEST*. Le temps est représenté avec une granularité d'un mois par une *timeline*, et par une *timewheel* selon une échelle cyclique d'une durée d'un an. L'interface de contrôle permet de modifier l'échelle cyclique représentée par la *timewheel*, pour différentes échelles calendaires.

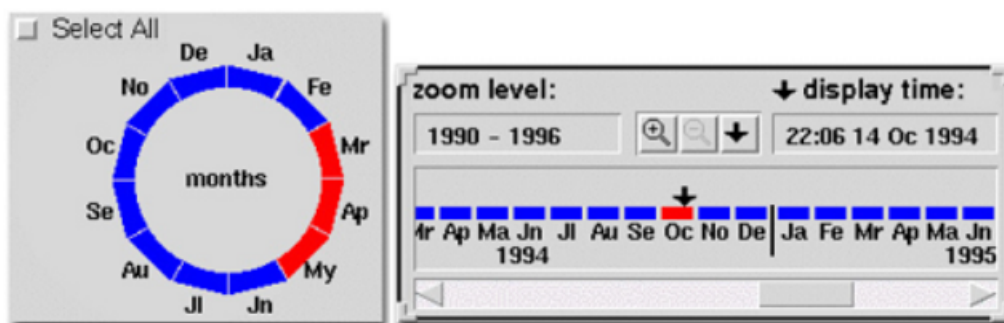


FIGURE 4.9 – Combinaison de la *timeline* et de la *timewheel* dans *TEMPEST* [Edsall and Peuquet, 1997]

La combinaison d'une *timeline* et d'une *timewheel* peut également permettre de représenter le temps selon deux granularités temporelles imbriquées :

- Le temps peut être représenté dans une *timeline* avec une granularité correspondant à une durée G .
- L'étendue temporelle G peut être représentée à une granularité plus fine de manière cyclique dans une *timewheel*. La durée de l'échelle cyclique représentée par la *timewheel* est égale à celle d'un granule utilisé dans la *timeline*.

Cette représentation permet de situer le moment d'apparition d'un événement dans le temps linéaire en utilisant la granularité la plus adaptée, tout en utilisant la *timewheel* pour situer le moment d'apparition de l'événement avec une granularité plus fine.

4.1.2.2 Identification de clusters relatifs au temps linéaire dans une *timeline*

L'identification d'une concentration temporelle d'événements selon le temps linéaire, dans une *timeline* prenant la forme d'une simple ligne de temps, revient à identifier des concentrations

d'objets graphiques le long de l'axe temporel (Figure 4.10). L'identification d'un cluster peut également s'effectuer sur un diagramme linéaire en nuage de point, par l'identification d'une concentration d'entités graphiques dans une tranche verticale du diagramme temporel.

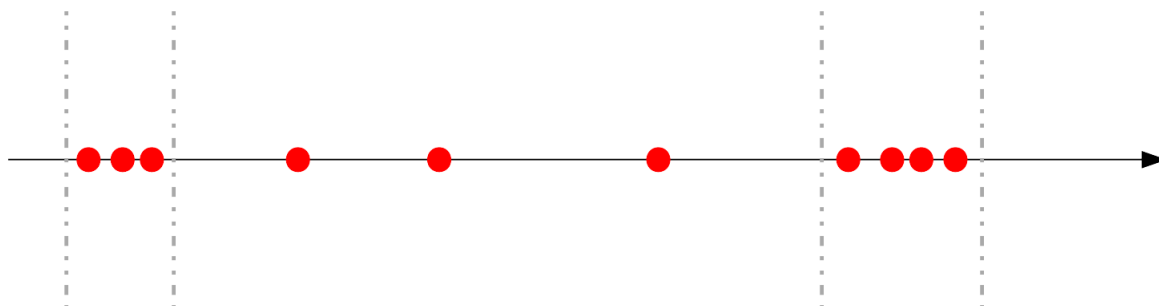


FIGURE 4.10 – Identification d'un cluster dans la distribution temporelle linéaire des événements dans une *timeline*

L'identification d'un cluster peut cependant devenir malaisée lorsque l'on représente un grand nombre d'événements. Les objets graphiques les représentant se superposent, et les frontières des intervalles temporels comportant un nombre élevé d'événements peuvent ne pas être visibles de façon nette. Deux solutions permettent de rendre plus visible l'existence d'un nombre élevé d'événements dans une section de temps restreinte.

La première solution consiste à agréger les événements temporellement proches, et à les représenter par une unique entité graphique :

- Les événements dont l'espacement temporel est inférieur à une valeur seuil sont agrégés entre eux.
- Chaque agrégat est représenté par une entité graphique.
- Une variable visuelle peut être utilisée pour représenter le nombre d'événements agrégés représentés par une entité graphique.

Dans l'hypothèse d'une agrégation d'événements ponctuels, les événements sont représentés par des entités ponctuelles dont la taille est fonction du nombre d'événements agrégés. Ce procédé permet l'identification de fortes concentrations d'événements grâce à la taille des entités graphiques (Figure 4.11).

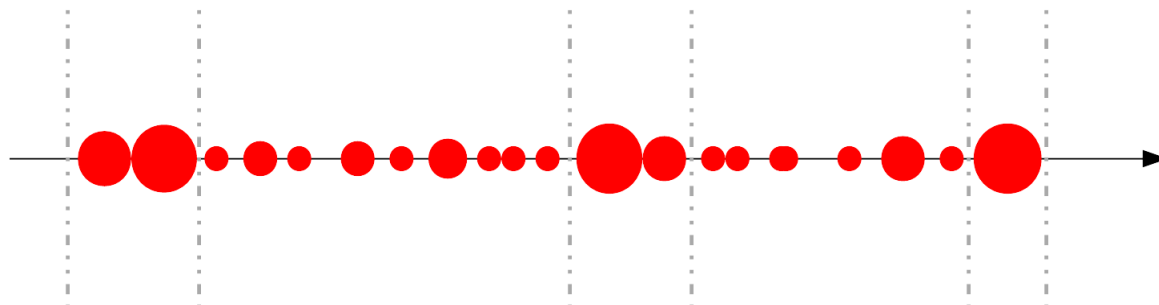


FIGURE 4.11 – Identification d'un cluster dans la distribution temporelle linéaire des événements dans une *timeline* avec agrégation des événements temporellement proches

La seconde solution consiste à mettre en forme la série événementielle de manière à créer une série temporelle représentant le nombre d'apparitions d'événements au cours du temps, et à

représenter cette dernière série dans un diagramme temporel.

Pour ce faire, les événements sont agrégés selon le temps linéaire, cependant l'espacement temporel entre deux événements n'est pas le critère pris en compte pour leur agrégation. Le temps est divisé en une succession d'intervalles temporels de même durée, correspondant à une fréquence d'échantillonnage. Pour chaque intervalle, les événements y apparaissant sont agrégés.

L'évolution du nombre d'événements apparaissant dans chaque intervalle constitue une série temporelle, ayant pour pas de temps la durée de l'intervalle d'échantillonnage. Cette série temporelle est alors représentée dans un diagramme temporel, au moyen de courbes temporelles ou d'histogrammes (Figure 4.12).

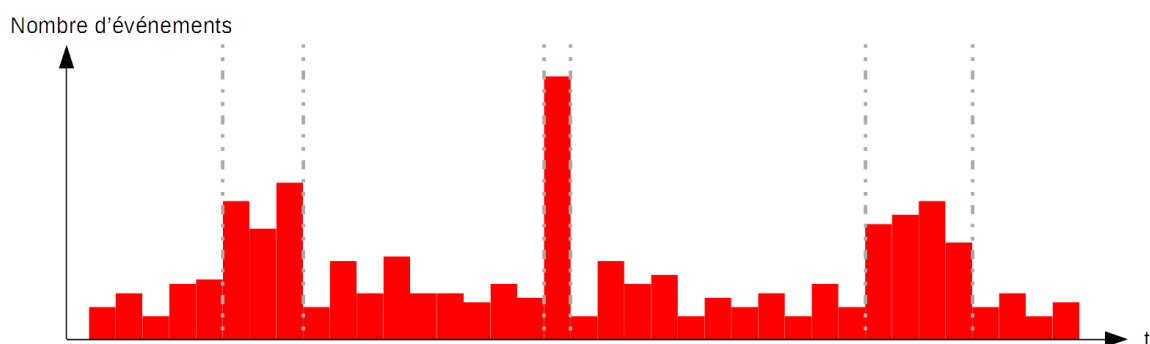


FIGURE 4.12 – Identification d'un cluster dans la distribution temporelle linéaire des événements dans une *timeline* avec agrégation des événements selon un pas d'échantillonnage

4.1.2.3 Identification de clusters relatifs à une échelle cyclique dans une *timewheel*

L'identification d'une concentration temporelle d'événements relative à une échelle cyclique peut s'effectuer en identifiant une concentration d'objets graphiques dans une section angulaire d'une *timewheel* représentant cette échelle cyclique (Figure 4.13). Le même procédé peut être utilisé pour une *timewheel* en nuage de point.

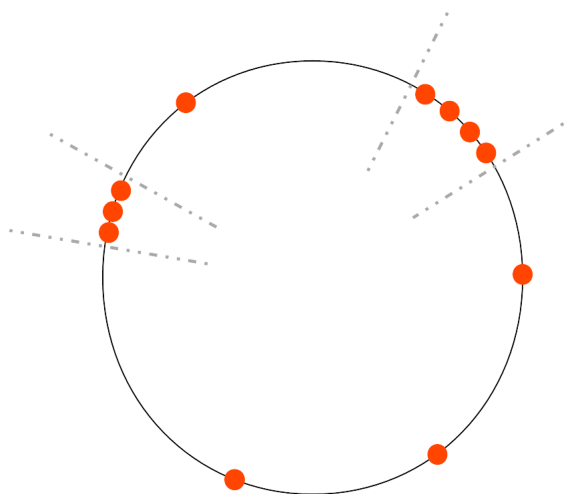


FIGURE 4.13 – Identification d'un cluster dans la distribution temporelle des événements selon une échelle cyclique dans une *timewheel*

Comme pour la *timeline*, l'identification d'un cluster dans la distribution temporelle des événements au sein d'une *timewheel* peut devenir malaisée lorsque l'on manipule un nombre important d'événements. Les mêmes procédés d'agrégation que ceux cités précédemment peuvent être utilisés pour identifier une forte concentration d'événements dans une section restreinte de l'échelle temporelle cyclique.

Les événements peuvent être agrégés selon leur proximité temporelle relative à l'échelle cyclique utilisée, et être représentés par des entités graphiques uniques, en utilisant une variable visuelle pour représenter le nombre d'événements agrégés. L'identification de clusters s'effectue en identifiant sur la *timewheel*, les entités graphiques représentant un important nombre d'événements agrégés.

Les événements peuvent également être agrégés selon un intervalle temporel cyclique d'échantillonnage, afin de recréer une série temporelle représentant le nombre d'apparitions d'événements en fonction du temps, selon l'échelle cyclique représentée. La série temporelle créée peut être représentée sur un diagramme circulaire s'inspirant de la *timewheel*.

La Figure 4.14 montre un exemple ancien de diagramme temporel circulaire représentant l'évolution d'une variable dans le temps cyclique. Ce type de diagramme est appelé *Rose Chart* ou *Polar Area Graph*. Ce diagramme, créé par Florence Nightingale, présente l'évolution du nombre de morts dans le corps expéditionnaire franco-britannique durant la guerre de Crimée, ayant eu lieu au milieu du XIXe siècle [Nightingale, 1858].

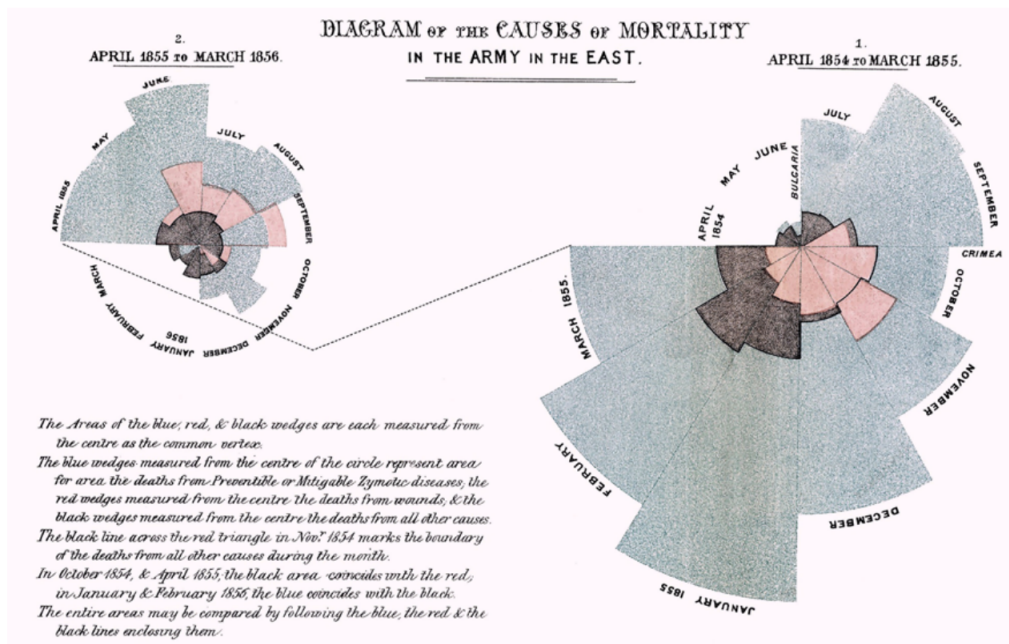


FIGURE 4.14 – Diagramme des causes de mortalité dans les forces expéditionnaires dans la Guerre de Crimée [Nightingale, 1858]

L'échelle cyclique représentée est d'une durée d'une année, et la granularité choisie est le mois. On peut considérer que les événements *décès d'un soldat* ayant eu lieu durant l'étendue de temps analysée ont été agrégés selon le mois. La Figure 4.14 présente deux diagrammes, chacun représentant une instance d'une échelle cyclique d'une année. Ces deux diagrammes représentent le nombre de décès survenus durant deux années du conflit, au travers d'une échelle cyclique d'un an. Le nombre de décès survenus durant ces deux années peut également être projeté sur un même diagramme.

Chaque diagramme se lit de la manière suivante :

- Chaque section angulaire du diagramme représente un mois de l'année.
- Le nombre de décès est représenté par l'axe radial du diagramme. Pour chaque section, le rayon de la section représente le nombre de décès ayant eu lieu durant le mois de l'année correspondant.
- Chaque section du diagramme est subdivisée en sous-sections radiales. Chaque sous-section représente une cause de décès différente. La largeur de chaque sous-section radiale indique le nombre de décès liés, pour le mois correspondant, à la cause de décès représentée. La proportion des sous-sections radiales à l'intérieur d'une section angulaire montre la proportion des différentes causes de décès dans le nombre de morts pour le mois correspondant.

Ces diagrammes sont également nommés (*Circular*) *Sector Graphs* [Harris, 2000] (Figure 4.15), ou *diagrammes polaires* [Bertin, 1973].

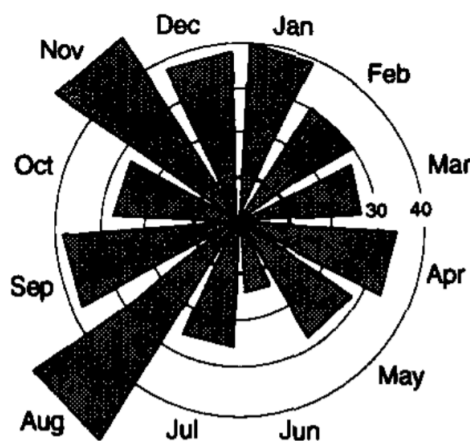


FIGURE 4.15 – *Sector Graph* de Harris [Harris, 2000]

Le diagramme *Sector Graph*, proposé par Harris [Harris, 2000], ne présente pas de subdivisions radiales des sections du diagramme, comme le *Rose Chart* de Florence Nightingale, mais utilise également chaque section angulaire du diagramme pour représenter un intervalle temporel relatif à une échelle cyclique.

L'utilisation de tels diagrammes permet de repérer les intervalles temporels cycliques présentant de fortes concentrations d'événements, symbolisées par un rayon plus important.

L'analyse de l'existence d'un cluster relatif à une échelle cyclique n'est rendue possible, par ces procédés, que pour l'échelle cyclique représentée par le cercle de la *timewheel*. Afin de pouvoir analyser l'existence de clusters relatifs à différentes échelles temporelles cycliques, il est nécessaire de pouvoir représenter le temps selon des échelles cycliques de durées différentes. Il faut donc proposer un environnement de géovisualisation combinant plusieurs *timewheels*, ou permettre à l'utilisateur de modifier dynamiquement l'échelle cyclique représentée dans la *timewheel*.

4.1.2.4 Identification d'une récurrence cyclique dans l'apparition des événements

Si l'identification de clusters dans la distribution temporelle des événements, selon le temps linéaire ou selon une ou plusieurs échelle(s) cyclique(s), est rendue possible dans les outils de types *timeline* et *timewheel*, l'identification d'une récurrence cyclique d'événements au travers de

ces outils est plus délicate.

Une récurrence cyclique dans l'apparition des événements peut prendre la forme :

- d'une réapparition périodique d'un événement ;
- d'une réapparition périodique d'un nombre plus important d'événements, ce qui peut être interprété comme une réapparition périodique d'un cluster dans le temps linéaire.

La réapparition périodique d'un événement ou d'un cluster peut être identifiée sur les diagrammes temporels basés sur les *timelines*. Cependant, ces derniers n'intégrant pas de représentation du temps selon une échelle cyclique, l'identification visuelle d'une récurrence cyclique implique que l'utilisateur identifie (Figure 4.16) :

- une série de motifs graphiques similaires le long de l'axe temporel de la *timeline* (concentration d'entités graphiques, entités graphiques de même taille pour le cas d'une agrégation des événements, hautes valeurs d'un histogramme ...);
- la séparation de chaque motif graphique similaire par une distance graphique régulière x .

L'utilisation d'une graduation du temps permet de projeter sur l'axe linéaire la succession d'instances d'une échelle cyclique (projection des mois de l'année sur l'axe linéaire), permettant d'aider l'utilisateur à identifier ces intervalles graphiques réguliers entre les concentrations récurrentes d'événements. Cependant cette identification visuelle au travers d'une *timeline* reste délicate.

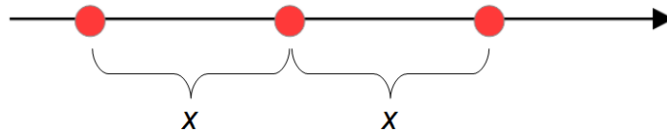


FIGURE 4.16 – Identification visuelle de récurrences cycliques dans les *timelines*

L'existence d'une récurrence cyclique dans les séries d'événements implique donc qu'il existe une échelle temporelle cyclique pour laquelle les événements récurrents apparaissent au voisinage d'un même moment temporel. Une récurrence cyclique dans les apparitions d'événements implique donc l'existence d'un cluster relatif à une échelle cyclique de durée égale à la période du cycle de réapparition des événements. Ce cluster peut être identifié au travers d'un diagramme de type *timewheel*.

Cependant, l'existence d'une récurrence cyclique dans les apparitions d'événements implique également que les événements récurrents réapparaissent pour chaque instance du cycle, donc pour chaque instance de l'échelle cyclique représentée par la *timewheel*.

Or, si une *timewheel* permet l'identification visuelle d'une concentration temporelle des événements selon une échelle cyclique, la *timewheel* ne permet pas une visualisation de chaque instance de l'échelle cyclique représentée, nécessaire pour vérifier si l'existence d'un cluster correspond à une récurrence cyclique des événements (Figure 4.17).

Les événements correspondants peuvent :

- n'apparaître que dans une seule instance de l'échelle cyclique représentée (Repère 1 de la Figure 4.17) ;

- apparaître dans des instances de l'échelle cyclique séparées entre elles par des durées irrégulières (Repère 2 de la Figure 4.17) ;
- présenter un comportement cyclique, mais d'une période correspondant à un harmonique de la période correspondant à l'échelle cyclique représentée dans la *timewheel* (Repère 3 de la Figure 4.17).

L'identification d'une récurrence cyclique dans les séries d'événements implique ainsi une analyse des événements au travers d'une représentation linéaire du temps.

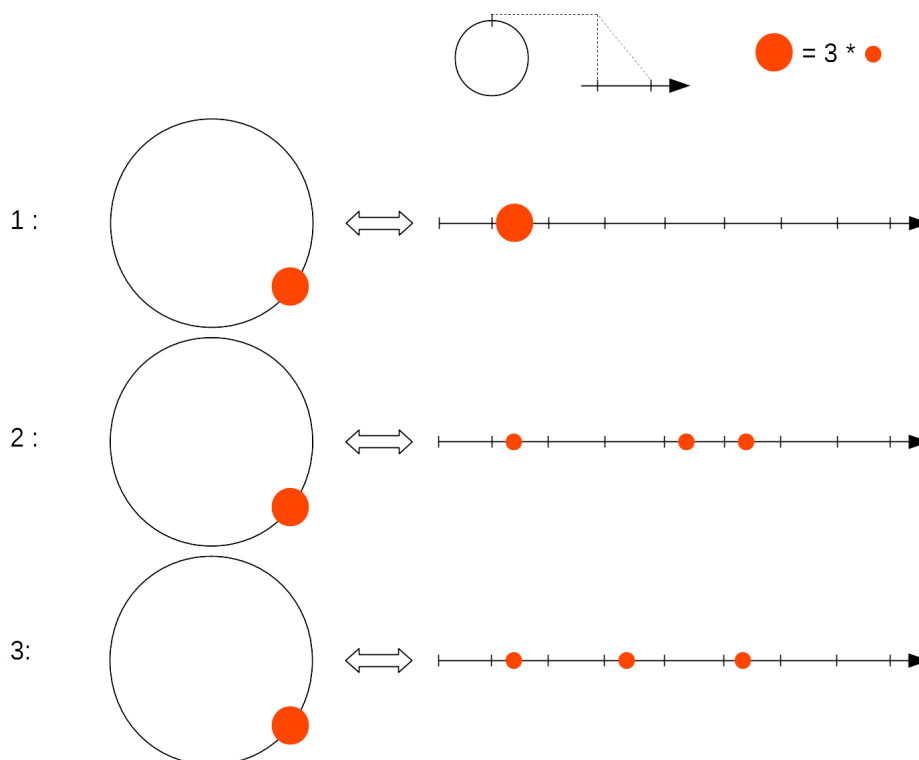


FIGURE 4.17 – Difficulté de l'identification d'une récurrence cyclique au travers d'une *timewheel*

L'identification de comportements récurrents dans les données temporelles peut être facilitée en soutenant l'analyse visuelle par des analyses quantitatives effectuées sur les données en arrière-plan.

Le principe consiste à intégrer dans l'environnement de géovisualisation des outils d'analyse mathématique effectuant, sans demander d'action de la part de l'utilisateur, une recherche des possibles récurrences dans les données temporelles, de leurs possibles fréquences, etc. Les résultats de ces analyses sont ensuite renvoyés à l'utilisateur, de manière graphique et intégrée à la représentation temporelle, afin d'orienter l'utilisateur dans sa recherche des possibles récurrences cycliques dans les données.

La Figure 4.18 montre une capture d'écran de l'environnement *Arc Diagram* [Wattenberg, 2002], dont le sujet d'étude est l'analyse de patterns dans les séries de valeurs séquentielles. Cet environnement combine :

- un outil d'analyse mathématique permettant d'identifier des répétitions cycliques dans des séries de valeurs dont le format de modélisation est proche de celui des séries temporelles ;
- une représentation graphique des données, proche de la *timeline*, permettant de représenter visuellement les récurrences cycliques identifiées ;

Les données d'une série de valeurs, numériques ou non, sont intégrées dans l'environnement et sont analysées mathématiquement afin de détecter des séquences de valeurs répétitives. Par exemple, dans une série de caractères alphanumériques, l'outil d'analyse peut détecter la répétition de la chaîne de caractères *AAE*.

La série de valeurs est ensuite projetée sur un axe linéaire, et la répétition de séquences semblables est représentée par des arcs de cercle reliant les séquences concernées sur le diagramme. La largeur d'un arc de cercle correspond à la durée de la séquence récurrente, et le rayon de l'arc à l'espacement temporel entre les deux séquences.

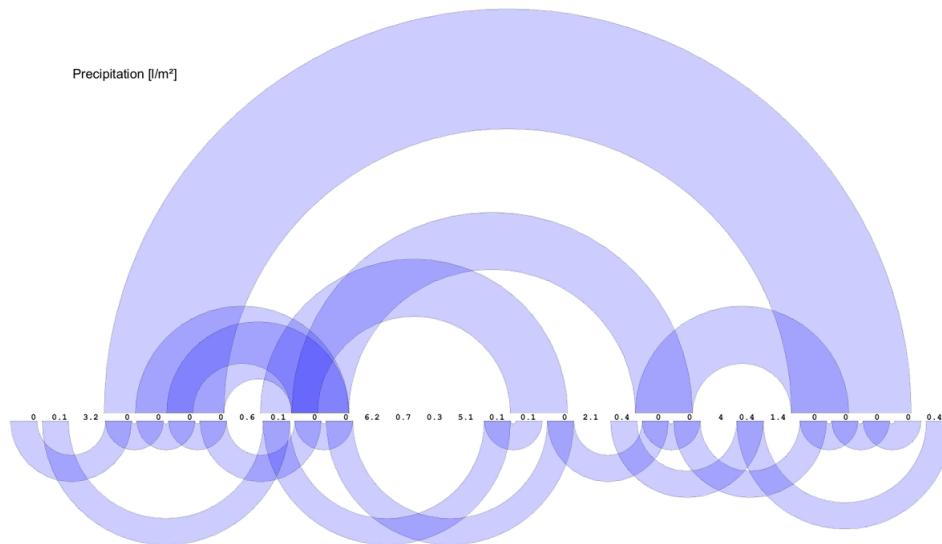


FIGURE 4.18 – *Arc Diagrams* : identification de récurrences cycliques dans les timelines, assistée par des analyses mathématiques [Wattenberg, 2002]

Appliquée à la représentation d'une série temporelle, cette application permettrait de visualiser des répétitions de séries de valeurs semblables. La présence d'une série d'arcs de cercle semblables et successifs, de même rayon et de même largeur, correspondrait à la répétition cyclique d'une série de valeurs selon une période temporelle correspondant au rayon des arcs de cercle.

Cette proposition pourrait être appliquée à des séries événementielles, ou appliquée à l'analyse d'une série temporelle représentant l'évolution du nombre d'événements au cours du temps. D'autres approches permettant d'intégrer des analyses mathématiques de récurrences cycliques dans les environnements de géovisualisation, et d'exporter les résultats de ces analyses dans les représentations du temps, pourraient également être développées.

4.1.3 La recherche de récurrences cycliques au travers des diagrammes temporels *mixtes*

4.1.3.1 Principe du diagramme temporel *mixte*

Une approche permettant d'identifier visuellement l'existence d'une récurrence cyclique de période P consiste à représenter graphiquement, dans un même diagramme, une série d'instances de la période P , et à repérer pour chaque instance l'existence d'un cluster localisé dans le même intervalle temporel, relativement à une échelle cyclique de durée P . L'utilisation de graduations pour identifier une récurrence cyclique dans un diagramme temporel linéaire repose sur ce principe, mais reste peu efficace. Cette approche peut cependant être effectuée au moyen de ce que

nous appelons des diagrammes temporels *mixtes*.

Ces derniers représentent le temps de manière linéaire, mais possèdent une structure permettant de situer temporellement les données au sein d'une ou de plusieurs échelles cycliques. Un diagramme temporel mixte s'inscrit dans une modélisation linéaire du temps, mais l'évolution linéaire du temps est représentée au travers de la succession d'instances d'une échelle temporelle cyclique. Les événements sont ainsi représentés dans ces diagrammes en fonction de leur moment d'apparition dans le temps linéaire, mais également en fonction de leur moment d'apparition selon une échelle temporelle cyclique.

Le principe de l'identification de récurrences cycliques au travers de ces diagrammes, consiste à identifier une échelle cyclique pour laquelle on peut observer, dans chaque instance de cette échelle, l'apparition d'un cluster durant le même intervalle temporel. Quand ce motif est observé sur un diagramme temporel mixte, il traduit l'existence d'une récurrence cyclique dans l'apparition des événements, de période égale à la durée de l'échelle cyclique représentée.

Pour permettre cette analyse, les diagrammes temporels mixtes doivent permettre de modifier la durée de l'échelle cyclique représentée, afin que l'utilisateur puisse rechercher pour quelle échelle cyclique il observe une réapparition régulière de clusters sur le diagramme.

Les diagrammes temporels mixtes peuvent se présenter sous différentes formes en fonction de l'approche adoptée pour combiner les échelles temporelles linéaire et cyclique. Les récurrences dans les apparitions d'événements vont alors se traduire par différents motifs graphiques selon la structure adoptée par le diagramme.

Nous présentons ici différents types de diagrammes temporels mixtes permettant de combiner dans la même représentation graphique le temps linéaire et le temps cyclique. Certains des exemples que nous présentons ne traitent pas de données événementielles, mais constituent de bons exemples des différentes structures adoptées par les diagrammes temporels mixtes. Ces derniers peuvent être appliqués, par la suite, à la représentation de données événementielles.

4.1.3.2 Projection du temps linéaire dans un diagramme temporel cyclique

Une première approche pouvant être utilisée dans les diagrammes temporels mixtes consiste à intégrer, dans un diagramme temporel cyclique, le temps linéaire comme une composante orthogonale aux axes du diagramme. Cette approche peut être illustrée par le diagramme *Cycle Plot* [Cleveland, 1993] (Figure 4.19).

Cycle Plot se présente sous la forme d'un diagramme à deux dimensions, dont l'ordonnée représente la valeur d'une variable portée par une série temporelle. L'abscisse est divisée en différentes sections, représentant les unités temporelles élémentaires d'une échelle cyclique définie. Sur le premier diagramme présenté dans la Figure 4.19 (Repère *A*), l'échelle temporelle utilisée est la semaine, et les unités temporelles élémentaires sont les jours de la semaine.

Pour chaque instance de l'échelle cyclique représentée par l'axe des abscisses, les valeurs de la série temporelle sont projetées dans la section correspondante de l'axe des abscisses. Dans ce diagramme, la première section de l'axe des abscisses présente une courbe temporelle représentant l'évolution des valeurs apparaissant le lundi. La section suivante présente une courbe temporelle représentant l'évolution des valeurs apparaissant le mardi, et ainsi de suite.

Une récurrence cyclique, de période égale à la durée représentée par l'axe des abscisses de ce diagramme, peut être identifiée lorsque l'on observe une stabilité des valeurs représentées dans

chaque section de l'axe des abscisses. Cette stabilité des valeurs, pour une section de l'échelle cyclique, se traduit graphiquement par un alignement de la courbe avec l'axe des abscisses.

On observe dans ce diagramme que les valeurs survenant les mêmes jours restent relativement proches au cours des différentes semaines, particulièrement le samedi et le dimanche. On peut en déduire que la série temporelle observe un comportement cyclique avec une période d'une semaine, non spécifique à un jour particulier de la semaine. Le lundi est le jour où ce comportement cyclique est le moins frappant, car on observe pour cette journée une plus forte variabilité dans les valeurs apparaissant durant les différentes semaines.

Cette interprétation est confirmée lorsque l'on observe le diagramme du repère *B* de la Figure 4.19, également issu de *Cycle Plot*. Ce diagramme représente les mêmes données que le diagramme du repère *A*, selon une échelle temporelle linéaire. On y observe le caractère périodique de la série temporelle avec une période d'une semaine. *Cycle Plot* peut ainsi être une source d'inspiration pour la représentation de l'évolution temporelle du nombre d'apparitions d'événements.

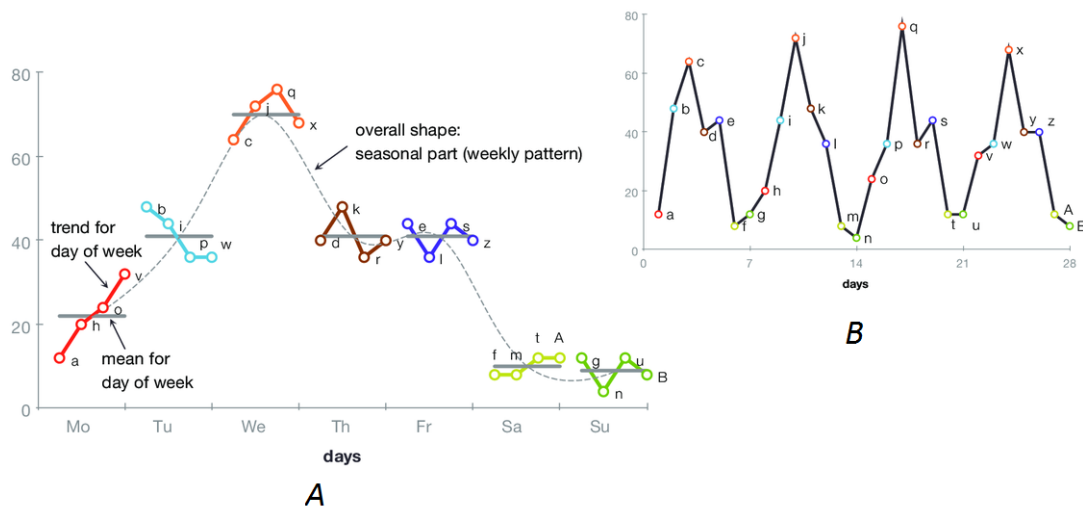


FIGURE 4.19 – Identification de récurrences cycliques dans Cycle Plot [Cleveland, 1993]

4.1.3.3 Représentation du temps au travers d'une succession de *timewheels*

Une autre approche consiste à représenter différentes instances successives d'une échelle temporelle cyclique à travers différentes *timewheels*, et à représenter ces différentes *timewheels* le long d'un axe linéaire orthogonal. Cette solution est illustrée par le diagramme *Parallel Glyphs* [Fanea et al., 2005] (Figure 4.20).

Ce diagramme représente l'évolution d'une variable portée par une série temporelle. Pour chaque instance d'une échelle cyclique considérée, la série temporelle est représentée dans un diagramme de type *timewheel*, dont le rayon représente la valeur de la variable au cours du temps. Les différentes *timewheels* sont placées les unes à la suite des autres sur un axe orthogonal, l'ensemble formant un diagramme en trois dimensions. La succession des *timewheels* sur cet axe représentant le temps linéaire.

Les *timewheels* sont placées sur un axe linéaire, de manière à ce qu'une droite, parallèle à cet axe, intersecte les différentes *timewheels* en des points représentant le même moment temporel relatif à l'échelle cyclique représentée.

Une récurrence cyclique de période égale à la durée de l'échelle représentée se traduit par la réapparition d'une même valeur, dans chaque instance de l'échelle cyclique, au même moment temporel. Chaque *timewheel* possède alors le même rayon dans la section angulaire correspondant à ce moment temporel. Les points représentant le même moment temporel sur chaque *timewheel* formant un alignement parallèle à l'axe linéaire, une récurrence cyclique se traduit par un alignement, parallèle à l'axe linéaire, de sections de *timewheel* de même rayon.

La recherche de récurrences cycliques au travers de ce type de diagramme consiste alors à chercher une échelle cyclique, pour laquelle on observe un alignement des mêmes motifs graphiques, parallèle à l'axe linéaire. Une récurrence cyclique peut ainsi se traduire par un alignement d'entités graphiques dans le cas d'une représentation de séries événementielles. Pour permettre cette recherche, ce diagramme doit intégrer la possibilité de modifier la durée de l'échelle cyclique représentée de manière synchronisée sur toutes les *timewheels*.

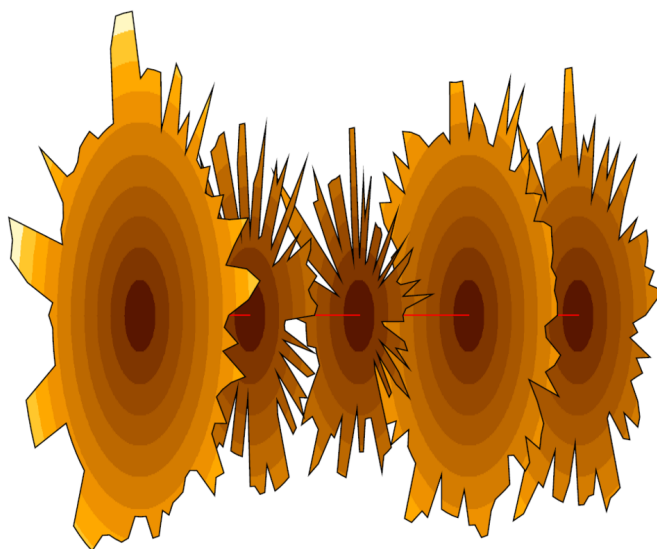


FIGURE 4.20 – Parallel Glyphs [Fanea et al., 2005]

4.1.3.4 Une version linéaire de la *timewheel*

Si dans les méthodes similaires au diagramme *Parallel Glyph*, le temps cyclique et le temps linéaire sont représentés par des variables graphiques, leurs représentations ne sont pas de même nature :

- Le temps cyclique est appréhendé au travers de la représentation d'un moment temporel. Pour un moment temporel, la valeur d'une variable ou la présence d'un ou de plusieurs événements est représentée sur le diagramme.
- Le temps linéaire est appréhendé au travers de la succession d'instances d'une échelle temporelle cyclique. Cette représentation utilise une granularité moins fine que la représentation du temps cyclique, la durée d'un granule correspondant à la durée de l'échelle cyclique. Cependant, cette représentation ne montre pas la valeur d'une variable, ou la présence d'un ou de plusieurs événements pour chaque granule temporel.

Une autre approche permettant de combiner le temps linéaire et le temps cyclique s'inspire de la forme des *timewheels* et consiste à transformer le cercle représenté dans les *timewheels* en spirale. La courbe du diagramme représente cette fois-ci le temps linéaire, mais l'enroulement de cette courbe autour d'un point central représente la répétition d'une échelle cyclique. Le temps

linéaire est alors appréhendé au travers de la représentation d'un moment temporel, symbolisé par la distance d'un point O sur la courbe du diagramme, au centre C du diagramme. Le temps cyclique est appréhendé au travers de la représentation d'un moment temporel, symbolisé par l'angle formé par l'axe $O-C$ avec l'axe vertical.

Plusieurs travaux ont porté sur ces diagrammes en spirale et sur la possibilité de représenter des données présentant des comportements périodiques [Carlis and Konstan, 1998, Weber et al., 2001, Dragicevic and Huot, 2002, Tominski and Schumann, 2008]. La Figure 4.21 montre un exemple des travaux de Weber [Weber et al., 2001]. Ce diagramme représente l'évolution d'une série temporelle au cours du temps. La valeur de variable portée par la série temporelle est représentée par l'intensité de la couleur utilisée dans le diagramme.

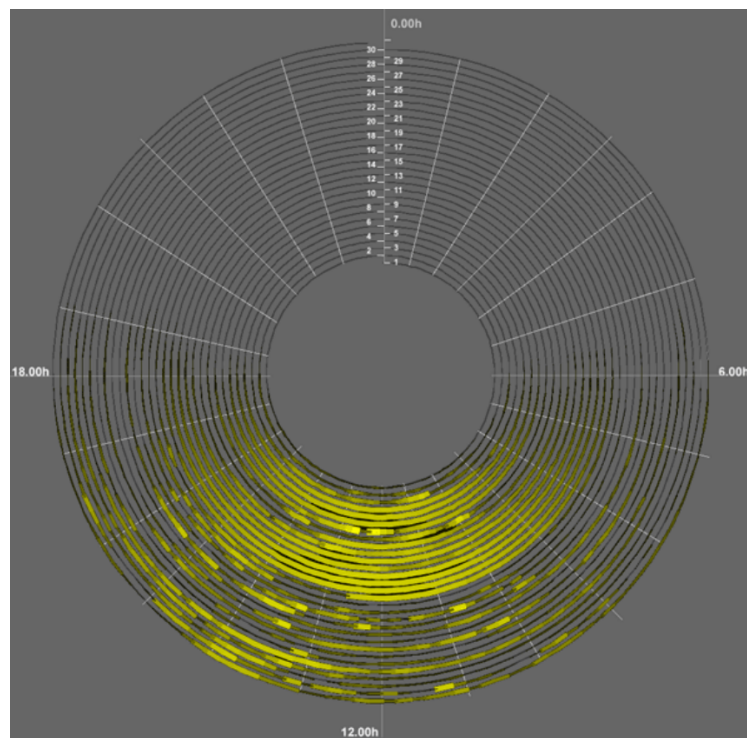


FIGURE 4.21 – Visualisation de données temporelles dans un diagramme en spirale [Weber et al., 2001]

Chaque enroulement de la spirale correspondant à une instance de l'échelle cyclique représentée, les intersections entre la courbe de la spirale et une droite partant du centre correspondent au même moment selon l'échelle cyclique représentée. Une récurrence cyclique, de période égale à la durée de l'échelle cyclique représentée, implique la répétition d'une même valeur au même moment temporel. Ceci se traduit par l'apparition du même motif graphique dans la même section angulaire du diagramme, pour chaque enroulement de la courbe de la spirale.

Une récurrence cyclique est alors visible par un alignement de motifs graphiques similaires le long d'une droite partant du centre de la spirale, comme le montre la Figure 4.21. On observe sur ce diagramme la réapparition périodique de fortes valeurs selon une période égale à la durée de l'échelle cyclique représentée.

La Figure 4.22 présente deux diagrammes, issus des mêmes travaux [Weber et al., 2001], représentant l'évolution de la même série temporelle.

On observe sur le premier diagramme, linéaire (Repère A de la Figure 4.22), une répétition

de hautes valeurs qui semble périodique. Cette périodicité est attestée dans le second diagramme (Repère B de la Figure 4.22), où la série est représentée dans un diagramme en spirale utilisant une échelle temporelle cyclique égale à la période de la récurrence cyclique observée.

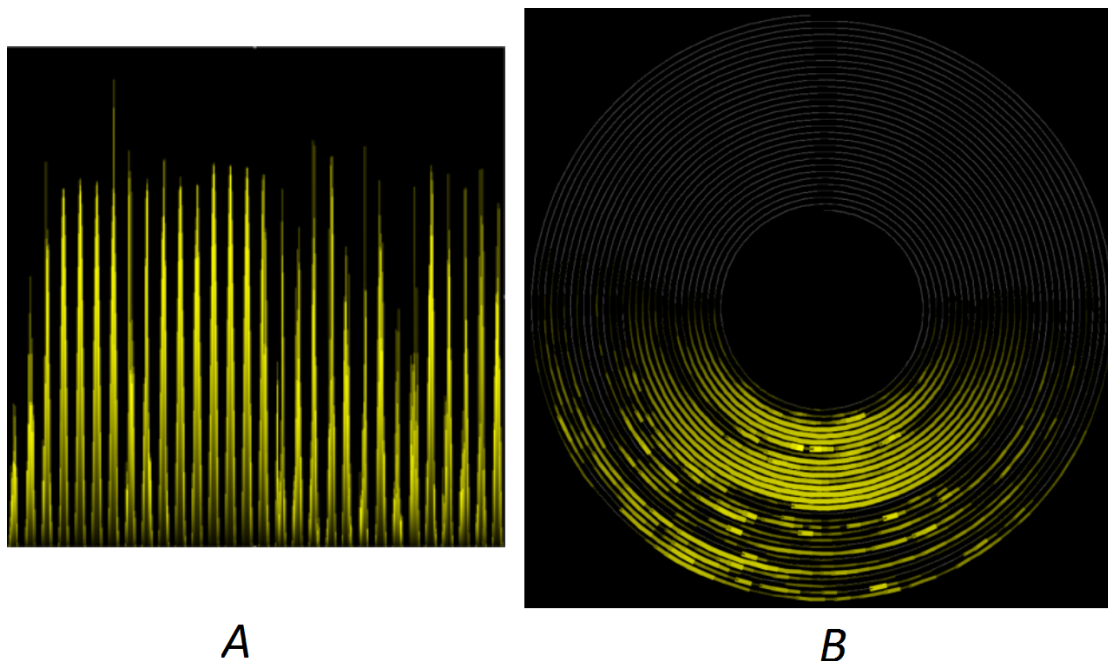


FIGURE 4.22 – Identification de récurrences cycliques dans un diagramme en spirale [Weber et al., 2001]

Par analogie avec la méthode illustrée par le diagramme *Parallel Glyph*, la recherche de récurrences cycliques sur les diagrammes en spirale consiste à chercher pour quelle échelle cyclique on observe un alignement des mêmes motifs graphiques le long d'un rayon de la spirale. Ces diagrammes doivent donc intégrer la possibilité de modifier l'échelle cyclique utilisée.

4.1.3.5 Les *tiles map*

La représentation du temps au sein de *tiles map* consiste à représenter la donnée temporelle, au sein d'un diagramme à deux dimensions dont les deux axes représentent le temps, le premier de manière linéaire et le second de manière cyclique, selon deux granularités différentes :

- L'axe des abscisses représente le temps linéaire selon une granularité temporelle a .
- L'axe des ordonnées représente le temps cyclique selon une granularité temporelle b .
- Un granule de l'axe des abscisses correspond à la durée temporelle de l'échelle cyclique représentée sur l'axe des ordonnées.

La représentation prend la forme d'un diagramme matriciel (d'où le nom de *tiles map*) où :

- Une cellule représente un moment temporel relatif au temps linéaire, selon la granularité b .
- La position d'une cellule au sein d'une colonne représente un moment temporel relatif à l'échelle temporelle cyclique utilisée, selon la granularité b .
- Une colonne représente une instance de l'échelle temporelle cyclique représentée, ainsi qu'un moment temporel relatif au temps linéaire selon la granularité a .

- Une ligne représente les différentes instances d'un même moment temporel relatif à l'échelle cyclique représentée, selon la granularité b .

Ces diagrammes sont notamment utilisés pour représenter l'évolution d'une variable portée par une série temporelle. Cette variable est représentée dans chaque cellule par une couleur, une valeur, ou encore l'utilisation d'une troisième dimension dans le cas d'une *tiles map* en trois dimensions, en fonction de la nature de la variable. Dans la Figure 4.23 [Mintz et al., 1997], une variable quantitative est représentée par une valeur appliquée à chaque cellule.

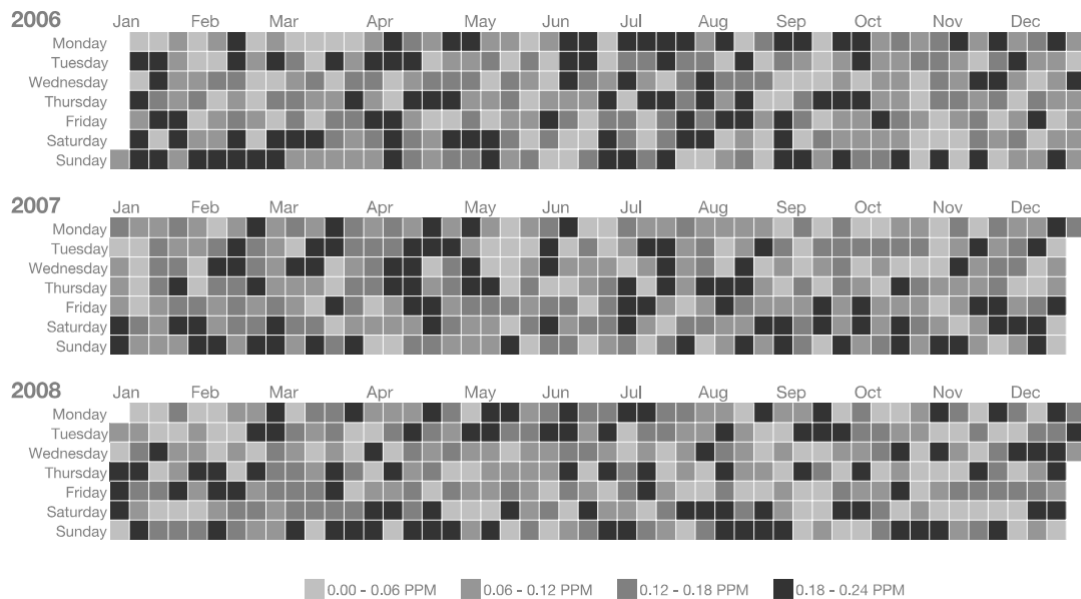


FIGURE 4.23 – Visualisation de données temporelles sur un diagramme en *tiles map* [Mintz et al., 1997, Aigner et al., 2011]

Une récurrence cyclique, de période égale à la durée de l'échelle cyclique représentée, se traduira par l'apparition de motifs graphiques similaires sur les cellules correspondant aux mêmes moments temporels relatifs à l'échelle temporelle cyclique. Ces motifs graphiques seront ainsi alignés sur une ou plusieurs lignes du diagramme, formant un alignement parallèle à l'axe des abscisses.

La recherche de récurrences cycliques sur les diagrammes en *tiles map* consiste donc à chercher pour quelle échelle cyclique on observe un alignement de motifs graphiques similaire parallèle à l'axe des abscisses. Ces diagrammes doivent également intégrer la possibilité de modifier l'échelle cyclique utilisée.

La Figure 4.24 présente une capture d'écran de l'environnement *Cluster Calendar* [Van Wijk and Van Selow, 1999]. Ce dernier utilise un diagramme de type *tiles map* en trois dimensions, où la variable est représentée, à la fois par une couleur, et à la fois par du relief. Ce diagramme représente l'évolution de la consommation électrique d'un lieu de vie dans le temps linéaire, et selon une échelle cyclique d'une journée. On observe un alignement de motifs graphiques symbolisant de fortes consommations d'énergie. Cet alignement s'interprète comme une récurrence cyclique, d'une période de vingt-quatre heures, dans la variation de la consommation électrique représentée. Cette récurrence correspond à une augmentation périodique de la consommation énergétique en pleine journée.

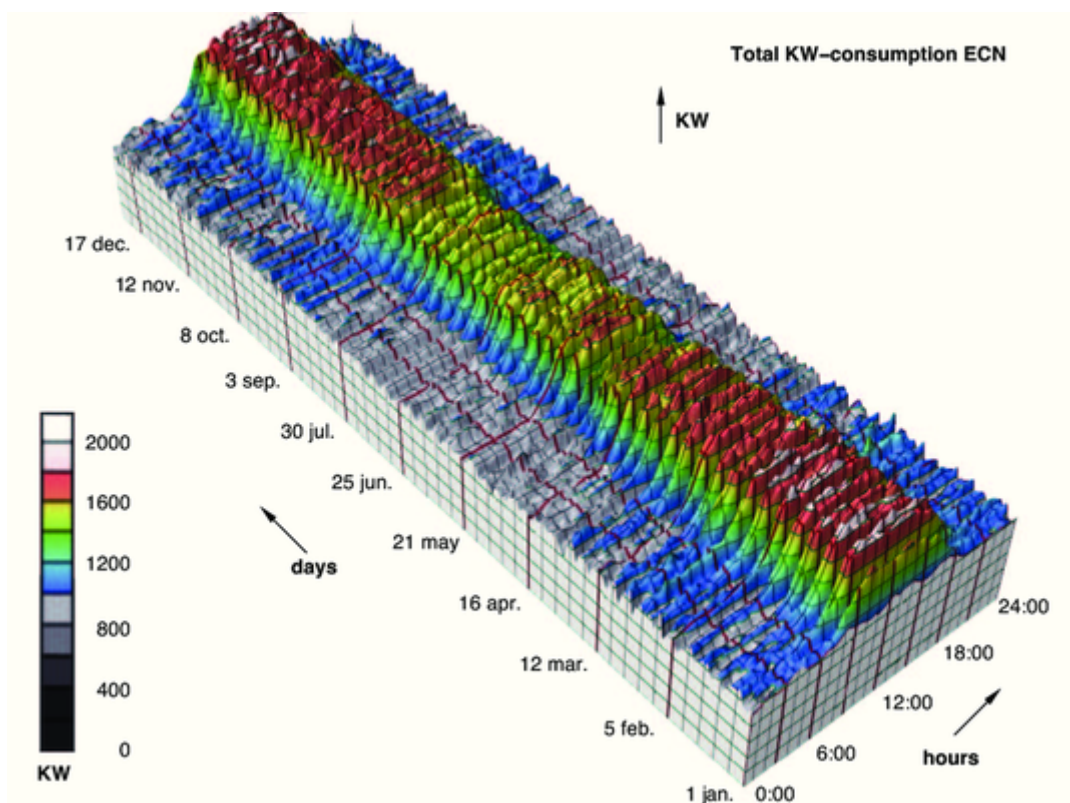


FIGURE 4.24 – Visualisation d’une récurrence cyclique dans l’environnement *Cluster Calendar* [Van Wijk and Van Selow, 1999]

4.1.3.6 La combinaison des structures de la *timeline* et de la *timewheel* au sein d’un même diagramme

Une dernière approche proposant des diagrammes temporels mixtes consiste à combiner les formes graphiques de la *timeline* et de la *timewheel* au sein d’un même diagramme temporel.

Time Coil. Un premier exemple de ces diagrammes est présenté par le diagramme *Time Coil* d’Edsall et Peuquet [Edsall and Peuquet, 1997] (Figure 4.25). Ce diagramme s’inspire de la figure d’une bobine (*Coil* en anglais) et présente la forme d’une courbe hélicoïdale en trois dimensions :

- L’axe x autour duquel la courbe hélicoïdale effectue une révolution représente le temps linéaire.
- Chaque révolution de la courbe autour de l’axe x représente une instance d’une échelle temporelle cyclique.

Un événement est représenté par une entité graphique placée sur la courbe du diagramme, en fonction du moment d’apparition de l’événement :

- Sa position le long de l’axe x représente son moment d’apparition selon le temps linéaire.
- Sa position sur un plan orthogonal à l’axe x représente son moment d’apparition selon l’échelle cyclique représentée.

Une récurrence cyclique, de période égale à la durée de l’échelle cyclique représentée, se traduit par la répétition d’un même motif graphique dans une même section angulaire pour chaque révolution de la courbe hélicoïdale. Ceci produit un alignement, parallèle à l’axe x , des motifs graphiques semblables.

Dans la Figure 4.25, chaque révolution de l'hélice correspond à une période de douze heures. Des entités graphiques de couleur bleue, pouvant représenter des événements, sont alignées parallèlement à l'axe de révolution du diagramme. Cet alignement peut être interprété comme une récurrence cyclique des événements avec une période de douze heures.

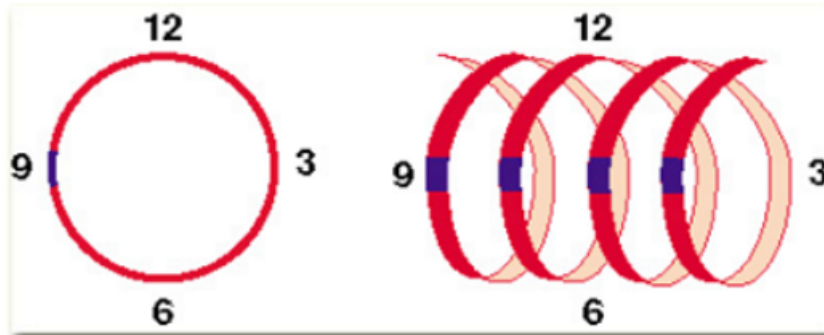


FIGURE 4.25 – Représentation du temps dans *Time Coil* [Edsall and Peuquet, 1997]

Time Wave. Un second exemple de ce type de diagramme est *Time Wave*, proposé par Li et Kraak [Li, 2010] (Figure 4.26). Le principe est ici de combiner les structures de la *timewheel* et de la *timeline* dans le même plan graphique. Le diagramme présente une courbe, en deux dimensions, formant une vague (*Wave* en anglais) effectuant des circonvolutions autour d'un axe central. Ces circonvolutions suivent la forme d'un cercle.

La courbe présente ainsi une structure périodique. L'axe central représente le temps linéaire, et chaque circonvolution de la courbe du diagramme représente une instance d'une échelle cyclique.

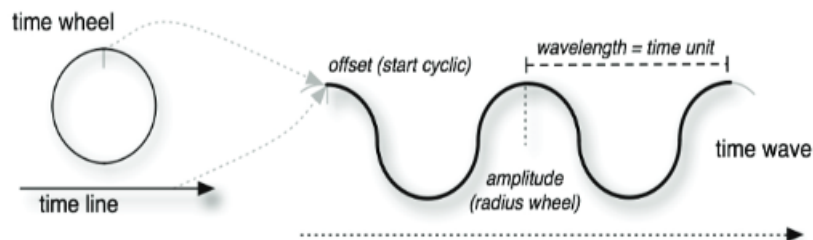


FIGURE 4.26 – Représentation du temps dans *Time Wave* [Li, 2010]

Un événement est représenté par une entité graphique placée le long de la courbe du diagramme en fonction de son moment d'apparition :

- Sa position le long de l'axe central représente son moment d'apparition selon le temps linéaire.
- Sa position à l'intérieur d'une circonvolution correspond à son moment d'apparition selon l'échelle temporelle cyclique.

Une récurrence cyclique, de période égale à la durée de l'échelle cyclique représentée, implique alors la répétition d'un même motif graphique dans la même section de la vague pour chaque circonvolution de la courbe. Ceci se traduit par un alignement, parallèle à l'axe central, de motifs graphiques semblables. Le mécanisme d'identification d'une récurrence cyclique sur un diagramme de type *Time Wave* suit ainsi le même procédé que pour un diagramme de type *Time*

Coil.

Cependant, le cas de *Time Wave*, proposant une représentation en deux dimensions, introduit la possibilité d'une erreur d'interprétation d'un alignement de motifs graphiques parallèle à l'axe central. Comme le montre la Figure 4.27, des objets graphiques peuvent former un alignement parallèle à l'axe central, et être positionnés sur des sections de la courbe ne correspondant pas au même moment temporel selon l'échelle cyclique représentée.

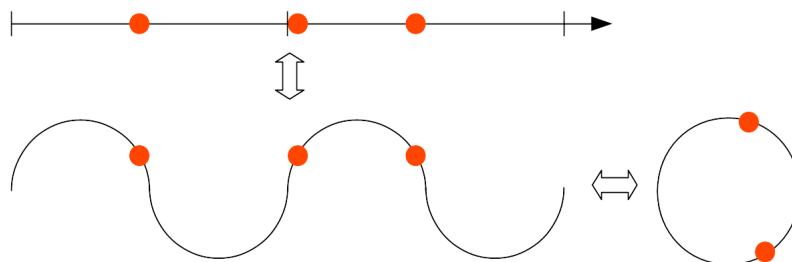


FIGURE 4.27 – Alignements sans récurrences cycliques dans un diagramme de type *Time Wave*

Si le cas de *Time Wave* demande une attention plus accrue de l'utilisateur pour ne pas effectuer cette erreur d'interprétation, ce diagramme offre une autre approche permettant d'identifier visuellement des récurrences cycliques dans les données événementielles.

L'utilisateur doit alors chercher pour quelle échelle temporelle cyclique la formation d'un alignement de motifs graphiques, parallèle à l'axe central du diagramme, est observable. Ces diagrammes doivent donc également intégrer la possibilité de modifier l'échelle cyclique représentée, ce que propose *Time Wave*.

4.1.3.7 La recherche d'une récurrence cyclique dans ces approches

Ces différentes approches de visualisation du temps, intégrant à la fois le temps linéaire et le temps cyclique, ont en commun de rendre observable un motif graphique particulier, lorsque les données temporelles analysées présentent un comportement cyclique de période égale à la durée de l'échelle cyclique représentée.

Ce motif prend la forme d'un alignement d'entités graphiques selon un axe, pouvant être différent selon l'approche suivie. La recherche d'une récurrence cyclique dans ces différents diagrammes se base ainsi sur le même procédé : l'utilisateur doit chercher pour quelle échelle cyclique un alignement d'entités graphiques similaires est observable.

Il est donc nécessaire, pour rendre possible une telle recherche, de permettre une modification dynamique et continue de la durée des échelles cycliques représentées. Cette modification graduelle doit permettre de repérer la formation d'alignements graphiques lorsque la durée de l'échelle cyclique se rapproche de la période d'une possible récurrence cyclique.

L'analyse effectuée est visuelle. C'est l'identification visuelle de tels alignements graphiques par l'utilisateur, lors de la modification de l'échelle cyclique, qui lui permet de détecter la présence de récurrences dans les données temporelles. Cependant, il est également possible d'intégrer à cette méthode d'analyse visuelle des outils mathématiques effectuant une analyse des données en arrière-plan, dans le but d'aider l'utilisateur dans son analyse exploratoire.

Ces analyses quantitatives peuvent identifier de possibles récurrences cycliques dans le jeu de données étudiées, et retourner les résultats trouvés à l'utilisateur de manière graphique, à l'instar d'*Arc Diagram* [Wattenberg, 2002] (Figure 4.18). Ces résultats permettent de suggérer à l'utilisateur, en guise de pistes d'exploration, des échelles temporelles cycliques à représenter dans les diagrammes temporels mixtes.

4.1.4 Synthèse

L'utilisation de certains diagrammes temporels permet l'identification des structures temporelles correspondant aux projections, sur un plan uniquement temporel, des patterns spatio-temporels que nous recherchons :

- Les diagrammes de type *timeline* permettent l'identification visuelle de concentrations d'événements dans le temps linéaire.
- Les diagrammes de type *timewheel* permettent l'identification visuelle de concentrations d'événements selon une échelle temporelle cyclique
- L'utilisation de diagrammes temporels mixtes permet d'identifier des clusters ou des récurrences cycliques dans la distribution temporelle des événements.

L'utilisation de tels diagrammes interactifs offre ainsi la possibilité d'une analyse exploratoire des données temporelles dans le but d'y identifier ces structures.

4.2 La visualisation du lien entre les composantes spatiale et temporelle pour l'identification de patterns

Intégrés au sein d'un outil de géovisualisation à multi-fenêtrage, ces outils graphiques peuvent offrir une analyse exploratoire permettant une identification des dynamiques spatio-temporelles des phénomènes. Il est cependant nécessaire de mettre en place des procédés visuels permettant à l'utilisateur de faire le lien entre les motifs d'intérêts observables dans la fenêtre temporelle, et ceux observables dans la fenêtre spatiale.

Des procédés de visualisation permettant d'allier les dimensions temporelle et spatiale au sein d'une même représentation graphique peuvent alors être utilisée de concert avec le principe du multi-fenêtrage. Des approches telles que l'utilisation de la symbologie, les *small-multiples*, l'utilisation de représentation en trois dimensions, ou l'animation peuvent être utilisés afin de représenter le temps et l'espace dans une même fenêtre dimensionnelle, spatiale ou temporelle.

L'analyse exploratoire s'effectue alors principalement au travers d'une première fenêtre. Les autres fenêtres de l'environnement de géovisualisation servent alors de légendes, temporelle ou spatiale, permettant d'avoir une vue d'ensemble de la composante transposée sur la première fenêtre, et permettant d'interagir avec les données selon cette dimension.

Parmi nos recherches sur les propositions existantes dans les environnements de géovisualisation, nous nous sommes intéressés aux environnements à multi-fenêtrage permettant une analyse visuelle des patterns spatio-temporels au travers du temps linéaire et du temps cyclique. Nous nous sommes notamment intéressés à deux environnements de géovisualisation en particulier :

- *Time Wave* [Li, 2010], proposé en 2010, dont est issu le diagramme présenté dans la section précédente.
- *PerSE* [Swedberg and Peuquet, 2017], présenté très récemment.

Dans la section suivante, nous présentons le principe de ces deux environnements de géovisualisation, comment y sont intégrés les diagrammes permettant une analyse visuelle des structures temporelles dans les données, et quels sont les procédés mis en place pour faire le lien entre ces diagrammes et la représentation spatiale.

4.2.1 L'analyse des données spatio-temporelles au travers du temps linéaire et du temps cyclique : les approches *TimeWave* et *PerSE*

4.2.1.1 Une structure commune

Les environnements *Time Wave* [Li, 2010] et *PerSE* [Swedberg and Peuquet, 2017] partagent quelques critères communs :

- ces environnements représentent les trois composantes de l'information spatio-temporelle au travers d'une interface à multi-fenêtrage respectant le modèle préconisé par Kraak [Kraak and Ormeling, 2011] ;
- ces environnements intègrent des représentations graphiques du temps permettant de représenter le temps linéaire et le temps cyclique ;
Ces diagrammes sont interactifs et permettent de modifier l'étendue temporelle linéaire ou l'échelle cyclique représentées.
- ces environnements intègrent, en parallèle du multi-fenêtrage, des procédés permettant d'intégrer les dimensions temporelle et spatiale dans la même représentation graphique ;
Time Wave utilise des règles de symbologie, tandis que *PerSE* utilise un dispositif *small-multiples*.
- les procédés utilisés pour effectuer le lien entre les composantes temporelle et spatiale des données se basent sur une approche orientée de la fenêtre spatiale vers la fenêtre temporelle : la représentation des données dans la fenêtre temporelle est déterminée par la composante spatiale de celles-ci.

4.2.1.2 *Time Wave*

L'environnement proposé par *Time Wave* a pour objectif de fournir une analyse de données temporelles rattachées à une composante spatiale. Ces données peuvent être spatio-temporelles, ou se présenter sous la forme de données temporelles liées à des objets spatiaux pérennes. L'analyse offerte par *Time Wave* est principalement *orientée temps*, et le diagramme temporel y occupe un rôle central.

Principe général de l'environnement.

Données représentées. Les données représentées dans *Time Wave* peuvent se présenter sous deux formes :

- sous la forme de séries d'événements spatio-temporels, possédant une composante spatiale, temporelle, et thématique.
- sous la forme de séries temporelles rattachées à des entités spatiales pérennes.

Time Wave permet ainsi de visualiser des séries temporelles rattachées à des stations de mesures localisées en différents lieux de l'espace. La solution présentée par *Time Wave* a l'avantage de combiner dans le même environnement de géovisualisation les modélisations durantes et occurrentes des phénomènes spatio-temporels.

Structure de l'environnement. L'environnement de *Time Wave* est structuré en trois fenêtres, conformément au modèle préconisé par Kraak [Kraak and Ormeling, 2011] (Figure 4.28) :

- La fenêtre temporelle est occupée par le diagramme temporel mixte présenté dans la section précédente (Figure 4.26).
- La fenêtre spatiale est occupée par une carte en deux dimensions.
- La fenêtre thématique est consacrée à la visualisation d'une variable attributaire rattachée aux données représentées.

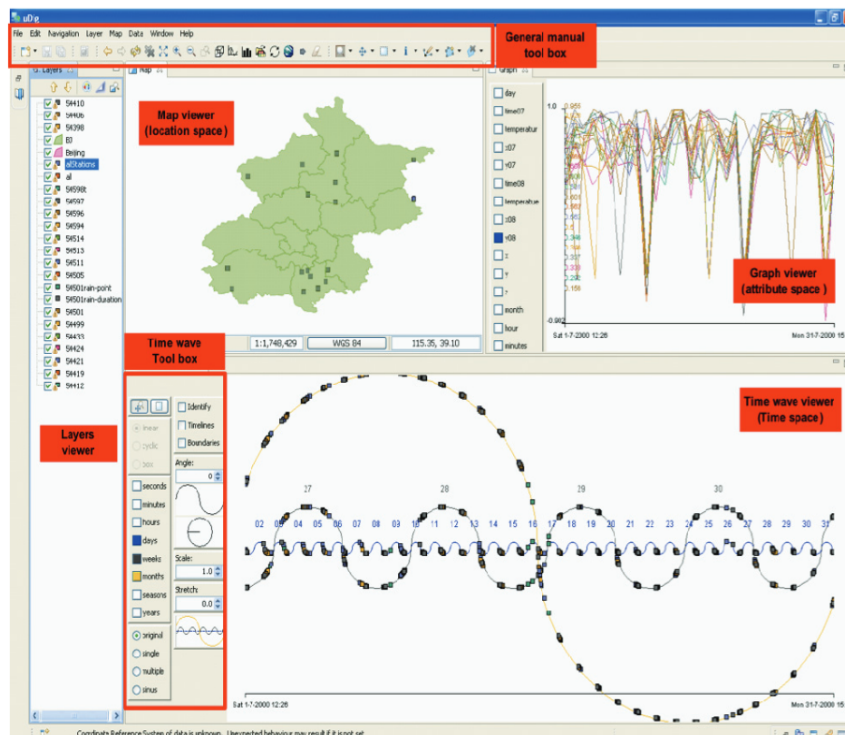


FIGURE 4.28 – Structure de Time Wave [Li, 2010]

Fenêtre thématique. La fenêtre thématique peut être occupée par des diagrammes temporels, représentant l'évolution temporelle d'une variable rattachée à des objets spatiaux, ou les valeurs d'attributs rattachées à des événements spatio-temporels.

Fenêtre spatiale. La carte de la fenêtre spatiale ne représente pas un état de l'espace pour un granule temporel donné, comme dans le cas d'une animation ou d'un dispositif *small-multiples* :

- dans le cas d'une modélisation endurante des phénomènes, les entités représentées sur la carte correspondent aux objets spatiaux pérennes auxquels sont rattachées des séries temporelles.
- dans le cas d'une modélisation occurrente des phénomènes, les entités représentées sur la carte correspondent à la composante spatiale des événements apparaissant durant l'étendue de temps représentée dans la fenêtre temporelle.

Fenêtre temporelle. La fenêtre temporelle est occupée par le diagramme temporel mixte présenté précédemment (Figure 4.26). Les données y sont représentées par des entités graphiques ponctuelles placées le long de la courbe. Ces entités représentent :

- soit des événements temporels créés à partir des séries temporelles représentées. Ces événements peuvent correspondre, par exemple, au franchissement d'une valeur seuil par une variable portée par une série temporelle.
- soit la composante temporelle d'événements spatio-temporels.

Les données sont ainsi toujours représentées sous un format événementiel. Les entités sont placées le long de la courbe du diagramme, en fonction de leur moment d'apparition relatif au temps linéaire et à l'échelle temporelle cyclique représentée.

L'échelle cyclique représentée est paramétrable au moyen d'une interface de contrôle. La forme suivie par les circonvolutions du diagramme est circulaire. L'utilisateur peut modifier le rapport entre les échelles graphiques utilisées sur les axes horizontaux et verticaux du diagramme, de manière à modifier l'amplitude verticale des *vagues* du diagramme.

Visualisation du lien entre les composantes temporelle et spatiale des données. *Time Wave* propose l'utilisation de règles de symbologie pour visualiser les relations entre les différentes composantes de la donnée.

Les données sont classées selon un critère spatial (appartenance à une zone, objet spatial de référence d'une série temporelle) ou thématique (valeur d'un attribut, type d'événement), et sont représentées en fonction de ce critère sur le diagramme temporel au moyen de règles de symbologie.

La Figure 4.29 présente deux exemples de représentation des données dans le diagramme temporel, en fonction de leurs autres composantes dimensionnelles :

- Dans le premier exemple (Repère A de la Figure 4.29), des événements sont classés en fonction de leur localisation spatiale, ramenée à une ville d'appartenance. La localisation est traitée comme une variable qualitative et représentée par de la couleur.
- Dans le second exemple (Repère B de la Figure 4.29), des événements sismiques sont classés en fonction de la magnitude du séisme. Cette variable est traitée comme une variable quantitative et représentée par la taille de l'entité graphique correspondante.

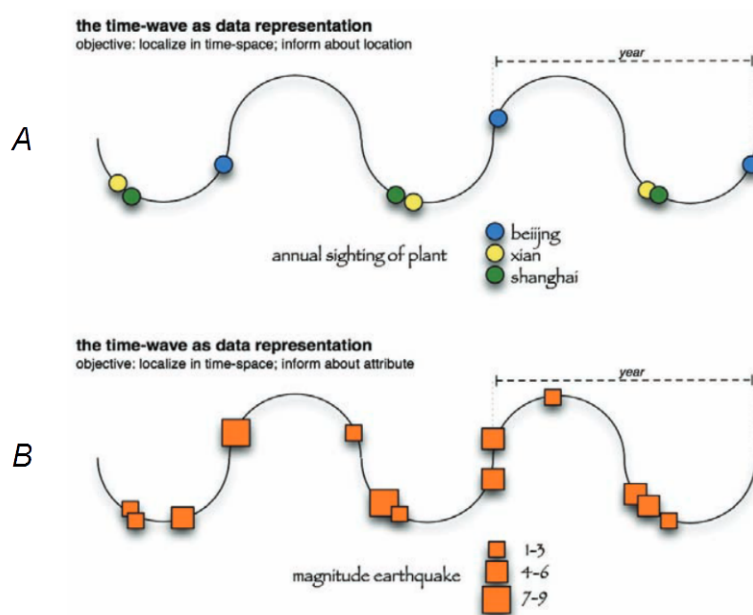


FIGURE 4.29 – Symbologie des événements dans *Time Wave* [Li, 2010]

La visualisation du lien entre les différentes composantes de la donnée, au moyen de règles de symbologie, est ainsi toujours orientée de la dimension spatiale (ou thématique) vers la dimension temporelle. Les données sont représentées pour différentes modalités de la dimension spatiale (ou thématique), et les différentes classes d'événements correspondantes sont représentées dans le diagramme temporel, au moyen de règles de symbologie (Figure 4.30).

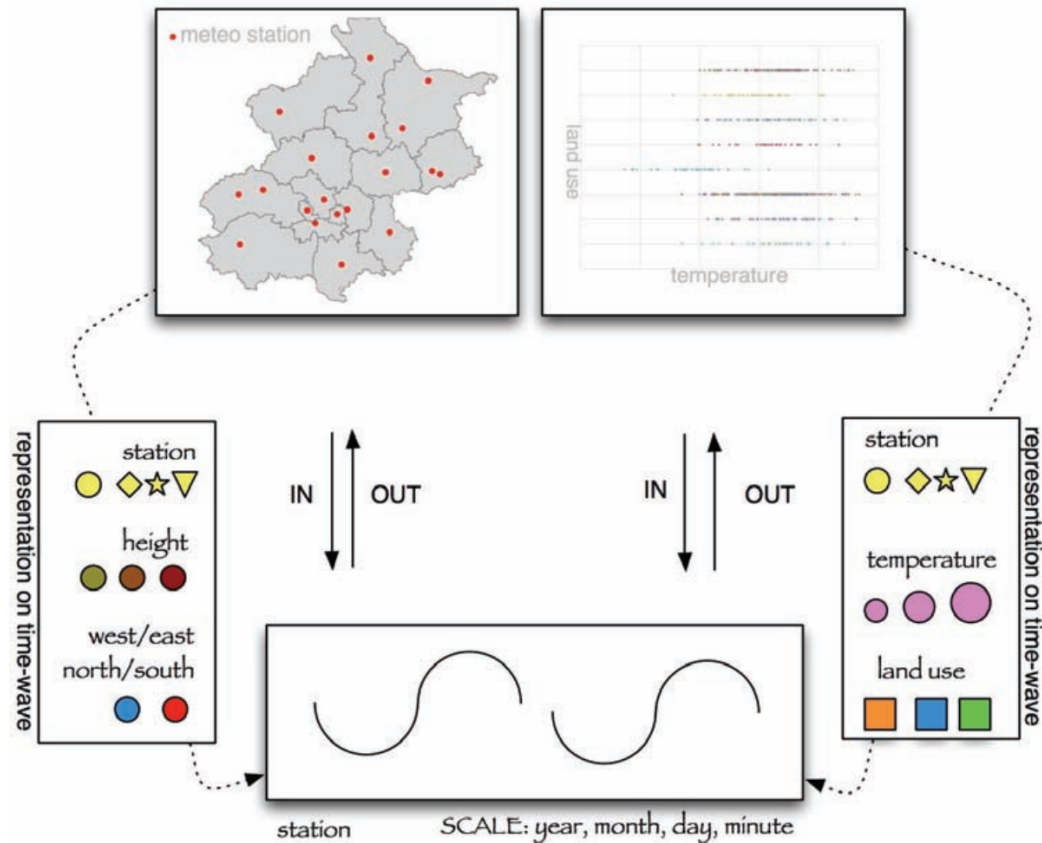


FIGURE 4.30 – Visualisation du lien entre les composantes temporelle et spatiale des données, une représentation orientée de la carte vers le diagramme temporel [Li, 2010]

Ce choix peut être dû à l'orientation donnée à l'analyse fournie par *Time Wave*, principalement centrée sur l'analyse de données temporelles, y compris de séries temporelles reliées à des objets spatiaux pérennes. Ces dernières peuvent ainsi être classées selon leur objet spatial de référence. Ce classement selon ce critère dimensionnel implique une représentation des données dans le diagramme temporel.

La visualisation du lien entre les composantes temporelle et spatiale des données s'effectue ainsi exclusivement au moyen de procédés orientés de la carte vers le diagramme temporel (Figure 4.30). Le diagramme temporel constitue l'élément central de l'environnement *Time Wave*, et l'analyse des structures spatio-temporelles des phénomènes s'effectue au travers du diagramme temporel.

4.2.1.3 *PerSE*

L'analyse proposée par *PerSE* vise à identifier les périodicités calendaires dans les séries d'événements spatio-temporels. Le terme de périodicité cité dans les travaux effectués autour de *PerSE* correspond plutôt à ce que nous désignons dans nos travaux par un cluster dans la distribution temporelle des événements, relatif à une échelle cyclique.

L'analyse proposée par *PerSE* cherche ainsi à visualiser, dans la distribution spatio-temporelle des événements, des structures particulières liées aux échelles temporelles calendaires. La particularité de *PerSE* est de s'intéresser à des échelles calendaires pouvant être basées sur d'autres calendriers que le calendrier géorgien.

Principe général de l'environnement.

Données représentées. À l'inverse de *Time Wave*, les données représentées dans *PerSE* se présentent toutes sous un format événementiel. Chaque événement possède une composante spatiale, une composante temporelle, et une composante thématique.

Structure de l'environnement. *PerSE* est également structuré en trois fenêtres, conformément au modèle préconisé par Kraak [Kraak and Ormeling, 2011] (Figure 4.31).

L'environnement est composé :

- d'une fenêtre temporelle, occupée par deux types de diagrammes temporels, l'un linéaire, l'autre cyclique ;
- d'une fenêtre spatiale, occupée par une carte en deux dimensions ;
- d'une fenêtre thématique, consacrée à la visualisation d'une variable attributaire.

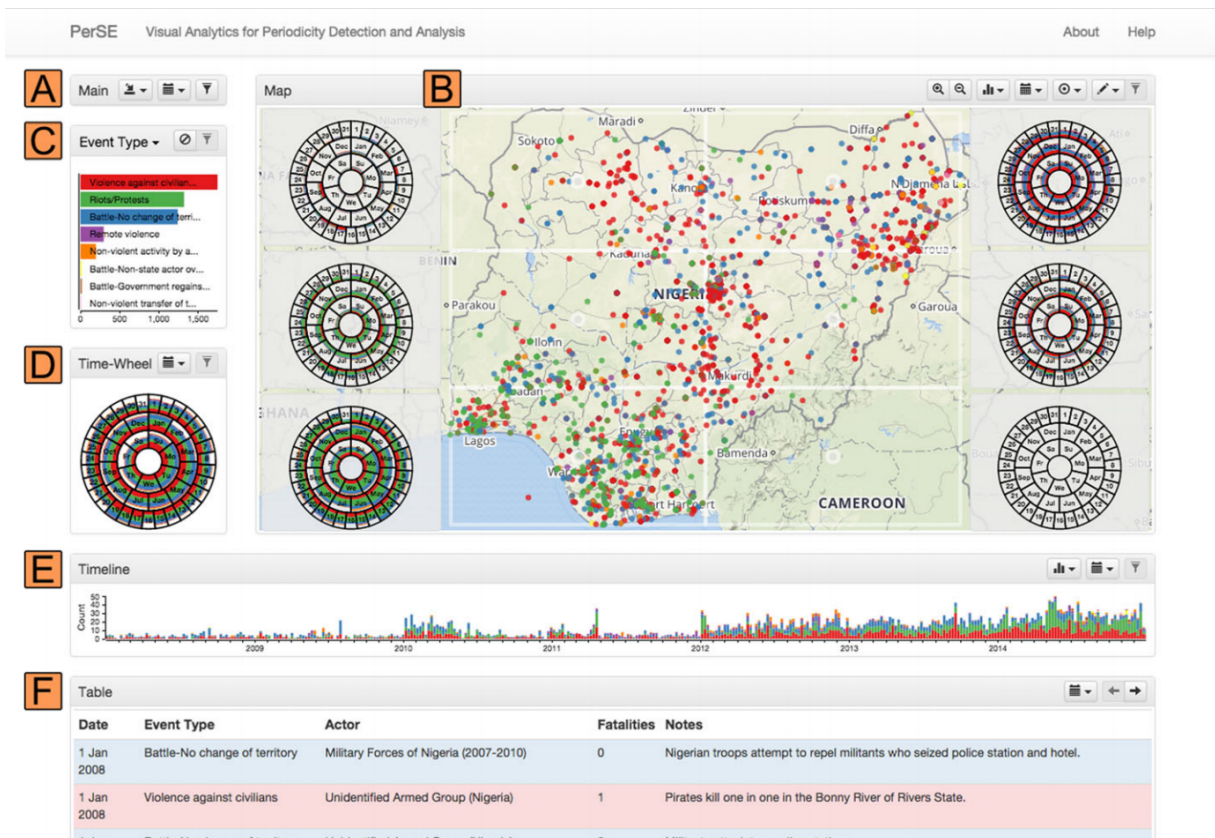


FIGURE 4.31 – Structure de *PerSE* [Swedberg and Peuquet, 2017]

Fenêtre thématique. La fenêtre thématique est divisée en deux sous-fenêtres :

- La première est occupée par un diagramme représentant l'effectif des événements représentés, en fonction de la valeur prise par une variable rattachée aux événements.

- La seconde est occupée par un tableau décrivant les informations thématiques rattachées à un événement que l'utilisateur peut sélectionner.

Fenêtre spatiale. Comme pour *Time Wave*, la carte ne représente pas l'état de l'espace pour un granule temporel donné, mais l'ensemble des événements apparaissant durant l'étendue de temps représentée dans la fenêtre temporelle. Chaque événement est représenté sur la carte par une entité graphique ponctuelle, positionnée au lieu d'apparition de l'événement.

Fenêtre temporelle. La fenêtre temporelle est occupée par des diagrammes temporels de type *timeline* et de type *timewheel*.

Ces diagrammes ne représentent pas la composante temporelle des événements, mais le nombre d'apparitions d'événements au cours du temps. Les événements sont agrégés selon un pas de temps régulier, et la série temporelle créée est représentée par les diagrammes temporels :

- La *timeline* représente le nombre d'apparitions d'événements selon le temps linéaire au travers d'un diagramme en bâtons.
- La *timewheel* représente le nombre d'apparitions d'événements selon une échelle cyclique au travers d'un diagramme en secteurs, prenant la forme du diagramme proposé par Nightingale [Nightingale, 1858] (Figure 4.14).

Le diagramme temporel cyclique proposé par *PerSE* consiste en une combinaison de plusieurs *timewheels* imbriquées, chacune représentant les événements selon une échelle cyclique et une granularité différente. Ces échelles et ces granularités temporelles correspondent à des durées calendaires, et sont paramétrables par l'utilisateur.

La Figure 4.32 représente un exemple de diagramme temporel cyclique utilisé dans *PerSE*. De l'extérieur vers l'intérieur du diagramme, le temps est représenté :

- selon une échelle mensuelle avec une granularité d'une journée ;
- selon une échelle annuelle avec une granularité d'un mois ;
- selon une échelle hebdomadaire avec une granularité d'une journée.

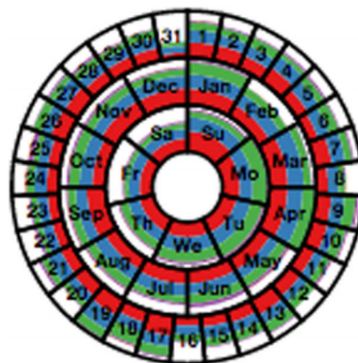


FIGURE 4.32 – Diagramme temporel cyclique de *PerSE* [Swedberg and Peuquet, 2017]

Visualisation des liens entre les composantes dimensionnelles de la donnée. *PerSE* propose l'utilisation d'un dispositif *small-multiples* pour visualiser les relations entre la composante spatiale et la composante temporelle des événements.

À l'instar de *Time Wave* les données sont classées selon un critère spatial, et sont représentées en fonction de ce dernier dans la fenêtre temporelle. Les événements sont ainsi classés en fonction

de l'appartenance de leur lieu d'apparition à différentes zones spatiales.

À la différence de *Time Wave*, les événements de chaque zone ne sont pas représentés au moyen de règles de symbologie, mais par un diagramme temporel cyclique distinct, à la manière d'un dispositif *small-multiples*. La fenêtre temporelle est ainsi composée :

- d'un diagramme linéaire ;
- d'un diagramme cyclique représentant la distribution temporelle de l'ensemble des événements ;
- d'une série de diagrammes cycliques représentant la distribution temporelle des événements de chaque zone.

PerSE offre également à l'utilisateur la possibilité d'échanger la série de diagrammes cycliques pour une série de diagrammes linéaire, permettant de représenter la distribution temporelle des événements de chaque zone au travers du temps linéaire.

Les zones spatiales permettant de classer les événements en fonction de leur lieu d'apparition peuvent correspondre à une division de l'étendue spatiale d'étude selon une grille régulière, comme c'est le cas dans la Figure 4.31.

PerSE propose également d'effectuer une classification basée sur des diagrammes de *Voronoi* paramétrables par l'utilisateur. Un diagramme de *Voronoi* consiste en un découpage d'un espace en deux dimensions, ici la zone spatiale représentée sur la carte, en cellules créées à partir d'un ensemble de points appelés *germes*. Chaque cellule contient un seul *germe*, et l'espace formé par une cellule représente l'ensemble des points de l'espace dont le *germe* correspondant est le plus proche. L'utilisateur peut choisir le nombre de *germes* utilisés pour créer le diagramme de *Voronoi*, et définir leur emplacement. Cette classification en diagrammes de *Voronoi* est illustrée dans la Figure 4.33.

En parallèle d'une représentation des événements selon un critère spatial, *PerSE* propose de représenter les événements simultanément selon un critère thématique. Les événements sont alors représentés, dans la carte et dans les diagrammes temporels, en fonction de la valeur d'un de leurs attributs thématiques, au moyen de règles de symbologie.

Dans la Figure 4.33, la variable visuelle de la couleur est utilisée pour représenter une variable thématique qualitative :

- La composante thématique de chaque événement est représentée dans la carte grâce à la couleur.
- Dans le diagramme temporel linéaire, la proportion des différentes valeurs de l'attribut, pour chaque granule temporel, est représentée par la proportion de la couleur correspondante dans chaque colonne du diagramme en bâton.
- Dans le diagramme temporel cyclique, la proportion des différentes valeurs de l'attribut, pour chaque granule temporel, est représentée par la proportion de la couleur correspondante dans chaque section du diagramme, à la manière du diagramme en secteur de Florence Nightingale [Nightingale, 1858] (Figure 4.14).

La visualisation du lien entre les composantes temporelle et spatiale des données s'effectue, à l'instar de *Time Wave*, au moyen de procédés orientés de la fenêtre spatiale vers la fenêtre temporelle. Les événements sont classés selon leur composante spatiale et représentés en fonction dans le diagramme temporel. L'utilisateur peut modifier les zones permettant de classer les événements, et ainsi modifier la représentation des événements dans le diagramme temporel.

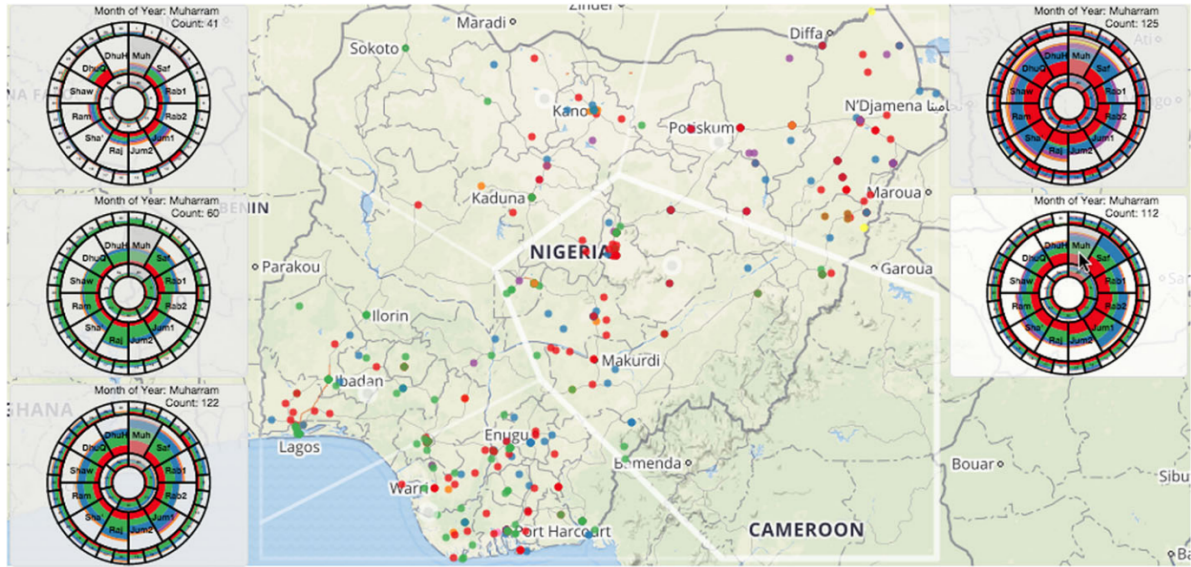


FIGURE 4.33 – Classification spatiale et représentation des événements selon les diagrammes de Voronoi [Swedberg and Peuquet, 2017]

4.2.1.4 Conséquence sur la recherche de pattern

Dans ces deux environnements, la visualisation de la relation entre la dimension spatiale et la dimension temporelle des événements s'effectue par le biais d'une représentation des événements dans les représentations graphiques du temps, selon différentes modalités de la dimension spatiale.

Il en résulte que le diagramme temporel devient l'outil central de l'analyse spatio-temporelle des données. Rechercher un pattern spatio-temporel au sein de ces environnements consiste :

- dans un premier temps, à représenter les événements en fonction de leur composante spatiale au sein de la fenêtre temporelle.
- dans un second temps,
 - dans le cas de *Time Wave*, à identifier un pattern temporel particulier impliquant des entités graphiques similaires (de même couleur ou de même valeur) dans le diagramme temporel ;
 - dans le cas de *PerSE*, à identifier un pattern temporel particulier apparaissant dans un même diagramme temporel.

Pour le cas de l'identification d'un cluster d'événements dans le temps et l'espace :

- Dans *Time Wave*, l'utilisateur doit identifier, dans le diagramme temporel mixte, un cluster temporel formé par des entités graphiques de même couleur représentant des événements d'une même zone spatiale.
- Dans le cas de *PerSE*, l'utilisateur doit repérer un cluster temporel apparaissant dans un seul diagramme temporel correspondant à une zone spatiale en particulier.

Le procédé est identique pour l'identification d'une récurrence cyclique spatialisée. L'utilisateur doit repérer, au moyen des diagrammes temporels, pour quelle échelle temporelle cyclique on observe une récurrence cyclique des événements apparaissant dans une même zone spatiale. Dans le cas de *Time Wave*, cela revient à rechercher pour quelle échelle cyclique on peut observer un alignement d'entités graphiques de même couleur sur le diagramme.

On peut supposer qu'avec une représentation adéquate des événements en fonction de leur lieu d'apparition, l'identification d'une propagation spatiale d'événements au cours du temps peut être observable à l'aide des diagrammes temporels. L'exemple suivant peut être un cas d'identification d'un tel pattern au moyen des diagrammes temporels :

- Des événements sont classés selon leur position le long d'un axe *nord-sud*.
- Les événements sont représentés sur un diagramme temporel linéaire, en fonction de leur position spatiale sur l'axe *nord-sud*, en utilisant la variable visuelle de la valeur.
- Une propagation des événements le long de cet axe *nord-sud* au cours du temps est identifiable par l'observation d'un gradient de valeur des entités graphiques le long de l'axe du diagramme temporel.

D'autres outils de classification spatiale des événements peuvent ainsi être développés afin d'améliorer la visualisation de ce pattern. Par exemple, la possibilité de classer les événements selon la distance à un point de référence ou selon un axe dimensionnel, pourrait permettre d'identifier une possible propagation spatiale des événements au cours du temps.

Les propositions apportées par ces deux environnements peuvent ainsi offrir la possibilité d'une analyse exploratoire des différents patterns spatio-temporels auxquels nous nous intéressons. Le procédé de représentation des événements dans les diagrammes temporels selon leur dimension spatiale, et les possibilités de représentation des événements selon le temps linéaire et selon plusieurs échelles temporelles cycliques, font de *Time Wave* et *PerSE* une source d'inspiration pour construire un environnement d'analyse exploratoire des dynamiques spatio-temporelles.

4.2.2 Les limites de *Time Wave* et de *PerSE*

Les environnements *Time Wave* et de *PerSE* proposent une première approche permettant une analyse exploratoire des dynamiques spatio-temporelles des phénomènes. Cependant, ces environnements présentent deux grandes limites pour l'identification des patterns auxquels nous nous intéressons :

- La première est due au fonctionnement des outils de modification des échelles temporelles cycliques représentées.
- La seconde est liée à l'orientation des procédés utilisés pour représenter les relations entre les composantes spatiale et temporelle des événements, uniquement tournée de la fenêtre spatiale vers la fenêtre temporelle.

4.2.2.1 Modification de l'échelle temporelle cyclique

L'échelle temporelle cyclique, représentée dans le diagramme temporel mixte de *Time Wave* et dans les timewheels de *PerSE*, est paramétrable dans ces deux environnements. Ceci permet d'analyser l'existence de clusters dans la distribution temporelle des événements, selon plusieurs échelles cycliques.

La possibilité de modifier la durée de l'échelle cyclique peut également être utilisée pour analyser l'existence de récurrences cycliques, de période égale à la durée des différentes échelles cycliques paramétrables :

- Cette récurrence cyclique est identifiable dans *Time Wave* par la formation d'un alignement d'entités graphiques parallèle à l'axe central.
- Dans *PerSE*, l'utilisateur doit identifier l'existence d'un cluster relatif à une échelle cyclique dans le diagramme *timewheel*, et vérifier, dans la *timeline*, que ce cluster correspond à une réapparition régulière des événements selon la période correspondante.

La modification de l'échelle temporelle cyclique dans *Time Wave*. Les échelles temporelles cycliques paramétrables dans *Time Wave* sont limitées à des durées usuelles, calendaires, telles que les jours, mois, années, etc.. (Figure 4.34). Ceci limite d'entrée les échelles cycliques pour lesquelles il est possible d'analyser l'existence d'un cluster relatif au temps cyclique, et les périodes pour lesquelles il est possible d'identifier une récurrence cyclique des événements.

Le caractère discontinu de la série d'échelles cycliques paramétrables (on passe sans transition d'une échelle cyclique d'un mois à une échelle cyclique d'un an) empêche également l'observation de la formation d'alignements graphiques, caractéristiques d'une récurrence cyclique, lors d'une modification graduelle de l'échelle cyclique du diagramme. Ceci rend plus difficile la recherche des récurrences cycliques dans les données événementielles pour l'utilisateur : l'observation de la formation d'un alignement permet au contraire de savoir si il approche d'une échelle pour laquelle une récurrence est identifiable.

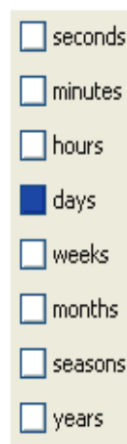


FIGURE 4.34 – Échelles temporelles cycliques paramétrables dans *Time Wave* (capture d'écran) [Li, 2010]

La limitation du nombre d'échelles cycliques paramétrables trouve son origine dans les objectifs de *Time Wave*, ce dernier pouvant être défini comme un environnement d'analyse des données au travers du temps linéaire et du temps cyclique, plutôt que comme un environnement d'analyse des récurrences cycliques existant dans les données.

Les travaux s'appuyant sur *Time Wave* [Li, 2010] ont conduit à des cas d'application sur différents phénomènes :

- l'analyse de phénomènes considérés d'emblée comme cycliques : *What are the seasonal movement patterns of the giant panda in the Qinling Mountains?*
- l'analyse de phénomènes n'étant pas considérés comme cycliques, mais pour lesquels l'échelle cyclique est utilisée comme repère temporel : *Did people migrate more in the 1980s than in the 1990s?*

Un cas d'application de *Time Wave* porte sur l'analyse de données météorologiques [Li, 2010]. Les données analysées sont des séries temporelles rattachées à des objets spatiaux. Ces derniers sont des stations météorologiques disposées à différentes localisations dans l'espace, et les séries temporelles rattachées correspondent à une mesure de l'évolution de la température pour chaque station.

Le diagramme temporel est utilisé pour représenter les moments temporels où chaque station atteint sa température quotidienne maximale (Figure 4.35). Le phénomène représenté sur

le diagramme temporel mixte, que l'on peut traduire comme *l'atteinte par une station de la température maximale quotidienne*, est cyclique de par sa construction (il existe une température maximale quotidienne par jour).

L'objectif de l'analyse est de représenter la réapparition d'événements relatifs à un même phénomène (ce qui nécessite la représentation du temps linéaire), mais en visualisant, pour chaque réapparition du phénomène, le moment d'apparition d'un événement selon échelle de temps cyclique. L'utilisateur doit analyser le déroulement du phénomène, les irrégularités dans la réapparition quotidienne du phénomène, les relations entre la localisation spatiale d'une station météorologique et le moment d'apparition, au cours d'une journée, de la température maximale quotidienne.

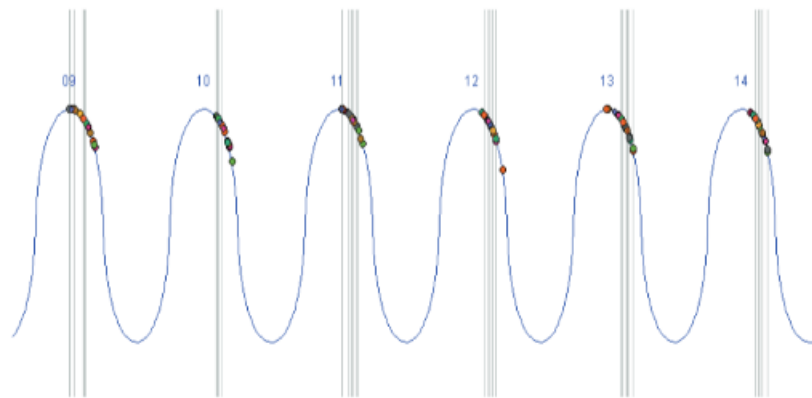


FIGURE 4.35 – Représentation de l'atteinte de la température maximale quotidienne dans *Time Wave* [Li, 2010]

L'objectif de *Time Wave* n'est pas de déterminer si un phénomène est cyclique ou non, mais de permettre une analyse d'un phénomène spatio-temporel, pouvant être déjà connu ou considéré comme cyclique, au travers d'une représentation combinée du temps linéaire et du temps cyclique. Le but est de visualiser les phénomènes au travers des cycles du temps, et non d'identifier d'éventuels comportements cycliques existant dans les phénomènes.

Cet objectif entraîne la limitation du nombre d'échelles cycliques paramétrables dans *Time Wave* aux échelles cycliques habituellement utilisées pour représenter les cycles du temps. Selon Hornsby et Egenhofer [Hornsby and Egenhofer, 2002], les échelles temporelles cycliques utilisées dans les représentations du temps sont le plus souvent basées sur les rythmes de la nature ou des activités humaines, et correspondent à des périodes usuelles et calendaires comme les jours, les mois, les heures, etc. Les échelles temporelles cycliques paramétrables dans *Time Wave* sont ainsi limitées à ces périodes usuelles.

Le cas de *PerSE*. L'objectif annoncé de *PerSE* est la recherche de périodicités (ici, l'existence de clusters relatifs à des échelles cycliques) correspondant à des périodes calendaires. Cependant, *PerSE* offre la possibilité de paramétrer n'importe quelle durée à l'échelle temporelle cyclique représentée, afin de pouvoir étendre l'analyse à des échelles calendaires n'étant pas basées sur un calendrier grégorien.

L'environnement permet par exemple de représenter la distribution temporelle des événements selon un calendrier grégorien, dont une année correspond à 364 ou 365 jours regroupés en 12 mois de 28 à 31 jours, ou selon un calendrier hégirien (également appelé calendrier musulman) dont une année correspond à 354 ou 355 jours, regroupés en 12 mois de 29 ou 30 jours (Figure 4.36).

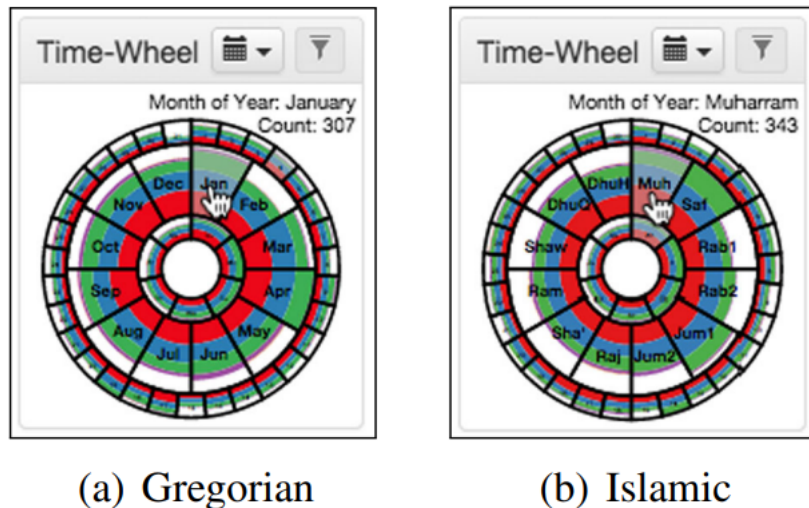


FIGURE 4.36 – Calendriers géorgien et hégirien dans *PerSE* [Swedberg and Peuquet, 2017]

La modification de l'échelle temporelle cyclique utilisée ne s'effectue pas au travers d'une possibilité de choix dans un ensemble plus large d'échelles calendaires, regroupant différents types de calendriers, mais représentant toujours un nombre fini d'échelles paramétrables. L'interface de *PerSE* propose au contraire de paramétrer directement les caractéristiques du calendrier utilisé dans la représentation du temps cyclique, permettant une représentation du temps au travers d'une année de x jours regroupés en y mois.

La possibilité d'appliquer n'importe quelle échelle cyclique à la représentation du temps semble ainsi répondre au problème posé, pour l'identification de clusters relatifs à n'importe quelle échelle cyclique, par la limitation des échelles cycliques paramétrables dans *Time Wave*. Cependant, la démarche de *PerSE* ne semble pas la plus optimale pour la recherche de récurrences cycliques.

L'analyse proposée dans *PerSE* consiste à repérer des clusters dans la distribution temporelle des événements, selon des échelles temporelles calendaires imbriquées hiérarchiquement entre elles. L'ensemble forme ainsi un ensemble d'échelles correspondant à un calendrier donné. Cet ensemble est censé être fixe le temps de l'analyse, un calendrier n'étant pas censé évoluer. L'objectif est d'analyser les événements au travers des échelles formant ce calendrier, afin d'identifier de potentielles périodicités calendaires correspondant à une des échelles de l'ensemble.

La possibilité d'affecter n'importe quelle valeur à la durée des échelles temporelles représentées, tire son origine d'une volonté de pouvoir étendre l'analyse offerte par *PerSE* à n'importe quelle échelle calendaire, quelle que soit son origine culturelle, plutôt que d'une volonté de pouvoir modifier de manière dynamique et graduelle l'échelle cyclique représentée afin d'identifier la formation de structures dans la distribution des événements lors de cette modification.

L'analyse des événements s'effectue toujours pour un ensemble fixe d'échelles temporelles cycliques, et c'est au travers de ces échelles que l'utilisateur cherche à identifier une distribution temporelle particulière. La modification de la durée des échelles cycliques utilisée s'effectue en amont de l'analyse, et non en cours d'analyse.

La modification interactive, dynamique et graduelle de l'échelle temporelle cyclique en cours d'analyse, n'est pas un critère retenu pour l'analyse des événements au travers de *PerSE*, rendant difficile une recherche visuelle de récurrences cycliques dans l'apparition des événements.

Une récurrence cyclique est visible, lorsque la durée de l'échelle cyclique du diagramme temporel correspond à la période du cycle. Pour rechercher ce dernier, l'utilisateur doit savoir, lorsqu'il modifie la durée de l'échelle cyclique, si celle-ci s'approche ou s'éloigne de la période du cycle. La possibilité de modifier la durée de l'échelle cyclique, de manière interactive, dynamique et graduelle est alors déterminante pour effectuer une recherche visuelle de récurrences cycliques, comme le montre l'exemple de l'environnement *Spiral Graph* [Weber et al., 2001].

Une possible source d'inspiration, *Spiral Graph*. L'environnement *Spiral Graph* [Weber et al., 2001] peut être une source d'inspiration pour palier aux limites de *Time Wave* et *PerSE* concernant la modification de l'échelle cyclique représentée dans les diagrammes temporels. Le diagramme temporel utilisé par *Spiral Graph* consiste en un diagramme en spirale, représentant une version linéaire de la *timewheel*. La courbe du diagramme représente le temps linéaire, et l'enroulement de cette courbe autour d'un point central représente la répétition d'une échelle cyclique.

Un des objectifs de *Spiral Graph* consiste à identifier visuellement des récurrences cycliques dans des séries temporelles. Pour cela, le diagramme temporel proposé par *Spiral Graph* offre la possibilité d'effectuer une modification dynamique et graduelle de l'échelle cyclique représentée. L'utilisateur peut faire évoluer la durée de l'échelle cyclique de manière progressive, par une série de modifications y ajoutant ou y soustrayant un pas de temps régulier, jusqu'à faire apparaître un alignement graphique dans le diagramme temporel.

La Figure 4.37 présente deux diagrammes temporels en spirale, représentant la même série temporelle.

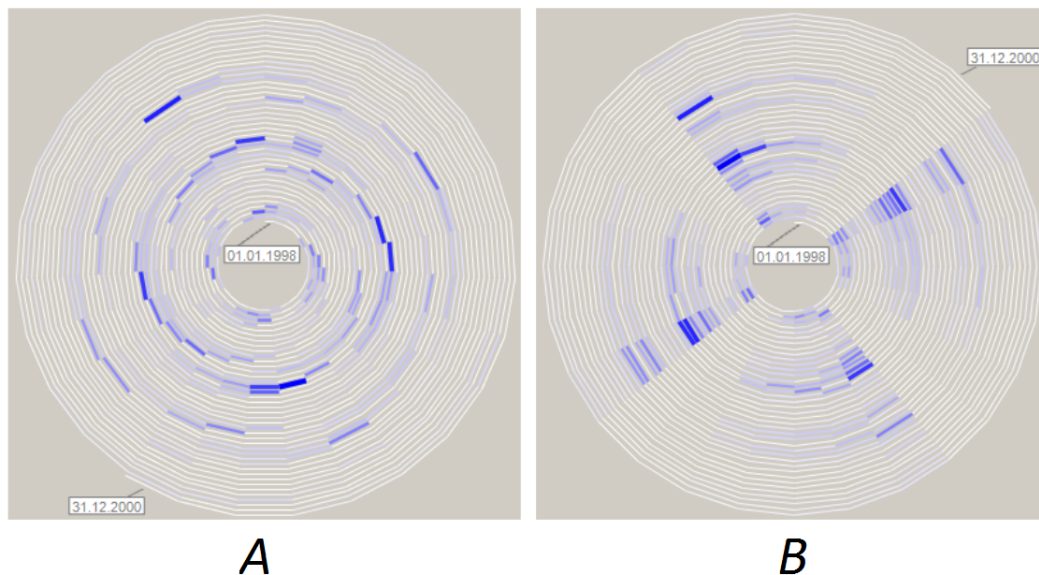


FIGURE 4.37 – La recherche de récurrences cycliques dans les données temporelles au moyen de *Spiral Graph* [Weber et al., 2001]

Chaque section de la courbe du diagramme représente une donnée dans la série temporelle. La valeur de la variable portée par la série temporelle est représentée par la saturation de la couleur bleue de chaque section du diagramme. Chaque diagramme représente les données au travers d'une échelle temporelle cyclique différente :

- Dans le premier diagramme (Repère A de la Figure 4.37), la durée de l'échelle cyclique est de 27 jours. On observe que les sections correspondant à de fortes valeurs commencent à

former un alignement en spirale.

- Dans le second diagramme (Repère B de la Figure 4.37), la durée de l'échelle cyclique est de 28 jours. Les sections de la courbe correspondant à de fortes valeurs forment alors quatre alignements, parallèles à des rayons de la spirale, et séparés entre eux par un même angle.

L'observation de quatre alignements, séparés régulièrement, pour une échelle cyclique de 28 jours traduit l'existence d'une récurrence cyclique des fortes valeurs de la série temporelle, d'une période égale à sept jours, soit le quart de l'échelle temporelle cyclique représentée.

Cette possibilité de modifier dynamiquement et graduellement la durée de l'échelle cyclique peut être appliquée à d'autres diagrammes temporels, comme celui proposé par *Time Wave*, afin d'identifier des récurrences cycliques dans les données d'événements. Le but est alors d'identifier la formation d'un alignement formé par les entités graphiques représentant les événements. Cet alignement prend d'abord la forme d'une vague (comme il prend d'abord la forme d'une spirale dans *Spiral Graph*). Cette vague prend progressivement la forme d'une ligne parallèle à l'axe central du diagramme, à mesure que la durée de l'échelle cyclique s'approche de la période du cycle recherché.

L'observation de la formation du motif graphique révélant une récurrence cyclique, au cours d'une modification graduelle de la durée de l'échelle temporelle, permet alors à l'utilisateur d'identifier si la durée de l'échelle temporelle s'approche ou s'éloigne de la durée du cycle recherché.

Dans le premier diagramme (Repère A de la Figure 4.37), l'alignement des sections représentant de fortes valeurs, sous la forme en spirale, ne paraît pas évident. Une solution pour améliorer l'observation de la formation de ces alignements consiste à réduire le pas de temps utilisé pour la modification graduelle de l'échelle cyclique. En effectuant une modification de l'échelle cyclique avec un pas de temps plus fin, comme une heure, la formation de l'alignement graphique apparaît de manière plus évidente.

4.2.2.2 Les limites de la représentation dans les diagrammes temporels en fonction de l'espace

Comme énoncé précédemment, le principe de l'analyse de structures spatio-temporelles dans les données d'événements, au sein des environnements *Time Wave* et *PerSE*, consiste à classer les événements selon un critère spatial, à les représenter en fonction au sein de la fenêtre temporelle, et à analyser la distribution temporelle de chaque groupe d'événements au travers du ou des diagramme(s) temporel(s) (Figure 4.38).

Ce procédé offre une première approche pour l'analyse des événements spatio-temporels, et permet notamment d'analyser la distribution spatio-temporelle des événements au travers de régions spécifiques déterminées par l'utilisateur. Celui-ci peut choisir de s'intéresser plus particulièrement à la distribution temporelle d'événements apparaissant dans une zone spécifique comme une vallée de montagne, le voisinage d'un axe routier, une zone urbaine, une forêt, etc.

Ces zones de classification spatiale peuvent être créées à partir de l'analyse d'autres données géographiques, afin de croiser des thématiques de recherche : les événements peuvent être classés selon des zones représentant différents niveaux de densité de population, permettant de croiser cette thématique avec l'analyse spatio-temporelle des événements.

Enfin, une classification des événements peut être effectuée en fonction de la distance de

l'événement à un objet spatial de référence, afin d'étudier une propagation spatiale des événements dans le temps depuis un point particulier dans l'espace, ou le long d'un axe directionnel particulier.

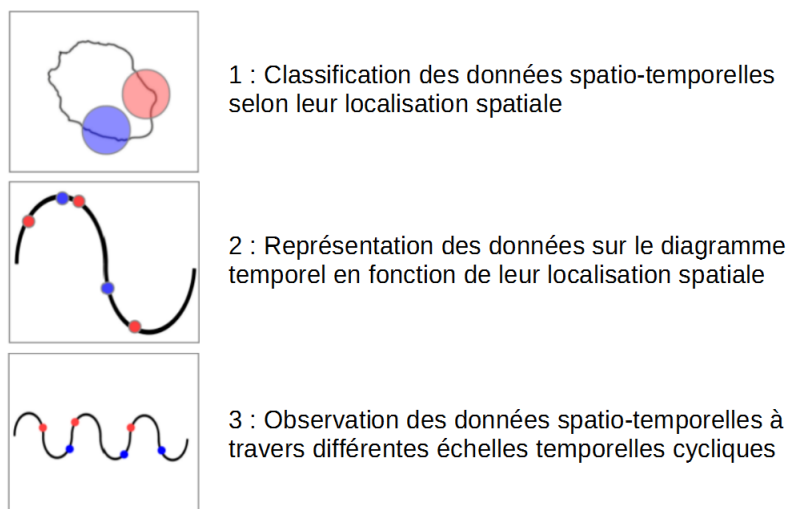


FIGURE 4.38 – Analyse des dynamiques spatio-temporelles par classification spatiale des événements dans un diagramme de type *Time Wave*

Ce procédé se base cependant sur l'hypothèse selon laquelle l'utilisateur, qui détermine les critères de classification spatiale des événements, suspecte à l'avance une certaine distribution spatio-temporelle des événements. Pour identifier des patterns spatio-temporels, les événements classés dans chacune des zones spatiales créées doivent présenter une distribution temporelle particulière.

On peut émettre l'hypothèse que la distribution temporelle d'événements proches spatialement les uns des autres est plus susceptible de présenter une structure particulière. Dans le cas de *PerSE*, la classification spatiale des événements effectuée grâce à des polygones de *Voronoi* permet de regrouper les événements spatialement proches en positionnant les *germes* sur les concentrations spatiales d'événements visibles sur la carte.

Cependant, des événements localisés dans une zone plus restreinte à l'intérieur d'un polygone de *Voronoi*, présentant une distribution temporelle différente de celle d'autres événements contenus dans le polygone, sont difficilement décelables au travers de cette approche. De même, des événements présentant une structure particulière dans leur distribution temporelle peuvent posséder une distribution spatiale qui ne serait pas englobée dans un polygone de *Voronoi*, comme dans le cas d'événements apparaissant le long d'un fleuve.

L'analyse spatio-temporelle des événements sur les diagrammes temporels, par le biais de la classification spatiale, reste ainsi très dépendante des critères de classification utilisés. La définition de ces critères par l'utilisateur nécessite que ce dernier suspecte une zone spatiale à analyser, ou une organisation particulière du territoire au travers de laquelle effectuer l'analyse. Ce besoin d'une hypothèse énoncée, *a priori*, sur la distribution spatio-temporelle des données limite les possibilités d'une analyse exploratoire pour déceler les dynamiques spatio-temporelles.

Une solution, pour répondre aux limites posées par la représentation temporelle des événements en fonction de l'espace, peut être d'inverser la représentation des événements, et d'orienter celle-ci de la fenêtre temporelle vers la fenêtre spatiale. Le principe est alors de représenter les événements dans la fenêtre spatiale en fonction de leur moment d'apparition.

4.2.3 L'intégration du temps dans les représentations cartographiques des environnements à multi-fenêtrage

L'identification d'un pattern spatio-temporel dans les approches *Time Wave* et *PerSE* consiste à identifier une structure temporelle, un cluster ou une récurrence cyclique, formée par des événements correspondant à une même modalité de la dimension spatiale.

À l'inverse, l'identification d'un pattern au travers d'une représentation des événements, orientée de la fenêtre temporelle vers la fenêtre spatiale, consiste à identifier une structure spatiale, un cluster, formé par des événements correspondant à une même modalité de la dimension temporelle. Cette modalité consiste en l'appartenance d'un événement à un intervalle temporel, relatif au temps linéaire ou à une échelle temporelle cyclique.

Les procédés utilisés doivent alors permettre de représenter les événements selon leur moment d'apparition, observé au travers d'une échelle temporelle linéaire ou cyclique.

Différentes approches, permettant d'intégrer le temps dans les représentations cartographiques, peuvent être associées à des environnements à multi-fenêtrage pour permettre cette représentation. Ces approches peuvent faire appel à l'utilisation de cubes spatio-temporels, de dispositifs *small-multiples*, ou de règles de symbologie.

4.2.3.1 L'intégration du temps par la *Time Geography*

Les environnements *Wakame* [Forlines and Wittenburg, 2010] (Repère *A* de la 4.39) et *Data Vases* [Thakur and Rhyne, 2009] (Repère *B* de la 4.39) offrent un exemple d'intégration des principes de la *Time Geography* au sein d'un environnement à multi-fenêtrage.

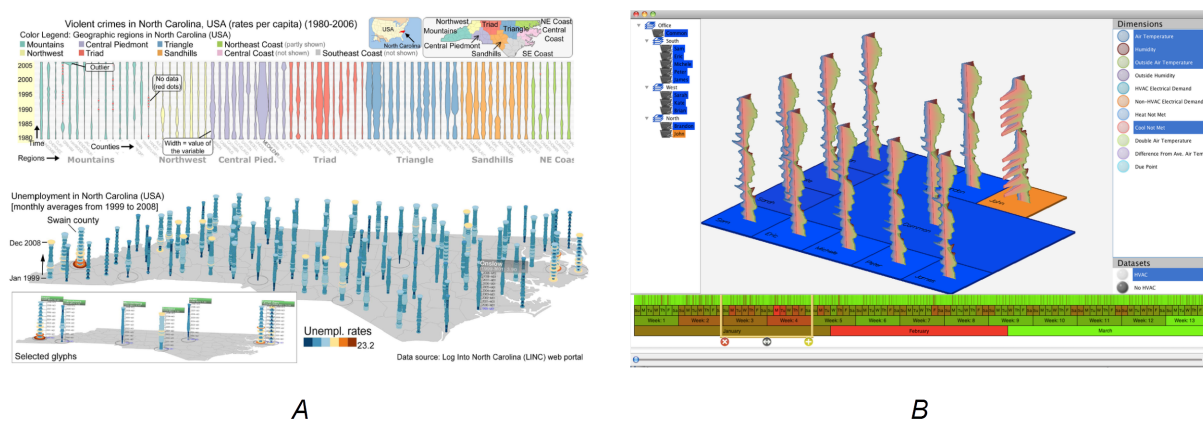


FIGURE 4.39 – Captures d'écran de *Wakame* [Forlines and Wittenburg, 2010] et de *Data Vases* [Thakur and Rhyne, 2009]

Ces environnements représentent l'évolution de séries temporelles rattachées à des entités spatiales pérennes :

- Ces entités correspondent à un découpage du territoire. Dans le cas de *Wakame*, ces entités représentent des bureaux, dans le cas de *Data Vase*, des divisions administratives.
- La fenêtre temporelle est occupée par un ou plusieurs diagrammes temporels.
- La fenêtre spatiale est occupée par une représentation en trois dimensions, composée d'une carte en deux dimensions, et d'un diagramme temporel linéaire. L'axe vertical de la fenêtre spatiale forme l'axe temporel de ce diagramme temporel. La valeur de la variable portée par la série temporelle correspondante est représentée par un axe orthogonal à l'axe linéaire

du diagramme, ainsi que par une couleur.

Ces environnements utilisent une modélisation endurente des phénomènes spatio-temporels, et représentent l'évolution temporelle d'entités spatiales pérennes immobiles (les entités spatiales représentées étant les éléments d'un découpage d'un territoire). Ainsi, si *Wakame* et *Data Vases* sont des exemples d'intégration du temps dans les représentations spatiales d'environnements à multi-fenêtrage, la représentation du temps est définie par l'espace, à l'image des approches proposées dans *Time Wave* et *PerSE*. Ces environnements ont en commun une représentation du temps en lien avec une division du territoire :

- *Time Wave* et *PerSE* représentent la distribution temporelle d'événements classés selon leur lieu d'apparition.
- *Wakame* et *Data Vases* représentent l'évolution temporelle d'entités spatiales réparties sur le territoire.

Ainsi, *Wakame* et *Data Vases* ne sont pas des exemples d'utilisation du cube spatio-temporel permettant une représentation orientée du temps vers l'espace.

En revanche, l'utilisation d'une modélisation occurrente dans des environnements similaires, permet d'effectuer le lien entre les composantes temporelle et spatiale d'une série d'événements au sein d'un cube spatio-temporel (Figures 3.8 et 4.40). Dans ces environnements, les événements peuvent être représentés au sein d'un cube spatio-temporel en fonction :

- de leur lieu d'apparition sur un plan horizontal du cube ;
- de leur moment d'apparition sur l'axe vertical du cube.

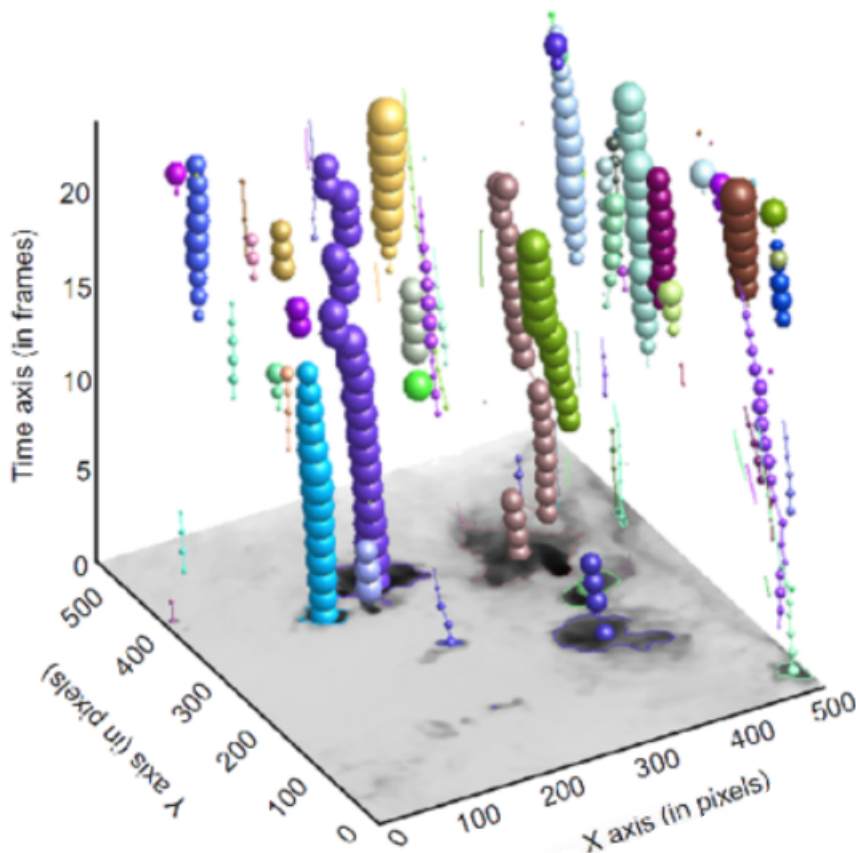


FIGURE 4.40 – Utilisation d'un cube spatio-temporel pour la représentation d'entités spatio-temporelles occurrentes [Gatalsky et al., 2004, Tominski et al., 2005, Turdukulov et al., 2007]

L'intégration de représentations en trois dimensions, dans la fenêtre spatiale d'un environnement à multi-fenêtrage, permet ainsi d'effectuer le lien entre le moment et le lieu d'apparition des événements, sans que ces derniers soient classés à l'avance selon un critère spatial.

Cependant, du fait du caractère linéaire de l'axe dédié à la représentation du temps dans un cube spatio-temporel, cette approche ne peut être la plus optimale pour la recherche de patterns relatifs à des échelles temporelles cycliques. Pour identifier une répétition cyclique d'événements, un utilisateur devra observer l'existence d'une même distance graphique, séparant les événements cycliques le long de l'axe vertical du cube spatio-temporel. Le problème rencontré ici est alors similaire à celui posé par l'identification de récurrences cyclique au moyen d'un diagramme de type *timeline* (Section 4.1.2.4).

4.2.3.2 L'intégration du temps par les dispositifs *small-multiples*

L'environnement *VIS-STAMP* [Guo et al., 2006] (Figure 4.41) offre un exemple d'intégration d'un dispositif *small-multiples* dans un environnement à multi-fenêtrage, orienté de la fenêtre temporelle vers la fenêtre spatiale, à l'inverse de *PerSE*.

VIS-STAMP représente l'évolution de séries temporelles rattachées à des entités spatiales pérennes :

- Ces entités représentent les différents États des États-Unis d'Amérique.
- La granularité temporelle utilisée est l'année.
- La fenêtre temporelle est occupée par plusieurs diagrammes temporels linéaires, représentant l'évolution d'une variable portée par une série temporelle pour chaque État.
- La fenêtre spatiale est occupée par une collection de cartes, chaque carte représentant la valeur portée par la série temporelle pour une année particulière.

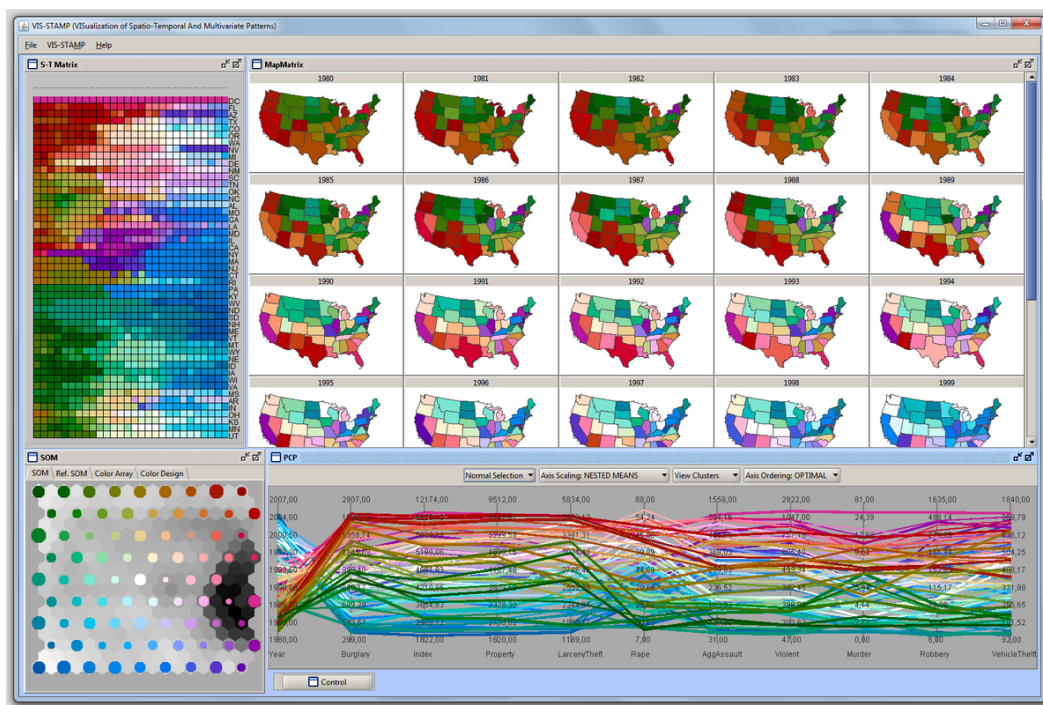


FIGURE 4.41 – Capture d'écran de *VIS-STAMP* [Guo et al., 2006]

La fenêtre spatiale de *VIS-STAMP* intègre le temps de manière linéaire, chaque carte représentant l'état de l'espace pour une succession de granules temporels dans le temps linéaire.

Une adaptation de ce dispositif peut être appliquée à la représentation de séries d'événements. Chaque carte représente alors la distribution spatiale d'événements apparaissant dans différents intervalles temporels successifs.

Un dispositif *small-multiples* peut également être adapté à la représentation d'une série d'événements au travers du temps cyclique. Une collection de cartes peut représenter la distribution spatiale d'événements apparaissant dans différents intervalles temporels, relatifs à une échelle temporelle cyclique. Pour une échelle cyclique d'une année, une collection de quatre cartes peut présenter la distribution spatiale des événements apparaissant durant les quatre saisons de l'année. La carte correspondant à l'hiver représente la distribution spatiale des événements apparaissant l'hiver, pour l'ensemble des années de l'étendue temporelle d'étude.

On peut également avancer que la collection de cartes de *VIS-STAMP* peut permettre une lecture du changement dans l'espace selon un point de vue linéaire, et également au travers d'une échelle temporelle cyclique :

- Horizontalement, les cartes se succédant représentent l'état de l'espace pour différentes années successives.
- Verticalement, les cartes se succédant représentent l'état de l'espace pour différentes années se succédant tous les cinq ans.

Une récurrence cyclique, selon une période de cinq ans, pourrait ainsi se traduire par l'apparition d'un même motif graphique sur les cartes d'une même colonne. L'utilisation d'un dispositif *small-multiples* peut ainsi permettre une intégration du temps cyclique et du temps linéaire dans la fenêtre spatiale des environnements à multi-fenêtrage.

Si les données représentées par *VIS-STAMP* correspondent à une modélisation endurente du changement, les *small-multiples* peuvent également être intégrés au sein d'environnements à multi-fenêtrage pour la représentation d'événements occurrents. La fenêtre spatiale de ces environnements peut alors représenter les événements selon leur moment d'apparition, à travers différentes représentations cartographiques, à l'image d'une collection de cartes de flux (Figure 4.42).

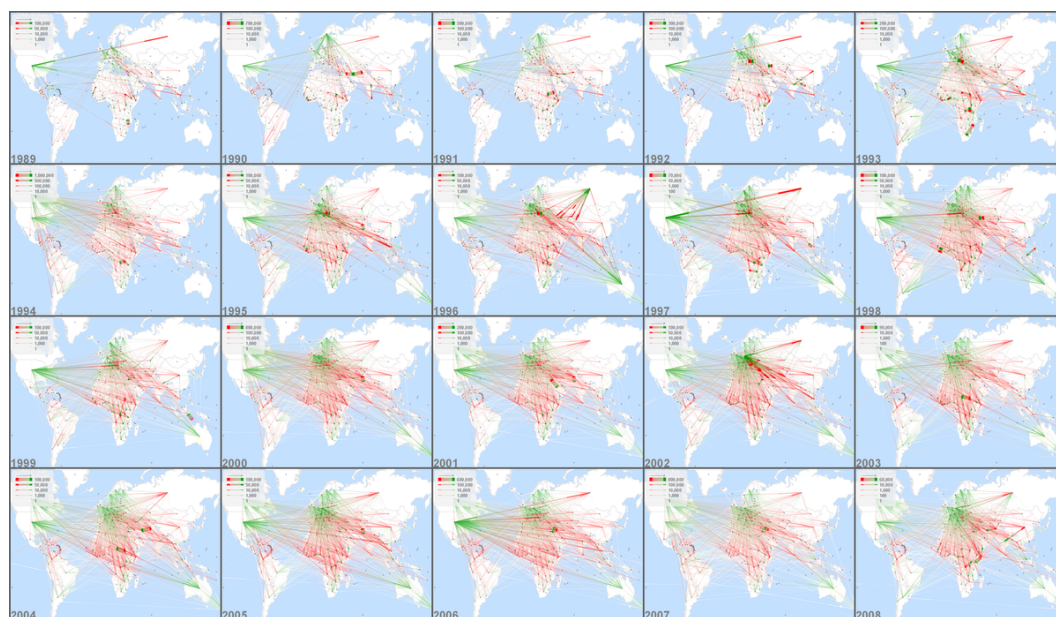


FIGURE 4.42 – Utilisation de dispositif *small-multiples* pour la représentation d'entités occurrentes [Tufte, 1983]

L'étendue temporelle étudiée est divisée en plusieurs intervalles temporels de durée égale. Les événements sont classés selon l'appartenance de leur moment d'apparition à un de ces intervalles temporels, et sont représentés dans la carte correspondant à cet intervalle, au sein d'une collection de cartes.

L'utilisation des dispositifs *small-multiples*, au sein des environnements de visualisation à multi-fenêtrage, permet alors de visualiser le lien entre les dimensions spatiale et temporelle des événements, sans devoir effectuer au préalable une classification de ces derniers selon un critère spatial.

Cependant, la difficulté de lecture dans le cas d'un trop grand nombre de cartes [Kaddouri et al., 2014] peut rendre difficile une analyse exploratoire des phénomènes spatio-temporels, surtout si l'utilisateur souhaite représenter les événements selon des intervalles temporels relativement fins. Par exemple, une représentation des événements au travers d'une échelle annuelle, en classant les événements selon leur mois d'apparition, nécessite une douzaine de cartes, ce qui peut rendre difficile l'analyse au travers d'une collection de cartes.

4.2.3.3 L'intégration du temps par la symbologie

La carte animée de l'environnement *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011] (Figure 4.43) propose une utilisation des règles de symbologie permettant d'intégrer le temps linéaire et le temps cyclique au sein d'une représentation cartographique. Associée à une représentation originale des événements spatio-temporels dans la carte, cet environnement permet l'identification de clusters dans la distribution spatio-temporelle des événements, selon le temps linéaire ou selon le temps cyclique, directement dans la représentation cartographique.

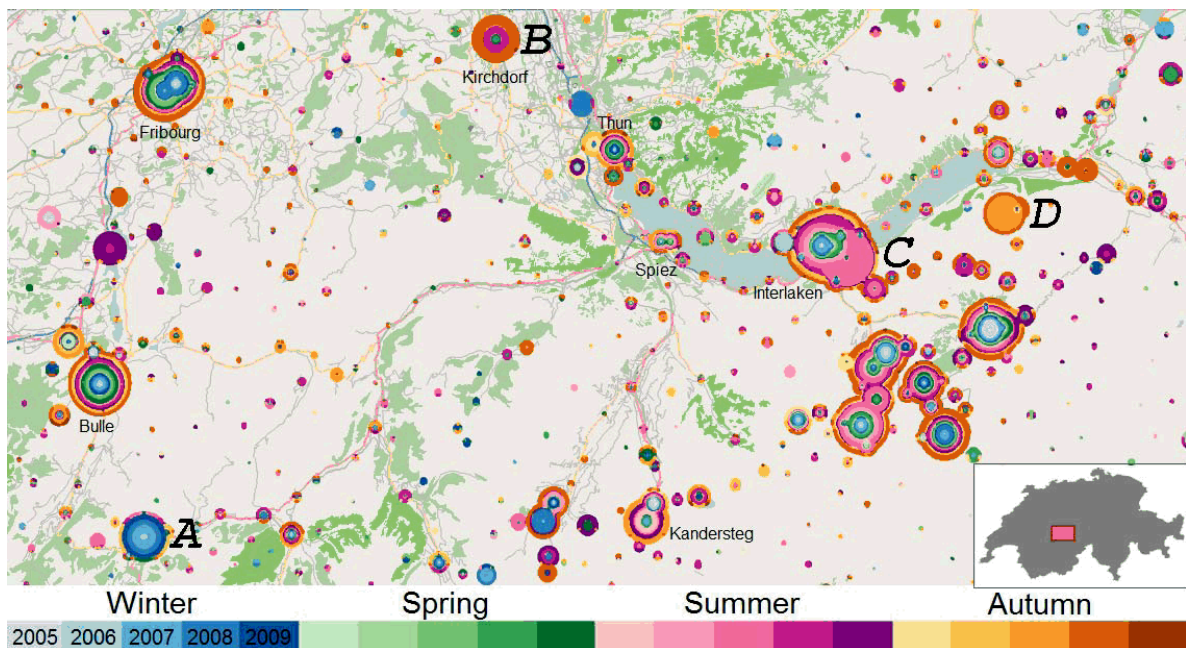


FIGURE 4.43 – Représentation d'événements au travers du temps linéaire et du temps cyclique dans *Growth Ring Map* [Bak et al., 2009, Andrienko et al., 2011]

Le principe de *Growth Ring Maps* est de représenter chaque événement, sur la carte, spatio-temporel par un pixel positionné en fonction du lieu d'apparition de l'événement. Les événements apparaissent au cours de l'animation en fonction de leur moment d'apparition. Les pixels représentant des événements restent visibles à l'écran une fois apparu. Lorsque des événements apparaissent au même endroit dans l'espace au cours du temps, les pixels correspondants ne se

superposent pas, mais s'agrègent, formant un amas circulaire de pixels dont la taille est fonction du nombre d'événements apparus au même endroit.

Chaque événement est représenté sur la carte :

- en fonction de son moment d'apparition relatif à une échelle cyclique d'une année, considéré comme une variable nominale pouvant prendre quatre valeurs ;
Ces valeurs correspondent aux quatre saisons de l'année. Ce moment d'apparition cyclique détermine la couleur du pixel correspondant.
- en fonction de son moment d'apparition, relatif à une échelle temporelle linéaire, considéré comme une variable ordonnée pouvant prendre cinq valeurs.
Ces valeurs correspondent aux cinq années de l'étendue temporelle étudiée. Ce moment d'apparition linéaire détermine la saturation de la couleur du pixel correspondant (la saturation de la couleur traduisant ici la variable visuelle de la valeur).

Lorsque de nouveaux pixels apparaissent dans un agrégat, les pixels de même couleur, correspondant aux événements apparaissant durant les mêmes saisons, se rassemblent, formant des cercles concentriques de couleurs différentes. Un agrégat de pixel peut ainsi comporter, au maximum, quatre anneaux de couleur différente, représentant les événements apparaissant durant les quatre saisons de l'année.

À l'intérieur de chaque anneau, les pixels représentant des événements apparaissant la même année se regroupent, formant à l'intérieur d'un anneau de couleur différents anneaux concentriques possédant la même valeur de saturation. Les pixels les plus clairs représentant les événements les plus anciens sont placés vers l'intérieur, et les pixels les plus sombres, représentant les événements les plus récents, sont placés vers l'intérieur.

Lors de l'animation, on observe la formation d'anneaux concentriques de différentes couleurs, devenant de plus en plus gros, d'où le nom *Growth Ring Maps* (*Cartes d'anneaux croissants*).

On observe alors dans la Figure 4.43 certains agrégats de pixels ne présentant qu'une seule couleur, ou bien présentant une couleur majoritaire.

- Le repère *A* présente un agrégat de pixels majoritairement bleus, avec différentes valeurs de saturation. Cet agrégat correspond à une concentration spatiale d'événements survenant majoritairement en hiver, pour différentes années. Cet agrégat indique ainsi la présence d'un cluster relatif à une échelle d'un an. Les événements apparaissant chaque hiver, on peut également conclure de cet agrégat l'existence d'une récurrence cyclique spatialisée.
- Le repère *B* présente un agrégat formé de pixels de trois couleurs, mais présentant la même valeur de saturation. Cet agrégat correspond à une concentration spatiale d'événements survenant durant le printemps, l'été, et l'automne de l'année 2008. Cet agrégat indique ainsi la présence d'un cluster relatif au temps linéaire, étendu sur une durée de trois saisons.
- Le repère *C* présente un agrégat de pixels majoritairement roses, avec différentes valeurs de saturation. Cet agrégat correspond à une concentration spatiale d'événements survenant majoritairement en été, pour différentes années. Cet agrégat indique ainsi la présence d'un cluster relatif à une échelle d'un an.
- Le repère *D* présente un agrégat de pixels oranges, possédant majoritairement la même valeur de saturation. Cet agrégat correspond à une concentration spatiale d'événements survenant majoritairement durant l'automne de l'année 2007. Cet agrégat indique ainsi la présence d'un cluster relatif au temps linéaire, étendu sur une durée d'une saison.

Cette représentation cartographique permet ainsi de repérer des structures dans la distribution spatio-temporelle des événements. La composante temporelle des événements peut être représentée à la fois selon le temps linéaire, et à la fois selon une échelle temporelle cyclique, contrairement au cas des cubes spatio-temporels. La représentation des événements dans une même représentation cartographique permet également une meilleure lisibilité que les dispositifs *small-multiples*, pour la comparaison des événements apparaissant durant les différents intervalles temporels, linéaires ou cycliques.

L'approche illustrée par cette carte peut alors fournir un outil d'analyse exploratoire des dynamiques spatio-temporelles, effectuée directement sur la représentation cartographique, et ne nécessitant pas l'énonciation d'hypothèses *a priori* sur la distribution spatio-temporelle des événements.

4.2.3.4 La combinaison de règles de symbologie et d'outils d'agrégation dans *Growth Ring Map*

La carte proposée par *Growth Ring Map* est animée, mais une représentation similaire peut être développée sans avoir recours à l'animation, et correspond plus à notre objectif. Notre analyse portant sur la distribution des événements dans l'espace et dans le temps, notre analyse doit s'effectuer simultanément sur l'ensemble des événements apparaissant durant l'étendue spatio-temporelle d'étude. L'utilisation d'une animation, masquant les événements apparaissant après le moment temporel donné, ne nous est pas utile.

L'approche de *Growth Ring Map* est basée sur trois procédés :

- l'utilisation de règles de symbologie permettant de représenter les événements sur la carte, en fonction de leur moment d'apparition, relatif au temps linéaire ou relatif à une échelle cyclique ;
- la représentation des événements en agrégats, permettant de repérer sur la carte les zones présentant de fortes concentrations d'événements ;
- La représentation graphique de ces agrégats en cercles concentriques, regroupant les événements selon la classe à laquelle ils appartiennent, ici l'intervalle temporel durant lequel ils apparaissent. En regroupant les événements "semblables" à l'intérieur d'un agrégat, cette représentation permet d'identifier la proportion des différentes classes d'événements au sein d'un agrégat. Ici, l'analyste peut identifier, au sein d'un agrégat, la proportion d'événements apparaissant durant l'hiver, l'été, etc.

L'identification de structures spatio-temporelles dans la carte au moyen de *Growth Ring Map* est permise par ces trois procédés. L'utilisation d'agrégats permet de repérer des concentrations spatiales. L'utilisation de règles de symbologie combinée à la représentation graphique des agrégats permet d'identifier, à l'intérieur d'un agrégat, une concentration d'événements dans le temps. Un environnement voulant offrir la même analyse exploratoire que permet *Growth Ring Map* doit également permettre d'observer les proportions des différents intervalles d'apparition dans un groupe d'événements spatialement proches.

4.2.3.5 L'utilisation de diagrammes pour représenter la proportion d'éléments agrégés

Si la représentation des événements en pixels, s'agrégeant en cercles concentriques selon leur classe d'appartenance, est la solution utilisée par *Growth Ring Map*, d'autres solutions existent, permettant de représenter simultanément, dans une carte, la proportion de différentes classes de valeurs rattachées à des entités spatiales.

Une solution pouvant être mise en oeuvre est illustrée par la Figure 4.44 [Andrienko and Andrienko, 2006]. Une série d'événements y est représentée dans un cube spatio-temporel sous la forme d'entités ponctuelles :

- Chaque événement possède un attribut quantitatif A .
- Chaque événement possède trois attributs qualitatifs B , dont la valeur est exprimée en pourcentage. La somme des valeurs des attributs B atteint 100%.

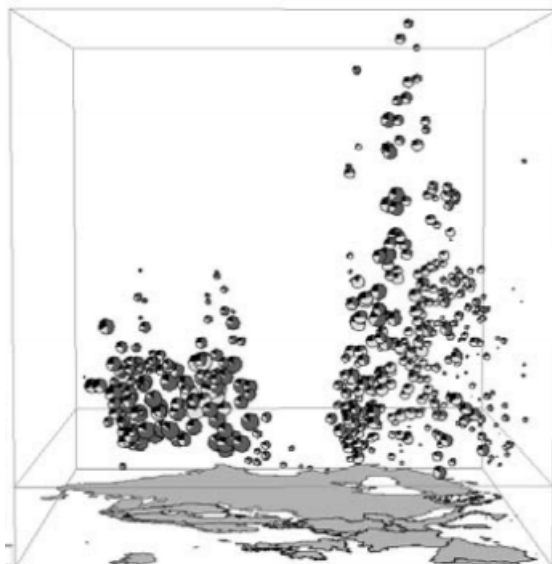


FIGURE 4.44 – Utilisation de *Pie Charts* pour représenter des événements dans un cube spatio-temporel [Andrienko and Andrienko, 2006]

Les événements sont alors représentés plus ou moins gros en fonction de l'attribut A . Les entités graphiques représentant les événements prennent alors la forme de diagrammes circulaires, ou *Pie Chart*. Les différentes sections des *Pie Chart* représentent les valeurs de chaque attribut B .

On peut imaginer une application de ces représentations en *Pie Chart*, à la représentation d'agrégats d'événements, classés selon leur moment d'apparition, dans le cadre d'une approche similaire à *Growth Ring Map* :

- La taille d'une entité *Pie Chart* représenterait le nombre d'événements agrégés.
- Les sections d'une entité *Pie Chart* représenteraient les proportions des différentes classes d'événements, selon l'intervalle temporel durant lequel ils apparaissent.

L'intégration de diagrammes dans les représentations cartographiques reste controversée en géographie. Selon Bertin [Bertin, 1973], les représentations graphiques utilisant des séries de diagrammes, que ceux-ci soient disposés au sein d'une collection de diagrammes ou au sein d'un cartogramme, sont inefficaces pour répondre à la majorité des questions que l'information représentée est susceptible de susciter.

Cette inefficacité tient au grand nombre d'images que l'oeil est obligé de sélectionner, de mémoriser, et de comparer, pour parvenir à analyser l'information. L'analyste ne peut pas effectuer une lecture d'ensemble d'une série de diagrammes. Il doit effectuer une succession de lecture élémentaire sur chaque diagramme, avant de comparer les informations perçues. Sans image d'ensemble réellement perceptible, l'utilisateur ne peut pas effectuer de comparaison efficace des informations portées par chaque diagramme.

Selon Bertin [Bertin, 1973], les collections de diagrammes peuvent être satisfaisantes pour tout lecteur se contentant de questions élémentaires, telles que la recherche d'une valeur représentée dans un des diagrammes, ou de questions ne portant que sur l'information portée par un seul de ces diagrammes. Celles-ci sont cependant inefficaces pour répondre à des questions nécessitant des niveaux de lecture moyens ou d'ensemble.

On peut cependant objecter à Bertin que l'analyse exploratoire n'a pas pour objectif de répondre à une question spécifique, en une seule fois, au moyen d'une lecture d'ensemble de la carte. Selon le procédé décrit dans l'*Information Seeking Mantra* [Shneiderman, 2003, Andrienko and Andrienko, 2006], l'analyse exploratoire passe d'abord par une observation globale de l'ensemble d'un jeu de données, avant d'approfondir l'analyse sur un sous-ensemble du jeu de données. Ce dernier est identifié lors de la première phase de l'analyse, selon des critères pouvant se révéler suffisamment élémentaires pour rendre pertinente l'utilisation de séries de diagrammes.

La Figure 4.45 est une représentation d'une série d'articles de l'encyclopédie en ligne *Wikipedia* [Bruggmann et al., 2013].

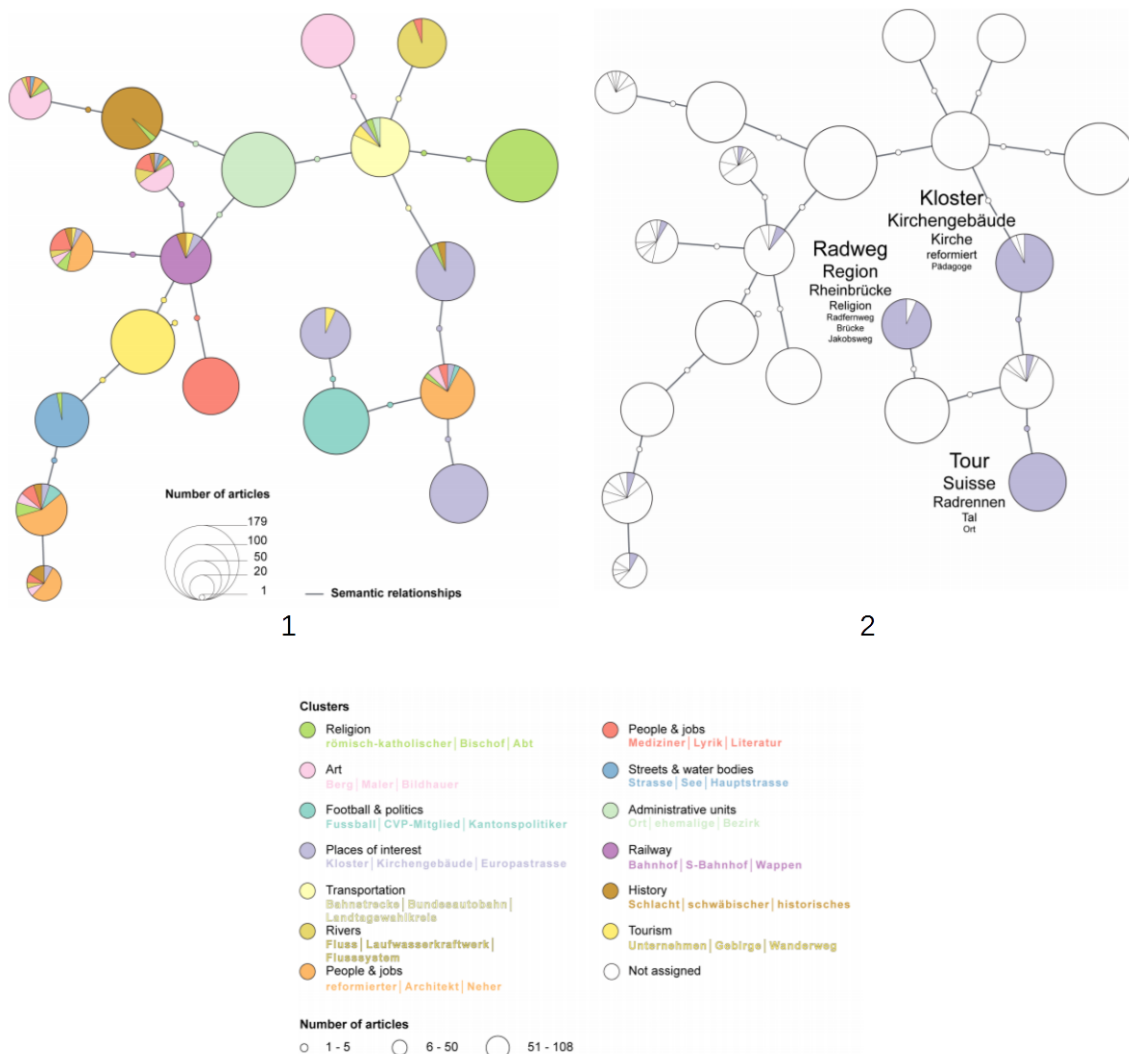


FIGURE 4.45 – Représentation d'articles *Wikipedia* à travers une série de *Pie Charts* [Bruggmann et al., 2013]

Les articles y sont représentés au travers d'une série de *Pie Charts* :

— Chaque *Pie Chart* représente un regroupement d'articles comportant des attributs sémant-

tiques similaires.

- La taille de chaque *Pie Chart* représente le nombre d'articles regroupés.
- Dans chaque *Pie Chart*, une couleur représente le thème du contenu de l'article.

Une analyse exploratoire au travers de cette représentation s'effectue ainsi selon le procédé décrit dans l'*Information Seeking Mantra* [Shneiderman, 2003, Andrienko and Andrienko, 2006] :

- Dans un premier temps, l'analyste effectue une observation globale de l'ensemble des *Pie Charts*, afin d'identifier les éléments sur lesquels approfondir l'analyse (Repère 1 de la Figure 4.45).
Le critère utilisé, pour déterminer ces éléments, peut être défini par une question : *Où sont présents les articles traitant de lieux d'intérêt?* Pour répondre à cette question élémentaire, l'analyste doit identifier les *Pie Charts* comportant un secteur de couleur mauve.
- Les éléments identifiés sont un sous-ensemble de *Pie Charts* possédant un secteur de couleur mauve (Repère 2 de la Figure 4.45).
L'analyse peut être approfondie, à travers une question supplémentaire, focalisée sur ce sous-ensemble : *Quel regroupement présente le plus d'articles?*
- L'objet d'analyse est alors réduit à un seul regroupement, représenté par un seul *Pie Charts*.
L'analyse peut alors devenir, de manière itérative, de plus en plus précise, à travers des questions de plus en plus détaillées, mais focalisées sur un objet d'analyse de plus en plus restreint.

Si l'intégration de diagrammes dans la carte reste décriée en géographie, son utilisation nous semble pertinente dans le cadre d'une analyse exploratoire. Les critiques énoncées à l'égard de ces représentations nous semblent être contrebalancées par le caractère itératif de l'analyse exploratoire, faisant intervenir une succession d'observations élémentaires, effectuées à différents niveaux de lecture.

L'utilisation de diagrammes nous semble ainsi être une solution pour représenter la proportion d'éléments agrégés.

4.2.4 Conclusion

Une représentation cartographique permettant :

- d'intégrer le temps, selon une échelle linéaire ou selon une échelle cyclique, au moyen de règles de symbologie ;
- de représenter les événements spatialement proches, en maintenant observable la proportion des différents intervalles d'apparition de ces événements ;

permet une approche d'analyse exploratoire de patterns dans la distribution spatio-temporelle des événements, sans que l'analyste ait à énoncer *a priori* une hypothèse sur la distribution des événements.

L'intégration de cette représentation dans un environnement de géovisualisation à multi-fenêtrage permet alors :

- d'intégrer à cette représentation des outils d'interaction permettant de modifier l'échelle cyclique représentée. D'autres outils d'interaction peuvent également permettre de modifier le nombre, et l'étendue temporelle, des intervalles utilisés pour représenter le temps dans la carte, afin d'affiner l'analyse exploratoire.
- d'associer cette représentation à des diagrammes représentant le temps linéaire et le temps cyclique, apportant le support d'une légende temporelle à la représentation cartographique.

Ces diagrammes peuvent être de type *timeline*, *timewheels* ou mixtes, et peuvent également offrir la possibilité d'une analyse exploratoire des patterns au sein de la fenêtre temporelle. Pour cela, ces diagrammes doivent être interactifs et notamment permettre une modification dynamique et graduelle des échelles cycliques représentées.

- de combiner à cette approche la possibilité de représenter les événements selon des zones spatiales spécifiques, à l'instar de *Time Wave*. La combinaison des deux approches peut, entre autres, permettre d'approfondir l'analyse sur des zones spatiales particulières, repérées en premier lieu lors d'une analyse effectuée à l'aide de cartes de type *Growth Ring Map*.

Conclusion

L'état de l'art présenté dans le chapitre 3 a permis de produire une synthèse des principales approches permettant d'intégrer les dimensions temporelle et spatiale au sein d'un même environnement de géovisualisation.

Certaines approches se basent sur l'intégration du temps dans les représentations cartographiques :

- Le temps peut être intégré dans des cartes en deux dimensions, au moyen de règles de symbologie [Bertin, 1973, Kaddouri et al., 2014].
- Le temps peut être intégré dans des représentations en trois dimensions au moyen de cubes spatio-temporels, utilisant une troisième dimension pour représenter le temps [Hägerstrand, 1970, Andrienko et al., 2003b, Gatalsky et al., 2004, Kaddouri et al., 2014].

L'approche *small-multiples* [Bertin, 1973, Kaddouri et al., 2014] est basée sur la multiplication de représentations graphiques de même nature, représentant l'état d'une dimension pour plusieurs modalités, fixées, d'une dimension particulière. Cette approche peut prendre la forme :

- d'une collection de cartes, représentant l'état de l'espace pour différents moments ou intervalles temporels ;
- d'une collection de diagrammes temporels, représentant l'historique de différentes localisations ou zones de l'espace.

L'utilisation de cartes animées constitue une autre approche [Antoni et al., 2004, Kaddouri et al., 2014], et permet d'intégrer le temps dans les représentations cartographiques, en représentant le temps réel par le temps de l'animation.

Enfin, une approche consiste à combiner différents supports de représentation de la donnée au sein d'une interface interactive, afin de représenter les différentes composantes de l'information spatio-temporelle au sein de fenêtres dédiées, et synchronisées entre elles [Kaddouri et al., 2014].

Davoine et Kraak [Davoine et al., 2001, Davoine et al., 2004, Kraak and Ormeling, 2011] proposent ainsi une formalisation d'environnements de géovisualisation à multi-fenêtrage, composés de trois fenêtres interactives et synchronisées :

- une fenêtre spatiale, occupée par une ou plusieurs représentations cartographiques ;
- une fenêtre temporelle, occupée par une ou plusieurs représentations de la dimension temporelle ;
- une fenêtre thématique.

Ces environnements à multi-fenêtrage permettent alors de représenter les différentes composantes de l'information spatio-temporelle, telles qu'elles sont décrites dans le *Triad Model* [Peuquet, 1994] et le *Pyramid Model* [Mennis et al., 2000].

Les environnements de géovisualisation à multi-fenêtrage semblent être la solution la plus adaptée pour permettre une analyse exploratoire des séries d'événements spatio-temporels, l'association de représentations cartographiques avec des diagrammes temporels permettant d'analyser la distribution des événements dans le temps et dans l'espace. L'avantage des environnements à multi-fenêtrage est confirmé par les travaux effectués par MacEachren [MacEachren, 1994] et Davoine [Davoine, 2014] sur les différents usages des environnements de géovisualisation. Les conclusions de ces travaux justifient le choix d'environnements proposant de nombreux supports de représentation de la donnée, combinée à de hauts niveaux d'interaction, pour effectuer une exploration de données inconnues.

Il a alors fallu définir quels procédés graphiques et quels outils d'interaction peuvent être intégrés dans ces environnements à multi-fenêtrage, afin de permettre à l'utilisateur d'explorer la distribution spatio-temporelle des événements et y identifier des structures.

Hormis la propagation d'événements dans l'espace au cours du temps, chacun des patterns recherchés se manifeste, sur un plan unidimensionnel, par une structure observable dans la distribution temporelle et dans la distribution spatiale des événements :

- Un cluster spatio-temporel se traduit par une concentration d'événements dans l'espace (cluster spatial), et dans le temps (cluster temporel).
- Une récurrence cyclique spatialisée se traduit par une concentration spatiale d'événements (cluster spatial), réapparaissant de manière cyclique dans le temps (récurrence cyclique).

Une concentration d'événements pouvant être identifiée au moyen d'une représentation cartographique, nous nous sommes intéressés, dans une première partie du chapitre 4, aux différentes méthodes permettant d'identifier visuellement des structures dans la distribution temporelle des événements, que celles-ci soient relatives au temps cyclique ou relatives au temps linéaire. Des concentrations d'événements, selon le temps linéaire et le temps cyclique, peuvent être identifiées au moyen de diagrammes de type *timeline* et *timewheel*, très représentés dans les outils de géovisualisation [Edsall et al., 1997, Edsall and Peuquet, 1997].

L'utilisation de procédés d'agrégation des événements, comme dans le diagramme en secteurs de Florence Nightingale [Nightingale, 1858], permet également de faciliter l'identification de concentrations temporelles d'événements lorsque l'analyse est effectuée sur un grand nombre de données.

L'identification visuelle de récurrences cycliques est plus difficile dans les *timelines* et les *timewheel*. D'autres diagrammes, que nous appelons *diagrammes mixtes*, intègrent à la fois le temps linéaire et le temps cyclique, en représentant graphiquement une succession linéaire d'instances d'une échelle cyclique donnée. Ces diagrammes, au travers de l'identification de motifs graphiques particuliers, permettent d'identifier une récurrence cyclique de période égale à la durée de l'échelle cyclique représentée. Combinée à la possibilité de modifier de manière dynamique et graduelle la durée de l'échelle cyclique représentée, ces outils offrent la possibilité d'une recherche visuelle des possibles récurrences cycliques dans les séries événementielles [Weber et al., 2001].

L'intégration de diagrammes de type *timeline*, *timewheel* et *mixte*, dans des environnements à multi-fenêtrage intégrant des représentations cartographiques, permet ainsi de visualiser des structures dans les distributions spatiale et temporelle des événements. Dans la seconde partie du chapitre 4, nous nous sommes interrogés sur les procédés permettant à l'utilisateur de faire le lien entre ce qu'il voit dans la fenêtre spatiale et ce qu'il voit dans la fenêtre temporelle, afin

d'identifier des patterns spatio-temporels.

L'identification de relations entre les composantes spatiale et temporelle des données représentées, est possible en intégrant, dans les environnements à multi-fenêtrage, les autres approches de représentation de la donnée spatio-temporelle, présentées dans le chapitre 3. Deux environnements à multi-fenêtrage, *Time Wave* [Li, 2010] et *PerSE* [Swedberg and Peuquet, 2017], utilisent ces procédés et intègrent des diagrammes temporels permettant d'identifier les structures temporelles citées précédemment :

- *Time Wave* utilise des règles de symbologie pour représenter, dans un diagramme temporel mixte, des événements apparaissant à différentes localisations de l'espace.
- *PerSE* utilise un dispositif *small-multiples* pour représenter, dans différents diagrammes *timeline* et *timewheel*, des événements apparaissant dans différentes zones de l'espace.

Ces environnements montrent deux principales limites :

- L'impossibilité de modifier, de manière dynamique et graduelle, l'échelle temporelle cyclique représentée, empêchant une recherche visuelle de récurrences cyclique dans les séries événementielles.
- Une visualisation du lien entre les composantes spatiale et temporelle de la donnée, uniquement effectuée par des procédés orientés de l'espace vers le temps. Les données sont classées selon différentes modalités de la dimension spatiale (différentes localisations ou zones d'apparition), et sont représentées en fonction de ces modalités dans la fenêtre temporelle.

Les différentes modalités de la dimension spatiale, permettant de classer les événements, sont déterminées par l'utilisateur. Cette approche a l'avantage de permettre une analyse au travers de zones spatiales particulières. Ce critère est intéressant, notamment si ces zones sont créées à partir de l'analyse d'autres données géographiques, ce qui permet de croiser les thématiques de recherche liées à ces données avec une analyse des dynamiques spatio-temporelle. Cependant, cette approche nécessite que l'utilisateur émette une hypothèse *a priori* sur la distribution spatio-temporelle des données, ce qui va à l'encontre du principe de l'analyse exploratoire.

Une solution pour résoudre ce problème peut être d'inverser l'approche de *Time Wave* et *PerSE*, en utilisant des méthodes similaires pour représenter le lien entre les fenêtres temporelle et spatiale, mais en orientant ces procédés de la fenêtre temporelle vers la fenêtre spatiale.

Il nous a fallu définir, parmi les approches disponibles, quelle était la plus adaptée pour effectuer une analyse de patterns relatifs au temps linéaire ou au temps cyclique directement dans la carte :

- Notre analyse devant être effectuée simultanément sur l'ensemble des événements, l'utilisation de l'animation, masquant certains événements, fut écartée.
- Le cube spatio-temporel ne représentant que le temps linéaire, son utilisation ne permet donc pas l'identification de patterns relatifs à une échelle cyclique.
- L'utilisation de collections de cartes apportant une difficulté de lecture dans le cas d'un trop grand nombre de représentations, nous avons écarté les dispositifs *small-multiples*.

L'utilisation de règles de symbologie est alors apparue comme l'approche la plus adaptée pour représenter, dans la carte, les événements selon leur moment d'apparition. Ce dernier peut être représenté au travers du temps linéaire, par l'utilisation de la variable visuelle de la valeur, ou au travers d'une échelle temporelle cyclique, par l'utilisation de la couleur. Combinée à des procédés d'agrégation permettant, à l'instar de *Growth Ring Map* [Bak et al., 2009, Andrienko et al.,

2011], de représenter la proportion de différents intervalles temporels d'apparition, au sein d'une concentration spatiale d'événements, cette approche doit permettre l'identification de clusters spatio-temporels directement sur la carte.

Nous émettons l'hypothèse, qu'un environnement de visualisation à multi-fenêtrage intégrant cette approche et :

- représentant le temps à travers des échelles linéaires et cycliques au moyen de diagrammes de type *timeline*, *timewheel*, et *mixte* ;
- permettant de modifier de manière dynamique et graduelle l'échelle cyclique représentée ;
- permettant de modifier le nombre et l'étendue temporelle des intervalles utilisés pour représenter le temps dans la carte ;
- permettant de combiner cette approche à celle présentée par *Time Wave* ;

peut offrir la possibilité d'une analyse exploratoire des phénomènes spatio-temporels en mesure d'identifier, dans les séries d'événements, les dynamiques présentées dans le chapitre 2.

Troisième partie

GrAPHiST

Introduction

Notre objectif est de proposer une méthode d'analyse permettant d'identifier les dynamiques spatio-temporelles des phénomènes géographiques dynamiques. Nous cherchons à analyser :

- la présence de pics d'intensité du phénomène, localisés dans le temps et l'espace.
- l'existence de tendances localisées dans la variation de l'intensité du phénomène au cours du temps.
- l'évolution spatiale d'un phénomène au cours du temps.
- l'existence de récurrences cycliques localisées du phénomène.

Notre problématique consiste à rechercher quelles sont les méthodes de modélisation, de représentation graphique de la donnée spatio-temporelle, et d'interaction devant être mises en oeuvre pour permettre une approche d'analyse exploratoire des dynamiques spatio-temporelles, applicable à des phénomènes de différentes natures. Notre objectif n'est pas de répondre à une problématique métier particulière, nécessitant une analyse de données spatio-temporelles, mais de proposer des méthodes permettant une analyse de données spatio-temporelles, pouvant être utilisées dans le cadre de différentes problématiques métier.

De l'analyse des dynamiques spatio-temporelles des phénomènes, à l'analyse exploratoire de *patterns* dans les séries événementielles

Dans une première partie I, nous avons opté pour une modélisation des données spatio-temporelles en séries d'*événements*, entités limitées dans le temps et l'espace (Chapitre 1). Les *événements* traitant de la même problématique sont regroupés en *phénomènes*. Cette modélisation occurrente [Galton, 2004] de la donnée spatio-temporelle suit les préconisations de Reitsma [Reitsma, 2005], et ne prend pas en compte les possibles impacts des événements sur l'évolution des entités pérennes de l'espace. Ce choix permet une modélisation générique du changement dans l'espace, applicable à différentes problématiques métiers.

Cette représentation des phénomènes en séries événementielles permet également de redéfinir les éléments à rechercher dans les données : le but n'est pas d'étudier l'évolution temporelle d'un événement dans l'espace, mais la distribution d'une série d'événements dans le temps et l'espace. L'analyse des dynamiques des phénomènes passe alors par l'identification de structures dans la distribution spatio-temporelle des séries d'événements : les *patterns* (Chapitre 2). Ceux-ci se composent :

- de clusters spatio-temporels, relatifs au temps linéaire ou relatifs à une échelle temporelle cyclique ;
- de propagations spatiales d'événements au cours du temps ;
- de récurrences cycliques spatialisées d'événements.

Dans un contexte d'augmentation croissante des données spatio-temporelles disponibles, nous nous positionnons dans le champ de l'analyse exploratoire [Andrienko and Andrienko, 2006], afin de permettre une analyse de jeux de données de différentes natures, pour lesquelles l'utilisateur peut ne pas connaître à l'avance les structures qu'il doit y identifier.

Le principe central de l'analyse exploratoire est de permettre à un analyste d'effectuer une exploration libre d'un jeu de données, afin d'y mettre en lumière des structures et des relations insoupçonnées [Tukey, 1977, Andrienko and Andrienko, 2006]. Cette approche est basée sur deux grands axes :

- L'utilisation d'outils de représentation, graphiques et interactifs, de la donnée [Tufte, 1983, Andrienko and Andrienko, 2006].
- La place centrale de l'analyste (l'utilisateur) dans le procédé de recherche [Tukey, 1977, Andrienko and Andrienko, 2006].

L'environnement visuel n'est plus un outil de retranscription visuel, permettant d'afficher les résultats d'une analyse statistique, mais un outil supportant le processus de recherche, effectué par l'utilisateur [Card et al., 1999, Spence, 2001, Andrienko and Andrienko, 2006]. Proposer une analyse exploratoire des phénomènes spatio-temporels passe alors par l'élaboration d'un environnement de visualisation permettant à un utilisateur d'explorer des séries événementielles et d'y identifier des patterns.

Les environnements d'analyse exploratoire sont cependant majoritairement développés de manière *ad hoc* dans le cadre de thématiques définies [Kaddouri et al., 2014]. Proposer une analyse générique des phénomènes spatio-temporels implique d'élaborer une approche d'analyse exploratoire indépendante du sujet traité.

Les méthodes disponibles pour la visualisation de données spatio-temporelles

Dans une seconde partie II, nous effectuons un état de l'art des différentes méthodes de représentations graphiques des données temporelles et spatio-temporelles. En établissant cet état de l'art, nous tentons de définir quelles sont les méthodes de représentation graphiques et interactives, pouvant être utilisées afin :

- de représenter visuellement des séries d'événements spatio-temporels.
- de permettre une recherche visuelle des patterns dans la distribution spatio-temporelle des séries d'événements.

À l'issue de cet état de l'art, il apparaît que les environnements de géovisualisation à multi-fenêtrage, dont la formalisation est proposée par Davoine et Kraak [Davoine et al., 2001, Davoine et al., 2004, Kraak and Ormeling, 2011], sont les plus adaptés pour proposer une analyse exploratoire des séries d'événements spatio-temporelles [MacEachren, 1994, Davoine, 2014], l'association de représentations cartographiques et de diagrammes temporels permettant d'analyser la distribution des événements dans le temps et dans l'espace.

Au sein de ces environnements à multi-fenêtrage :

- l'utilisation de cartes d'inventaires au sein d'une fenêtre spatiale, à la manière de l'environnement *PerSE* [Swedberg and Peuquet, 2017], permet la visualisation de concentration d'événements dans l'espace.

- l'utilisation de diagrammes temporels de type *timeline* et *timewheel* [Edsall et al., 1997, Edsall and Peuquet, 1997], au sein d'une fenêtre temporelle, permet la visualisation de concentration d'événements dans le temps, selon une échelle linéaire ou cyclique.
- l'utilisation de diagrammes temporels mixte, représentant le temps de manière linéaire et cyclique, combinée à la possibilité d'une modification dynamique et graduelle de l'échelle temporelle cyclique [Weber et al., 2001], permet une recherche de récurrences cycliques dans les données temporelles.

Pour permettre une identification des patterns spatio-temporels, une approche rendue possible par plusieurs environnements [Li, 2010, Bak et al., 2009, Andrienko et al., 2011, Swedberg and Peuquet, 2017] consiste à intégrer une dimension supplémentaire dans une des fenêtres d'un environnement à multi-fenêtrage (intégration de l'espace dans la fenêtre temporelle ; intégration du temps dans la fenêtre spatiale). La méthode la plus adaptée pour effectuer cette approche semble être l'utilisation de règles de symbologie. Cette approche permet de représenter les événements :

- en fonction de leur composante spatiale dans la fenêtre temporelle [Li, 2010] ;
- en fonction de leur composante temporelle, selon une échelle linéaire ou cyclique, dans la fenêtre spatiale [Bak et al., 2009, Andrienko et al., 2011].

Dans ce second cas, l'utilisation de règles de symbologie combinée à des procédés d'agrégation permettant, à l'instar de *Growth Ring Map* [Bak et al., 2009, Andrienko et al., 2011], de représenter la composante temporelle d'événements concentrés dans les mêmes zones dans l'espace, permet l'identification de patterns spatio-temporels.

Enjeux de notre proposition

Pour fournir une méthode d'analyse générique des dynamiques spatio-temporelles des phénomènes, notre proposition doit répondre à trois enjeux :

- L'établissement d'une modélisation générique des phénomènes spatio-temporels sous un format événementiel, permettant l'analyse de données relatives à différentes problématiques métier, au sein d'un même environnement de géovisualisation.
- L'intégration des différentes méthodes d'interaction et de représentation graphique, identifiées au cours de notre état de l'art, au sein d'un même environnement de géovisualisation.
- La formalisation de l'utilisation de ces différentes méthodes, afin de décrire les différentes actions que devrait effectuer un utilisateur au moyen d'un tel environnement de géovisualisation, pour identifier des patterns dans les séries d'événements spatio-temporelles.

La structure de *GrAPHiST*

Dans un premier chapitre (5), nous répondons aux deux premières questions en présentant *GrAPHiST* (*(environnement de) Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels*), un environnement de géovisualisation devant permettre une analyse exploratoire des dynamiques spatio-temporelles des phénomènes.

Nous présentons tout d'abord (Section 5.1) le modèle de données utilisé par *GrAPHiST*, permettant de modéliser les phénomènes relatifs à différentes problématiques métier sous la forme de séries d'événements spatio-temporels.

Puis (Section 5.2), nous présentons la structure de *GrAPHiST*, et nous montrons comment nous intégrons, au sein d'un même environnement de géovisualisation les différentes approches

de visualisation identifiées lors de notre état de l'art (Section 5.3, Section 5.4 et Section 5.5).

Le développement de *GrAPHiST* s'est effectué de manière continue tout au long de la thèse, depuis le commencement de l'état de l'art jusqu'aux phases d'expérimentation, selon une approche *bottom up*¹. L'implémentation de *GrAPHiST* n'a pas fait l'objet d'une étape unique, précédée d'une identification de l'ensemble des fonctionnalités à intégrer, et d'une modélisation de l'ensemble du prototype avant son développement.

Au contraire, l'implémentation des différentes fonctionnalités de *GrAPHiST* a été effectuée de manière simultanée avec l'élaboration de notre état de l'art. L'intégration des différentes méthodes de représentation graphique et d'interaction s'est effectuée de manière empirique. D'anciennes fonctionnalités jugées peu probantes, ont été implémentées puis supprimées du prototype, certaines fonctionnalités ont été conceptualisées, mais n'ont pas été implémentées, tandis que certaines fonctionnalités actuelles de l'environnement n'ont été conceptualisées et implémentées que très tardivement.

L'analyse exploratoire au moyen de *GrAPHiST*

Dans un second chapitre 6, nous présentons comment notre environnement peut fournir la possibilité d'une analyse exploratoire des phénomènes spatio-temporels, et comment l'utilisateur doit interagir avec l'interface de *GrAPHiST* pour identifier différents patterns dans les séries d'événements spatio-temporels. Nous présentons comment l'utilisateur peut, à travers l'utilisation des différents composants de l'environnement, construire son procédé d'analyse et identifier chacune des structures recherchées. Pour cela, nous formalisons différents scénarios d'analyse exploratoire rendus possibles par *GrAPHiST*, et répondant à différents cas de figure.

1. Dans le cadre de la recherche, l'approche ascendante (ou *bottom up*) tente de faire émerger des théories de la pratique. L'implémentation de *GrAPHiST* a fait l'objet d'un développement empirique, plutôt que de l'implémentation d'un environnement entièrement théorisé au préalable

Chapitre 5

Structure de *GrAPHiST*

Sommaire

5.1	Modélisation des événements	172
5.1.1	Entités <i>SPAN</i> , et <i>intervalles (spatio-)temporels</i>	172
5.1.2	Notre modèle de données	173
5.2	Structure générale de l'interface de <i>GrAPHiST</i>	175
5.2.1	Un environnement de géovisualisation à multi-fenêtrage	175
5.2.2	La représentation de la dimension thématique	177
5.2.3	Limitation des données à représenter aux événements ponctuels	179
5.3	Représentation de la composante temporelle des événements	180
5.3.1	Le diagramme temporel mixte	180
5.3.2	Représentation des événements dans le diagramme mixte	184
5.3.3	Intérêts et limites des deux approches	188
5.3.4	Diagramme temporel circulaire	189
5.3.5	Relations entre le diagramme mixte et le diagramme circulaire	191
5.4	Représentation de la composante spatiale des événements	192
5.4.1	Principe	192
5.4.2	Représentation des événements dans la carte	193
5.5	Représentation de la composante spatio-temporelle des événements	196
5.5.1	Sélection et <i>brushing</i> des événements selon un critère dimensionnel	197
5.5.2	Représentation des événements selon leur lieu d'apparition	199
5.5.3	Représentation des événements selon leur moment d'apparition relatif à une échelle linéaire	203
5.5.4	Représentation des événements selon leur moment d'apparition relatif à une échelle cyclique	206
5.6	Implémentation de <i>GrAPHiST</i>	209
5.6.1	Implémentation de la base de données	209
5.6.2	Implémentation de l'interface de géovisualisation	210
5.7	Conclusion	211

5.1 Modélisation des événements

5.1.1 Entités *SPAN*, et *intervalles (spatio-)temporels*

Nous considérons l'*événement* comme une entité spatio-temporelle occurrente :

- limitée dans le temps et l'espace, s'inscrivant dans le modèle $4D$ décrit par Galton [Galton, 2004] ;
- n'ayant pas d'influence sur la trajectoire de vie d'entités spatiales perdurantes, conformément au modèle de Reitsma [Reitsma, 2005]. Les objets à représenter dans notre environnement de géovisualisation se rapportent aux événements et non aux objets spatiaux qu'ils impactent dans le monde réel.

Pour le cas des données que nous évoquons dans le premier chapitre (1), nous disposons d'une géométrie unique pour chaque événement, correspondant :

- soit à la localisation de l'événement dans le cas d'une implantation géométrique ponctuelle ;
- soit à la zone d'extension maximale de l'événement dans le cas d'une implantation géométrique zonale.

Cependant, une conceptualisation des phénomènes spatio-temporels en $4D$ doit répondre aux cas où l'on dispose de données sur les différentes étendues spatiales d'un événement au cours du temps.

Dans la définition de l'objet occurrent par Grenon et Smith [Grenon and Smith, 2004], un objet *SPAN* est caractérisé par un *début*, une *fin*, et une succession d'*intervalles* de temps. Chaque intervalle correspond à une modification de l'environnement spatial, survenant durant cet intervalle de temps, dans une zone de l'espace.

Soit l'exemple d'un événement *feu de forêt*, ce dernier peut être modélisé par une entité *SPAN*. Il possède :

- une date de début ;
- une date de fin ;
- une série d'intervalles temporels. Chacun de ces intervalles correspond à une géométrie représentant l'étendue spatiale du feu de forêt durant l'intervalle de temps correspondant.

Ces intervalles forment des intervalles spatio-temporels, car chacun est relié à une composante spatiale. Chaque entité *SPAN* est ainsi la somme de sous-entités spatio-temporelles.

Transposé au monde réel, cela revient à considérer qu'un événement spatio-temporel duratif est composé d'une infinité d'intervalles spatio-temporels. Dans le cas d'une modélisation de l'événement, le nombre d'intervalles représenté est défini par les données à disposition. Si l'on possède des données représentant trois extensions spatiales successives d'un feu de forêt, cet événement peut être modélisé par un objet *SPAN*, ayant un début, une fin, et trois intervalles de temps. Chaque intervalle correspond alors à une extension spatiale différente de l'événement au cours du temps.

Le même raisonnement peut être appliqué pour la représentation de l'évolution d'un attribut de l'événement au cours du temps. Soit un événement *feu de forêt*, d'une durée de x jours, pour lequel les données disponibles sont le nombre de pompiers déployés chaque jour sur la zone de l'événement. Cet événement peut être représenté par une entité *SPAN* ayant x intervalles de temps d'une durée d'une journée. Chaque intervalle permet de représenter le nombre de pompiers déployés chaque jour dans la zone de l'événement.

Dans le cas de l'étude des trajectoires d'objets mobiles, on s'intéresse à la trajectoire complète d'un objet plutôt qu'à la position de cet objet au cours du temps. L'événement correspond au trajet complet du véhicule, tandis que chaque intervalle de temps de l'événement correspond à une section de ce trajet. On remarque ici la différence avec la façon de représenter le changement comme une évolution de la position du mobile, à la manière d'une succession de *snapshots* temporels. Un intervalle spatio-temporel ne correspond pas à la position spatiale d'un objet, mais au trajet de l'objet. En passant d'un intervalle à un autre, on ne présente pas le changement de position du mobile, mais les deux trajets effectués durant deux intervalles temporels différents.

5.1.2 Notre modèle de données

Notre modèle s'articule autour de trois entités spatio-temporelles :

- le *phénomène* ;
- l'*événement* ;
- la *sous-partie spatio-temporelle* d'un *événement*.

Ces trois entités spatio-temporelles sont reliées entre elles par des relations de composition. Un *phénomène* est composé d'*événements*, tandis qu'un *événement* est composé de *sous-parties spatio-temporelles*. L'élément central de notre modèle est l'*événement*.

Un événement composé de plusieurs sous-parties spatio-temporelles correspond à la modélisation d'un événement étendu dans le temps pour lequel on possède des données correspondant à plusieurs états spatio-temporels de l'événement : on peut ainsi disposer de données sur l'étendue spatiale d'un feu de forêt pour un jour J , puis pour un jour $J+1$, puis pour un jour $J+2$ (Figure 5.1).

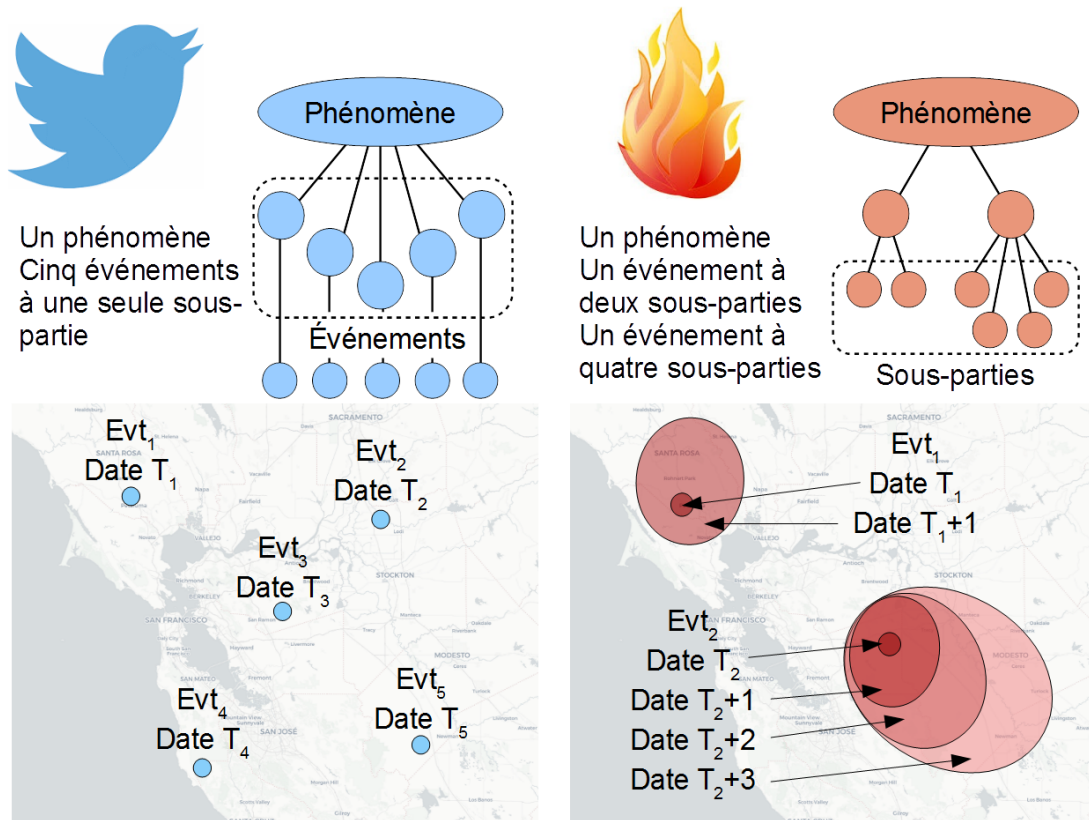


FIGURE 5.1 – Phénomènes, événements, et sous-parties spatio-temporelles

La construction d'une entité occurrente, au travers du modèle $4D$ de Galton [Galton, 2004], ne permettant pas de variation dans le temps et l'espace au sein d'une même entité, ces trois phases du feu de forêt sont modélisées en trois entités spatio-temporelles distinctes : les *sous-parties* de l'*événement*. L'*événement* est alors une entité spatio-temporelle de rang supérieur, formé par ces trois *sous-parties*. Ce feu de forêt est donc modélisé par un événement formé de trois sous-parties spatio-temporelles.

À l'inverse, un événement, duratif ou ponctuel, pour lequel on ne possède des informations que sur une phase spatio-temporelle (une date ou un intervalle de temps associé à une emprise spatiale), est modélisé par un événement n'ayant qu'une seule sous-partie spatio-temporelle.

L'incendie d'un bâtiment, par exemple, peut être représenté par un événement à une seule sous-partie :

- dont la composante spatiale représente la localisation et la géométrie du bâtiment ;
- dont la composante temporelle représente le moment d'apparition et la durée de l'incendie.

Ces deux entités spatio-temporelles possèdent leurs propres composantes thématiques, temporelles et spatiales :

- Chaque sous-partie spatio-temporelle possède une géométrie, la composante spatiale d'un événement étant une union des géométries des sous-parties le composant.
- Chaque sous-partie possède une date de début et de fin. Les sous-parties d'un événement étant contingentes dans le temps, les bornes de l'étendue temporelle formée par cet ensemble sont les dates de début et de fin de l'événement.
- Chaque sous-partie d'un événement, ainsi que ce dernier, possède ses propres attributs thématiques.

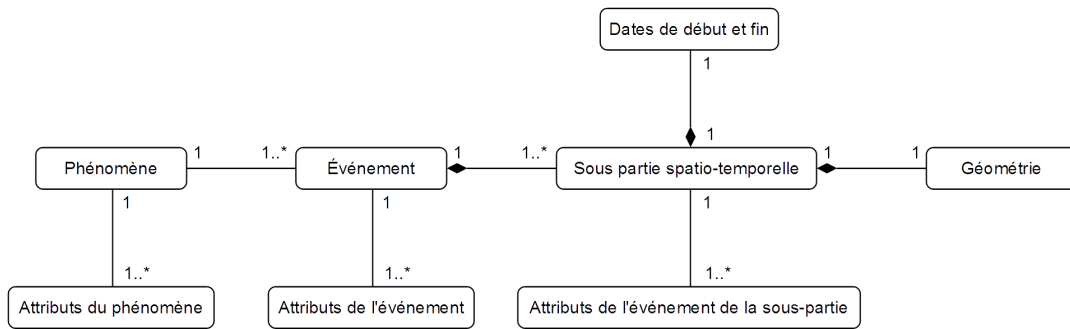
Les *événements* de même nature composent une *méta-entité* spatio-temporelle supérieure, le *phénomène*. Un feu de forêt en particulier sera modélisé par un événement, tandis que l'ensemble des feux de forêt sera modélisé par un phénomène. En tant qu'entité spatio-temporelle à part entière, le phénomène possède :

- une composante thématique, formée d'attributs qui lui sont propres ;
- une composante spatiale, correspondant à l'étendue spatiale englobant l'ensemble des événements composant le phénomène ;
- une composante temporelle, formée par l'enveloppe temporelle des événements le composant.

La relation de composition entre un phénomène et les événements le composant est légèrement différente de la relation de composition entre un événement et ses sous-parties :

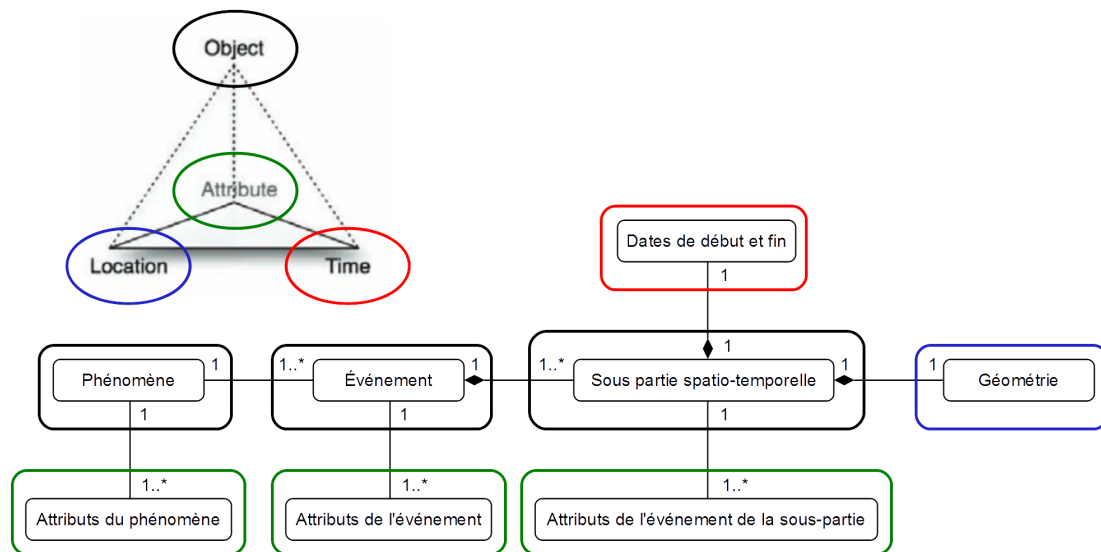
- Les sous-parties d'un événement forment une suite chronologique d'entités contingentes sur le plan temporel.
- Les événements regroupés dans un même phénomène ne sont pas liés entre eux par une relation spatio-temporelle, mais par une relation thématique étant donné qu'ils traitent de la même problématique.

Ainsi formalisée, notre approche de modélisation se présente sous la forme décrite dans la Figure 5.2.

FIGURE 5.2 – Modélisation des événements dans *GrAPHiST*

Cette modélisation peut être mise en parallèle avec la décomposition de l'entité spatio-temporelle en composante spatiale, temporelle, et thématique, telle qu'elle est décrite dans la pyramide de Mennis [Mennis et al., 2000]. Le *phénomène*, l'*événement*, et la *sous-partie spatio-temporelle* d'un *événement* sont des objets spatio-temporels (Encadrés noirs dans la Figure 5.3), possédant chacun :

- une composante thématique (Encadrés verts dans la Figure 5.3) ;
- une composante spatiale (Encadrés bleus dans la Figure 5.3) ;
- une composante temporelle (Encadrés rouges dans la Figure 5.3).

FIGURE 5.3 – Modélisation des événements dans *GrAPHiST*, lien avec le Pyramid Model de Mennis [Mennis et al., 2000]

5.2 Structure générale de l'interface de *GrAPHiST*

5.2.1 Un environnement de géovisualisation à multi-fenêtrage

Pour visualiser de manière simultanée la distribution des événements dans le temps et l'espace, nous les représentons au sein d'un environnement de géovisualisation à multi-fenêtrage, respectant la formalisation énoncée par Kraak [Kraak and Ormeling, 2011].

5.2.1.1 Structure des différentes fenêtres

Notre environnement de visualisation doit permettre, non pas une analyse des composantes dimensionnelles d'un événement, mais une identification des potentielles structures dans la distribution spatio-temporelle d'une série d'événements. Soit l'exemple des feux de forêt, le but n'est pas d'analyser l'évolution de chaque feu de forêt au cours du temps, mais comment l'ensemble des feux de forêt est réparti dans le temps et l'espace. Les fenêtres de notre environnement doivent ainsi permettre une visualisation des distributions temporelles et spatiales de l'ensemble des événements formant un phénomène.

Dans la fenêtre spatiale, nous utilisons une carte d'inventaire, ne représentant pas l'état de l'espace à un instant donné, mais les lieux d'occurrence de l'ensemble des événements apparaissant durant une étendue temporelle étudiée. Une représentation spatiale de ce type est utilisée dans PerSE [Swedberg and Peuquet, 2017].

Dans la fenêtre temporelle, nous utilisons une série de diagrammes temporels, permettant de représenter la distribution temporelle d'une série d'événements, selon le temps linéaire ou selon une ou plusieurs échelles temporelles cycliques.

5.2.1.2 Composition de l'environnement de géovisualisation

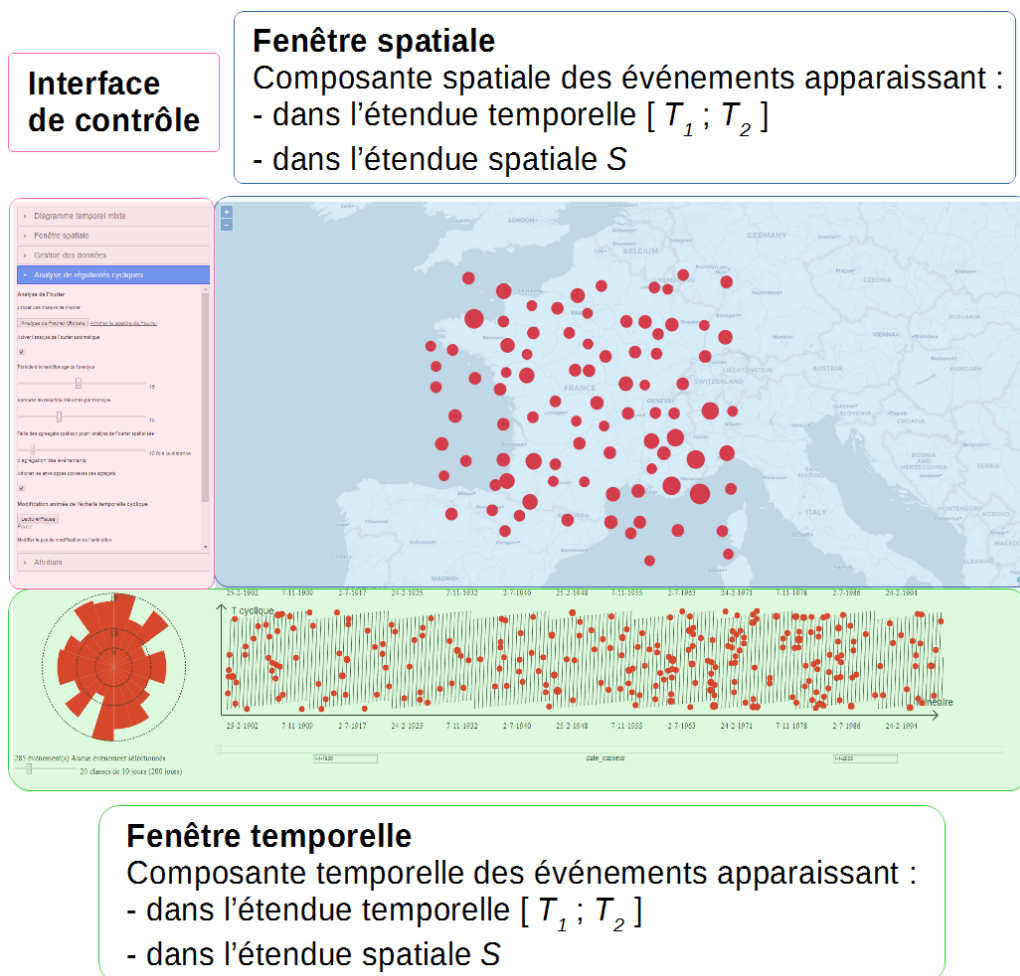


FIGURE 5.4 – Structure générale de *GrAPHiST*

Notre environnement comporte (Figure 5.4) :

- une fenêtre spatiale représentant une étendue spatiale S au travers d'une carte ;
- une fenêtre temporelle représentant une étendue temporelle linéaire $[T_1 ; T_2]$ au travers de diagrammes temporels ;
- une interface de contrôle permettant d'interagir avec l'environnement.

Les événements visibles à l'écran apparaissent durant l'étendue temporelle et dans l'emprise spatiale représentées dans ces deux fenêtres. Les événements apparaissant en dehors de l'étendue temporelle représentée dans les diagrammes temporels ne sont pas représentés dans la carte. Les événements apparaissant en dehors de l'emprise spatiale de la carte ne sont pas représentés dans les diagrammes temporels.

L'environnement ne possède pas de fenêtre dédiée à la représentation de la composante thématique des événements. Le traitement particulier de cette dernière est présenté dans la section 5.2.2.

5.2.1.3 Synchronisation des fenêtres et interactions

Les deux fenêtres sont synchronisées entre elles. Les événements représentés sur le diagramme temporel mixte correspondent à ceux représentés dans la carte de la fenêtre spatiale. Chaque événement représenté dans la fenêtre temporelle est également représenté dans la fenêtre spatiale, et inversement.

Les deux fenêtres de l'environnement sont également interactives. La sélection d'éléments ou le *brushing* (le fait de limiter la représentation à certains éléments) font partie des interactions permises par les environnements à multi-fenêtrage [Davoine, 2016]. Le procédé de synchronisation repose alors sur le principe selon lequel une interaction effectuée au sein d'une fenêtre de l'environnement a une incidence sur les autres fenêtres. Ainsi :

- Une interaction effectuée sur un événement au travers de la fenêtre temporelle a une incidence sur sa représentation spatiale, et inversement. Par exemple, une sélection d'événements au travers d'une fenêtre de l'environnement entraîne la sélection de ces événements dans l'ensemble des fenêtres de l'environnement.
- Une modification de l'étendue temporelle linéaire représentée dans la fenêtre temporelle impacte la représentation spatiale. Les nouveaux événements apparaissant durant la nouvelle emprise temporelle sont ajoutés à la représentation spatiale. Les événements n'apparaissant plus dans la nouvelle emprise temporelle sont supprimés de la représentation spatiale.
- Une modification de l'étendue spatiale représentée dans la fenêtre spatiale impacte la représentation temporelle. Les nouveaux événements apparaissant dans la nouvelle emprise de la fenêtre spatiale sont ajoutés dans la fenêtre temporelle. Les événements n'apparaissant plus dans la nouvelle emprise la fenêtre spatiale sont supprimés de la fenêtre temporelle.

5.2.2 La représentation de la dimension thématique

5.2.2.1 Concept

Comme précisé auparavant, nous considérons le phénomène spatio-temporel comme un ensemble d'événements spatio-temporels semblables (Figure 5.5). Les critères de similitude des événements peuvent être temporels ou spatiaux : on peut considérer des événements similaires, mais n'apparaissant pas aux mêmes lieux ou aux mêmes époques comme appartenant à des phénomènes distincts. Cependant, les principaux critères définissant l'ensemble *phénomène* sont d'ordre thématique. On considère des événements comme faisant partie d'un phénomène, car ils

sont de même type, ou étant rattachés à la même problématique.

Dans notre approche, les phénomènes sont, par construction, des regroupements d'événements proches sur le plan thématique. Lors de l'analyse d'un phénomène spatio-temporel en particulier, la dimension thématique des événements étudiés est plus ou moins *fixée*. C'est sur les dimensions temporelle et spatiale, et sur les relations entre ces deux dimensions, que s'effectue l'analyse exploratoire. Notre proposition se focalise ainsi sur la représentation graphique de ces deux dimensions. À l'inverse, la composante thématique des événements ne faisant pas l'objet d'une analyse exploratoire, celle-ci ne fait pas l'objet d'une représentation graphique. Notre interface de représentation ne comporte ainsi pas de fenêtre dédiée à la représentation de la composante thématique des événements représentés.

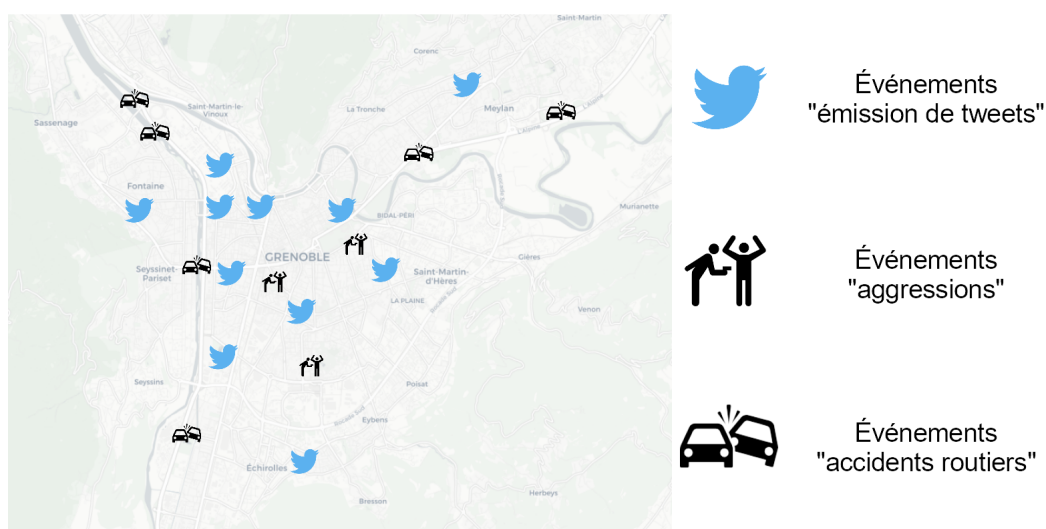


FIGURE 5.5 – Classification des événements par phénomène

Ce choix d'omettre la représentation graphique de la composante thématique est plus lié à une volonté de privilégier l'espace graphique de l'interface pour la représentation des composantes temporelles et spatiales, plutôt qu'à un choix délibéré de ne pas prendre en compte la dimension thématique. Cette dernière est représentée, mais de manière non graphique, dans l'environnement de visualisation.

5.2.2.2 Représentation, non graphique, de la dimension thématique

La prise en compte de cette dimension s'effectue :

- d'une part, au travers d'outils permettant de définir les événements à représenter dans l'environnement, en fonction de critères thématiques ;
- d'autre part, au travers d'outils de sélection et de représentation, sous la forme de tableaux, de la composante thématique des événements sélectionnés.

Définition des événements à représenter en fonction de critères thématiques. La sélection des événements à représenter en fonction de leur composante thématique peut s'effectuer de deux manières :

- en sélectionnant les phénomènes à représenter, les événements d'un même phénomène étant proches sur le plan thématique (Figure 5.5) ;
- en sélectionnant, au sein d'un phénomène, les événements à représenter en fonction de la valeur prise par les attributs formant leur composante thématique. La représentation peut ainsi être limitée à des événements possédant une certaine valeur pour un attribut défini.

Représentation tabulaire de la composante thématique des événements. Pour éviter la représentation d'un trop grand nombre de données sous la forme de tableaux, nous limitons la description tabulaire des événements à une sélection d'entités, effectuée selon un critère temporel ou spatial.

Lors d'une sélection d'événements, une liste des événements sélectionnés est représentée dans l'interface de contrôle (Figure 5.6). Chaque événement est représenté par un élément de la liste, sous la forme d'un menu déroulant, décrivant les attributs thématiques de l'événement en question.

Attributs	
5 événements sélectionnés	
▼ id_st_event: 1000145	
Date d'apparition	2008 01 02 00:00:00
Date de disparition	2008 01 04 00:00:00
nom	Fortes pluies sur les Cévennes et tempête
inondation	inconnu
regimHydro	null
periodR	0
▶ id_st_event: 1000149	
▶ id_st_event: 1000161	
▶ id_st_event: 1000170	
▶ id_st_event: 1000172	

FIGURE 5.6 – Représentation tabulaire de la composante thématique des événements

5.2.3 Limitation des données à représenter aux événements ponctuels

La modélisation des phénomènes spatio-temporels dans notre base de données permet de gérer les cas de séries d'événements, pour lesquels nous aurions accès à des données suffisamment détaillées pour représenter différentes phases spatio-temporelles des événements (comme les différentes extensions spatiales d'un événement au cours de son existence). Les données auxquelles nous avons eu accès, lors de l'implémentation de notre environnement, se limitaient cependant à une structure spatio-temporelle simple.

Nous disposions pour chaque événement d'une date d'apparition, parfois d'une durée, d'une liste de données attributaires, d'une localisation spatiale, et parfois d'une géométrie polygonale correspondant à l'étendue spatiale maximale de l'événement. Cependant, la majorité des données utilisées présentaient une dimension temporelle et une dimension spatiale ponctuelle. La dimension temporelle des événements était limitée à une date d'apparition, et la dimension spatiale à un point dans l'espace.

Ainsi, si les premières implémentations du prototype ont été effectuées en utilisant des événements possédant une composante temporelle et spatiale étendue (possédant une durée dans le temps et occupant une surface dans l'espace), la suite du développement de notre prototype a été en grande partie focalisée sur la représentation de données spatio-temporelles ponctuelles.

Ce choix a conditionné les méthodes utilisées pour la représentation graphique des événements au sein de chaque fenêtre de l'environnement, ainsi que pour la représentation du lien entre chaque fenêtre. En effet, les approches de représentation graphique devant être utilisées, dans le cas d'une agrégation des événements, ou dans le cas d'une représentation des événements, selon un critère temporel dans la carte ou selon un critère spatial dans le diagramme temporel, sont radicalement différentes selon que les données représentées sont ponctuelles ou étendues.

Durant nos travaux, certaines propositions de représentations graphiques, permettant d'intégrer dans notre environnement des événements zonaux et duratifs, ont été avancées. Cependant, dans le cadre de notre travail, nous avons préféré mettre l'accent sur la représentation d'événements ponctuels, et garder ces propositions à l'état de perspectives. Les méthodes de représentations graphiques, permettant de représenter les entités étendues dans l'espace et le temps, n'ont pas été implémentées dans le prototype et pourront faire l'objet de développements futurs.

5.3 Représentation de la composante temporelle des événements

La fenêtre temporelle de notre environnement de géovisualisation doit permettre la visualisation de la distribution temporelle des événements, selon le temps linéaire et selon plusieurs échelles cycliques. Elle doit également permettre une identification de clusters temporels et de récurrence cycliques. La fenêtre est scindée en deux parties, chacune comportant un diagramme temporel différent, ces deux diagrammes étant synchronisés 5.3.5.

Le premier diagramme temporel 5.3.1 prend la forme d'un diagramme temporel mixte combinant une échelle temporelle linéaire, et une échelle temporelle cyclique. Ce diagramme représente la dimension temporelle de chaque événement, ces derniers y étant représentés par des entités graphiques, dont la position dépend de leur moment d'apparition.

Le second diagramme temporel 5.3.4 prend la forme d'un diagramme temporel cyclique, utilisant la même échelle temporelle cyclique que le premier diagramme. Ce diagramme représente le nombre d'événements apparaissant au cours du temps. L'échelle temporelle cyclique représentée est divisée en plusieurs sections de temps, et le diagramme présente le nombre d'événements apparaissant dans chaque section de temps cyclique.

5.3.1 Le diagramme temporel mixte

5.3.1.1 Principe

Le diagramme temporel principal est un diagramme temporel mixte. Le temps y est représenté de manière linéaire, au travers d'une succession d'instances d'une échelle cyclique. Ce diagramme est caractérisé par cinq critères, pouvant être modifiés dynamiquement au cours d'une analyse exploratoire :

- **L'étendue temporelle du diagramme** définit la période de temps linéaire étudiée et représentée dans le diagramme linéaire, pour laquelle les événements spatio-temporels sont représentés dans l'interface. Cette étendue est définie par une date de début et une date de fin, toutes deux paramétrables. Une modification de l'étendue temporelle modifie les événements représentés dans l'interface.
- **La durée de l'échelle temporelle cyclique représentée** définit la durée des instances de temps cycliques au travers desquelles l'étendue temporelle étudiée est représentée dans le diagramme temporel mixte. Une modification de ce critère n'influe pas sur le nombre d'événements représentés, ou sur l'étendue temporelle représentée, mais influe sur le moment d'apparition des événements représentés dans le temps cyclique.
- **Le déphasage de l'échelle de temps cyclique** définit l'origine de l'échelle cyclique, et la date de début de chaque instance, dans le temps linéaire. Comme pour la durée de l'échelle cyclique, une modification de ce critère influe sur le moment d'apparition des événements représentés dans le temps cyclique.
- **La granularité temporelle** utilisée définit la granularité temporelle utilisée dans le diagramme temporel mixte, ainsi que la granularité utilisée pour définir les bornes temporelles du diagramme et la durée de l'échelle cyclique. La granularité temporelle est paramétrable

parmi une série de durées usuelles : seconde, minute, heure, jour, semaine, mois, année. Une modification de la granularité temporelle utilisée peut impacter la distribution temporelle des événements dans le diagramme temporel. Des événements distincts dans le temps peuvent être représentés comme apparaissant au même moment si la granularité choisie n'est pas assez fine.

- **La forme prise par le diagramme temporel mixte.** Nous choisissons de nous baser sur deux structures différentes, toutes deux inspirées des fonctions périodiques. Le diagramme peut prendre la forme d'une courbe sinusoïdale inspirée de *TimeCoil* et *TimeWave*, ou bien prendre la forme d'une courbe périodique en dents de scie, inspiré des diagrammes en *tiles map*.

5.3.1.2 Diagramme sinusoïdal

Le premier diagramme s'inspire de la *Time Coil* d'Edsall et Peuquet [Edsall and Peuquet, 1997] et de la *Time Wave* de Li [Li, 2010] et représente le temps au travers d'une hélice, mais en deux dimensions. Le diagramme prend ainsi la forme d'une courbe sinusoïdale, chaque convolution de la courbe correspondant à une instance d'une échelle temporelle cyclique de durée définie (Figure 5.8).

Plusieurs propositions ont été envisagées au cours de l'implémentation du diagramme sinusoïdal. Au départ, ce diagramme s'inspire plus largement de la *Time Coil*. La première implémentation du diagramme prenait la forme d'une hélice représentée en trois dimensions. La Figure 5.7 présente une capture d'écran de cette première version du diagramme mixte. Les entités graphiques situées sur la courbe du diagramme représentent des événements étendus dans le temps.

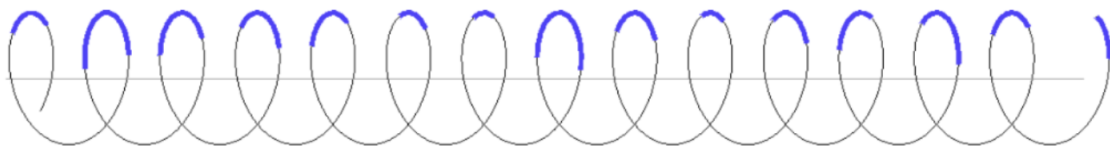


FIGURE 5.7 – Première implémentation du diagramme temporel mixte [Gautier et al., 2016]

Nous avons par la suite opté pour sur une version "plane" de ce diagramme, la représentation en perspective du diagramme en trois dimensions gênant la lecture de l'information dans nos premières représentations :

- Il était difficile d'identifier le moment d'apparition d'un événement selon le temps linéaire.
- Quand le nombre de révolutions de la courbe, représentant les instances de l'échelle cyclique du diagramme, devenait trop important, il était difficile d'identifier durant quelle instance de l'échelle cyclique un événement apparaissait.

Cette version "plane" du diagramme se rapproche graphiquement plus de *Time Wave*. Cependant, la formule mathématique décrivant la courbe de notre diagramme est plus proche de celle décrivant la courbe du modèle *Time Coil*.

La courbe de ce diagramme est modélisée par le système d'équations suivant (5.1).

$$\begin{cases} x &= \alpha \times (t - t_0) \\ y &= \beta \times \sin \left((t - t_0) \times \left(\frac{2 \times \pi}{T} \right) + \phi \right) \end{cases} \quad (5.1)$$

Dans ce système d'équations :

- x et y sont les coordonnées graphiques d'un point P sur la courbe du diagramme ;
- t représente le moment temporel représenté par le point P ;
- t_0 représente la date correspondant à l'origine graphique du diagramme ;
- T représente la durée de l'échelle temporelle cyclique représentée ;
- α et β sont des constantes définissant le pas graphique du diagramme ;
- ϕ est une constante permettant de définir le déphasage de la courbe sinusoïdale du diagramme à son origine.

La représentation graphique de cette formule forme la courbe présentée dans la Figure 5.8.

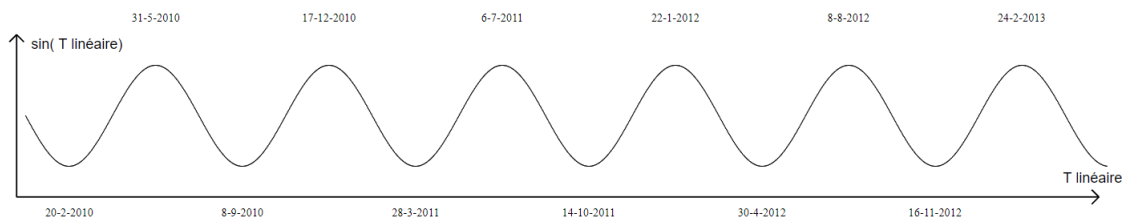
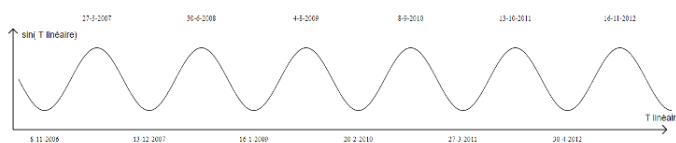


FIGURE 5.8 – Diagramme temporel mixte sinusoïdal

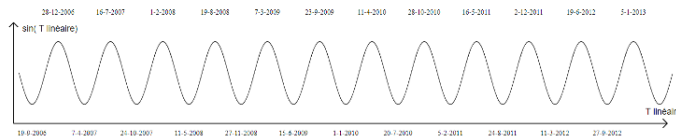
Ce diagramme ressemble visuellement à la représentation temporelle proposée par *Time Wave*, mais diffère dans sa façon de représenter le temps selon différentes échelles cycliques.

Les *vagues* de *Time Wave* observent une courbe strictement circulaire. Plus la durée représentée par une *vague* de *Time Wave* augmente, plus l'amplitude de cette dernière augmente (même s'il est possible de modifier les échelles graphiques dans *Time Wave*, afin d'obtenir des courbes de même amplitude, représentant des échelles temporelles différentes).

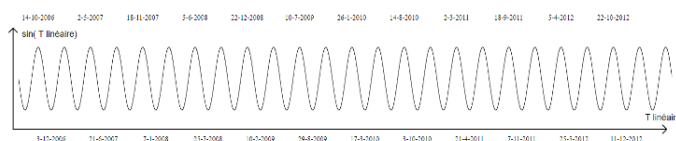
La courbe de notre diagramme, étant créée à partir d'une fonction sinusoïdale, la modification de la durée représentée par une circonvolution donne un effet de *compression* ou de *tension* de la courbe, tel un ressort (Figure 5.9).



Durée de l'échelle cyclique:
400 jours



Durée de l'échelle cyclique:
200 jours



Durée de l'échelle cyclique:
100 jours

FIGURE 5.9 – Changement d'échelle dans le diagramme sinusoïdal

5.3.1.3 Diagramme en dents de scie

Le second diagramme prend la forme d'une fonction périodique en dents de scie¹. Ce diagramme est en partie inspiré des diagrammes de type *tiles map* [Mintz et al., 1997][Aigner et al., 2011]. Le temps est représenté sur deux axes, l'axe des abscisses représentant le temps de manière linéaire, l'axe des ordonnées représentant le temps selon une échelle temporelle cyclique (Figure 5.10).

Le temps y est cependant représenté dans les deux axes en utilisant la même granularité temporelle. Dans un diagramme de type *tiles map*, deux moments temporels distincts, mais contenus dans le même granule sur l'échelle temporelle linéaire, peuvent être représentés au même niveau sur l'axe horizontal et à deux différents niveaux sur l'axe vertical (Section 4.1.3.5). En résulte une représentation matricielle : ces moments temporels distincts sont représentés sur deux lignes différentes de la matrice du diagramme, dans une même colonne.

Dans notre diagramme, deux moments temporels distincts, représentés à deux différents niveaux sur l'axe vertical, seront positionnés à deux niveaux différents sur l'axe horizontal. Chaque instance de l'échelle cyclique, représentée dans les diagrammes *tiles map* par une colonne, est représentée ici par un axe linéaire oblique. La représentation d'une suite d'instances de l'échelle cyclique prend la forme d'une suite de lignes brisées, rappelant la forme d'un diagramme en dents de scie.

La courbe de ce diagramme est modélisée par le système d'équations suivant (5.2).

$$\begin{cases} x &= \alpha \times (t - t_0) \\ y &= \beta \times ((t - t_0) \pmod{T}) + \phi \end{cases} \quad (5.2)$$

Dans ce système d'équations :

- x et y sont les coordonnées graphiques d'un point P sur la courbe du diagramme ;
- t représente le moment temporel représenté par le point P ;
- t_0 représente la date correspondant à l'origine graphique du diagramme ;
- T représente la durée de l'échelle temporelle cyclique représentée ;
- α et β sont des constantes définissant le pas graphique du diagramme ;
- ϕ est une constante permettant de définir le déphasage de la courbe en dents de scie du diagramme à son origine.

La représentation graphique de cette formule forme la courbe présentée dans la Figure 5.10.



FIGURE 5.10 – Diagramme temporel mixte en dents de scie

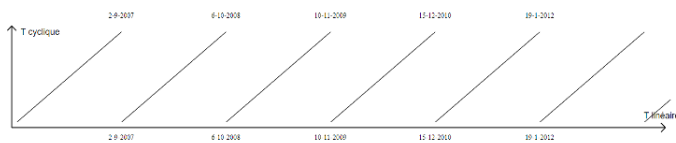
Lors d'une modification de l'échelle temporelle cyclique représentée dans le diagramme, la longueur graphique utilisée pour représenter la durée de l'échelle cyclique sur l'axe des ordonnées

1. Un signal en dents de scie est un signal périodique non sinusoïdal, tirant son nom de sa représentation graphique qui se rapproche des dents d'une scie [Wikipédia, 2018b].

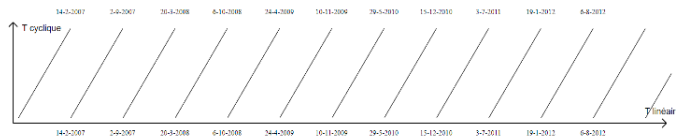
ne change pas, et le ratio (*étendue temporelle représentée*) \div (*distance graphique*) reste constant sur l'axe des abscisses.

Ainsi (Figure 5.11) :

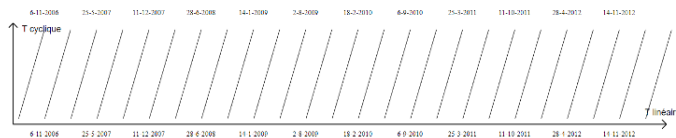
- Lorsque la durée de l'échelle cyclique diminue, les lignes brisées représentant les instances de l'échelle cyclique deviennent plus nombreuses, se rapprochent, se redressent et deviennent de plus en plus verticales.
- Lorsque la durée de l'échelle cyclique augmente, les lignes brisées représentant les instances de l'échelle cyclique deviennent moins nombreuses, s'éloignent, et deviennent de plus en plus horizontales.



Durée de l'échelle cyclique:
400 jours



Durée de l'échelle cyclique:
200 jours



Durée de l'échelle cyclique:
100 jours

FIGURE 5.11 – Changement d'échelle dans le diagramme en dents de scie

5.3.2 Représentation des événements dans le diagramme mixte

Les événements sont représentés dans le diagramme temporel mixte en fonction de leur moment d'apparition. Ils peuvent y être représentés individuellement, ou de manière agrégée.

5.3.2.1 Position des entités graphiques dans le diagramme

Dans les deux formes du diagramme temporel mixte, les événements sont représentés par des entités graphiques placées le long de la courbe du diagramme. La position de chaque entité sur la courbe est liée à un moment temporel. Plusieurs cas de figure peuvent se poser :

- Dans le cas d'une représentation de chaque événement par une entité graphique ponctuelle, la position de chaque entité le long de la courbe représente le moment d'apparition de l'événement.
- Dans le cas où des entités graphiques ponctuelles représentent des agrégats d'événements, la position de chaque entité le long de la courbe représente un moment temporel calculé à partir des moments d'apparition des événements agrégés.

La position x_{EG} d'une entité graphique EG le long de l'axe des abscisses représente un moment temporel t_{EG} dans le temps linéaire, selon la formule $x_{EG} = \alpha \times (t_{EG} - t_0)$, provenant du système d'équations 5.1.

Cas du diagramme sinusoïdal. Du fait de la nature sinusoïdale de la courbe du diagramme, la position y_{EG} de l'entité graphique EG sur l'axe des ordonnées ne représente pas le moment temporel t_{EG} au travers du temps cyclique, mais le résultat de la formule $y_{EG} = \beta \times \sin((t_{EG} - t_0) \times (\frac{2 \times \pi}{T}) + \phi)$, provenant du système d'équations 5.1 (Figure 5.12).

Le moment d'apparition t_{EG} est représenté, selon l'échelle temporelle cyclique du diagramme, par la position de l'entité graphique EG au sein d'une sinusoïde de la courbe du diagramme.

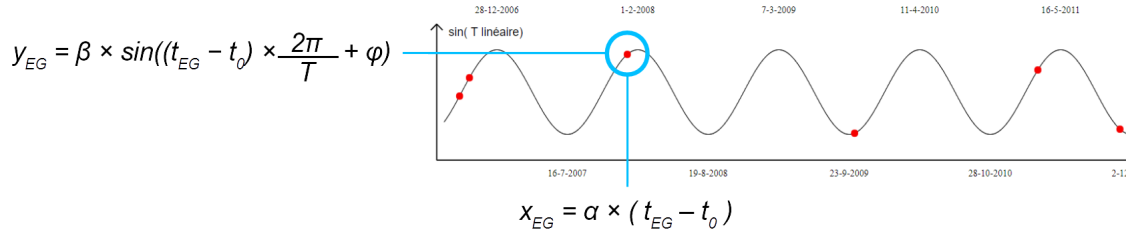


FIGURE 5.12 – Position des entités graphiques dans le diagramme sinusoïdale

Cas du diagramme en dents de scie. À l'inverse, la position y_{EG} de l'entité graphique EG , sur l'axe des ordonnées du diagramme, est la représentation directe du moment temporel t_{EG} au travers de l'échelle cyclique représentée (Figure 5.13).

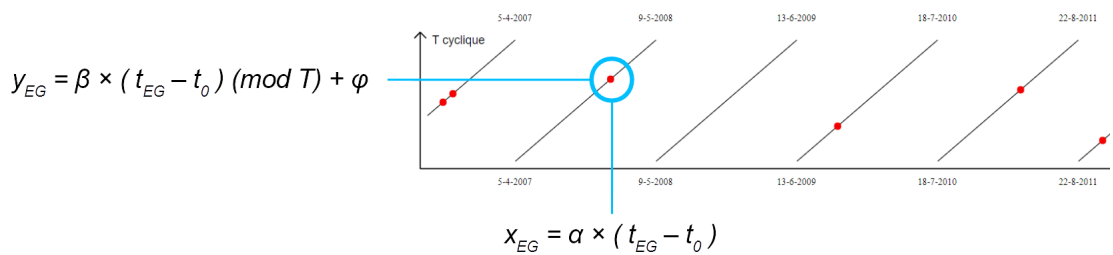


FIGURE 5.13 – Position des entités graphiques dans le diagramme en dents de scie

5.3.2.2 Représentation d'événements et d'agrégats d'événements

Comme énoncé précédemment, les événements peuvent être représentés individuellement par une entité ponctuelle, ou bien ces entités peuvent représenter des agrégats d'événements, afin de traiter un grand nombre d'événements :

- Dans le premier cas, la position de l'entité graphique représente le moment d'apparition de l'événement correspondant. Toutes les entités se présentent sous la forme d'un même symbole ponctuel, de même taille. Les événements apparaissant au même moment sont représentés par des entités ponctuelles superposées.
- Dans le second cas, le moment représenté par la position de l'entité graphique est calculé à partir des moments d'apparition des événements agrégés. La taille de chaque entité est calculée à partir du nombre d'événements agrégés.

Se pose alors le problème de la création des agrégats d'événements, et de leur représentation graphique.

Création des agrégats et étendue temporelle. Les agrégats d'événements sont créés de manière récursive à partir de la liste d'événements à représenter :

- Une valeur seuil, paramétrable, définit un espacement chronologique, en deçà duquel deux événements doivent être agrégés (Figure 5.14).
- Si l'espacement chronologique entre deux événements est inférieur à cette valeur, ils forment un premier agrégat (Repères 1 et 2 de la Figure 5.14).
- Cet agrégat est associé à un moment temporel, correspondant au barycentre des moments d'apparition des deux événements (Repère 2 de la Figure 5.14).
- Ce moment est utilisé pour calculer l'espacement chronologique entre l'agrégat et l'événement le plus proche temporellement. Si la distance temporelle entre ce troisième événement et l'agrégat est inférieure à la valeur seuil, l'événement est agrégé à l'ensemble (Repère 2 et 3 de la Figure 5.14).
- Le moment associé à l'agrégat est recalculé à partir des moments d'apparition des trois événements (Repère 3 de la Figure 5.14).

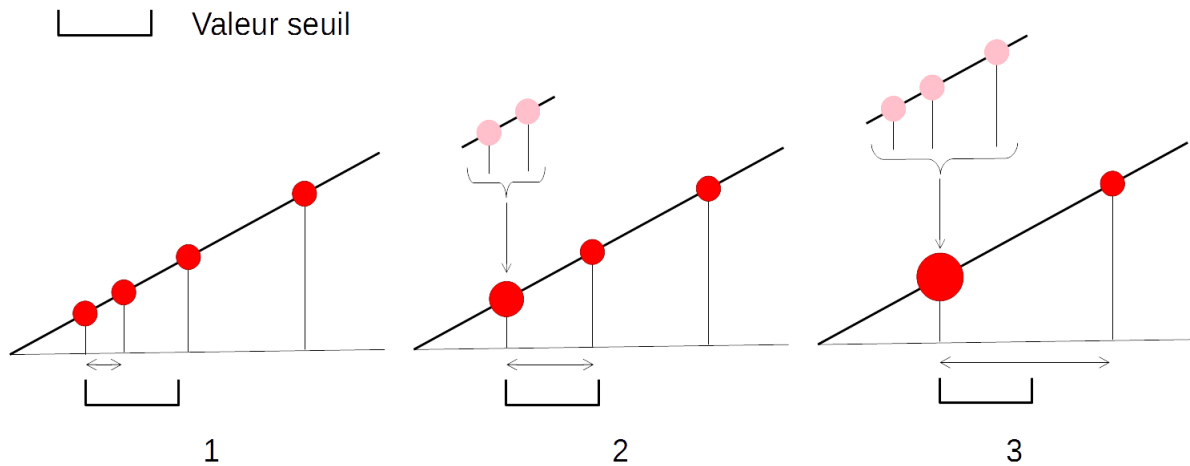


FIGURE 5.14 – Procédé de création des agrégats temporels

La valeur seuil influe sur l'étendue temporelle des agrégats d'événements. Une valeur seuil élevée entraîne l'agrégation d'événements sur des étendues temporelles plus larges. L'étendue temporelle couverte par les événements agrégés peut être représentée sur le diagramme temporel, lors d'un survol des entités graphiques par le curseur de la souris (Figure 5.15).

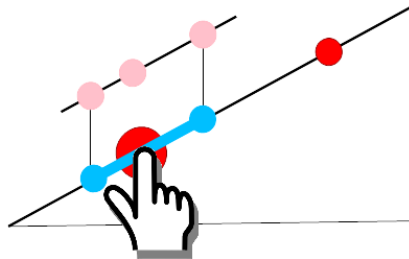


FIGURE 5.15 – Affichage de l'étendue temporelle des agrégats lors d'un survol

La valeur seuil ne désigne pas directement une étendue temporelle, mais une distance graphique entre deux positions sur l'axe des abscisses du diagramme. Elle définit une distance graphique, en deçà de laquelle les représentations ponctuelles de deux événements doivent être agrégées en une seule entité graphique. Il s'agit néanmoins d'une agrégation temporelle, car cette distance graphique critique correspond à un espacement chronologique. La durée de celui-ci est

dépendante de l'étendue temporelle représentée dans le diagramme (soit du *niveau de zoom* du diagramme). Plus celle-ci est petite, plus l'espacement chronologique en deçà duquel deux événements sont agrégés est petit. Deux événements agrégés, pour une étendue donnée couverte par le diagramme, peuvent être représentés par deux entités distinctes si l'étendue couverte par le diagramme diminue (si le *niveau de zoom* du diagramme est plus élevé).

Dans le cas d'un diagramme en dents de scie, les événements apparaissant en fin et en début de période cyclique ne sont pas agrégés, car ils sont suffisamment distants graphiquement pour être discernés. À l'inverse, ces mêmes événements peuvent être agrégés dans un diagramme sinusoïdal.

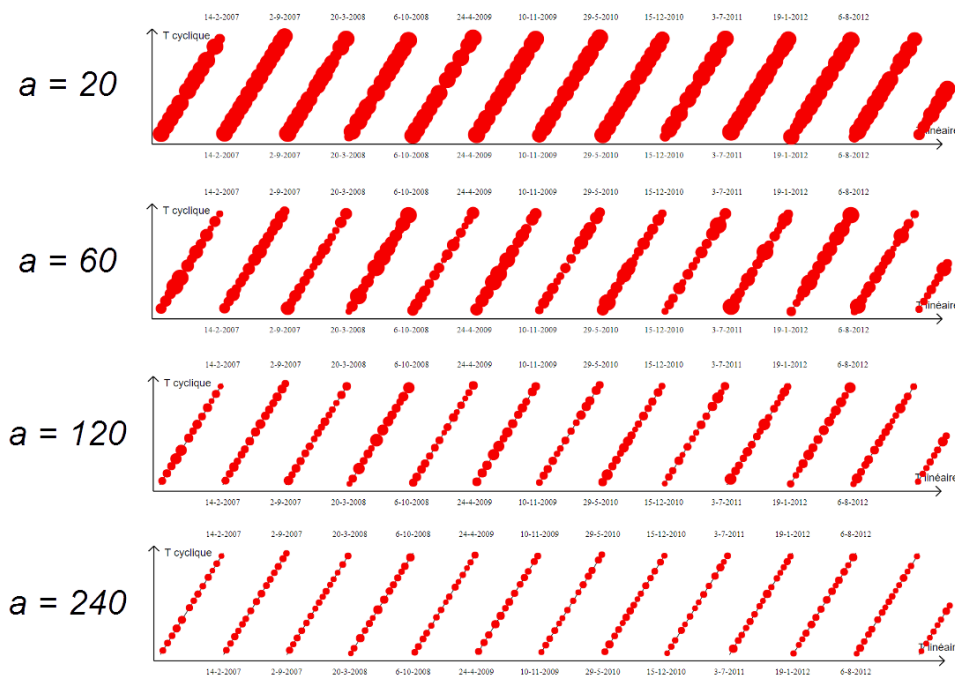
Représentation de l'agrégat. Chaque agrégat est représenté par une entité ponctuelle, dont la position sur la courbe du diagramme correspond au "barycentre" des moments d'apparition des événements agrégés.

La taille Ta_{EGagg} des entités graphiques $EGagg$ est calculée en fonction du nombre d'événements agrégés n_{evt} , selon le système d'équations (5.3).

$$\begin{cases} Ta_{EGagg} = Ta_0 + Ta * (n_{evt}/a) & \text{si } n_{evt} < a \text{ ou } n_{evt} = a \\ Ta_{EGagg} = Ta_0 + Ta + \log(n_{evt} - (a - 1)) & \text{si } n_{evt} > a \end{cases} \quad (5.3)$$

Dans ce système d'équations :

- Ta_0 représente la taille minimum des entités graphiques sur le diagramme ;
- Ta une constante graphique ;
- a une constante définissant le nombre d'événements au-delà duquel la taille d'une entité est calculée de manière logarithmique, et non plus de manière proportionnelle.



*Homicides sur le territoire de Chicago du 31-7-2006 au 4-6-2013
 Nombres d'événements: 3150
 Source: City of Chicago Data Portal*

FIGURE 5.16 – Représentation des agrégats temporels selon différents paramètres pour le calcul de la taille des entités

L'utilisation d'une fonction logarithmique pour calculer la taille des entités au-delà de a événements agrégés permet de représenter des agrégats rassemblant de nombreux événements. La valeur de la constante a peut être paramétrée selon le nombre d'événements à représenter (Figure 5.16). Affecter une forte valeur à la constante a permet de mieux visualiser les écarts entre des agrégats rassemblant de nombreux événements.

5.3.3 Intérêts et limites des deux approches

L'intégration de ces deux diagrammes dans la fenêtre temporelle, permet de contrebalancer les inconvénients inhérents à chacun de ces diagrammes.

Du fait de sa forme continue, le diagramme sinusoïdal permet une meilleure appréciation de la continuité dans le temps des différentes instances de l'échelle temporelle cyclique représentée. À l'inverse, des événements apparaissant en fin et en début de période cyclique, dans le diagramme en dents de scie, sont plus difficilement décelables comme temporellement proches, car les entités graphiques correspondantes sont éloignées l'une de l'autre dans ce diagramme.

Dans la Figure 5.17, les événements représentés sont séparés par des durées de temps égales.

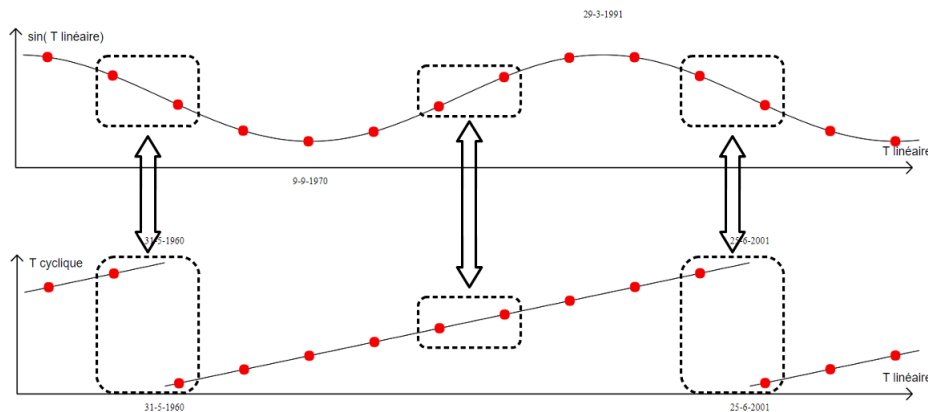


FIGURE 5.17 – Visualisation de la continuité du temps linéaire dans les deux types de diagrammes temporels mixte

s

Le diagramme sinusoïdal permet une visualisation du caractère continu de l'apparition des événements. Cette visualisation est plus difficile dans le diagramme en dents de scie à cause de la plus grande distance graphique séparant les événements arrivant en fin et en début d'instance de période cyclique.

Cependant, la nature sinusoïdale du premier diagramme rend plus difficile l'identification du moment d'apparition d'un événement dans l'échelle cyclique représentée. Cette lecture est rendue plus aisée dans le diagramme en dents de scie.

Dans la Figure 5.18, des entités graphiques possédant la même position sur l'axe des ordonnées dans le diagramme sinusoïdal semblent correspondre à des événements apparaissant durant le même intervalle temporel cyclique. Le diagramme en dents de scie montre cependant que ces événements apparaissent à des instants cycliques différents.

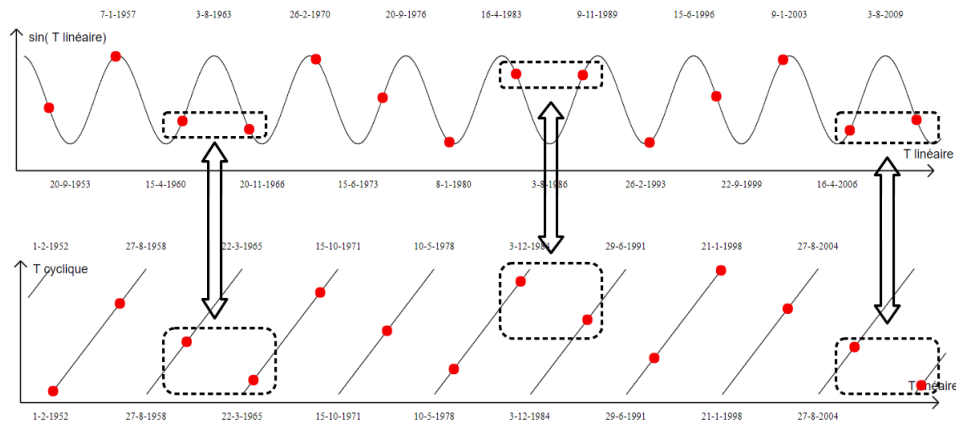


FIGURE 5.18 – Visualisation des moments d'apparition dans le temps cyclique dans les deux types de diagrammes temporels mixte

S

Enfin, le modèle sinusoïdal de la courbe du premier diagramme fausse légèrement la lecture des espacements temporels entre les événements, car une même durée temporelle ne correspond pas à la même longueur graphique tout au long de la courbe. Un même intervalle temporel t sera représenté par une longueur graphique moins importante aux extrémités de la courbe qu'en son milieu (Figure 5.19).

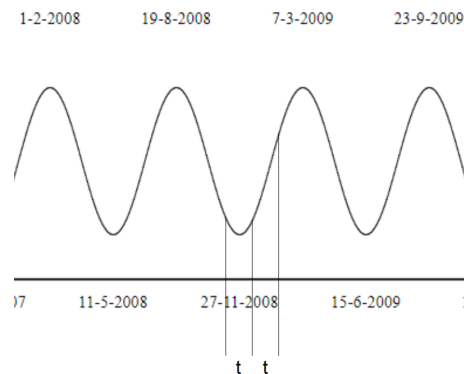


FIGURE 5.19 – Caractère non constant du rapport entre intervalle temporel représenté et longueur graphique dans le diagramme sinusoïdal

5.3.4 Diagramme temporel circulaire

5.3.4.1 Principe

Le second diagramme inclus dans la fenêtre temporelle de notre proposition est un diagramme en secteur, inspiré du diagramme de Florence Nightingale [Nightingale, 1858] et du *Sector Graph* de Harris [Harris, 2000]. Il représente l'évolution du nombre d'apparitions d'événements selon l'échelle cyclique représentée dans le diagramme temporel mixte. Le nombre d'apparitions d'événements, au cours du temps, est calculé par une agrégation des événements selon une division de l'échelle temporelle cyclique.

Ce diagramme prend une forme basée sur un disque découpé en sections d'angles égaux, chaque section représentant un intervalle temporel cyclique pour laquelle les événements sont agrégés (Figure 5.20). Le nombre de sections divisant l'échelle cyclique est paramétrable.

Le diagramme mixte et le diagramme circulaire représentent le temps au travers de la même échelle cyclique. L'étendue temporelle cyclique représentée par le diagramme circulaire correspond à la projection des étendues temporelles linéaires représentées par les sinusoïdes ou les lignes brisées du diagramme mixte.

Le sommet du diagramme circulaire représente le moment temporel cyclique T_O , marquant le début de l'échelle cyclique. Ce moment temporel est représenté par le début de chaque sinusoïde ou ligne brisée du diagramme mixte.

L'intervalle temporel cyclique représenté par une section du diagramme circulaire correspond à la projection de portions répétitives de la courbe du diagramme mixte, situées au même niveau sur chaque sinusoïde ou chaque ligne brisée du diagramme.

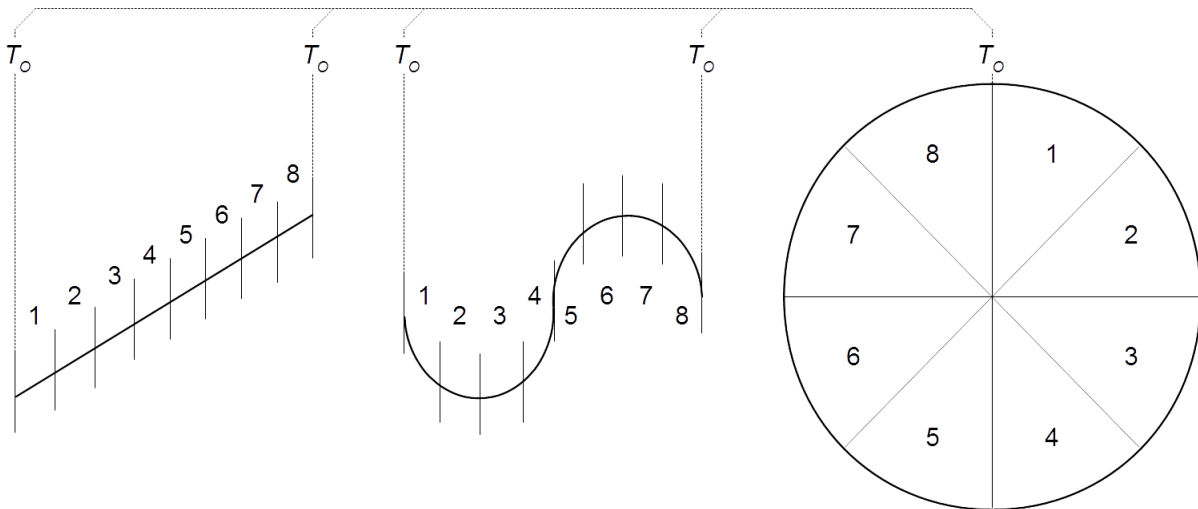


FIGURE 5.20 – Division de l'échelle temporelle cyclique en différents intervalles dans les diagrammes temporels

Le rayon d'une section du diagramme circulaire représente le nombre d'événements, apparaissant durant l'intervalle de temps cyclique correspondant (Figure 5.21). Ainsi, le rayon d'une section représente la somme des événements représentés dans les portions périodiques correspondantes de la courbe du diagramme temporel mixte. Ce rayon est calculé relativement au nombre total d'événements apparaissant durant l'étendue de temps linéaire étudiée.

Le rayon r_S , d'une section S du diagramme, est calculée selon l'équation suivante (5.4).

$$\left\{ r_S = R_{max} \times \frac{n_S}{n_{S-MAX}} \right. \quad (5.4)$$

Dans cette équation :

- R_{max} représente une constante graphique définissant le rayon maximal des sections du diagramme ;
- n_S représente le nombre d'événements apparaissant durant l'intervalle temporel correspondant à la section S ;
- n_{S-MAX} représente le nombre maximal d'événements, apparaissant durant un intervalle temporel représenté par une section du diagramme.

Ainsi :

- si l'ensemble des événements apparaît durant un seul intervalle temporel, le rayon de la section correspondante est égal à R_{max} ;
- si les événements apparaissent équitablement dans deux intervalles temporels, les deux sections correspondantes ont pour rayon R_{max} ;
- si un tiers des événements apparaît dans un intervalle représenté par une section de rayon r_{S1} , et que deux tiers des événements apparaissent dans un intervalle représenté par une section de rayon r_{S2} , alors $r_{S1} = R_{max}/2$ et $r_{S2} = R_{max}$.

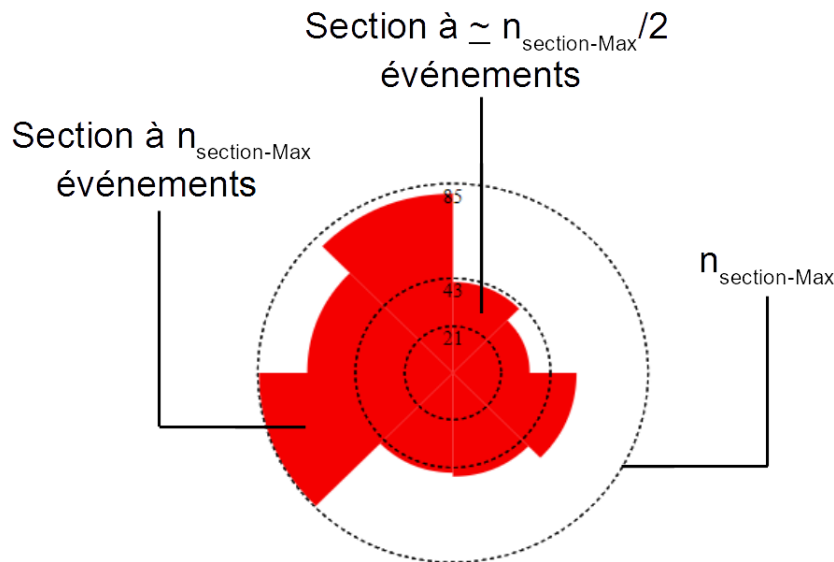


FIGURE 5.21 – Diagramme circulaire d'effectif

5.3.5 Relations entre le diagramme mixte et le diagramme circulaire

5.3.5.1 Synchronisation des diagrammes

Les diagrammes mixte et circulaire sont représentés conjointement dans l'interface de visualisation, et sont synchronisés entre eux. Ainsi une modification de la représentation des événements dans le diagramme mixte se répercute dans le diagramme circulaire.

Lors d'une modification de la durée de l'échelle cyclique, du déphasage de la courbe du diagramme, ou de l'emprise temporelle étudiée, le nombre de sections du diagramme circulaire reste constant. Les étendues temporelles linéaires projetées sur les sections du diagramme circulaire sont cependant modifiées. Le nombre d'événements apparaissant durant la somme de ces étendues temporelles linéaires évolue, entraînant une modification du rayon des sections correspondantes dans le diagramme circulaire.

5.3.5.2 Représentation des événements et représentation du nombre d'événements

Ces diagrammes diffèrent dans leur approche de la représentation de la distribution temporelle des événements :

- Le diagramme temporel mixte représente la distribution temporelle des événements, dans le temps linéaire et cyclique, à travers la représentation de la composante temporelle des événements sur le diagramme.
- Le diagramme temporel circulaire représente la distribution temporelle des événements dans le temps cyclique, à travers la représentation du nombre d'apparitions d'événements

dans le temps cyclique.

Ces diagrammes répondent ainsi à des usages différents, que nous présentons dans le chapitre 6 :

- Nous privilégions le diagramme temporel cyclique pour l'identification de concentrations temporelles des événements selon une échelle temporelle cyclique.
- Nous privilégions le diagramme temporel mixte pour analyser la composante temporelle des événements dans le temps linéaire, et pour effectuer une analyse visuelle des récurrences cycliques dans les données spatio-temporelles.

L'utilisation des diagrammes polaires est critiquée par Bertin [Bertin, 1973]. Selon lui, ces diagrammes ne permettent pas une comparaison efficace, ni d'un secteur angulaire par rapport à l'ensemble du diagramme, ni de plusieurs secteurs entre eux. Or, pour identifier une concentration temporelle des événements, nous devons identifier un ou plusieurs secteurs présentant un plus fort rayon que la moyenne des secteurs du diagramme.

Cependant, nous ne cherchons pas à identifier précisément la proportion d'événements représentée par chaque secteur. Nous cherchons à identifier, grossièrement, quels secteurs du diagramme possèdent les plus importants rayons, ces secteurs correspondant aux plus importants nombres d'apparitions d'événements. Nous émettons ainsi l'hypothèse qu'un diagramme polaire permet d'identifier des concentrations temporelles d'événements relatives à une échelle cyclique.

Ce diagramme présente également l'avantage d'effectuer une analogie entre sa structure circulaire, et la structure de l'échelle temporelle cyclique représentée, à la manière d'une horloge : chaque secteur correspond à un intervalle temporel cyclique, à la manière d'une *timewheel*. Cette analogie permet d'identifier facilement à quels moments temporels cycliques correspondent les concentrations d'événements identifiées dans ce diagramme.

5.4 Représentation de la composante spatiale des événements

La fenêtre spatiale de notre environnement de géovisualisation doit permettre la visualisation de la distribution spatiale des événements, et permettre une identification de clusters spatiaux. Pour cela, nous utilisons une carte d'inventaire en deux dimensions.

5.4.1 Principe

Notre analyse porte sur la distribution des événements dans l'espace, et doit permettre une visualisation de l'ensemble des événements apparaissant durant l'étendue temporelle étudiée, afin d'identifier des concentrations d'événements dans l'espace.

À l'instar de la fenêtre spatiale de *PerSE* [Li, 2010], et d'une version sans animation de la carte proposée par *Growth Ring Map* [Bak et al., 2009, Andrienko et al., 2011], notre fenêtre spatiale est composée d'une carte en deux dimensions, représentant la composante spatiale de l'ensemble des événements apparaissant durant l'étendue temporelle étudiée, et couverte par le diagramme temporel mixte.

Les événements sont projetés sur un fond de carte, pouvant être une carte physique, une orthophoto, une carte routière ... etc. Par défaut, le fond de carte affiché provient d'*OpenStreetMap*. Le choix du fond de carte peut entraîner un paradoxe temporel entre les objets décrits dans le fond de carte et les événements représentés. Un fond de carte *OpenStreetMap* édité de 2017 peut

représenter des routes ou des bâtiments n'existant pas aux mêmes époques que les événements représentés si ceux-ci sont antérieurs à 2017.

Cependant nous focalisons notre analyse sur les événements spatio-temporels, sans nous intéresser à l'évolution des objets spatiaux impactés par ces événements. Les objets géographiques présentés dans le fond de carte jouent uniquement un rôle de contextualisation spatiale, afin d'aider l'utilisateur à localiser ces événements dans l'espace. Sauf s'ils sont liés à la thématique du phénomène observé, ces objets n'influencent pas davantage dans l'analyse exploratoire. La représentation d'une forêt sur un fond de carte sert à localiser les événements par rapport à la position de cette forêt. Si les événements observés concernent des cambriolages, il n'est pas nécessaire que la représentation de cette forêt soit conforme à sa réelle étendue spatiale au moment de l'apparition des événements.

5.4.2 Représentation des événements dans la carte

Les événements sont représentés sur la carte en fonction de leur lieu d'apparition. Ils peuvent y être représentés individuellement, ou de manière agrégée.

5.4.2.1 Position des entités cartographiques

Les événements sont représentés par des entités graphiques placées sur la carte. Les événements peuvent être représentés individuellement par une entité ponctuelle. Afin de traiter un grand nombre de données, les événements proches peuvent être agrégés, et être représentés par une seule entité ponctuelle.

- Dans le cas d'une représentation de chaque événement par une entité ponctuelle, la position d'une entité est définie par les coordonnées géographiques de l'événement représenté. Toutes les entités se présentent sous la forme du même symbole ponctuel, de même taille. Les événements apparaissant au même lieu sont représentés par des entités ponctuelles superposées.
- Dans le cas d'une représentation d'agrégats d'événements par des entités ponctuelles, la position d'une entité sur la carte est calculée à partir des coordonnées géographiques des événements agrégés. La taille de chaque entité est calculée à partir du nombre d'événements agrégés.

Se pose alors la question de la création des agrégats spatiaux et de leur représentation dans la carte.

5.4.2.2 Création des agrégats et étendue spatiale

Les événements sont agrégés spatialement en utilisant la fonction de *clusterisation* spatiale de la librairie *JavaScript Openlayers 3*. Cette fonction permet d'agréger les entités ponctuelles représentant les événements de manière récursive, à l'instar du procédé que nous utilisons pour créer les agrégats temporels du diagramme temporel mixte.²

Une valeur seuil, paramétrable, définit une distance spatiale critique entre deux événements, en deçà de laquelle ces deux événements sont agrégés (Figure 5.22). La valeur seuil influe sur l'étendue spatiale des agrégats d'événements. Une valeur seuil élevée entraîne l'agrégation d'événements sur des étendues spatiales plus larges.

2. Nous précisons que dans le cas d'*Openlayers 3*, la *clusterisation* correspond à l'action d'agréger des entités spatiales proches. Elle diffère de ce que nous appelons un *cluster spatial*, qui définit une concentration d'entités dans l'espace, sans que cette proximité implique une agrégation.

L'agrégation des événements dans la carte suit alors ce procédé :

- Si la distance entre les lieux d'apparition de deux événements est inférieure à la valeur seuil, ils forment un premier agrégat (Repère 1 de la Figure 5.22).
- Cet agrégat est associé à des coordonnées géographiques, correspondant au barycentre des lieux d'apparition des deux événements (Repère 2 de la Figure 5.22).
- Ces coordonnées sont utilisées pour calculer la distance spatiale entre l'agrégat et l'événement le plus proche. Si la distance entre ce troisième événement et l'agrégat est inférieure à la valeur seuil, l'événement est agrégé à l'ensemble (Repère 3 de la Figure 5.22).
- Les coordonnées géographiques de l'agrégat sont alors recalculées. C'est ici que la démarche d'agrégation d'*Openlayers 3* diffère de la démarche que nous utilisons pour créer les agrégats temporels. Les coordonnées géographiques du nouvel agrégat ne correspondent pas au barycentre des lieux d'apparition des trois événements agrégés, mais au barycentre de l'ensemble formé par les coordonnées de l'ancien agrégat et celles du nouvel événement agrégé (Repère 4 de la Figure 5.22). Les coordonnées du nouvel agrégat sont ainsi plus proches de celles du dernier événement agrégé.

Le choix d'utiliser la fonction d'agrégation, proposée *Openlayers 3*, a été effectué pour obtenir un gain de temps lors de l'implémentation de l'environnement. La représentation des événements sous la forme d'agrégats dans la carte a pour objectif d'identifier grossièrement la position de regroupement d'événements dans l'espace. Nous avons ainsi considéré comme négligeable la distance entre la position d'un agrégat dans la carte, et la position du barycentre des événements agrégés.

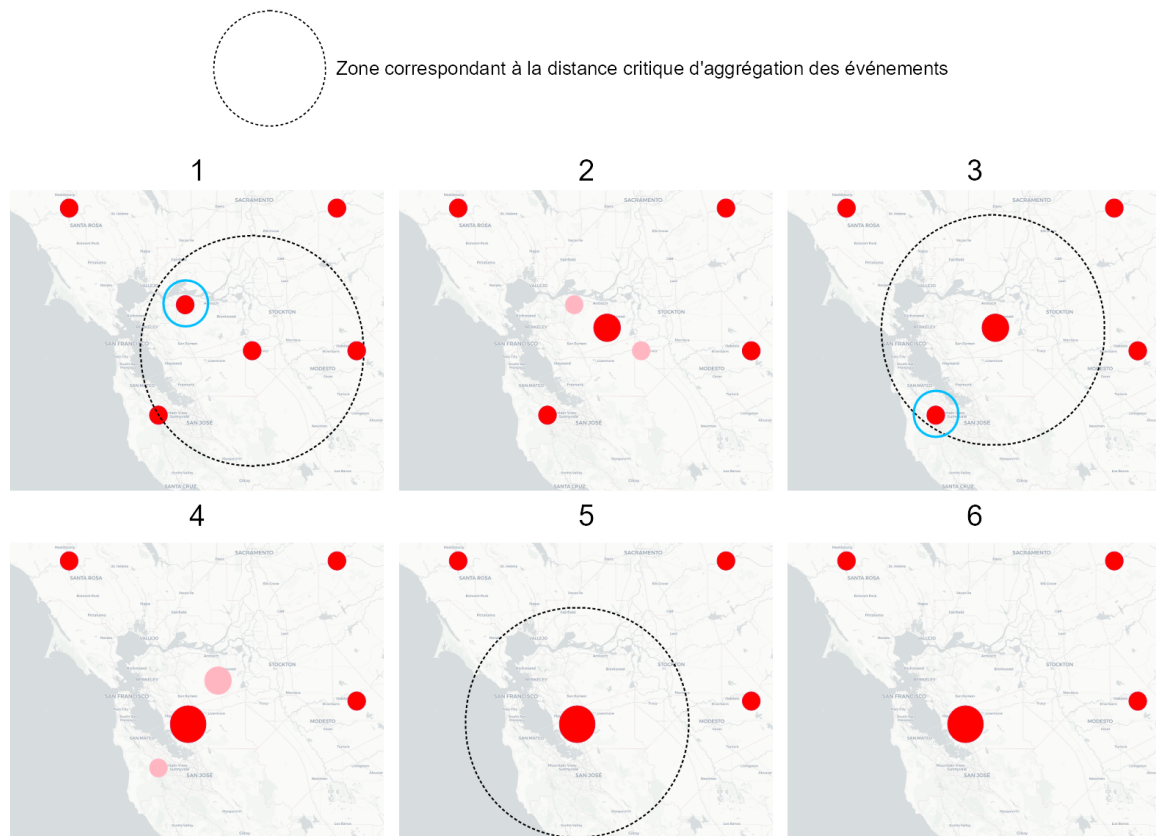


FIGURE 5.22 – Procédé de création des agrégats spatiaux

L'étendue spatiale couverte par un agrégat peut être affichée dans la carte, au moyen d'un survol de l'entité correspondante par le curseur de la souris (Figure 5.23). L'étendue représentée correspond à l'enveloppe convexe des lieux d'apparition des différents événements agrégés.

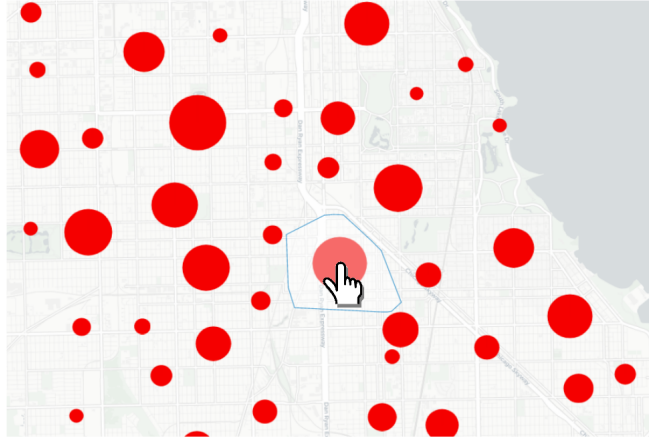


FIGURE 5.23 – Affichage de l'étendue spatiale des agrégats lors d'un survol

De la même manière que pour l'agrégation temporelle, la valeur seuil ne désigne pas directement une distance spatiale, mais une distance graphique entre deux points sur la carte, en deçà de laquelle les représentations ponctuelles de deux événements doivent être agrégées en une seule entité graphique. Il s'agit d'une agrégation spatiale, car cette distance graphique critique représente une distance spatiale. La longueur de cette dernière dépend cependant de l'emprise spatiale de la carte (soit du *niveau de zoom* de la carte). Plus celle-ci est petite, plus la distance spatiale en deçà de laquelle deux événements sont agrégés est petite. Deux événements agrégés pour une emprise spatiale donnée pourront être représentés par deux entités distinctes si l'emprise de la carte diminue (si le *niveau de zoom* de la carte augmente).

5.4.2.3 Représentation de l'agrégat

Chaque agrégat est représenté par une entité ponctuelle, dont la position sur la carte est calculée à partir des coordonnées des événements agrégés.

La taille des entités cartographiques t_{EC} est calculée en fonction du nombre d'événements agrégés n_{evt} , selon un système d'équations similaire à celui utilisé pour les agrégats temporels (Équation (5.5)).

$$\begin{cases} t_{EC} = TC_0 + TC * (n_{evt}/b) & \text{si } n_{evt} < b \text{ ou } n_{evt} = b \\ t_{EC} = TC_0 + TC + \log(n_{evt} - (b - 1)) & \text{si } n_{evt} > b \end{cases} \quad (5.5)$$

Dans ce système d'équations :

- TC_0 représente la taille minimum des entités graphiques sur la carte ;
- TC une constante graphique ;
- b une constante définissant le nombre d'événements au-delà duquel la taille d'une entité est calculée de manière logarithmique, et non plus de manière proportionnelle.

L'utilisation d'une fonction logarithmique pour calculer la taille des entités au-delà de b événements agrégés répond aux mêmes besoins que pour l'agrégation temporelle (Figure 5.24). Affecter une forte valeur à la constante b permet de mieux visualiser les écarts entre des agrégats rassemblant de nombreux événements.

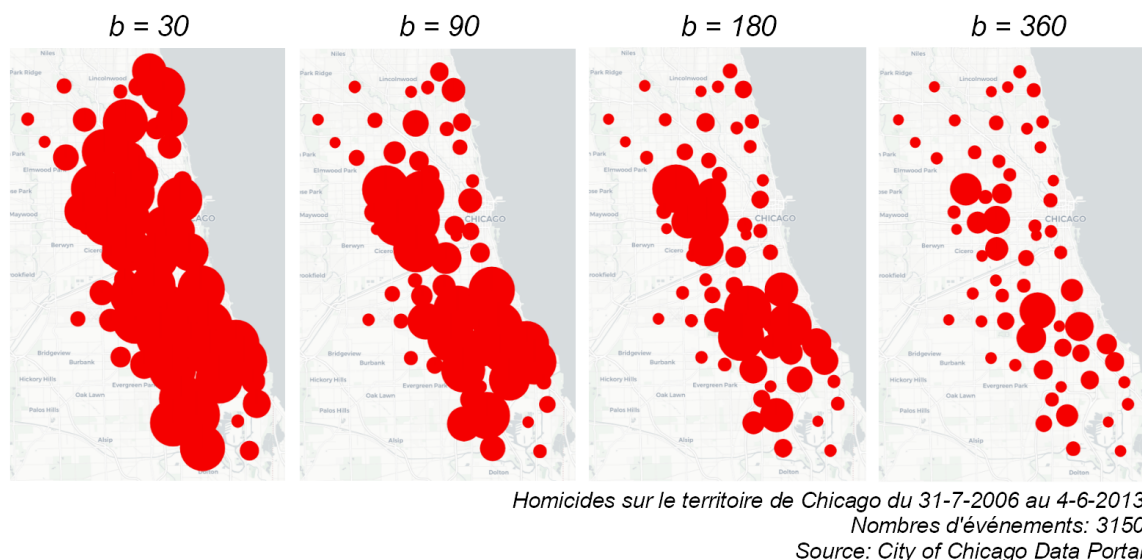


FIGURE 5.24 – Représentation des agrégats spatiaux selon différents paramètres pour le calcul de la taille des entités

5.5 Représentation de la composante spatio-temporelle des événements

Les parties 5.3 et 5.4 de ce chapitre sont focalisées sur la représentation des distributions temporelle et spatiale des événements. Nous y présentons des approches de représentation graphique permettant d'identifier des structures dans ces distributions, rendant possible l'analyse de patterns spatiaux ou temporels. Dans la partie suivante, nous nous focalisons sur l'association de ces représentations, dans le but d'analyser la distribution spatio-temporelle des événements.

Pour permettre une identification de patterns spatio-temporels, il faut pouvoir faire le lien entre ce qui est représenté dans la fenêtre temporelle et ce qui est représenté dans la fenêtre spatiale, afin de visualiser des relations entre la distribution spatiale et la distribution temporelle des événements. Dans l'état de l'art (Section 4.2), nous montrons que ces relations peuvent être représentées grâce à la combinaison du multi-fenêtrage avec d'autres procédés graphiques, intégrant les dimensions spatiale et temporelle. Ces procédés permettent, en intégrant une dimension supplémentaire dans les fenêtres de l'environnement, d'effectuer le lien entre les deux composantes des événements représentés.

Dans cet état de l'art, nous montrons également que l'utilisation de règles de symbologie permet d'effectuer le lien entre les fenêtres spatiale et temporelle d'un environnement à multi-fenêtrage, en intégrant les aspects linéaire et cyclique du temps. L'utilisation de règles de symbologie permet ainsi de représenter les événements :

- en fonction de leur composante spatiale dans la fenêtre temporelle, à l'instar de *Time Wave* [Li, 2010] ;
- en fonction de leur composante temporelle, au travers d'une échelle linéaire ou cyclique, dans la fenêtre spatiale, à l'instar de *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011].

L'utilisation de règles de symbologie pour la représentation d'agrégats d'événements en fonction d'un critère dimensionnel, pose alors, comme dans *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011], la question de la représentation de ces agrégats en fonction des composantes

dimensionnelles des événements agrégés.

Cependant, si l'utilisation de règles de symbologie permet d'effectuer le lien entre le contenu des différentes fenêtres d'un environnement à multi-fenêtrage, ce lien peut également être effectué au moyen d'outils de sélection et de *brushing*, reposant sur le principe de synchronisation des fenêtres.

5.5.1 Sélection et *brushing* des événements selon un critère dimensionnel

L'outil permet une sélection graphique des événements selon un critère dimensionnel (Repère A de la Figure 5.25) :

- soit temporel, au moyen d'une sélection d'entités graphiques dans le diagramme temporel mixte ;
- soit spatial, au moyen d'une sélection d'entités cartographiques dans la carte.

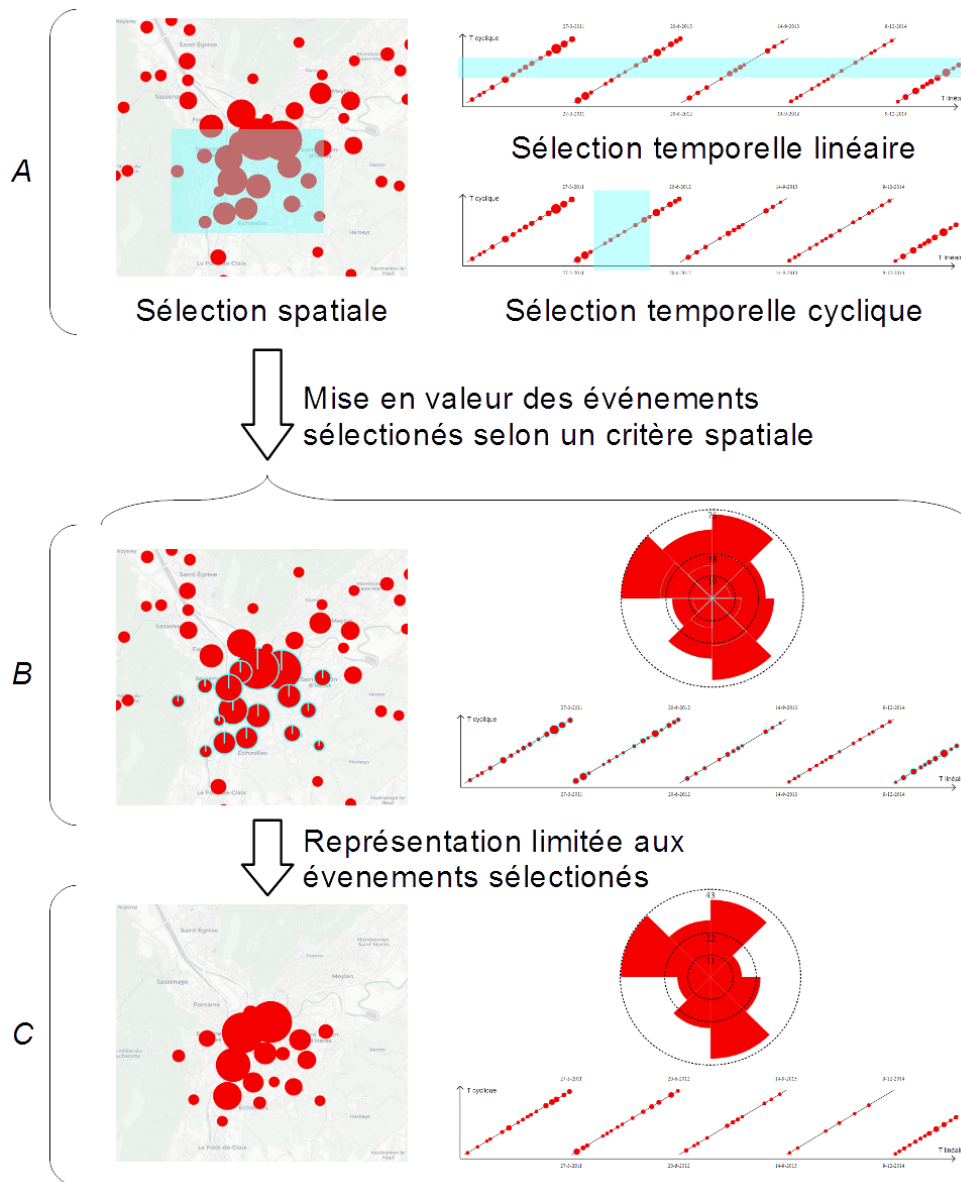


FIGURE 5.25 – Sélection et *brushing* des événements selon un critère dimensionnel

La sélection d'événements au sein d'une fenêtre dimensionnelle, combinée au principe de synchronisation des fenêtres spatiales et temporelles, permet de mettre en valeur les événements sélectionnés dans l'ensemble des fenêtres de l'interface (Repère *B* de la Figure 5.25). La sélection d'un événement appartenant à un agrégat entraîne la mise en valeur de l'entité graphique représentant l'agrégat. Cependant seul l'événement est sélectionné.

La représentation peut alors être limitée aux événements sélectionnés, les événements non sélectionnés étant masqués dans les fenêtres temporelle et spatiale (Repère *C* de la Figure 5.25). Ce procédé est appelé *brushing* [Davoine, 2016].

Ces outils de sélection et de *brushing*, liés à la synchronisation des fenêtres spatiales et temporelles, permettent de relier les composantes temporelle et spatiale des événements sélectionnés. Ils rendent possible une visualisation :

- de la dimension spatiale d'événements sélectionnés sur des critères temporels ;
- de la dimension temporelle d'événements sélectionnés sur des critères spatiaux.

Une possible corrélation entre la distribution spatiale et la distribution temporelle des événements peut alors être mise en évidence de manière interactive.

Ce procédé n'est cependant pas optimal pour identifier de potentiels patterns spatio-temporels. L'analyse de patterns s'effectue de manière individualisée sur les différents groupes d'événements sélectionnés, l'interface ne permettant pas une visualisation globale de la distribution spatio-temporelle des événements.

De plus, l'absence de connaissances *a priori* sur la distribution spatio-temporelle des événements peut rendre difficile une sélection d'événements présentant une distribution spatio-temporelle significative. Les critères de sélection optimaux, permettant de mettre en évidence un pattern spatio-temporel, sont inconnus de l'analyste.

L'utilisation de règles de symbologie, telles que celles utilisées dans Time Wave [Li, 2010] et *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011], permet une approche plus efficace pour identifier des patterns. En permettant d'intégrer une composante dimensionnelle supplémentaire dans une fenêtre de l'interface, ces outils offrent la possibilité de faire le lien entre les composantes temporelle et spatiale de l'ensemble des événements représentés.

Par défaut, les entités graphiques représentant les événements possèdent la même symbologie dans les fenêtres spatiale et temporelle. Les événements peuvent cependant être représentés, au sein des deux fenêtres, en fonction d'un critère dimensionnel supplémentaire en suivant les principes de la sémiologie graphique. Les événements peuvent alors être représentés :

- dans la fenêtre temporelle selon leur lieu d'apparition ;
Nous désignons cette représentation comme ayant une **entrée spatiale**, et étant **orientée de la carte vers les diagrammes temporels**.
- dans la fenêtre spatiale selon leur moment d'apparition relatif à une échelle linéaire ;
Nous désignons cette représentation comme ayant une **entrée temporelle linéaire**, et étant **orientée du diagramme temporel mixte vers la carte et le diagramme temporel circulaire**.
- dans la fenêtre spatiale selon leur moment d'apparition relatif à une échelle cyclique.
Nous désignons cette représentation comme ayant une **entrée temporelle cyclique**, et étant **orientée du diagramme temporel circulaire vers la carte et le diagramme temporel mixte**.

5.5.2 Représentation des événements selon leur lieu d'apparition

5.5.2.1 Principe : une représentation orientée de la carte vers les diagrammes temporels

Les événements sont représentés en fonction de leur lieu d'apparition, symbolisé par la position des entités graphiques correspondantes dans la carte. L'emprise spatiale de la carte est divisée en différentes zones spatiales. Les événements sont ensuite représentés en fonction de l'appartenance de leur lieu d'apparition à ces zones, de la même manière que dans *PerSE* [Swedberg and Peuquet, 2017].

Nous proposons d'appliquer un découpage du territoire pouvant :

- soit être créé à la volée dans la fenêtre spatiale (Repère *A* de la Figure 5.26) ;
- soit correspondre à un découpage du territoire préalablement établi, et importé dans l'environnement de géovisualisation (Repère *B* de la Figure 5.26). Utiliser ce second procédé permet de représenter les événements selon un découpage du territoire correspondant à une problématique particulière (découpage administratif, bassins versants, régions présentant différents niveaux de pollution, etc.), permettant de croiser cette thématique de recherche avec l'analyse de la distribution spatio-temporelle des événements.

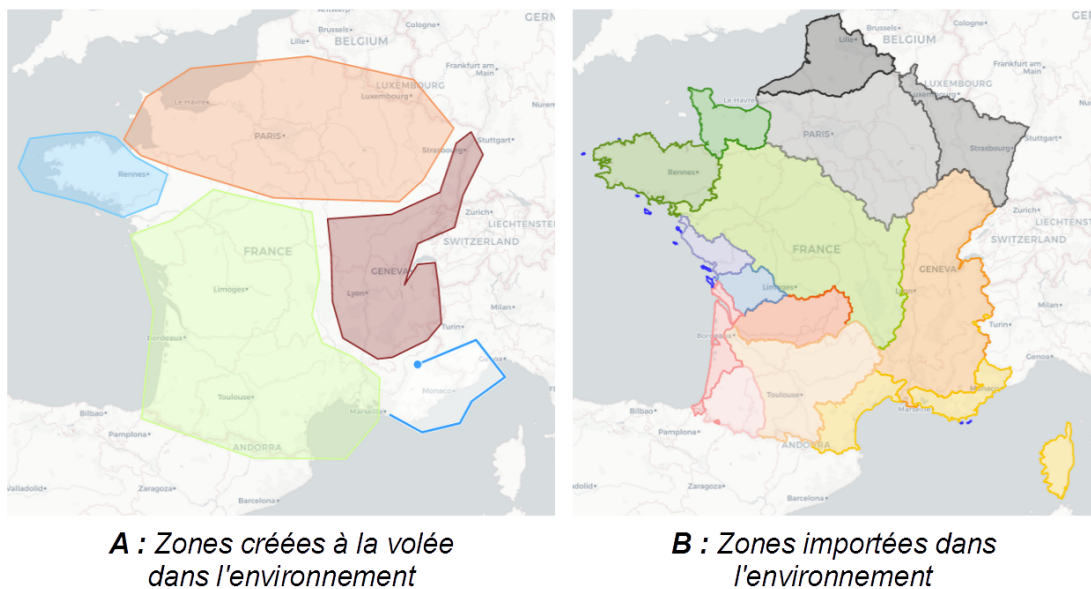


FIGURE 5.26 – Création des zones de classification spatiale

L'appartenance d'un événement à une zone peut être traitée comme une variable qualitative nominale rattachée à l'événement : aucune relation hiérarchique ne lie les zones spatiales entre elles. Nous utilisons alors la variable visuelle de la couleur pour représenter les événements selon leur lieu d'apparition. Une couleur, paramétrable, est attribuée à chaque zone, et est utilisée pour représenter les événements y apparaissant. Une couleur est également définie pour représenter les événements apparaissant en dehors des zones créées. Ces derniers sont désignés comme apparaissant dans la zone spatiale *par défaut*, regroupant l'ensemble des points de la carte n'étant pas contenus dans le découpage du territoire utilisé.

La classification des événements en différentes zones spatiales peut être utilisée pour effectuer un *brushing* des événements : la visualisation peut être limitée aux événements apparaissant dans une zone spatiale particulière, en masquant les événements apparaissant dans les autres zones du

découpage. Les fenêtres spatiale et temporelle étant synchronisées, ces événements sont masqués dans la carte et dans les diagrammes de la fenêtre temporelle.

5.5.2.2 Représentation des événements dans la carte et dans le diagramme temporel mixte

La couleur associée à une zone de classification spatiale est affectée à l'ensemble des événements dont le lieu d'apparition est inclus dans cette zone. Cette couleur est alors affectée aux entités graphiques correspondantes dans la carte et dans le diagramme temporel mixte :

- Dans le cas d'une entité graphique représentant un seul événement, l'entité graphique prend la couleur affectée à l'événement.
- Dans le cas d'une entité graphique représentant un agrégat d'événements apparaissant dans une même zone spatiale, l'entité graphique prend la couleur affectée à la zone où apparaissent les événements.
- Dans le cas d'une entité graphique représentant un agrégat d'événements apparaissant dans des zones spatiales différentes, la question de la représentation de l'agrégat, en fonction des zones spatiales où apparaissent les événements agrégés, se pose. L'agrégat doit être représenté de manière à permettre l'identification de la proportion, au sein de l'agrégat, de chaque groupe formé par les événements apparaissant dans la même zone spatiale.

Pour répondre à une problématique semblable à ce troisième cas de figure (Chapitre 4), *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011] utilise un réagencement des pixels représentant les événements spatio-temporels dans la carte, de manière à ce que les pixels représentant les événements apparaissant dans les mêmes intervalles temporels soient regroupés dans l'agrégat. Cette représentation permet ainsi d'identifier la proportion de chaque groupe formé par les événements apparaissant au même moment, dans chaque agrégat.

Nos agrégats d'événements étant représentés par une entité graphique unique, nous ne pouvons utiliser ce procédé. En revanche, l'exemple de la Figure 4.44 [Andrienko and Andrienko, 2006], dans l'état de l'art 4, montre que l'utilisation d'entités graphiques ponctuelles sous la forme de *diagrammes camemberts* (ou *Pie Chart*) permet une représentation de la proportion des différentes valeurs d'une variable, rattachée à une même entité graphique ponctuelle.

Ainsi, dans le cas d'une entité graphique représentant un agrégat d'événements apparaissant dans différentes zones spatiales, cette entité est représentée sous la forme d'un *Pie Chart* :

- Le nombre de secteurs (ou sections angulaires) du *Pie Chart* définit le nombre de zones de classification spatiale dans lesquelles apparaissent les événements agrégés.
- Chaque secteur du *Pie Chart* correspond à une zone spatiale, et se voit affecter la couleur correspondante.
- L'ordre des différents secteurs du *Pie Chart* correspond à l'ordre dans lequel les zones spatiales correspondantes sont créées dans l'environnement de géovisualisation.
- L'angle au centre de chaque secteur est fonction de la proportion d'événements agrégés apparaissant dans la zone spatiale correspondante, par rapport au nombre total d'événements agrégés.

Par exemple, si une entité graphique représente un agrégat de huit événements apparaissant :

- pour deux d'entre eux (un quart), dans une première zone *A*, à laquelle est affectée la couleur bleue ;
- pour deux d'entre eux (un quart), dans une seconde zone *B*, à laquelle est affectée la couleur verte ;

- pour quatre d’entre eux (la moitié), dans une troisième zone *C*, à laquelle est affectée la couleur orange ;

l’entité graphique correspondante prend la forme d’un *Pie Chart* à trois sections (Repère * de la Figure 5.27) :

- la première, bleue, couvrant un quart du diagramme, et représentant la proportion d’événements apparaissant dans la zone *A* ;
- la seconde, verte, couvrant un quart du diagramme, et représentant la proportion d’événements apparaissant dans la zone *B* ;
- la dernière, orange, couvrant la moitié du diagramme, et représentant la proportion d’événements apparaissant dans la zone *C*.

La position et la taille des agrégats sont calculées de la même manière que précédemment, la taille représentant le nombre d’événements agrégés.

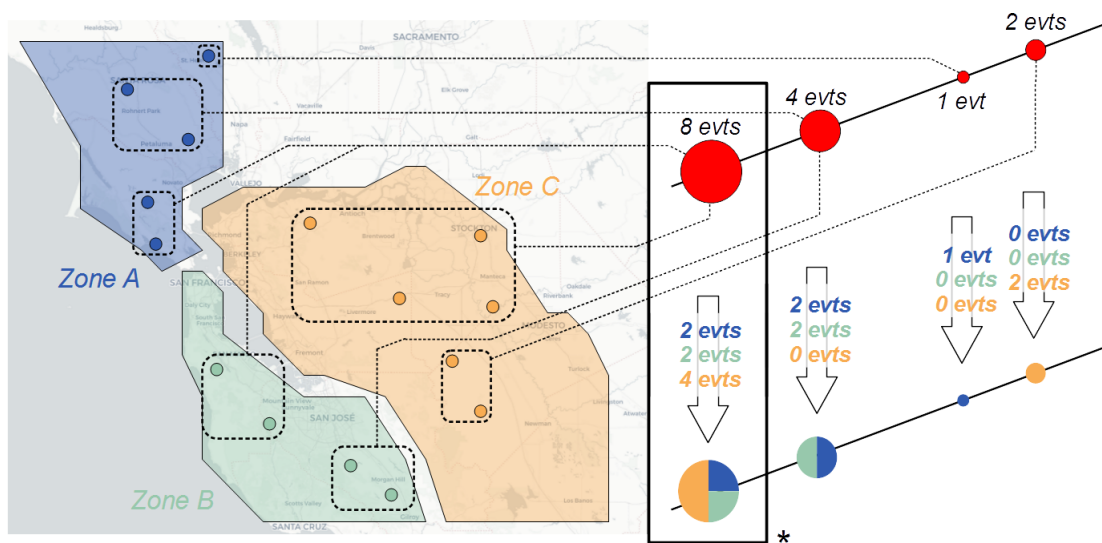


FIGURE 5.27 – Représentation d’agrégats d’événements selon leur lieu d’apparition

Cette représentation en *Pie Chart* permet ainsi de représenter les événements sous une forme agrégée, tout en représentant les proportions de chaque classe d’événements dans les agrégats.

Cette représentation est appliquée de la même manière dans la carte et dans le diagramme temporel mixte. Dans la carte, elle permet la représentation d’agrégats dont les événements sont situés de part et d’autre de la frontière entre deux zones de classification spatiale. Un agrégat comportant 25% d’événements apparaissant dans une zone *bleue*, et 75% d’événements apparaissant dans une zone *rouge*, est représenté dans la carte par un *Pie Chart* à deux sections :

- l’une, bleue, couvrant le premier quart du diagramme ;
- l’autre, rouge, couvrant les trois quarts restants.

L’utilisation des diagrammes de type *Pie Chart* pour représenter l’information est critiquée. Selon Bertin [Bertin, 1973], ces diagrammes permettent :

- une comparaison de la proportion des différentes classes représentées par rapport à l’ensemble du diagramme ;

mais ne permettent pas :

- une comparaison de la proportion des différentes classes entre elles ;

- une représentation des classes nulles, pour lesquelles il n'existe pas d'événements.

Cependant, comme nous le présentons dans le chapitre suivant (6), notre objectif n'est pas d'observer, pour chaque agrégat d'événements, la distribution des événements agrégés selon les différentes classes utilisées. Notre objectif est d'observer grossièrement l'ensemble des agrégats, afin d'identifier lesquels présentent une concentration d'événements agrégés dans une, ou quelques-unes, des classes. Nous ne cherchons pas à comparer entre eux les secteurs de chaque *Pie Chart*, mais à identifier la surreprésentation d'un secteur dans les *Pie Charts*.

Ainsi, nous nous intéressons à la possibilité, donnée par les *Pie Charts*, de comparer la proportion des différentes classes par rapport à l'ensemble du diagramme. Nous considérons la difficulté de comparer les secteurs entre eux, et l'impossibilité de représenter les classes nulles, comme négligeables.

5.5.2.3 Représentation des événements dans le diagramme temporel circulaire

Afin de représenter les événements selon leur appartenance aux différentes zones de classification spatiale, le diagramme temporel circulaire adopte la structure du *Rose Charts* de Florence Nightingale [Nightingale, 1858] (Figure 4.14).

Chaque section angulaire du diagramme circulaire, représentant le nombre d'événements apparaissant durant un intervalle temporel cyclique, est divisée en sous-sections selon l'axe radial :

- Pour chaque section angulaire, le nombre de sous-sections radiales correspond au nombre de zones spatiales comprenant des événements apparaissant dans l'intervalle temporel cyclique représenté.
- L'ordre de représentation des sous-sections radiales, du centre vers l'extérieur du diagramme, correspond à l'ordre dans lequel les zones spatiales correspondantes sont créées dans l'environnement de géovisualisation.
- La largeur radiale de chaque sous-section est fonction de la proportion d'événements apparaissant dans la zone spatiale correspondante, par rapport au nombre total d'événements apparaissant dans l'intervalle temporel représenté par la section angulaire du diagramme.

Par exemple, si un intervalle temporel cyclique, représenté par une section angulaire du diagramme circulaire, comporte huit événements apparaissant :

- pour deux d'entre eux (un quart), dans une première zone *A*, à laquelle est affectée la couleur bleue ;
- pour deux d'entre eux (un quart), dans une seconde zone *B*, à laquelle est affectée la couleur verte ;
- pour quatre d'entre eux (la moitié), dans une troisième zone *C*, à laquelle est affectée la couleur orange ;

la section angulaire correspondante est découpée en trois sous-sections radiales (Repère * de la Figure 5.28) :

- la première en partant du centre, bleue, couvrant un quart du rayon de la section, et représentant la proportion d'événements apparaissant dans la zone *A* ;
- la seconde, verte, couvrant un quart du rayon de la section, et représentant la proportion d'événements apparaissant dans la zone *B* ;
- la dernière, orange, couvrant la moitié du rayon de la section, et représentant la proportion d'événements apparaissant dans la zone *C*.

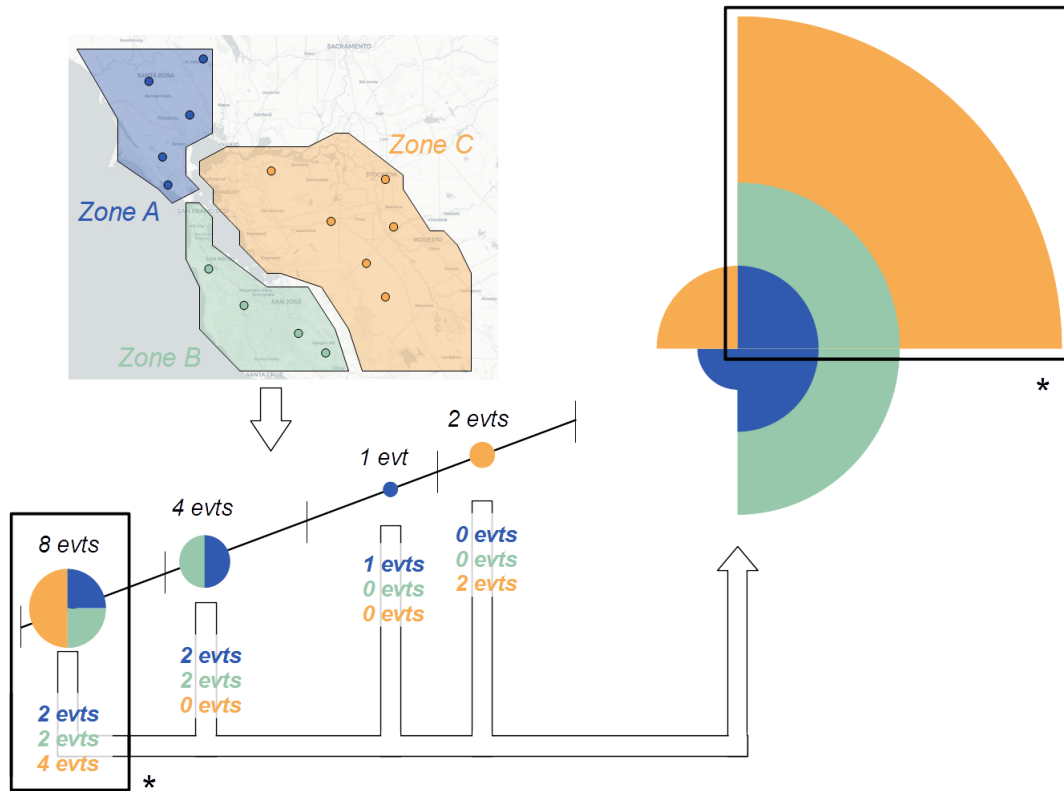


FIGURE 5.28 – Représentation de l’effectif des différentes classes d’événements, selon leur intervalle d’apparition relatif à une échelle cyclique

Dans le cas d’un *brushing* appliqué aux événements selon leur zone d’apparition, les sous-sections radiales correspondant aux zones masquées sont supprimées du diagramme. Le rayon des sections angulaires et la proportion des sous-sections radiales restantes sont recalculés.

5.5.3 Représentation des événements selon leur moment d’apparition relatif à une échelle linéaire

5.5.3.1 Principe : une représentation orientée du diagramme temporel mixte vers la carte et le diagramme temporel circulaire

Les événements sont représentés en fonction de leur moment d’apparition selon une échelle linéaire, symbolisé par la position des entités graphiques correspondantes le long de l’axe des abscisses du diagramme temporel mixte. Cet axe est divisé en plusieurs sections de largeur égale, dont le nombre est paramétrable. Les différentes sections représentent des intervalles temporels linéaires, successifs et de durée égale, découpant l’étendue temporelle du diagramme. Les événements sont ensuite représentés en fonction de l’appartenance de leur moment d’apparition à ces intervalles temporels linéaires.

L’appartenance d’un événement à un intervalle temporel peut être traitée comme une variable qualitative ordonnée rattachée à l’événement. Les différents intervalles temporels linéaires se succédant chronologiquement, il existe une relation d’ordre entre eux. Nous utilisons alors la variable visuelle de la valeur pour représenter l’appartenance d’un événement à un intervalle temporel linéaire. Dans notre cas, cette variable visuelle se traduit par la valeur de saturation d’une couleur définie.

La valeur affectée à un intervalle linéaire est choisie entre deux valeurs extrêmes, affectées

aux bornes du diagramme temporel mixte (Figure 5.29). La position d'un intervalle temporel linéaire par rapport aux bornes du diagramme définit la valeur qui lui est affectée. Celle-ci est définie automatiquement en fonction des valeurs extrêmes affectées aux bornes, du nombre d'intervalles linéaires utilisés pour classer les événements, et de leur ordre d'apparition. L'entièreté du diagramme est couverte par les intervalles temporels linéaires créés. Il n'existe pas de zone de classification temporelle linéaire *par défaut*, regroupant les moments n'étant pas couverts par un des intervalles linéaires créés.

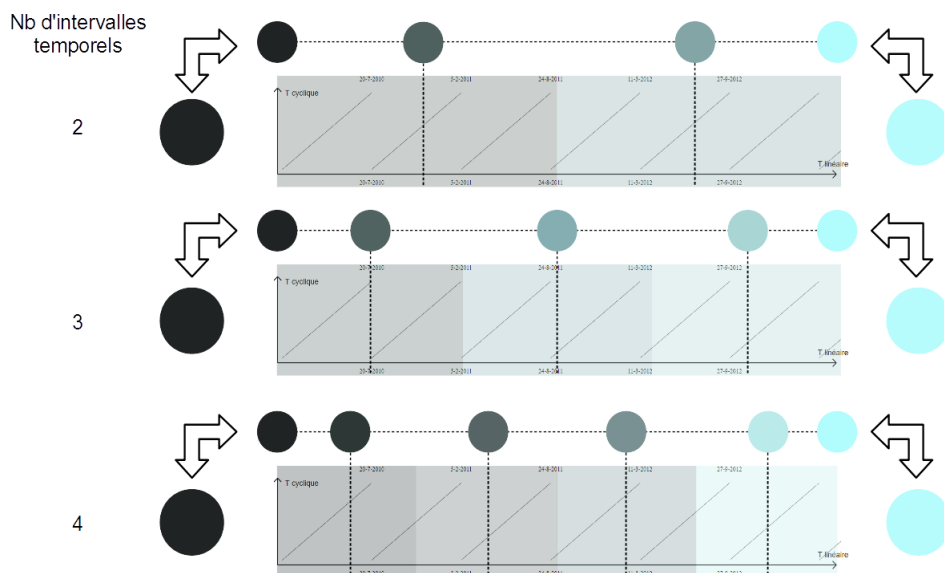


FIGURE 5.29 – Division de l'étendue temporelle étudiée en intervalles temporels linéaires

5.5.3.2 Représentation des événements dans la carte et dans le diagramme temporel mixte

La valeur associée à un intervalle temporel linéaire est affectée à l'ensemble des événements dont le moment d'apparition est inclus dans cet intervalle. Cette valeur est alors affectée aux entités graphiques correspondantes dans la carte et dans le diagramme temporel mixte, selon un procédé analogue à celui utilisé pour représenter les événements selon leur lieu d'apparition :

- Dans le cas d'une entité graphique représentant un seul événement, l'entité graphique prend la valeur affectée à l'événement.
- Dans le cas d'une entité graphique représentant un agrégat d'événements apparaissant durant le même intervalle temporel, l'entité graphique prend la valeur affectée à l'intervalle durant lequel les événements apparaissent.
- Dans le cas d'une entité graphique représentant un agrégat d'événements apparaissant durant différents intervalles temporels, l'entité graphique est représentée sous la forme d'un *Pie Chart*.

Dans ce troisième cas de figure (Figure 5.30) :

- Le nombre de secteurs du *Pie Chart* définit le nombre d'intervalles temporels durant lesquels apparaissent les événements agrégés.
- Chaque secteur du *Pie Chart* correspond à un intervalle temporel, et se voit affecter la valeur correspondante.
- L'ordre des différents secteurs dans le *Pie Chart* correspond à l'ordre d'apparition chronologique des différents intervalles temporels.

- L'angle au centre de chaque secteur est fonction de la proportion d'événements agrégés apparaissant durant l'intervalle de temps correspondant, par rapport au nombre total d'événements agrégés.

La position et la taille des agrégats dans la carte et le diagramme temporel sont calculées de la même manière que précédemment.

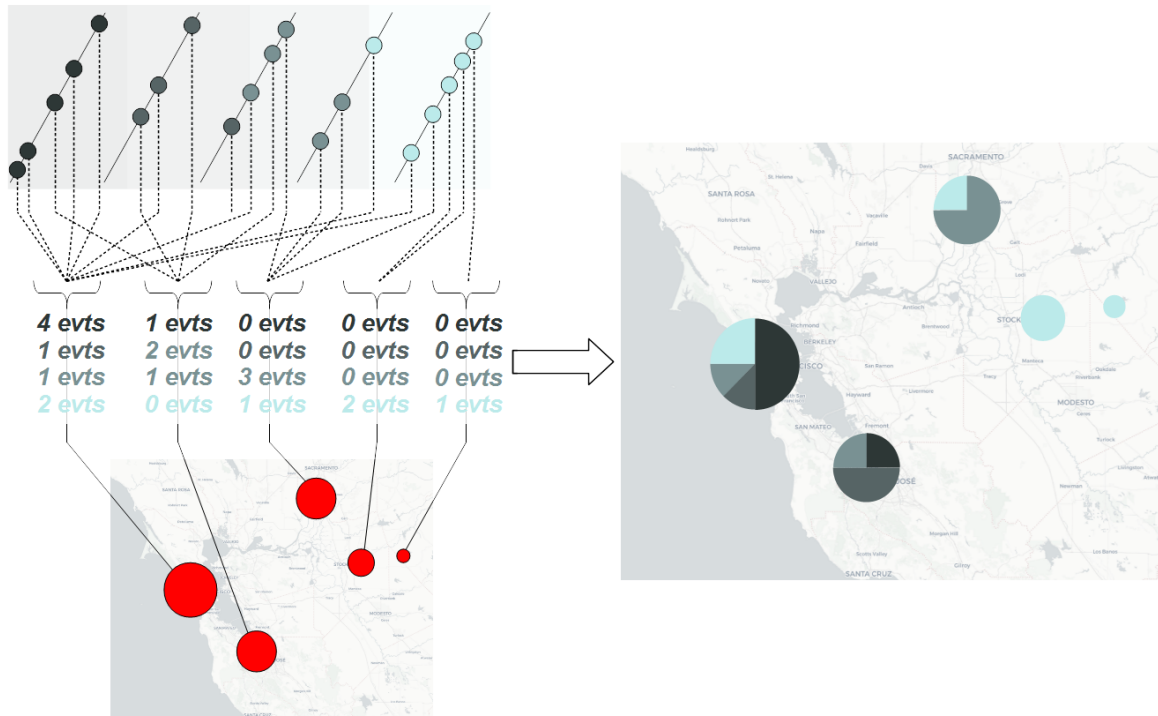


FIGURE 5.30 – Représentation d'agrégats d'événements selon leur moment d'apparition linéaire

Cette représentation en *Pie Chart* des agrégats d'événements est appliquée de la même manière dans la carte et dans le diagramme temporel mixte. Dans le diagramme, elle permet la représentation d'agrégats dont les événements sont situés de part et d'autre de la limite entre deux intervalles temporels.

5.5.3.3 Représentation des événements dans le diagramme temporel circulaire

Afin de représenter les événements selon leur appartenance aux différents intervalles temporels linéaires, le diagramme temporel circulaire adopte la même structure que pour la représentation des événements selon leur lieu d'apparition.

Chaque section angulaire du diagramme circulaire est divisée en sous-sections selon l'axe radial (Figure 5.31) :

- Pour chaque section angulaire, le nombre de sous-sections radiales correspond au nombre d'intervalles temporels linéaires comprenant des événements apparaissant dans l'intervalle temporel cyclique représenté.
- L'ordre de représentation des sous-sections radiales, du centre vers l'extérieur du diagramme, correspond à l'ordre d'apparition chronologique des différents intervalles temporels linéaires.
- La largeur radiale de chaque sous-section, est fonction de la proportion d'événements apparaissant durant l'intervalle temporel linéaire correspondant, par rapport au nombre total

d'événements apparaissant dans l'intervalle temporel cyclique représenté par la section angulaire du diagramme.

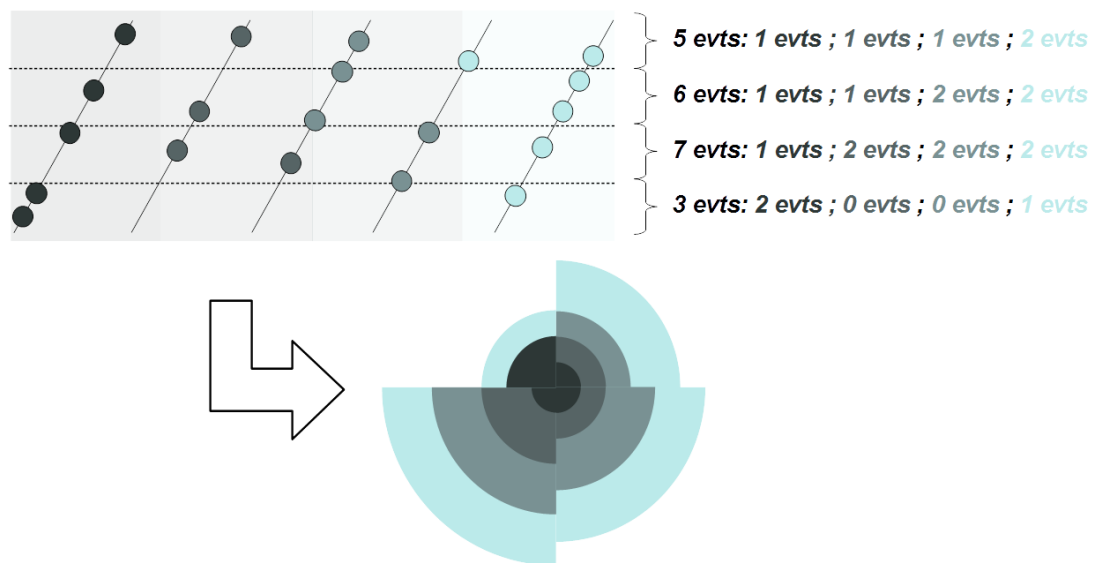


FIGURE 5.31 – Représentation de l'effectif des différentes classes temporelles d'événements, selon leur intervalle d'apparition relatif à une échelle cyclique

5.5.4 Représentation des événements selon leur moment d'apparition relatif à une échelle cyclique

5.5.4.1 Principe : une représentation orientée du diagramme temporel circulaire vers la carte et le diagramme temporel mixte

Les événements sont représentés en fonction de leur moment d'apparition relatif à l'échelle cyclique utilisée dans les diagrammes temporels. Celui-ci est symbolisé par la position des entités graphiques correspondantes, au sein d'une section périodique de la courbe du diagramme mixte. Il est également symbolisé par la section du diagramme temporel circulaire pour laquelle l'événement est pris en compte. Les événements sont représentés en fonction de l'appartenance de leur moment d'apparition à des intervalles temporels, de durée égale, relatifs à l'échelle cyclique représentée dans les diagrammes temporels.

Pour définir ces intervalles temporels, nous utilisons le découpage temporel représenté dans le diagramme temporel circulaire. Les événements sont ainsi classés selon la section angulaire du diagramme temporel circulaire pour laquelle ils sont comptabilisés. La classification temporelle des événements peut alors être modifiée par :

- une modification de la durée de l'échelle cyclique représentée ;
- une modification du déphasage du diagramme temporel mixte ;
- une modification du nombre de sections du diagramme temporel circulaire.

Les intervalles temporels utilisés étant relatifs à une échelle temporelle cyclique, il n'est pas possible d'établir une relation d'ordre entre ces intervalles. Pour représenter les événements selon leur appartenance à un intervalle temporel cyclique, nous utilisons la variable visuelle de la couleur. Une couleur est affectée à chaque intervalle temporel, et donc à chaque section du diagramme temporel circulaire, ainsi qu'à l'ensemble des événements apparaissant durant cet

intervalle.

Afin de définir la couleur affectée à chaque intervalle temporel, nous effectuons une analogie entre le diagramme temporel circulaire et un cercle chromatique (Figure 5.32).

Notre approche utilise deux, quatre ou six couleurs, qui sont affectées à des bornes temporelles relatives à une échelle temporelle cyclique représentée, de durée T . Ces bornes sont disposées à une égale distance temporelle l'une de l'autre (Figure 5.32) :

- dans le cas d'une représentation à deux couleurs, une couleur représente le début et la fin de l'échelle temporelle cyclique, et une couleur représente le milieu de l'échelle temporelle cyclique ;
- dans le cas d'une représentation à quatre couleurs, une couleur représente le début et la fin de l'échelle temporelle cyclique, et trois autres couleurs représentent les moments temporels cycliques situés aux $1/4$, $1/2$, $3/4$ de l'échelle cyclique ;
- enfin, dans le cas de six couleurs, une couleur représente le début et la fin de l'échelle temporelle cyclique, et les cinq autres couleurs représentent les moments temporels cycliques situés aux $1/6$, $1/3$, $1/2$, $2/3$, et $5/6$ de l'échelle cyclique.

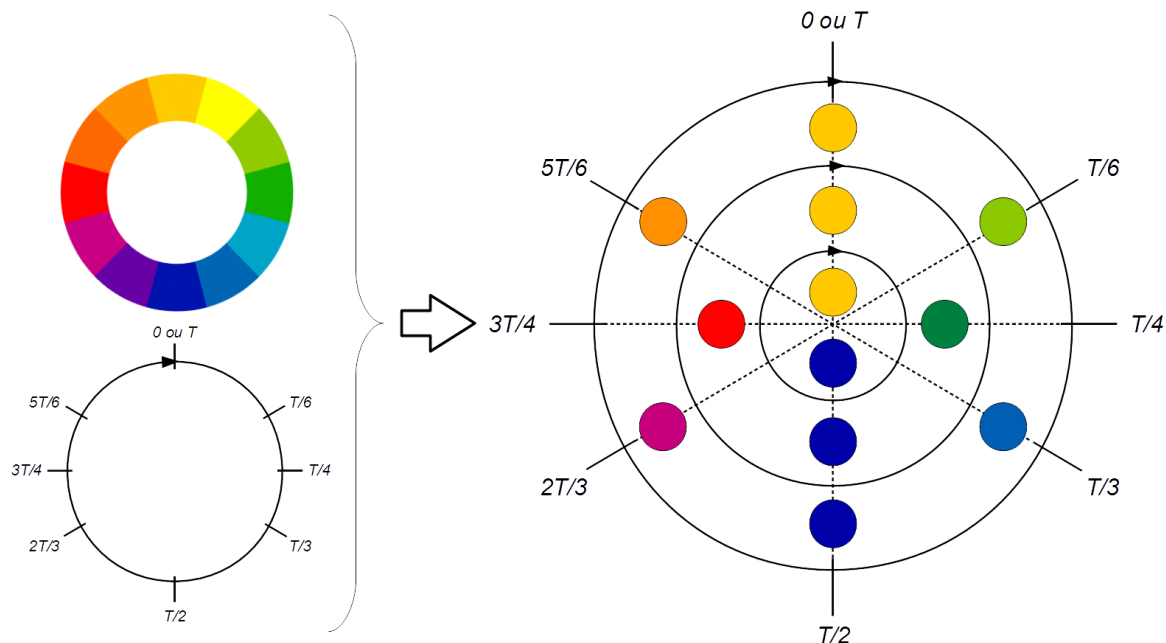


FIGURE 5.32 – Représentation du moment temporel cyclique selon deux, quatre ou six couleurs

La couleur affectée à chaque intervalle temporel, et donc à chaque section du diagramme circulaire, est calculée à partir :

- du nombre de sections du diagramme ;
- aux couleurs de base utilisées ;
- de la position de la section, par rapport aux bornes temporelles auxquelles sont affectées les couleurs de base.

L'entièreté du diagramme étant couverte par les intervalles temporels cycliques créés, il n'existe pas d'intervalle temporel *par défaut*.

5.5.4.2 Représentation des événements dans le diagramme temporel circulaire

Le diagramme temporel circulaire prend la forme d'un cercle chromatique à deux, quatre, ou six couleurs de base. Le nombre de couleurs utilisées pour représenter les événements correspond aux nombres de sections utilisées pour diviser l'échelle cyclique dans le diagramme (Figure 5.33). Chaque section du diagramme est affectée à une seule couleur. Pour chaque section du diagramme, le rayon de la section représente le nombre d'événements apparaissant durant l'intervalle temporel cyclique correspondant. Le diagramme représente donc l'effectif de chaque classe utilisée pour représenter les événements.

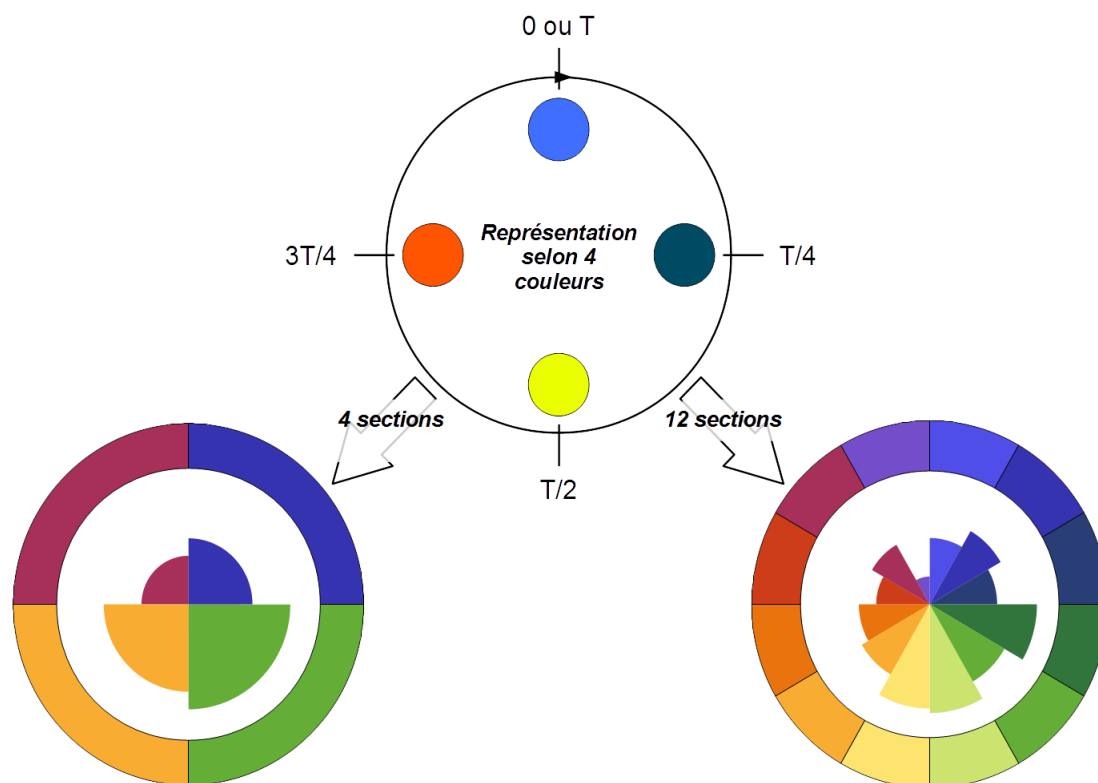


FIGURE 5.33 – Affection des couleurs aux sections du diagramme temporel circulaire, dans le cas de quatre couleurs de base

5.5.4.3 Incidence sur la représentation des événements dans la carte et dans le diagramme temporel mixte

La couleur associée à un intervalle temporel cyclique est affectée à l'ensemble des événements dont le moment d'apparition est inclus dans cet intervalle. Cette valeur est alors affectée aux entités graphiques correspondantes dans la carte et dans le diagramme temporel mixte, selon un procédé analogue à celui utilisé pour représenter les événements selon leur lieu d'apparition ou selon leur moment d'apparition dans le temps linéaire :

- Dans le cas d'une entité graphique représentant un seul événement, l'entité graphique prend la couleur affectée à l'événement.
- Dans le cas d'une entité graphique représentant un agrégat d'événements apparaissant durant le même intervalle temporel, l'entité graphique prend la valeur affectée à l'intervalle durant lequel les événements apparaissent.
- Dans le cas d'une entité graphique représentant un agrégat d'événements apparaissant durant différents intervalles temporels, l'entité graphique est représentée sous la forme d'un

Pie Chart.

Dans ce troisième cas de figure (Figure 5.34) :

- Le nombre de secteurs du *Pie Chart* définit le nombre d’intervalles temporels durant lesquels apparaissent les événements agrégés.
- Chaque secteur du *Pie Chart* correspond à un intervalle temporel, et se voit affecter la couleur correspondante.
- L’ordre des différents secteurs dans le *Pie Chart* correspond à l’ordre de représentation des sections correspondantes dans le diagramme temporel circulaire, dans le sens des aiguilles d’une montre depuis le sommet du diagramme.
- L’angle au centre de chaque secteur est fonction de la proportion d’événements agrégés apparaissant durant l’intervalle de temps correspondant, par rapport au nombre total d’événements agrégés.

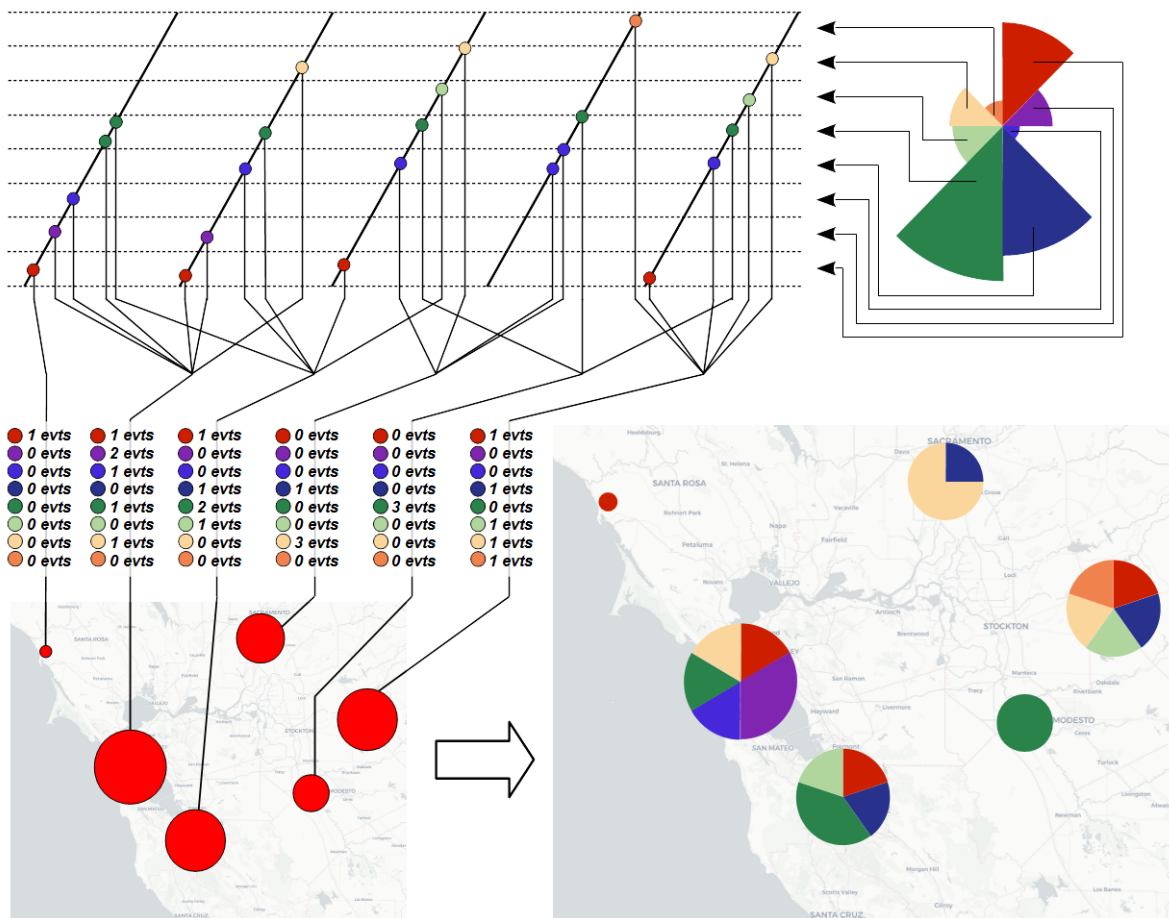


FIGURE 5.34 – Représentation des événements selon huit intervalles temporels cycliques

La position et la taille des agrégats dans la carte et le diagramme temporel mixte sont calculées de la même manière que précédemment.

5.6 Implémentation de *GrAPHiST*

5.6.1 Implémentation de la base de données

Les données spatio-temporelles sont stockées dans une base de données *PostgreSQL*, utilisant l’extension *PostGIS*.

La base de données est structurée de manière à suivre la modélisation de la donnée spatio-temporelle présentée au début de ce chapitre. Une modélisation de la base de données est ainsi présentée dans la Figure 5.35.

Les liens entre les entités phénomènes, événements, sous-parties spatio-temporelles, et leur composante thématique respective, s'effectue grâce aux identifiants de chaque objet stocké dans la base. Par exemple, une sous-partie spatio-temporelle est reliée à l'événement dont elle est le composant par l'identifiant de l'événement. Cet événement est relié au phénomène dont il est une instance par l'identifiant du phénomène.

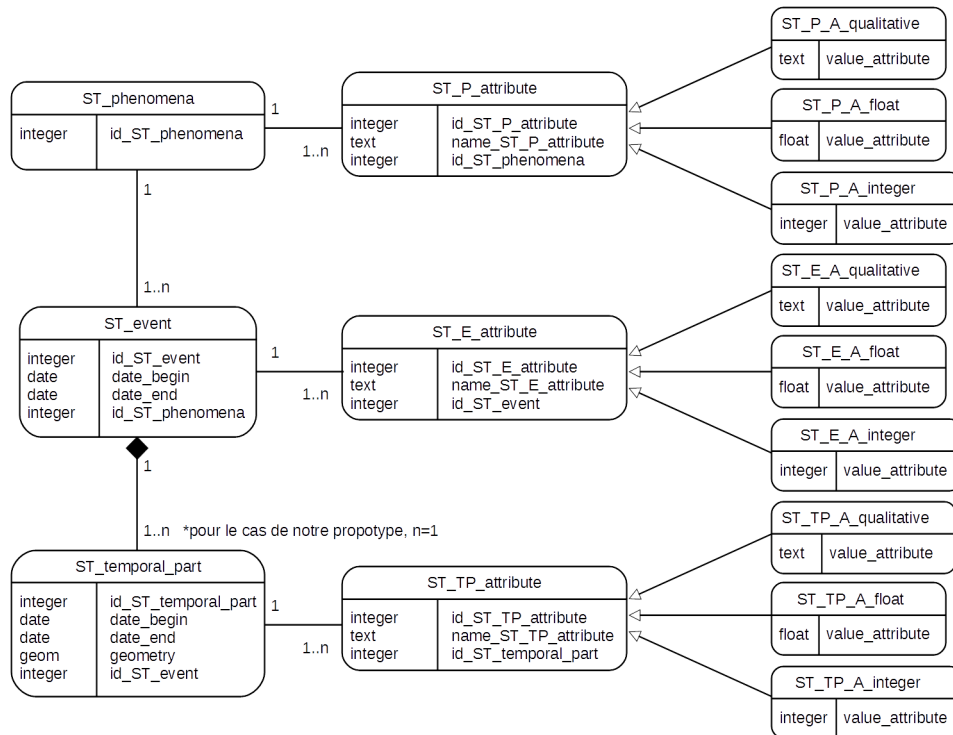


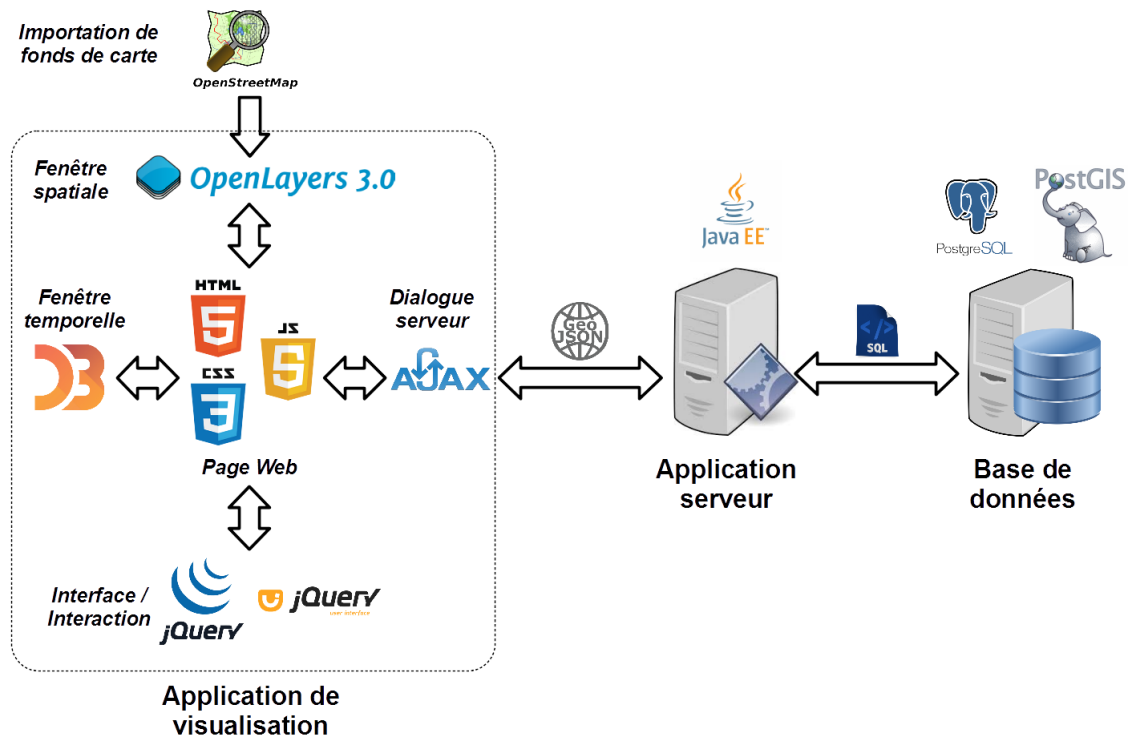
FIGURE 5.35 – Structure de la base de données

5.6.2 Implémentation de l'interface de géovisualisation

GrAPHiST a été implémenté sous la forme d'une application *Web* (Figure 5.36) :

- L'interface de géovisualisation se présente sous la forme d'une page *HTML*.
- Les diagrammes et les entités graphiques de la fenêtre temporelle sont créés et manipulés en utilisant les fonctions de la bibliothèque *JavaScript D3*.
- La carte et les entités cartographiques de la fenêtre spatiale sont créées et manipulées en utilisant la bibliothèque *Openlayers 3*. Les fonds de carte sont fournis par *OpenStreetMap*.
- Le reste des fonctions d'interaction, ainsi que la fenêtre de contrôle, utilisent les bibliothèques *JQuery* et *JQueryUI*.
- La majorité des processus de l'application sont effectués côté client au moyen de fonctions *JavaScript*. Les processus implémentés côté serveur ont pour but d'interroger la base de données afin d'importer les événements en fonction des critères choisis par l'utilisateur (type de phénomène, étendues temporelle et spatiale étudiées, filtre attributaire).

- Le dialogue entre la page *Web* et l'application côté serveur s'effectue au moyen de requête *Ajax*. Les données envoyées et reçues par la requête sont au format *JSON*, ou *GeoJSON* s'il s'agit d'un import d'événements.
- Les processus côté serveur sont implémentés en *Java EE*. Le dialogue avec la base de données s'effectue au moyen de requêtes *SQL*.

FIGURE 5.36 – Fonctionnement de *GrAPHiST*

5.7 Conclusion

Dans ce chapitre, nous présentons la structure générale de *GrAPHiST*, ainsi que les différents procédés de modélisation, de représentation graphique et d'interaction que nous mettons en place dans le but de permettre une analyse exploratoire des dynamiques spatio-temporelles des phénomènes.

L'approche que nous proposons s'appuie sur :

- une modélisation généraliste des phénomènes en séries d'événements spatio-temporels.
- l'utilisation d'un environnement de géovisualisation à multifenêtrage.

Cet environnement doit permettre une analyse des dynamiques spatio-temporelles des phénomènes, à travers l'identification visuelle de structures dans la distribution spatio-temporelle d'une série événementielle. Nous nous intéressons ainsi à quatre types de patterns :

- des clusters spatio-temporels relatifs au temps linéaire ;
- des clusters spatio-temporels relatifs au temps cyclique ;
- des propagations spatiales d'événements au cours du temps ;
- des récurrences cycliques spatialisées.

L'analyse proposée par *GrAPHiST* est basée sur :

- une représentation graphique des composantes temporelle et spatiale des séries événementielles, permettant une identification de structure dans les distributions temporelle et spatiale des événements ;
- l'utilisation de règles de symbologie et de procédé d'agrégation permettant d'établir un lien entre les contenus des fenêtres temporelle et spatiale de l'interface de *GrAPHiST*.

Les fenêtres de l'interface utilisent des représentations graphiques, inspirées d'approches présentées dans notre état de l'art, permettant d'identifier visuellement des patterns :

- La fenêtre temporelle s'appuie sur la représentation de la distribution temporelle des événements au travers de deux diagrammes temporels :
 - Le premier est un diagramme mixte, pouvant se présenter sous la forme d'une courbe sinusoïdale inspirée des approches *Time Coil* [Edsall and Peuquet, 1997] et *Time Wave* [Li, 2010], ou d'une courbe en dents de scie inspirée des diagrammes en *tiles map* [Mintz et al., 1997][Aigner et al., 2011]. Ce diagramme doit permettre une visualisation de la composante temporelle des événements selon le temps linéaire et selon le temps cyclique, et de permettre la visualisation d'une récurrence cyclique des événements, de période égale à la durée de l'échelle cyclique représentée.
 - Le second est un diagramme circulaire inspiré du diagramme *Rose Charts* de *Florence Nightingale* [Nightingale, 1858] et des diagrammes en secteurs (*Sector Graphs* de Harris [Harris, 2000]). Ce diagramme doit permettre une visualisation de la distribution temporelle des événements selon une échelle temporelle cyclique.
- La fenêtre spatiale s'appuie sur la représentation de la composante spatiale de l'ensemble des événements apparus durant l'étendue temporelle étudiée, au sein d'une carte d'inventaire, à la manière de *PerSE* [Swedberg and Peuquet, 2017]. Cette carte doit permettre une visualisation de la distribution spatiale des événements.

Afin de faire le lien entre le contenu des fenêtres temporelle et spatiale de l'interface de *GrAPHiST*, nous proposons des règles de symbologie, permettant de représenter les événements selon un critère dimensionnel supplémentaire au sein des fenêtres de l'environnement. Les événements peuvent alors être représentés :

- dans la fenêtre temporelle selon leur lieu d'apparition, à l'instar des représentations proposées par *PerSE* [Swedberg and Peuquet, 2017] et *Time Wave* [Li, 2010] ;
Nous désignons cette représentation comme ayant une **entrée spatiale**, et étant **orientée de la carte vers les diagrammes temporels**. Pour cela, nous classons les événements selon des zones spatiales et nous les représentons en fonction à l'aide de la variable visuelle de la couleur.
- dans la fenêtre spatiale selon leur moment d'apparition relatif à une échelle linéaire, à l'instar de *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011] ;
Nous désignons cette représentation comme ayant une **entrée temporelle linéaire**, et étant **orientée du diagramme temporel mixte vers la carte et le diagramme temporel circulaire**. Pour cela, nous classons les événements selon des intervalles temporels linéaires et nous les représentons en fonction à l'aide de la variable visuelle de la valeur.
- dans la fenêtre spatiale selon leur moment d'apparition relatif à une échelle cyclique, à l'instar de *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011] ;
Nous désignons cette représentation comme ayant une **entrée temporelle cyclique**, et étant **orientée du diagramme temporel circulaire vers la carte et le diagramme temporel mixte**. Pour cela, nous classons les événements selon des intervalles temporels

cycliques et nous les représentons en fonction à l'aide de la variable visuelle de la couleur. Ces deux dernières propositions s'inscrivent dans la continuité des travaux de l'équipe *STeamer* du *Laboratoire d'informatique de Grenoble* sur l'utilisation des variables de la couleur et de la valeur pour représenter le temps dans la carte [Davoine et al., 2012b, Saint-Marc et al., 2014, Saint-Marc, 2017, Saint-Marc et al., 2017, Saint-Marc et al., 2018].

Conjointement à l'utilisation de règles de symbologie, nous proposons un procédé d'agrégation des événements spatialement ou temporellement proches. Nous proposons de représenter ces agrégats sous la forme de *Pie Chart*, afin de représenter la proportion des différents groupes d'événements similaires au sein d'un agrégat. Combinée aux outils de symbologie, cette représentation permet d'identifier des structures dans la distribution spatio-temporelle des événements, et permet une identification de patterns spatio-temporels.

Si ces procédés de représentation graphique et d'interaction rendent possible l'identification visuelle d'un pattern, l'analyse exploratoire repose sur une suite d'interactions entre l'utilisateur et l'interface, au cours de laquelle l'utilisateur modifie les critères de représentation des événements, jusqu'à mettre en valeur un pattern au travers des fenêtres de l'environnement de géovisualisation.

Le chapitre suivant (refchapter :6) présente ainsi différents scénarios d'utilisation de *GrAPHiST*, permettant à l'utilisateur de rechercher et d'identifier des patterns spatio-temporels, au cours d'une analyse exploratoire de séries événementielles.

Chapitre 6

Scénarios d'analyse exploratoire pour l'identification de patterns au moyen de *GrAPHiST*

Sommaire

6.1	Analyse de patterns unidimensionnels	217
6.1.1	Identification de concentrations d'événements dans le temps et l'espace .	218
6.1.2	Identification d'une récurrence cyclique d'événements	222
6.1.3	Proposition de pistes d'exploration à l'utilisateur par suggestion de valeurs à affecter à l'échelle temporelle cyclique	228
6.2	Structure des scénarios d'analyse	232
6.3	Analyse de patterns spatio-temporels, au travers de zones spécifiques	234
6.3.1	Procédé général	234
6.3.2	Analyse de clusters relatifs au temps linéaire, selon une entrée spatiale .	236
6.3.3	Analyse de clusters relatifs à une échelle temporelle cyclique, selon une entrée spatiale	238
6.3.4	Analyse d'une propagation spatiale d'événements au cours du temps, selon une entrée spatiale	240
6.3.5	Analyse de récurrences cycliques spatialisées, selon une entrée spatiale .	242
6.4	Analyse de patterns spatio-temporels, sans hypothèse <i>a priori</i> sur la distribution spatiale des événements	246
6.4.1	Procédé général	246
6.4.2	Analyse de clusters relatifs au temps linéaire, selon une entrée temporelle	248
6.4.3	Analyse de clusters relatifs à une échelle temporelle cyclique, selon une entrée temporelle	250
6.4.4	Analyse d'une propagation spatiale d'événements au cours du temps, selon une entrée temporelle	252
6.4.5	Analyse de récurrences cycliques spatialisées, selon une entrée temporelle	253
6.5	Synthèse	259

L'analyse que nous cherchons à fournir est exploratoire. *GrAPHiST* est un outil de recherche, et non un outil de retranscription graphique d'un résultat obtenu en amont de la visualisation [Card et al., 1999, Spence, 2001, Andrienko and Andrienko, 2006].

GrAPHiST doit ainsi permettre d'explorer la distribution spatio-temporelle des événements, d'émettre des hypothèses sur sa structure, d'y identifier des patterns, au moyen des outils de représentation graphique et des possibilités d'interaction [Andrienko and Andrienko, 2006]. Ben Shneiderman [Shneiderman, 2003, Andrienko and Andrienko, 2006] définit ainsi trois étapes dans le déroulement général d'une analyse exploratoire, pouvant être répétée de manière itérative :

- la première, *Overview First*, consistant en une observation globale de l'ensemble d'un jeu de données ;
- la seconde étape, *zoom and filter*, consistant en une focalisation de l'analyse sur des objets d'intérêt ;
- la dernière étape, *details on demand*, consistant en une exploration approfondie des caractéristiques des objets identifiés lors de la seconde étape.

Dans chacune de ces étapes, l'utilisateur occupe une place centrale :

- En fonction des données étudiées et de ses objectifs de recherche, l'utilisateur choisit les critères définissant les objets d'intérêts à identifier [Andrienko and Andrienko, 2006].
- Ces objets d'intérêt sont identifiés par l'utilisateur, au travers de l'interprétation de ce qu'il voit dans les représentations graphiques [Tukey, 1977, Andrienko and Andrienko, 2006].
- À partir de cette interprétation, l'utilisateur décide quelles sont les actions à effectuer pour approfondir l'analyse sur objets d'intérêts, dans le but de mettre en lumière d'autres informations [Tukey, 1977, Andrienko and Andrienko, 2006].
- L'utilisateur interprète le retour visuel de ces actions, et définit dans quelle direction approfondir l'analyse.

L'analyse exploratoire de patterns dans la distribution spatio-temporelle des événements repose ainsi sur une suite d'interactions entre l'utilisateur et l'interface. Au cours de celles-ci :

- L'utilisateur interprète le contenu des trois espaces de représentation des données (carte, diagramme mixte, diagramme circulaire).
- Il modifie, en fonction de ce qu'il observe dans ces espaces, les critères de représentation des événements. Il peut ainsi choisir de modifier les étendues spatiales et temporelles de la zone d'étude, modifier l'échelle temporelle cyclique, modifier les critères de la représentation graphique, classer les événements selon un critère dimensionnel.

L'utilisateur répète alors ce procédé de manière itérative jusqu'à mettre en valeur un pattern dans les fenêtres de l'interface.

Notre proposition d'analyse exploratoire ne repose alors pas seulement sur des procédés de représentation graphique et interactive de la donnée spatio-temporelle, mais également sur leur utilisation. Si l'environnement de géovisualisation proposé dans le chapitre 5 est développé dans le but d'identifier des patterns, se pose alors la question suivante : comment une analyse exploratoire peut amener à la découverte de patterns au sein de cet environnement.

Si la formalisation d'une méthode universelle d'analyse exploratoire reste difficile [Andrienko and Andrienko, 2006], la limitation de notre objectif de recherche à l'identification de patterns définis nous permet de proposer, pour chaque pattern, un scénario d'analyse exploratoire.

Chaque cas d'analyse nécessite que l'utilisateur explore librement un jeu de données, dans le but de devenir *familier* avec celui-ci, avant d'approfondir son analyse [Andrienko and Andrienko, 2006]. L'analyse d'un type de pattern particulier, au moyen de *GrAPHiST*, suit néanmoins un patron général définissant les grandes étapes de la recherche, chacune faisant intervenir des outils interactifs et visuels différents.

Ce chapitre est ainsi consacré aux différents scénarios d'analyse exploratoire pouvant être effectués au moyen de *GrAPHiST*. Nous présentons comment l'utilisateur peut, au travers d'une suite d'interprétations visuelles et d'interactions, construire son procédé d'analyse et identifier chacune des structures recherchées. Chaque scénario décrit ainsi les grandes étapes permettant de passer d'une représentation de données brutes, à une représentation graphique mettant en valeur un pattern spatio-temporel particulier (Figure 6.1).

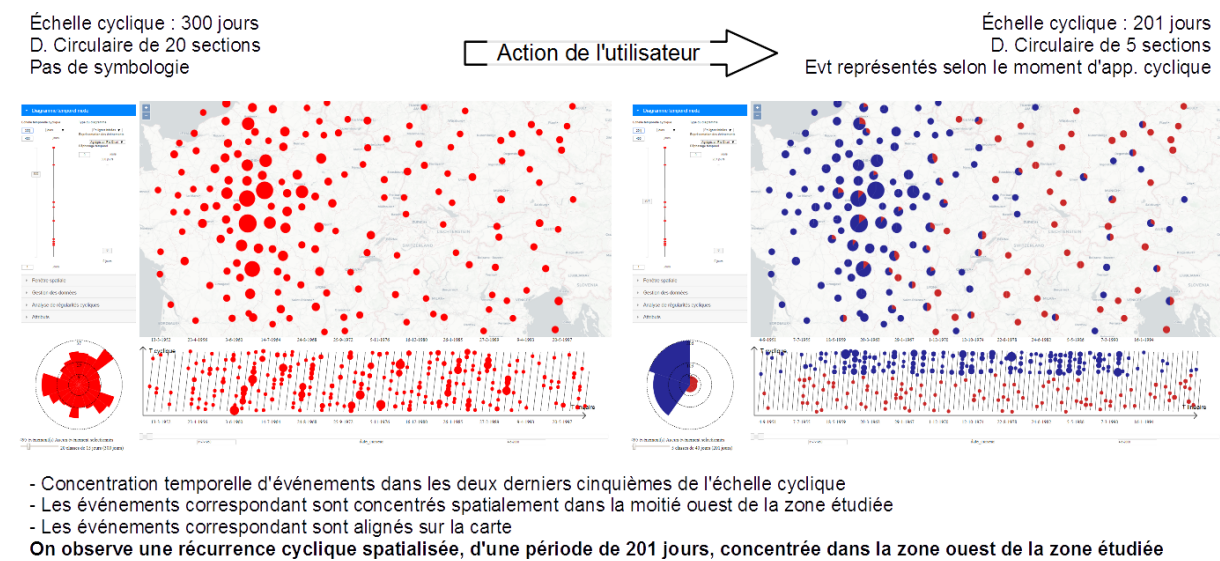


FIGURE 6.1 – Résultats des interactions de l'utilisateur avec l'interface de *GrAPHiST*, dans le cadre d'une recherche d'un pattern spatio-temporel

Nos scénarios d'analyse se divisent en deux groupes, correspondant aux deux approches d'analyse de la distribution spatio-temporelle des événements permises par *GrAPHiST* :

- Le premier groupe suit une approche partant d'une entrée spatiale, à la manière de *Time Wave* [Li, 2010] et *PerSE* [Swedberg and Peuquet, 2017], et se base sur une classification des événements selon leur composante spatiale.
- Le second groupe suit une approche partant d'une entrée temporelle, à la manière de *Growth Ring Maps* [Bak et al., 2009, Andrienko et al., 2011], et se base sur une classification des événements selon leur composante temporelle, appréhendée de manière linéaire ou cyclique.

Cependant, comme nous le précisons dans le chapitre 4, l'analyse de patterns spatio-temporels passe par une identification de structures dans les distributions spatiale et temporelle des événements. Aussi, en préambule aux différents scénarios d'analyse de patterns spatio-temporels, nous présentons comment l'utilisation des représentations graphiques et interactives de *GrAPHiST* permet une analyse de patterns unidimensionnels, temporels ou spatiaux.

6.1 Analyse de patterns unidimensionnels

Comme nous l'évoquons au second chapitre (2), parmi les quatre patterns auxquels nous nous intéressons, trois d'entre eux peuvent se traduire par la combinaison de deux patterns

unidimensionnels, identifiables dans les fenêtres temporelles et spatiales :

- Un cluster spatio-temporel, relatif au temps linéaire ou cyclique, se traduit par un cluster spatial, et par un cluster temporel.
- Une récurrence cyclique spatialisée se traduit par un cluster spatial, et par une récurrence cyclique des événements dans le temps.

L'identification d'une propagation spatiale des événements nécessite, quant à elle, de traiter les dimensions spatiale et temporelle de manière simultanée.

Nous présentons ici comment l'utilisation des représentations graphiques permet l'identification de patterns unidimensionnels, dans les fenêtres de *GrAPHiST*.

6.1.1 Identification de concentrations d'événements dans le temps et l'espace

6.1.1.1 Concentration temporelle selon le temps linéaire

Une telle structure est observable sur le diagramme temporel mixte.

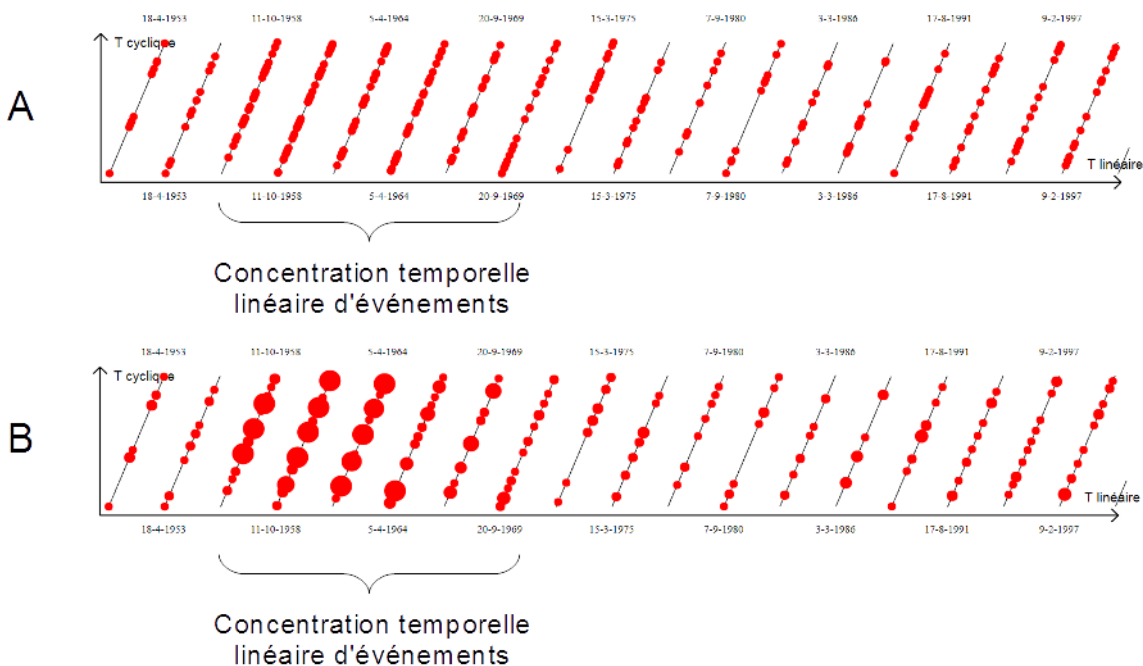


FIGURE 6.2 – Identification d'une concentration temporelle selon le temps linéaire

Dans le cas d'une représentation ponctuelle des événements, cette concentration d'événements se manifeste par une concentration d'entités graphiques le long de l'axe des abscisses du diagramme (Repère A de la Figure 6.2). Dans le cas d'une représentation agrégée, elle se manifeste également par la concentration d'entités graphiques, de taille importante, le long de l'axe des abscisses (Repère B de la Figure 6.2).

La représentation d'un grand nombre d'événements peut, dans cette situation, rendre plus difficile l'identification visuelle des écarts entre le nombre d'événements de chaque agrégat. L'utilisateur peut alors modifier la taille des entités graphiques en fonction du nombre des événements agrégés, afin de mettre en valeur les agrégats regroupant le plus d'événements (Figure 6.3).

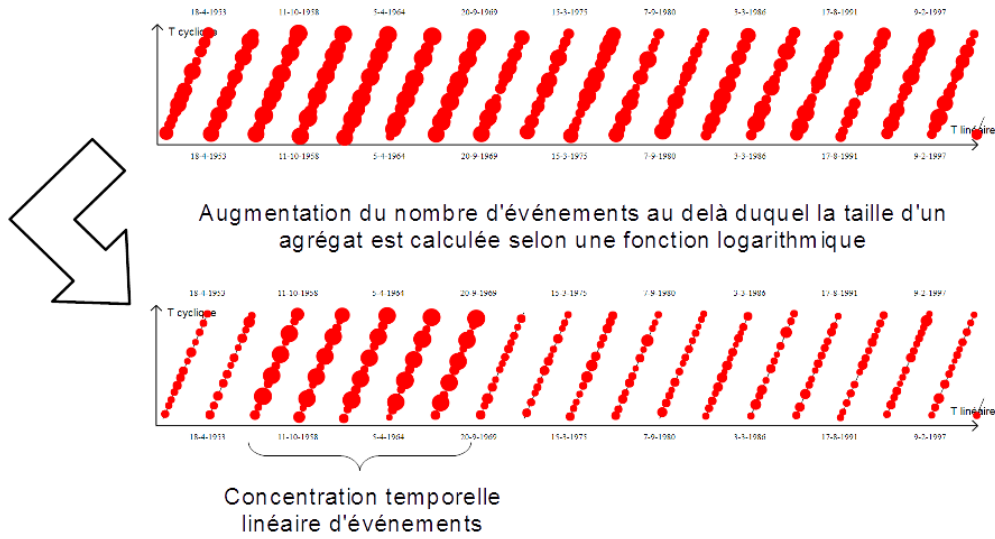


FIGURE 6.3 – Modification de la représentation des agrégats dans le diagramme mixte pour l'identification de concentrations temporelles selon le temps linéaire

6.1.1.2 Concentration temporelle selon une échelle cyclique

L'observation d'une telle structure nécessite d'avoir fixé à l'avance l'échelle cyclique considérée. Cette structure se manifeste par une concentration d'entités graphiques (ou d'entités de taille importante, dans le cas d'une représentation agrégée) le long de l'axe des ordonnées du diagramme temporel mixte, comme le montre la Figure 6.4.

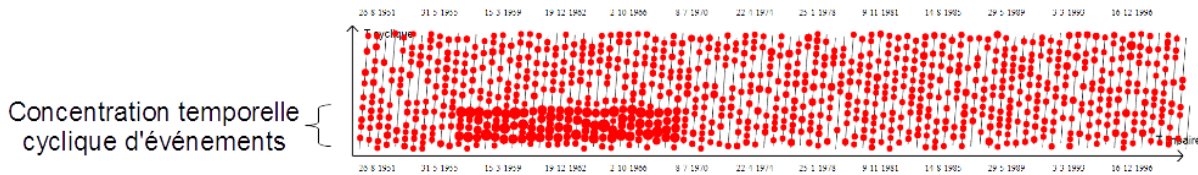


FIGURE 6.4 – Identification de concentrations temporelles selon le temps cyclique dans le diagramme mixte

Une concentration temporelle d'événements selon une échelle cyclique est cependant plus facilement observable dans le diagramme circulaire. Une concentration temporelle cyclique se manifeste sur ce diagramme par l'apparition d'une ou de plusieurs sections circulaires du diagramme présentant un rayon plus important que les autres sections (Figure 6.5).

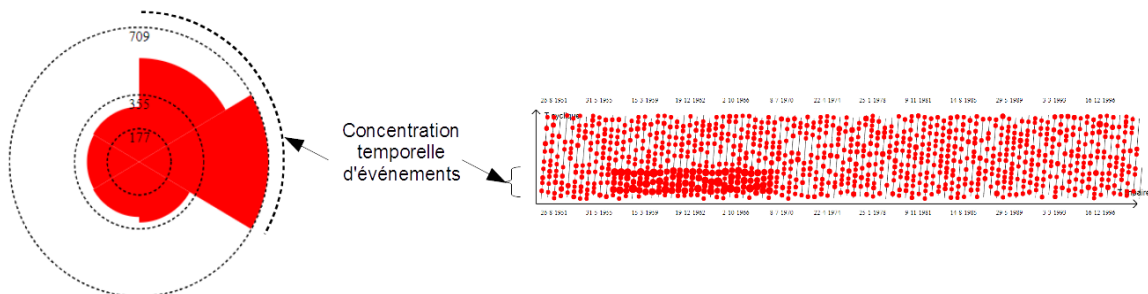


FIGURE 6.5 – Identification de concentrations temporelles selon le temps cyclique dans le diagramme circulaire

L'utilisateur peut modifier le nombre de sections du diagramme circulaire, afin d'affiner l'identification d'une concentration d'événements (Figure 6.6). L'utilisateur peut, par exemple, passer de la représentation d'une échelle d'un an divisée en douze mois, à la représentation d'une échelle d'un an divisée en cinquante-deux semaines, ce qui permet d'identifier des concentrations temporelles s'étendant sur moins d'un mois.

Augmenter le nombre de sections du diagramme circulaire permet ainsi de mieux définir les limites d'un cluster, et d'analyser si la concentration d'événements est homogène sur l'étendue temporelle du cluster. Dans la Figure 6.6, un diagramme à soixante sections permet d'observer une concentration d'événements sur un intervalle moins étendu que celui observé pour un diagramme à six sections. De plus, un diagramme à soixante sections permet d'observer que la concentration d'événements n'est pas homogène sur cet intervalle temporel, le plus grand nombre d'événements apparaissant au début et à la fin de l'intervalle.

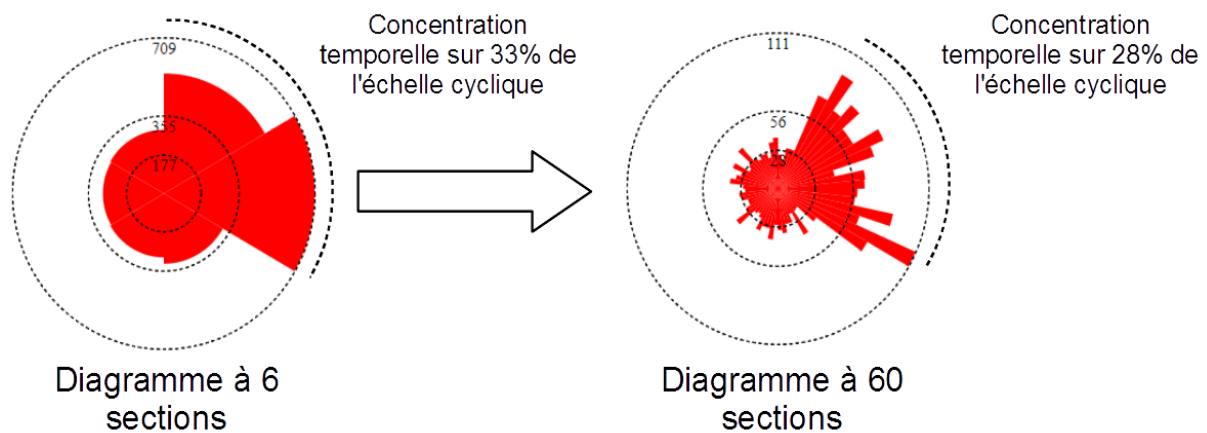


FIGURE 6.6 – Modification du nombre de sections dans le diagramme circulaire pour l'identification de concentrations temporelles

L'étendue d'une concentration temporelle peut également être à la limite entre deux intervalles temporels cycliques représentés par le diagramme circulaire. L'utilisateur peut alors modifier le déphasage du diagramme temporel mixte, afin que le nombre d'événements correspondant soit représenté par une seule section du diagramme circulaire.

6.1.1.3 Concentration spatiale

Une telle structure est observable sur la carte par une concentration d'entités graphiques (Figure 6.7) ou d'entités graphiques de grande taille dans le cas d'une représentation agrégée des événements.

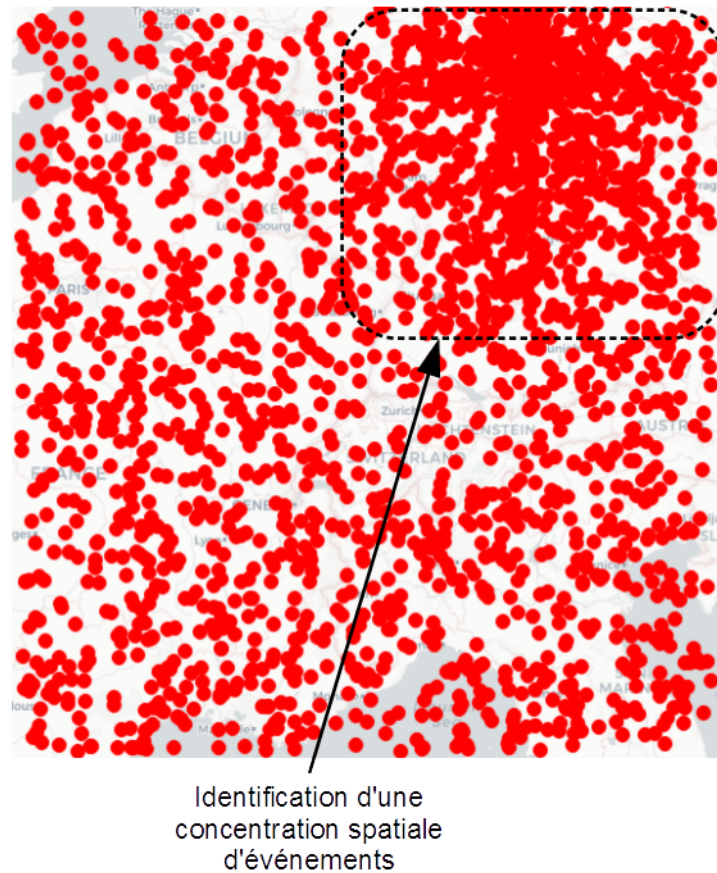


FIGURE 6.7 – Identification d'une concentration spatiale d'événements, représentés par des entités ponctuelles

Comme pour le cas d'une représentation agrégée des événements dans le diagramme temporel mixte, la représentation d'un grand nombre d'événements peut rendre plus difficile l'identification des concentrations spatiales d'événements dans la carte. L'utilisateur peut également modifier la taille des entités graphiques afin de mettre en valeur les agrégats regroupant le plus d'événements (Figure 6.8).

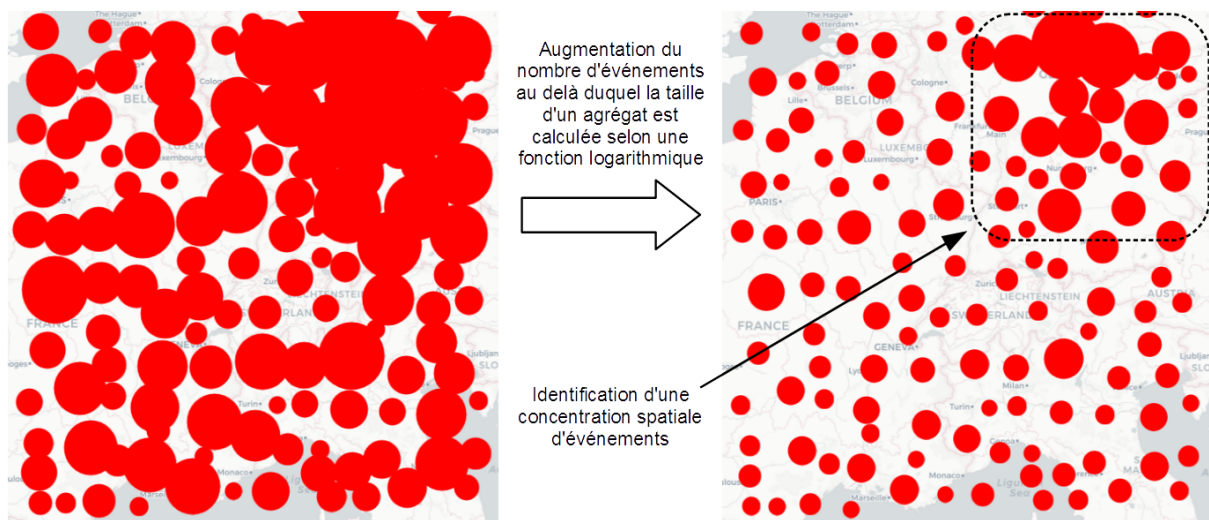


FIGURE 6.8 – Modification de la représentation des agrégats dans la carte pour l'identification de concentrations spatiales

6.1.2 Identification d'une récurrence cyclique d'événements

6.1.2.1 Visualisation d'une récurrence cyclique d'événements dans les diagrammes temporels

Comme nous l'avons évoqué au quatrième chapitre (4), l'utilisation d'un diagramme temporel mixte, combinant une échelle temporelle linéaire et une échelle temporelle cyclique, permet la visualisation d'une récurrence cyclique d'événements, si la période P du cycle est égale à la durée T de l'échelle cyclique représentée.

Cette récurrence cyclique est observable sur le diagramme temporel mixte lorsque $T = P$, par un alignement régulier, parallèle à l'axe des abscisses, des entités graphiques représentant les événements de la série.

Un phénomène présentant un caractère strictement cyclique sera ainsi observable par un alignement strict et régulier de toutes les entités représentant les occurrences du phénomène (Figure 6.9).

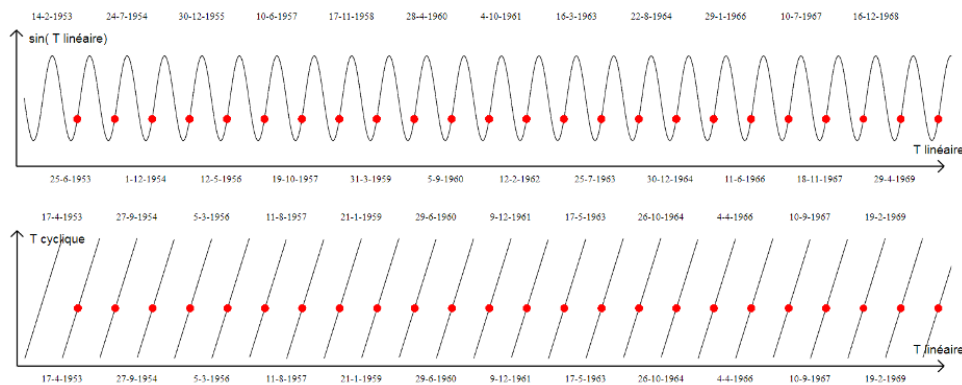


FIGURE 6.9 – Identification d'un cycle strict sur le diagramme temporel mixte

Un phénomène dont le caractère cyclique se manifeste par l'apparition régulière d'un intervalle temporel, où apparait un nombre variable d'événement, sera également observable sur le diagramme temporel mixte. Le caractère cyclique entraînant une surreprésentation d'entités graphiques dans la même section cyclique du diagramme mixte. La récurrence cyclique de l'intervalle d'apparition des événements sera observable par un alignement d'entités graphiques, réparties sur une section plus large de l'axe des ordonnées (Figure 6.10).

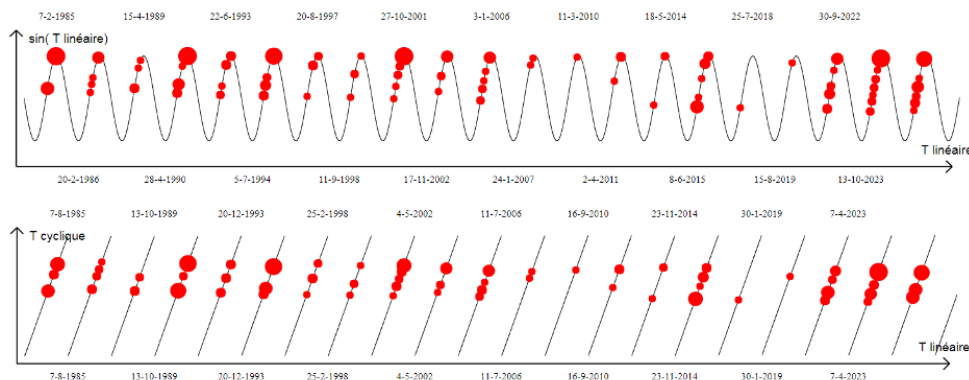


FIGURE 6.10 – Identification d'un cycle, apparaissant durant un intervalle temporel étendu, sur le diagramme temporel mixte

Cette surreprésentation des entités graphiques dans la même section cyclique du diagramme mixte implique généralement une mise en valeur de la section correspondante dans le diagramme circulaire (Figure 6.11).

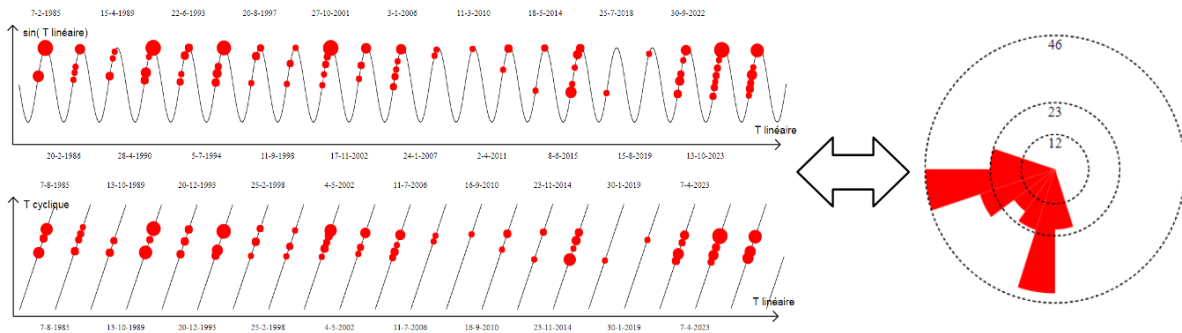


FIGURE 6.11 – Identification d'un cycle sur le diagramme temporel mixte et de la concentration temporelle cyclique correspondante sur le diagramme circulaire

L'identification de cette concentration temporelle dans le diagramme circulaire, étant une conséquence du caractère cyclique d'un phénomène, est un indice supplémentaire pour permettre l'identification d'un cycle par l'utilisateur (Repère *A* de la Figure 6.12).

Elle n'est cependant pas un critère suffisant pour conclure à l'existence d'un cycle. Les événements correspondants peuvent tous apparaître dans quelques-unes ou dans une seule instance de l'échelle temporelle représentée (Repère *B* de la Figure 6.12).

Ces événements peuvent également correspondre à un cycle dont la fréquence est un harmonique de la période correspondant à l'échelle cyclique représentée (Repère *C* de la Figure 6.12).

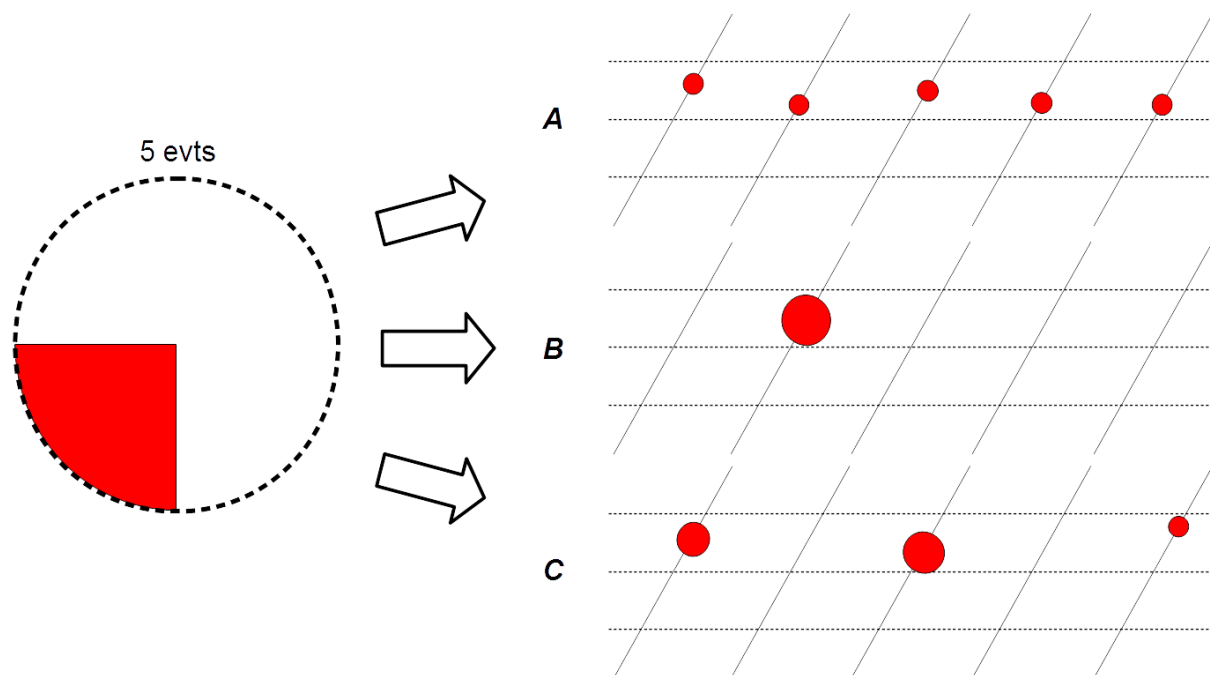


FIGURE 6.12 – Différentes manifestations d'une concentration temporelle identifiée sur le diagramme circulaire, dans le diagramme mixte

Pour conclure à l'existence d'un cycle, l'identification d'une surreprésentation d'apparition d'événements dans le diagramme circulaire doit donc être suivie de l'identification d'un alignement formé des entités correspondantes sur le diagramme temporel mixte.

6.1.2.2 Recherche visuelle de récurrences cycliques d'événements par modification de l'échelle temporelle cyclique

Tout comme dans *SpiralGraph* [Weber et al., 2001], la recherche visuelle de récurrences cycliques s'effectue en modifiant graduellement la durée T de l'échelle temporelle cyclique¹, dans le but d'observer la formation d'un alignement d'entités graphiques sur le diagramme mixte. La formation de cet alignement est observable lorsque la valeur de T approche celle de la période P de réapparition d'un phénomène cyclique.

Dans le diagramme sinusoïdal, cet alignement d'entités graphiques prend la forme d'une sinusoïde (Repère B de la Figure 6.13). Au fur et à mesure que la valeur de T approche celle de P , le nombre de pulsations de la sinusoïde diminue, et la forme de l'alignement se rapproche de celle d'une droite (Repères B à H de la Figure 6.13). Les entités sont alors alignées parallèlement à cet axe lorsque $T = P$ (Repère H de la Figure 6.13). On observe ici un cycle ayant une période de 402 jours.

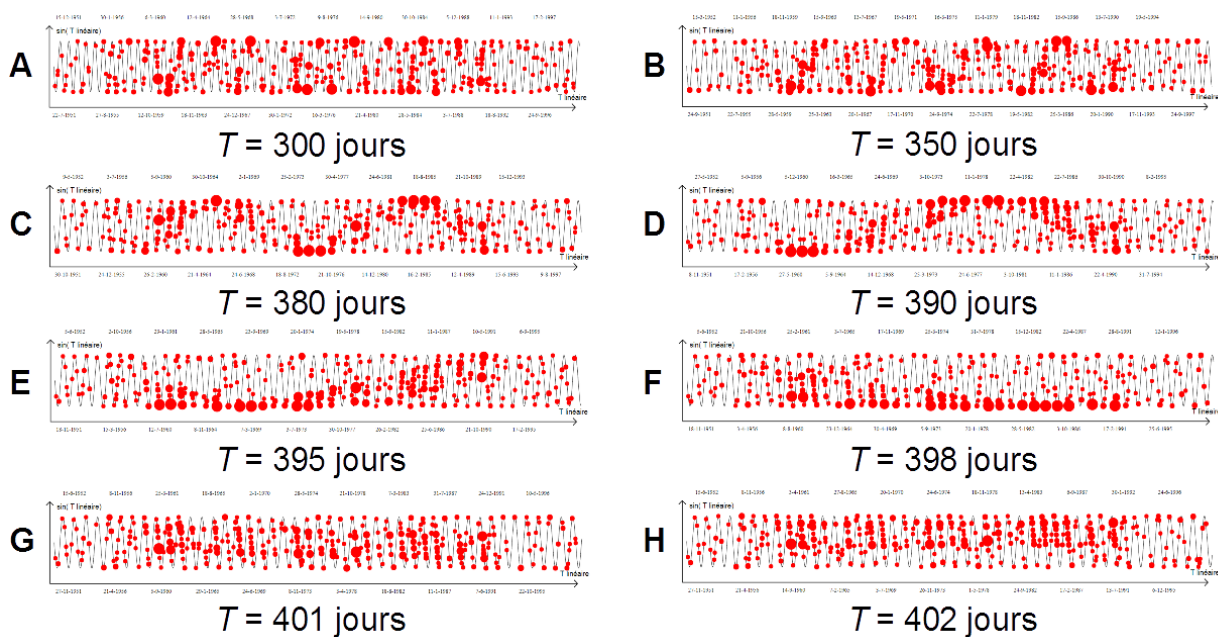


FIGURE 6.13 – Recherche d'une récurrence cyclique au moyen du diagramme temporel sinusoïdal

Dans le diagramme en dents de scie, l'alignement des entités graphiques forme d'abord une, ou plusieurs droites obliques. Lorsque les entités représentant des événements cycliques forment un tel alignement pour une échelle cyclique de durée T , l'orientation des droites formées indique si la période P est inférieure ou supérieure à T (Figure 6.14).

Si $P < T$, chaque ligne brisée du diagramme mixte représente une durée supérieure à celle séparant deux instances du cycle. Les entités graphiques alignées selon une droite oblique seront placées, successivement, plus bas sur l'axe des ordonnées que l'entité précédente. La ou les droites

1. La modification de T est effectuée, manuellement, de manière continue au moyen d'un curseur dans l'interface de contrôle de l'environnement.

formées par l'alignement des entités graphiques sont ainsi descendantes.

À l'inverse, si $P > T$, chaque ligne brisée du diagramme mixte représente une durée inférieure à celle séparant deux instances du cycle. Les entités graphiques alignées selon une droite oblique seront placées, successivement, plus haut sur l'axe des ordonnées que l'entité précédente. La ou les droites formées par l'alignement des entités graphiques sont ainsi ascendantes.

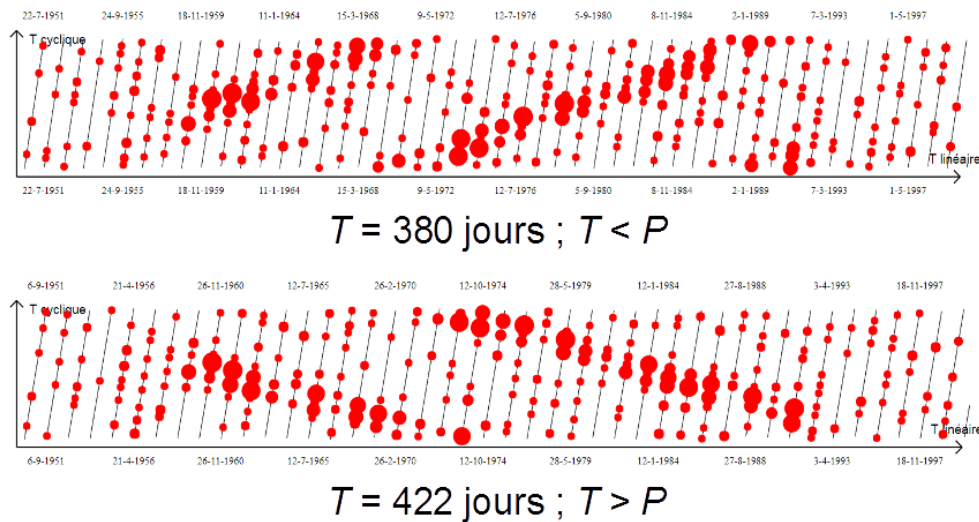


FIGURE 6.14 – Influence de l'orientation des alignements graphiques dans la recherche d'une récurrence cyclique au moyen du diagramme temporel en dents de scie

Au fur et à mesure que la valeur de T approche de celle de P , l'orientation de la ou des droites formées se rapproche de celle de l'axe des abscisses (Repères B à H de la Figure 6.15). Les entités sont alignées, parallèlement à cet axe, lorsque $T = P$ (Repère H de la Figure 6.15). On identifie également sur ce diagramme un cycle ayant une période de 402 jours.

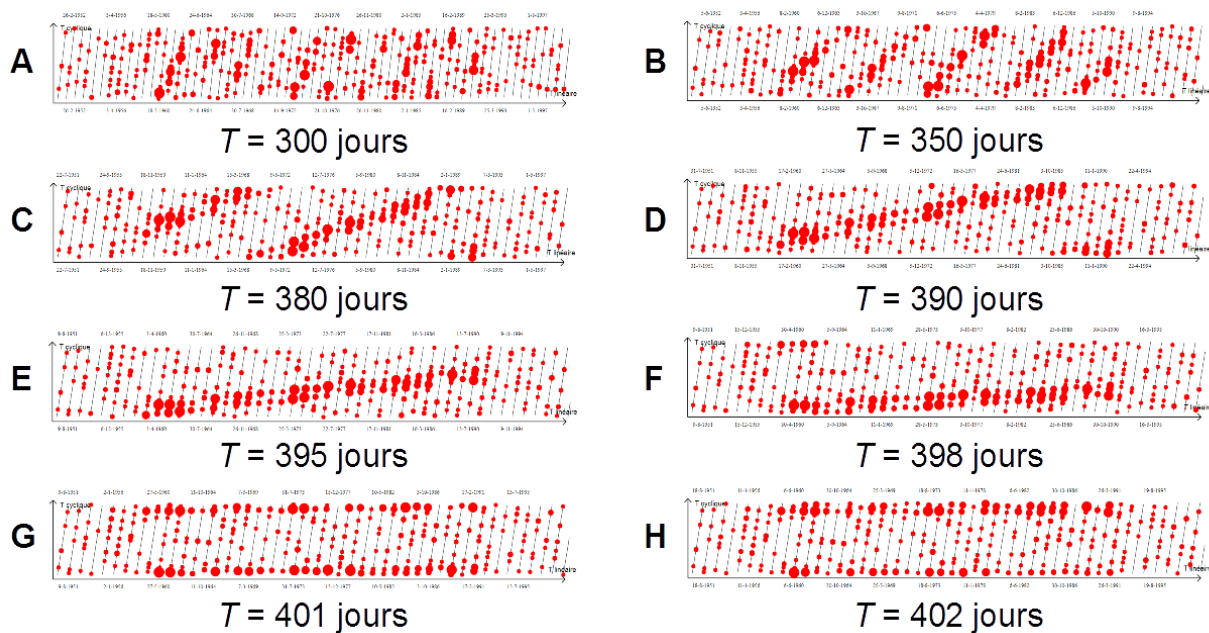


FIGURE 6.15 – Recherche d'une récurrence cyclique au moyen du diagramme temporel en dents de scie

La formation d'un alignement des entités graphiques représentant des événements cycliques dans le diagramme mixte, est également observable, dans une moindre mesure, par la formation d'un motif graphique particulier dans le diagramme circulaire (Figure 6.16).

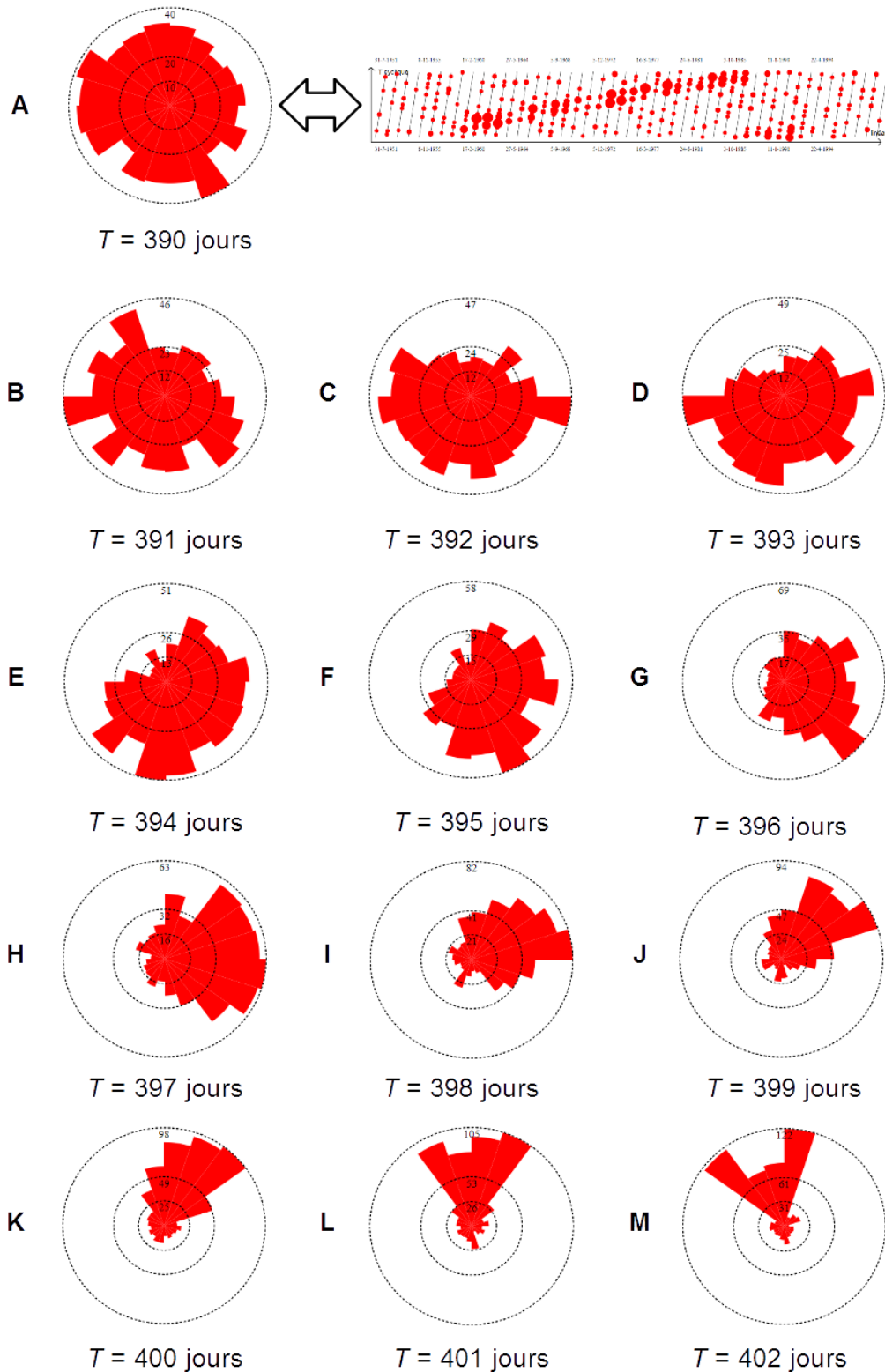


FIGURE 6.16 – Identification de la formation d'une concentration temporelle sur le diagramme circulaire, dans le cadre de la recherche d'une récurrence cyclique

Cette observation est possible à partir du moment où T est suffisamment proche de P , pour qu'un alignement des entités graphiques correspondantes dans le diagramme mixte prenne la forme :

- d'une sinusoïde à une seule pulsation dans le diagramme sinusoïdal.
- d'une seule droite oblique (ou de deux droites ne se recouvrant pas sur l'axe des ordonnées) dans le diagramme en dents de scie (Repère A de la Figure 6.16).

À partir de l'instant où cette échelle cyclique est représentée, l'alignement progressif des entités graphiques avec l'axe des abscisses va s'accompagner d'une diminution du nombre de sections cycliques de la courbe du diagramme occupées par ces entités.

Le nombre de sections angulaires correspondantes dans le diagramme circulaire va ainsi diminuer, jusqu'à atteindre son minimum lorsque T atteint la valeur de P .

L'approche de la valeur P par l'échelle cyclique va alors être observable par l'accroissement du rayon d'un nombre de plus en plus restreint de sections du diagramme circulaire (Repères A à M de la Figure 6.16). On identifie ici une concentration temporelle relative à une échelle cyclique de 402 jours.

6.1.2.3 Déplacement général des entités selon l'axe des ordonnées du diagramme mixte lors de la modification de l'échelle temporelle cyclique

Au cours de l'augmentation de la durée de l'échelle temporelle cyclique dans la Figure 6.16, on observe un déplacement de l'alignement d'événements, en cours de formation, vers le bas du diagramme. Ce déplacement est dû au fait qu'à chaque augmentation de la durée de l'échelle cyclique, un événement apparaît plus "tôt" dans l'instance d'échelle cyclique dans laquelle il se trouve.

Soit un événement e dont le moment d'apparition dans le temps linéaire M_l est de 150 jours (par rapport à une date θ). Soit M_c son moment d'apparition relatif à une échelle cyclique :

- au travers d'une échelle cyclique d'une durée de 100 jours, M_c est égal à 50 jours (sur 100) ;
- au travers d'une échelle cyclique d'une durée de 120 jours, M_c est égal à 30 jours (sur 120) ;
- au travers d'une échelle cyclique d'une durée de 149 jours, M_c est égal à 1 jour (sur 149).

Ainsi, plus la durée de l'échelle cyclique augmente, plus l'événement e apparaîtra "tôt" à l'intérieur d'une instance de l'échelle temporelle cyclique. Ceci jusqu'à ce que la durée de l'échelle dépasse 150 jours : pour une échelle cyclique d'une durée de 151 jours, l'événement e apparaît à la fin de l'instance d'échelle précédente, et M_c est égal à 150 jours (sur 151).

Ainsi, lors d'une augmentation de la durée de l'échelle cyclique, l'ensemble des événements (hormis ceux dont le moment d'apparition dans le temps cyclique est proche de θ) apparaissent plus tôt, relativement à cette échelle, qu'avant modification. Le phénomène inverse est observable lorsque l'on diminue l'échelle temporelle cyclique.

Le moment d'apparition cyclique des événements étant représenté par l'axe des ordonnées dans le diagramme temporel en dents de scie, une augmentation de la durée de l'échelle cyclique entraîne ainsi un déplacement des événements vers le bas, tandis qu'une diminution de la durée de l'échelle cyclique entraîne un déplacement des événements vers le haut.

De la même manière, les événements formant un alignement graphique dans la Figure 6.15 se déplacent vers le bas lors de l'augmentation de la durée de l'échelle, entraînant un déplacement, vers le bas, de l'alignement formé.

Pour cette même raison, lors de la formation d'une concentration temporelle dans le diagramme temporel circulaire, au cours d'une augmentation de la durée de l'échelle cyclique, les sections du diagramme correspondant subissent une rotation dans le sens antihoraire (Figure 6.16).

Lorsque l'échelle cyclique est égale à 402 jours (Repère H de la Figure 6.15), le moment d'apparition dans le temps cyclique des événements formant l'alignement graphique étant proche de θ , une partie des événements apparaît en haut du diagramme en dents de scie. Il est alors possible de modifier la phase du diagramme temporel mixte, afin de représenter l'alignement, formé par les événements, au centre du diagramme (Section 5.3.1).

6.1.3 Proposition de pistes d'exploration à l'utilisateur par suggestion de valeurs à affecter à l'échelle temporelle cyclique

L'identification de récurrences cycliques dans les événements est ainsi basée sur une recherche visuelle et exploratoire au travers des diagrammes temporels.

Cependant, il convient de guider l'utilisateur dans sa recherche visuelle, et de lui suggérer des pistes d'exploration, car les échelles de valeurs possibles sont trop nombreuses.

Une proposition peut être de combiner l'analyse exploratoire avec l'utilisation de méthodes quantitatives, analysant les données en arrière-plan et transcrivant de manière graphique leurs résultats, à l'instar d'*Arc Diagram* [Wattenberg, 2002] (Figure 4.18). Ces résultats sont alors proposés à l'utilisateur en guise de pistes d'exploration.

Notre environnement effectue ainsi, en arrière-plan, une analyse mathématique de la composante temporelle des données événementielles, dans le but d'y détecter de possibles comportements cycliques. Les périodes des cycles identifiés sont alors retranscrites de manière graphique à l'utilisateur, et constituent une suggestion d'échelles temporelles cycliques "remarquables" que l'utilisateur peut "inspecter". Pour identifier ces possibles comportements cycliques, nous appliquons aux séries événementielles des analyses harmoniques.

6.1.3.1 Principe de l'analyse harmonique

Recherche des composantes périodiques d'un signal temporel. L'analyse harmonique est une branche des mathématiques étudiant la représentation des signaux, ou des fonctions mathématiques, comme des superpositions de signaux ou de fonctions sinusoïdales élémentaires, de fréquence et d'amplitude différentes [Bony, 2001, Katznelson, 2004].

Les signaux ou fonctions sinusoïdales élémentaires sont appelés les *harmoniques* du signal ou de la fonction analysée, d'où le nom d'*analyse harmonique*.

Les analyses harmoniques ont connu de nombreuses applications en physique sous le nom d'analyse spectrale, et interviennent dans de nombreux domaines tels que le traitement du signal, l'analyse prévisionnelle, etc. Ces analyses font appel à des outils mathématiques, tels que les séries de Fourier et les transformées de Fourier, permettant d'obtenir les harmoniques d'un signal [Gasquet and Witomski, 2000].

Séries et transformée de Fourier. Les séries de Fourier visent à décomposer un signal périodique, de forme quelconque et de fréquence f , en une somme de fonctions sinusoïdales dont les fréquences sont des multiples de f . La fréquence f est appelée la fréquence *fondamentale* du signal, et ses multiples sont appelés les *harmoniques* de f .

La transformée de Fourier est une extension, pour les fonctions non périodiques, du développement des fonctions périodiques en séries de Fourier. La transformée de Fourier s'exprime comme une "somme infinie" des fonctions trigonométriques de toutes fréquences. La transformée de Fourier permet ainsi de décomposer un signal, combinant plusieurs fonctions périodiques de fréquence f , en une somme de sinusoïdes dans les fréquences sont les différentes fréquences fondamentales f et leurs sinusoïdes [Bony, 2001, Katznelson, 2004].

Spectre de Fourier. Lorsqu'une fonction mathématique représente un phénomène physique, on appelle cette fonction *signal*, et sa transformée de Fourier s'appelle son *spectre*. Le spectre permet de recenser les fréquences, fondamentales et harmoniques, de toutes les fonctions sinusoïdales dont la somme permet de recomposer le signal originel.

Utilisation de la transformée de Fourier pour l'analyse des récurrences cycliques d'un phénomène. En permettant de chercher quelles fonctions sinusoïdales permettent de recomposer au mieux le signal représentant un phénomène physique, les transformées et séries de Fourier sont un bon outil pour tenter d'identifier les caractéristiques cycliques de ces phénomènes.

Dans le cas de phénomènes physiques observés au travers de capteurs, les données à disposition ne se présentent généralement pas sous une forme continue, mais plutôt sous la forme d'un signal échantillonné. Ce dernier consiste en une suite de valeurs discrètes séparées par une unité de temps constante correspondant à une fréquence d'échantillonnage. Un signal échantillonné prend ainsi la forme d'une série temporelle. Un équivalent discret de la transformée de Fourier permet de traiter ces signaux échantillonnés et se nomme la transformation de Fourier discrète (*TFD*).

Exemple d'analyse d'un phénomène cyclique par les séries de Fourier : les marées

Le phénomène de marée est la variation de la hauteur d'eau du niveau des mers, causée par l'effet des forces de gravitation dues à la Lune et au Soleil, et également causée par l'effet de la force d'inertie due à la révolution de la Terre autour du centre de gravité du système Terre-Lune.

Le phénomène de marée présente plusieurs comportements cycliques :

- Le plus visible consiste en la variation bi journalière du niveau de la mer entre les états que l'on appelle *marée haute*, et *marée basse*.
- Des marées plus fortes ont cependant lieu lors d'un alignement de la Lune et du Soleil lors des nouvelles lunes et des pleines lunes. Ces marées sont nommées *vives-eaux*. À l'inverse, des marées plus faibles, appelées *mortes-eaux* sont observées lors des premiers et derniers quartiers de Lune. Cette évolution entre *vives-eaux* et *mortes-eaux* se répète périodiquement avec une période d'environ 14 jours.
- Enfin, les marées les plus fortes apparaissent au moment des équinoxes, tandis que les marées les plus faibles apparaissent au moment des solstices, ce qui se traduit par une évolution périodique du phénomène de marée avec une période de six mois.

Les phénomènes de marée sont étudiés, entre autres, au travers de l'utilisation de marégraphes disposés le long des côtes. Ces appareils permettent, en un point donné, de mesurer l'évolution du niveau de la mer, au cours du temps, par rapport à un point de référence. Ce type de capteur permet de retranscrire le phénomène de marée en un point, en un signal temporel représentant l'évolution du niveau de la mer. Ce signal prend la forme d'un signal échantillonné.

L'utilisation de la transformée de Fourier discrète sur ce signal permet d'obtenir le spectre de ce signal. La visualisation graphique du spectre permet alors de retrouver les trois fréquences fondamentales, et leurs harmoniques, correspondant aux marées bi journalières, au cycle des *vive-eaux* et *mortes-eaux*, et au cycle des marées d'équinoxe et de solstices.

L'utilisation de la transformée de Fourier discrète permet ainsi d'identifier les possibles com-

portements cycliques d'un phénomène, si l'on dispose de données permettant de le décrire sous la forme de séries temporelles.

6.1.3.2 Mise en oeuvre

Principe : analyser le spectre issu de l'analyse harmonique de nos séries d'événements. Le principe de l'outil d'aide à la recherche visuelle de récurrences cycliques consiste à produire le spectre de Fourier du signal correspondant à nos données d'événements. Puis les fréquences les plus significatives du signal sont identifiées à partir du spectre obtenu, et proposées à l'utilisateur comme suggestion de valeur pour l'échelle cyclique représentée dans le diagramme mixte. L'utilisateur peut alors assigner les périodes correspondantes à l'échelle cyclique, et vérifier si, à l'approche de ces valeurs d'échelles, la formation d'un alignement d'entités graphiques est observable.

Passage d'une série événementielle à une série temporelle. La transformée de Fourier n'est utilisable que sur un signal temporel, soit continu, soit échantillonné à intervalle constant. Afin de pouvoir utiliser la transformée de Fourier sur nos données, la première étape est de recréer une série temporelle à partir de nos données événementielles.

Nous définissons un intervalle d'échantillonnage (*10 jours*), paramétrable par l'utilisateur. À partir de cet intervalle d'échantillonnage, nous effectuons une discrétisation de l'étendue temporelle étudiée (*1000 jours*). Nous obtenons ainsi une série temporelle (de *100* individus), chacun correspondant à un intervalle de temps d'une durée égale à l'intervalle d'échantillonnage (*10 jours*).

Pour chaque individu de cette série est calculé le nombre d'événements apparaissant durant l'intervalle correspondant (nous disposons ainsi d'une série temporelle de *100* individus, décrivant le nombre d'événements apparaissant tous les *10* jours sur une période de *1000* jours).

La représentation graphique du signal obtenu prend la forme d'un diagramme en bâton, rappelant une fonction de *Dirac*² (Repères *A* et *B* de la Figure 6.17).

Analyse harmonique de la série temporelle et renvoi des résultats. Nous appliquons une transformée de Fourier discrète sur la série temporelle obtenue. Notre environnement sélectionne alors les fréquences correspondant aux pics significatifs du spectre obtenu (Repère *C* de la Figure 6.17).

La liste des périodes significatives correspondantes est alors projetée sur le curseur permettant de modifier la durée de l'échelle cyclique dans l'interface de contrôle (Repère *D* de la Figure 6.17). L'utilisateur peut alors voir sur le curseur vers quelle valeur il doit faire évoluer l'échelle cyclique afin de vérifier s'il identifie, pour ces échelles de représentation du temps, des alignements d'objets graphiques sur le diagramme temporel mixte (Repère *E* de la Figure 6.17).

2. La distribution de Dirac peut être informellement considérée comme une fonction prenant une valeur infinie en 0, et la valeur zéro partout ailleurs. L'expression *un Dirac* désigne une fonction ou une courbe *piquée* en une valeur donnée.[Wikipédia, 2018a]

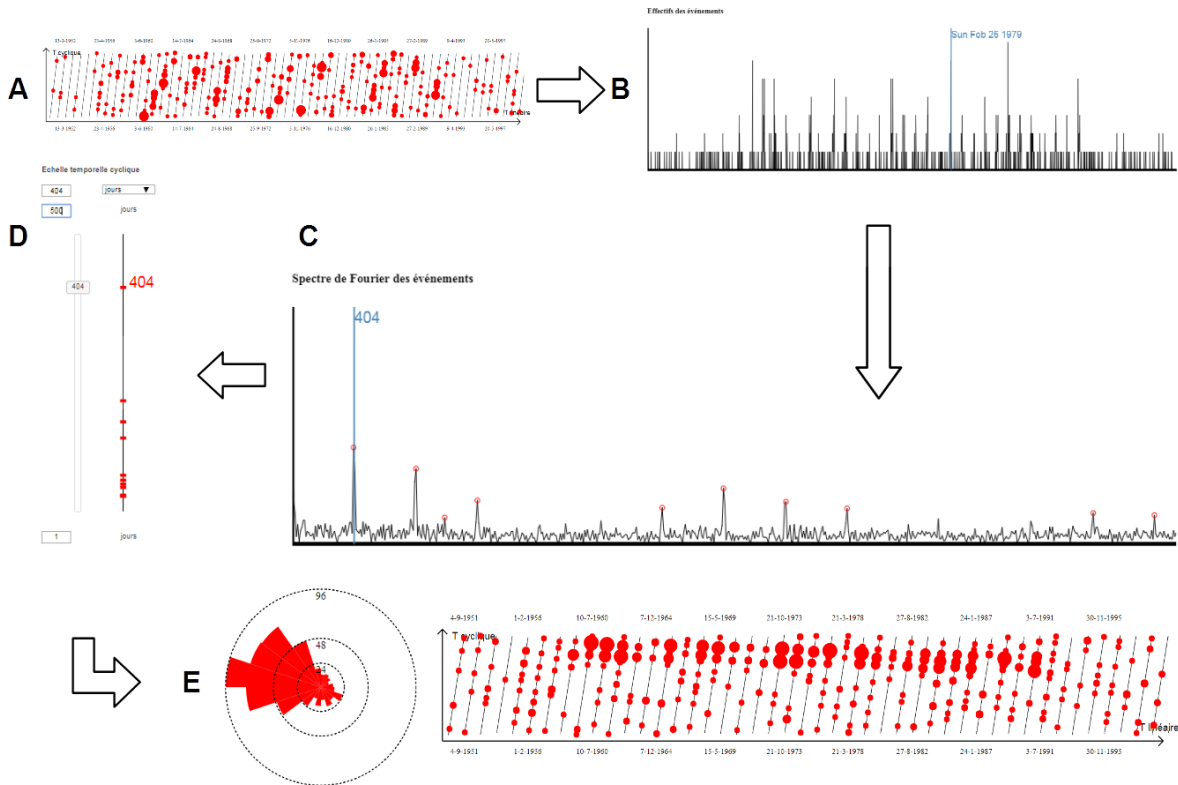


FIGURE 6.17 – Analyse de cycles assistée par les transformées de Fourier

6.1.3.3 Utilisation

La transformée de Fourier est effectuée en arrière-plan lors de chaque import d'événements dans l'interface de visualisation. Par défaut, l'analyse est effectuée de manière automatique, avec un intervalle d'échantillonnage de trois jours, et porte sur l'ensemble des événements. L'analyse peut cependant être effectuée à la demande de l'utilisateur. Le nombre de résultats projetés sur le curseur de modification de l'échelle cyclique peut également être modifié.

L'utilisateur peut visualiser le spectre de Fourier obtenu, et l'explorer plus exhaustivement à l'aide d'outils de navigation et de zoom, afin d'identifier les échelles cycliques les plus susceptibles de permettre l'identification d'un cycle dans les données événementielles.

Dans le cas de séries d'événements présentant une part trop importante d'événements aléatoires, la transformée de Fourier du signal obtenu peut donner un spectre ne permettant pas d'identifier facilement des fréquences significatives. L'utilisateur peut alors tenter de "lisser" l'aspect du signal en entrée de la transformée de Fourier, en diminuant la fréquence d'échantillonnage.

Ce "lissage" permet de mettre en valeur les composantes cycliques de faible fréquence du phénomène dans la courbe du signal obtenu. Les composantes cycliques ayant une courte période de réapparition sont cependant plus difficiles à détecter, voir impossibles si cette période est plus courte que l'intervalle d'échantillonnage choisi.

6.2 Structure des scénarios d'analyse

Nous présentons différents scénarios sous la forme de diagrammes de séquence. Ces diagrammes font intervenir différents *objets*, correspondant aux différents *acteurs* intervenant dans l'analyse (l'utilisateur, seul véritable *acteur*, et les composants de l'environnement de géovisualisation), et aux différents liens entre ces *acteurs* (Figure 6.18).

Les diagrammes décrivent les séries d'actions devant être effectuées par l'utilisateur (Repère *A*), au moyen de *GrAPHiST* (Repère *B*), pour identifier un type de pattern particulier.

Les actions de l'utilisateur dans *GrAPHiST* s'effectuent au travers de l'interface visuelle (Repère *C*). Celle-ci est composée :

- d'une interface de contrôle (Repère *D*) ;
- des trois espaces de représentation de la donnée :
 - la carte (Repère *E*) ;
 - le diagramme temporel mixte (Repère *F*) ;
 - le diagramme temporel circulaire (Repère *G*).

Les actions effectuées par l'utilisateur dans l'interface (Repère *H*) regroupent :

- des actions d'interaction (Repère *I*), par lesquelles l'utilisateur effectue une sélection d'événements, modifie leur représentation dans les fenêtres de l'interface, effectue une classification des événements, etc ;

Le déroulement d'une interaction s'effectue en quatre temps :

- l'utilisateur effectue une requête (Repère *J*) :
 - soit uniquement au moyen de l'interface de contrôle (changement de la durée de l'échelle cyclique) ;
 - soit en deux temps, au moyen de l'interface de contrôle afin de sélectionner un type de requête, puis au moyen des espaces de représentation de la donnée (création de zones de classification spatiale : l'utilisateur sélectionne l'action à effectuer dans l'interface de contrôle, puis dessine les zones dans la fenêtre spatiale) ;
- chaque requête est traitée par *GrAPHiST* au travers de routines de calcul (Repère *K*) effectuées en arrière-plan ;
- ces routines renvoient un résultat (Repère *L*), retranscrit dans les fenêtres de l'interface ;
- l'affichage des modifications dans l'interface permet un retour visuel à l'utilisateur (Repère *M*) ;
- des interprétations visuelles (Repère *N*), par lesquelles l'utilisateur identifie un motif d'intérêt dans les représentations graphiques de l'interface.

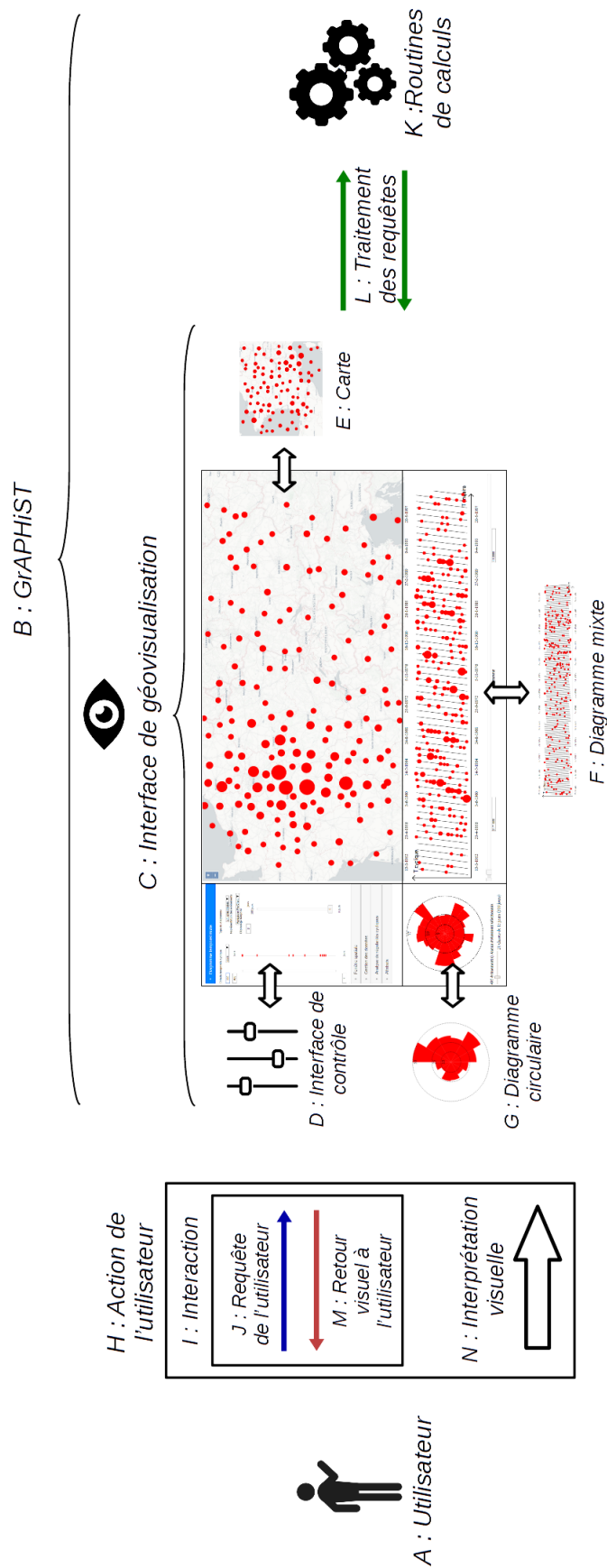


FIGURE 6.18 – Composantes des diagrammes de séquences décrivant les scénarios d’analyse

Ces différents *objets* sont retranscrits dans les diagrammes de séquences (Figure 6.19) :

- Un diagramme représente les différents *acteurs* intervenants dans l'analyse (Repère *O*).
- Chaque scénario est divisé en étapes (Repères *P* et *Q*), elles-mêmes divisées en sous-étapes (Repère *R*). Le descriptif de chaque étape est explicité pour chaque scénario au cours de ce chapitre.
- Les différentes séquences de chaque action de l'utilisateur sont représentées par des rectangles sur les diagrammes (Repère *S*).

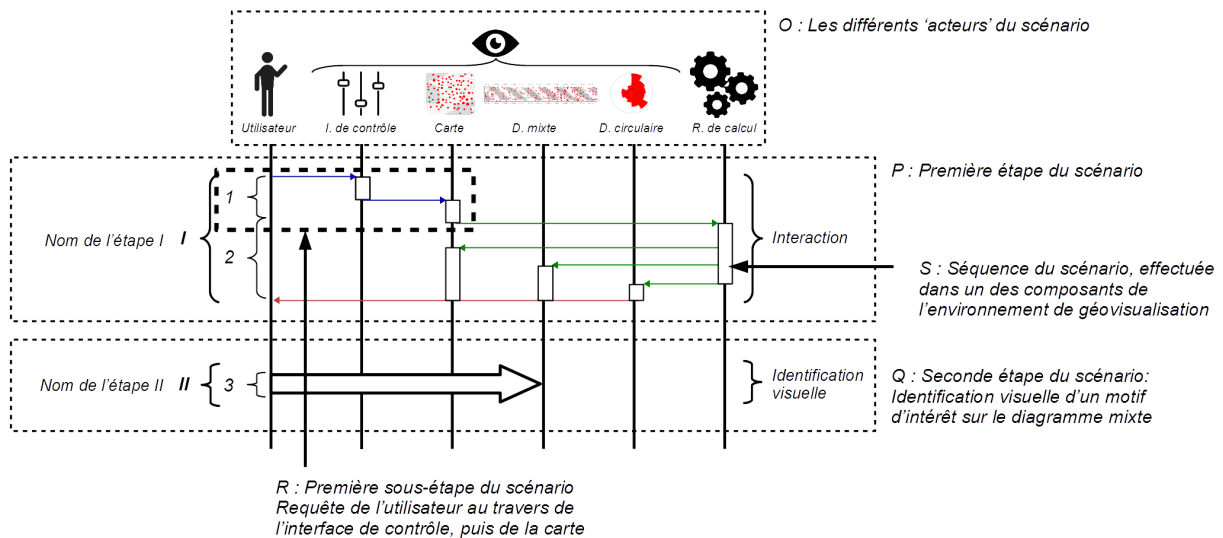


FIGURE 6.19 – Exemple d'un diagramme de séquence

6.3 Analyse de patterns spatio-temporels, au travers de zones spécifiques

6.3.1 Procédé général

6.3.1.1 Classification spatiale et analyse au travers des diagrammes temporels

À l'instar des approches illustrées par *Time Wave* [Li, 2010] et *PerSE* [Swedberg and Peuquet, 2017], l'analyse de patterns au travers de zones spécifiques consiste à classer et à représenter les événements en fonction de leur lieu d'apparition, puis à identifier des structures dans la distribution temporelle des événements appartenant aux classes spatiales créées (Figure 6.20).

Cette approche utilise une représentation des événements en fonction de leur appartenance à des zones spatiales créées à la volée ou importées par l'utilisateur.

Les événements de chaque classe spatiale étant représentés distinctement au moyen d'une variable visuelle, l'analyse des structures temporelles de chaque classe s'effectue à l'aide des diagrammes temporels. Les événements de chaque classe peuvent être masqués afin de focaliser l'analyse sur les événements d'une seule zone spatiale.

6.3.1.2 Création des zones de classification

Une création de zones à main levée permet à l'utilisateur d'analyser de manière plus approfondie des groupes d'événements observables sur la carte, dont il suspecte une distribution temporelle particulière, ou bien d'analyser la distribution temporelle d'événements situés dans

une zone d'intérêt connue de l'utilisateur.

Utiliser un découpage préalable du territoire permet en revanche de classer les événements selon les caractéristiques administratives (découpage selon des pays ou communes), physiques, topologiques (découpage selon des plages d'altitude), environnementales (découpage selon la végétation, un indice de pollution), ou sociétales de leur zone d'apparition. Cette approche permet de croiser les thématiques de recherche correspondant au découpage du territoire utilisé, avec l'analyse des phénomènes spatio-temporels.

6.3.1.3 L'utilisation des outils de sélection et de *brushing* pour affiner l'analyse

L'observation, dans les diagrammes temporels, de motifs d'intérêt formés par des entités graphiques similaires permet de conclure à l'existence d'un pattern spatio-temporel s'étendant dans la zone de classification spatiale correspondante.

Une sélection graphique (pouvant être suivie d'un *brushing*) des entités formant ce motif d'intérêt dans le diagramme temporel, permet de mettre en valeur la composante spatiale des événements correspondants sur la carte. Il est alors possible de définir plus précisément l'étendue spatiale de la structure spatio-temporelle identifiée, à l'intérieur des frontières de la zone de classification spatiale.

Les outils de sélection forment ainsi un outil supplémentaire pour l'analyse exploratoire, permettant d'améliorer la définition de l'étendue spatio-temporelle du pattern recherché.

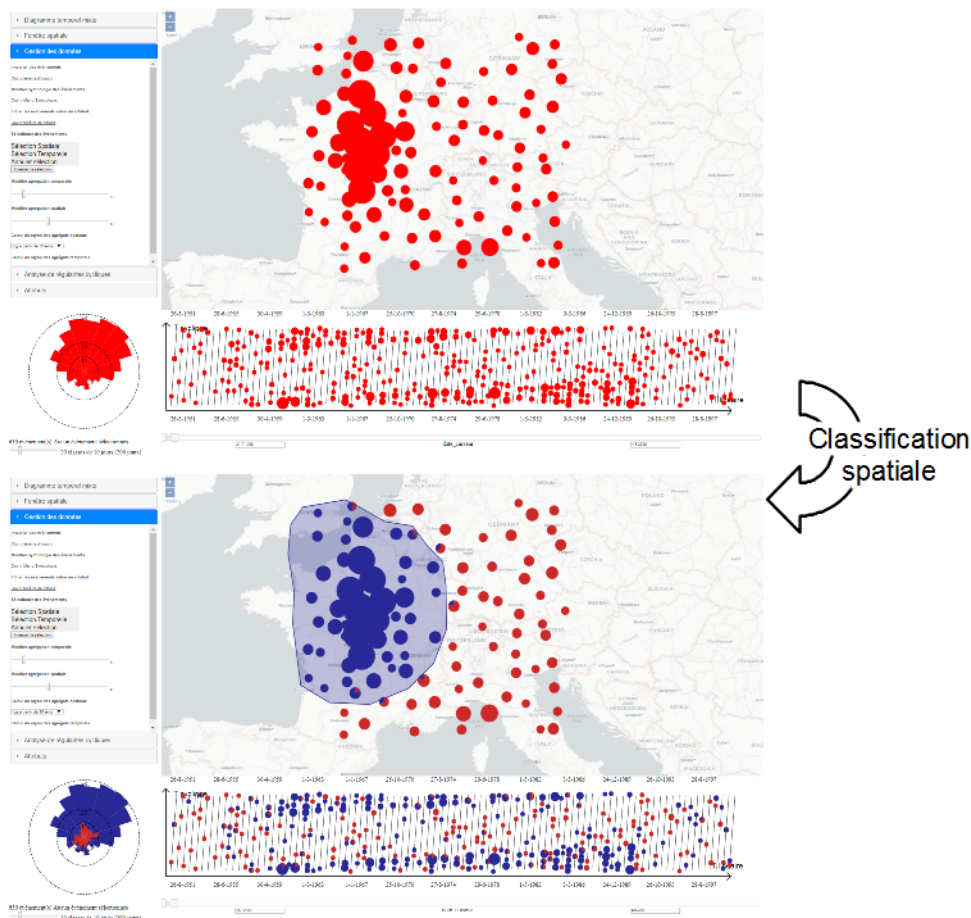


FIGURE 6.20 – Classification spatiale des événements en deux zones, créées à main levée

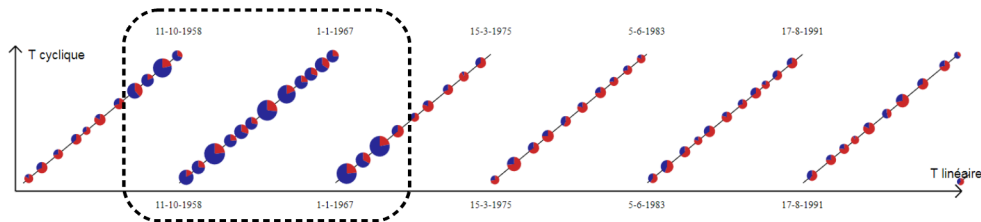
6.3.2 Analyse de clusters relatifs au temps linéaire, selon une entrée spatiale

Cette analyse consiste à identifier des concentrations d'événements, dans l'espace et dans le temps linéaire, au travers d'un découpage du territoire prédéfini. Un exemple peut être l'analyse des accidents routiers sur une période de dix ans, au travers des régions administratives françaises. Le but de l'analyse est d'identifier, durant cette étendue temporelle de dix ans, des pics d'apparition d'événements dans certaines régions.

Cette analyse suit un scénario en sept étapes, décrites dans la Figure 6.21, et explicitées dans l'encadré suivant.

Identification de clusters relatifs au temps linéaire, selon une entrée spatiale

- I.** Classification spatiale des événements :
 1. Définition des zones de classification spatiales au travers de la carte, et choix des couleurs correspondantes.
 2. Classification des événements. Application de la couleur correspondante sur les entités graphiques dans la carte et les diagrammes temporels.
- II.** Focalisation sur une ou plusieurs classes d'événements :
 3. Définition des zones spatiales à masquer au travers de la carte.
 4. Identification des entités graphiques à masquer dans la carte et les diagrammes temporels.
- III.** Modification de la représentation temporelle des événements :
 5. Modification des critères de représentation des événements dans le diagramme temporel mixte (événements *vs* agrégats; rayon des agrégats; étendue temporelle des agrégats).
 6. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans le diagramme temporel mixte.
- IV.** Identification d'un cluster spatio-temporel sur le diagramme mixte :
 7. Identification d'une concentration, selon l'axe des abscisses, d'entités ponctuelles de même couleur, ou d'entités *Pie Chart* présentant une couleur majoritaire (Figure ci-dessous).



- V.** Sélection des événements du cluster :
 8. Sélection graphique, dans le diagramme temporel mixte, des événements identifiés dans la séquence *IV*.
 9. Identification des événements à sélectionner. Mise en valeur des entités graphiques correspondantes dans la carte et les diagrammes temporels.
- VI.** Modification de la représentation spatiale des événements :
 10. Modification des critères de représentation des événements dans la carte (événements *vs* agrégats; rayon des agrégats; étendue spatiale des agrégats).
 11. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans la carte.
- VII.** Identification de l'étendue spatiale du cluster :
 12. Identification, dans la carte, de l'étendue spatiale couverte par les événements sélectionnés.

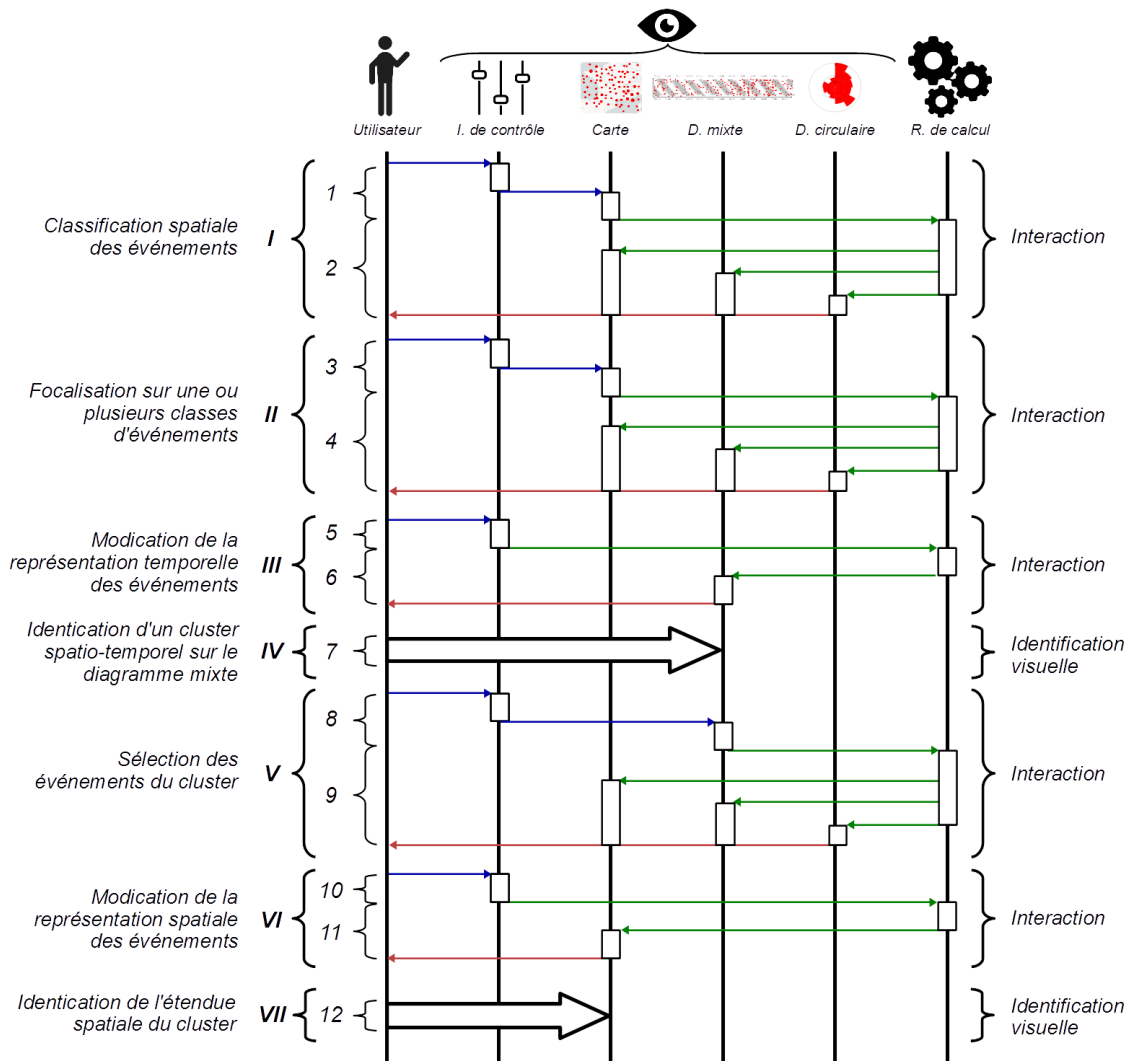


FIGURE 6.21 – Identification de clusters relatifs au temps linéaire, selon une entrée spatiale

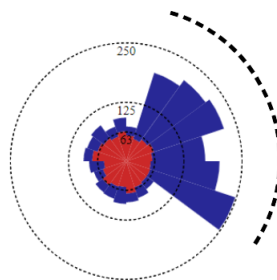
6.3.3 Analyse de clusters relatifs à une échelle temporelle cyclique, selon une entrée spatiale

Cette analyse consiste à identifier des concentrations d'événements, dans l'espace et dans certaines plages de temps relatives à une échelle cyclique, au travers d'un découpage du territoire prédéfini. Un exemple peut être l'analyse de l'apparition d'accidents routiers selon une échelle d'une année, au travers des régions administratives françaises. Le but de l'analyse est d'identifier si des pics du nombre d'événements apparaissent dans certaines régions à certains moments de l'année (plus d'accidents durant l'hiver en *Rhône-Alpes*, plus d'accidents au mois de septembre en Bretagne, etc.).

Cette analyse suit un scénario en huit étapes, décrites dans la Figure 6.22, et explicitées dans l'encadré suivant.

Identification de clusters relatifs à une échelle cyclique, selon une entrée spatiale

- I.** Choix de l'échelle temporelle cyclique :
 1. Modification de la durée de l'échelle temporelle cyclique dans l'interface de contrôle.
 2. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans les diagrammes temporels.
- II.** Classification spatiale des événements :
 3. Définition des zones de classification spatiales au travers de la carte, et choix des couleurs correspondantes.
 4. Classification des événements. Application de la couleur correspondante sur les entités graphiques dans la carte et les diagrammes temporels.
- III.** Focalisation sur une ou plusieurs classes d'événements :
 5. Définition des zones spatiales à masquer au travers de la carte.
 6. Identification des entités graphiques à masquer dans la carte et les diagrammes temporels.
- IV.** Modification de la représentation temporelle des événements :
 7. Modification des critères de représentation des événements dans le diagramme temporel circulaire (Modification du nombre de sections du diagramme temporel circulaire).
 8. Calcul de la nouvelle représentation graphique. Modification du diagramme temporel circulaire.
- V.** Identification d'un cluster spatio-temporel sur le diagramme circulaire :
 9. Identification d'une, ou de plusieurs sections angulaires du diagramme circulaire présentant une couleur majoritaire (Figure ci-dessous).



- VI.** Sélection des événements du cluster :
 10. Sélection graphique, dans le diagramme temporel mixte, des événements identifiés dans la séquence V.
 11. Identification des événements à sélectionner. Mise en valeur des entités graphiques correspondantes dans la carte et les diagrammes temporels.
- VII.** Modification de la représentation spatiale des événements :
 12. Modification des critères de représentation des événements dans la carte (événements *vs* agrégats; rayon des agrégats; étendue spatiale des agrégats).
 13. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans la carte.
- VIII.** Identification de l'étendue spatiale du cluster :
 14. Identification, dans la carte, de l'étendue spatiale couverte par les événements sélectionnés.

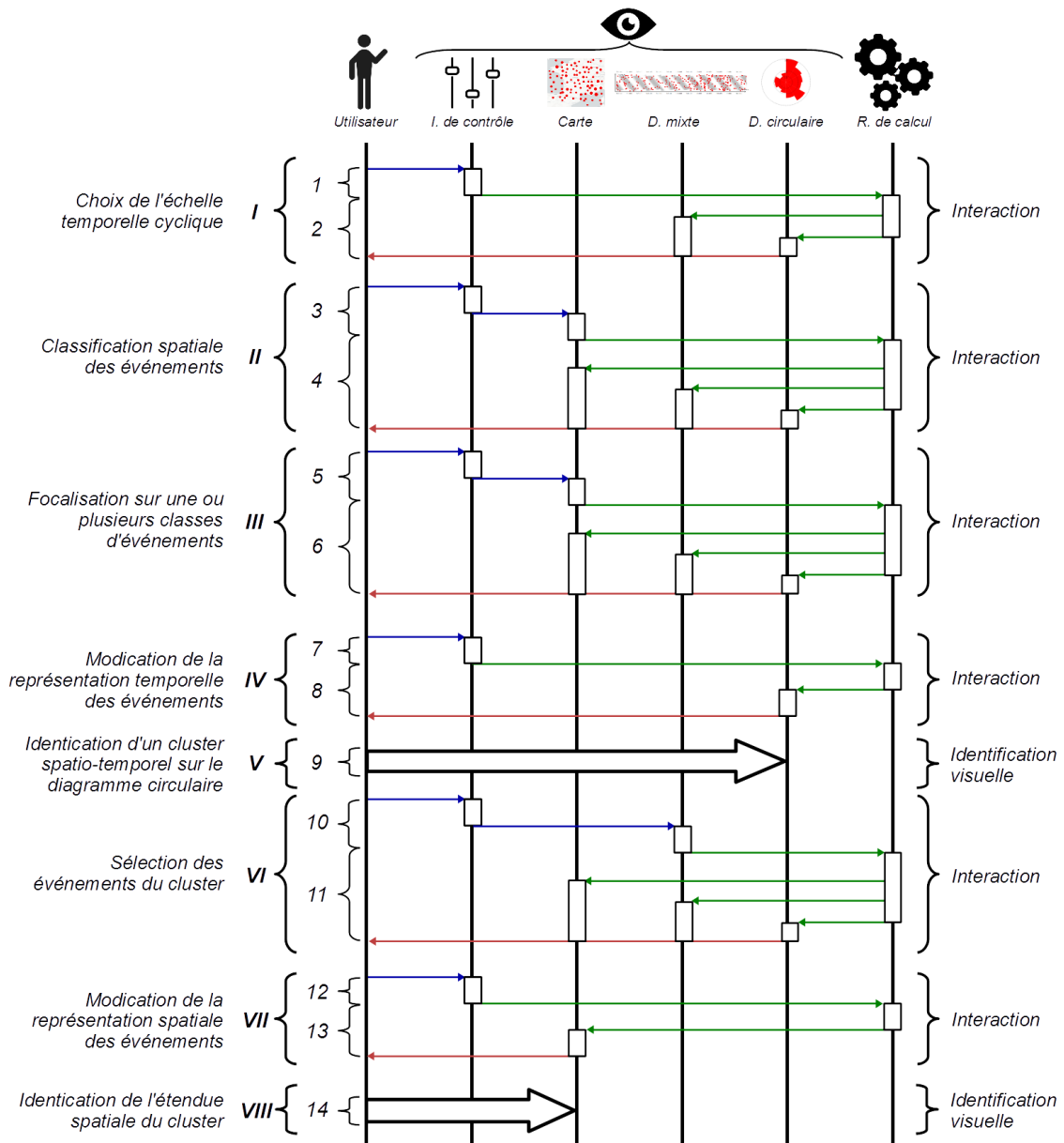


FIGURE 6.22 – Identification de clusters relatifs à une échelle cyclique, selon une entrée spatiale

6.3.4 Analyse d'une propagation spatiale d'événements au cours du temps, selon une entrée spatiale

Cette analyse consiste à identifier une propagation d'événements dans l'espace au cours du temps, au travers des plages de distances disposées le long d'un axe directionnel. Un exemple peut être l'analyse de la succession des apparitions de cas d'épidémie dans le temps, en fonction de leur localisation le long d'un axe de communication (route ou autre...) défini, et divisé en plusieurs tronçons. Le but de l'analyse est d'identifier si la localisation des cas d'épidémie le long de l'axe de communication correspond à leur ordre d'apparition dans le temps.

6.3.4.1 Procédé spécifique

L'identification d'une propagation spatiale, à travers une classification des événements selon leur lieu d'apparition, nécessite d'utiliser un découpage particulier de l'espace pour définir les zones de classification.

L'analyse étant effectuée au travers d'une approche spatiale, le point de départ de l'analyse est la classification des événements selon leur appartenance à des plages de distance, situées le long d'un axe directionnel, et partant d'un point d'origine. L'étape suivante consiste à étudier la distribution temporelle des événements apparaissant dans les différentes plages de distance.

Le découpage territorial utilisé doit donc refléter les différentes plages de distance classant les événements. Les zones de classification doivent être créées relativement à un axe et un point d'origine de référence (Figure 6.23).

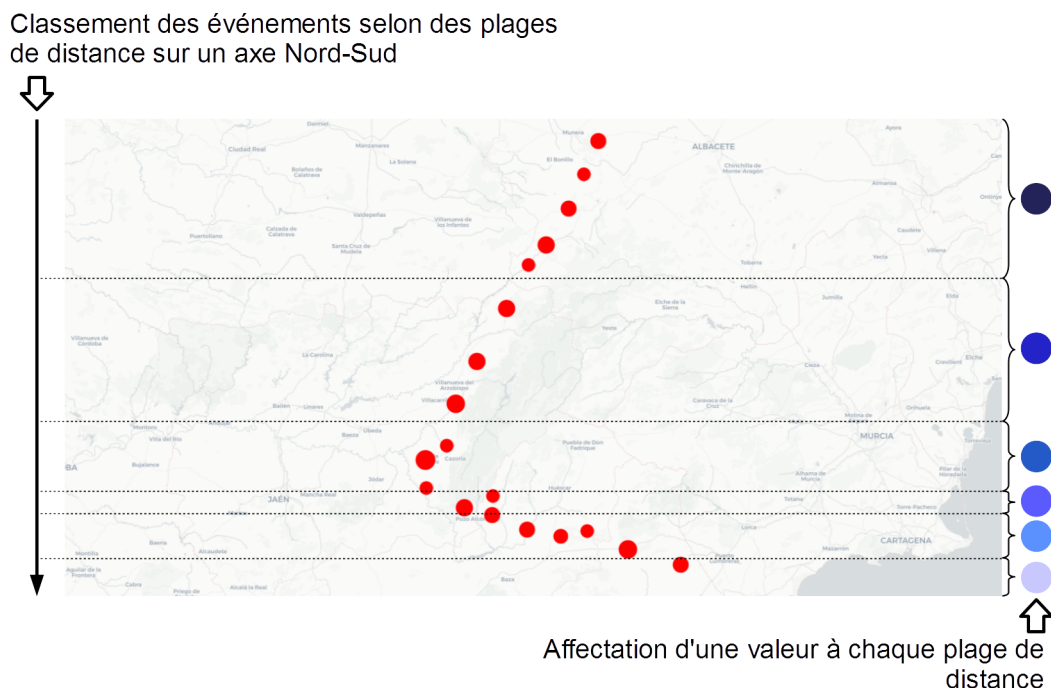


FIGURE 6.23 – Classification des événements selon des plages de distance sur l'axe nord-sud

La variable représentée au travers de cette classification n'est plus l'appartenance à une zone dans l'espace, mais l'appartenance à une plage de distance par rapport à un point de référence, que l'on peut rapprocher d'une variable quantitative. La variable visuelle affectée à la représentation des événements au travers de cette classification est donc celle de la valeur.

6.3.4.2 Scénario d'analyse

Cette analyse suit un scénario d'analyse en trois étapes, décrites dans la Figure 6.24, et explicitées dans l'encadré suivant.

Identification d'une propagation d'événements, selon une entrée spatiale

I. Classification spatiale des événements :

1. Définition de zones de classification spatiales au travers de la carte. Les zones spatiales sont créées selon un axe directionnel et un point d'origine. Une valeur est définie pour représenter chaque zone.
2. Classification des événements. Application de la valeur correspondante sur les entités graphiques dans la carte et les diagrammes temporels.

II. Modification de la représentation temporelle des événements :

3. Modification des critères de représentation des événements dans le diagramme temporel mixte (événements *vs* agrégats; rayon des agrégats; étendue temporelle des agrégats).
4. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans le diagramme temporel mixte.

III. Identification d'une propagation spatiale sur le diagramme mixte :

5. Identification d'un gradient de valeur, selon l'axe des abscisses, formé par les entités graphiques du diagramme (Figure ci-dessous).

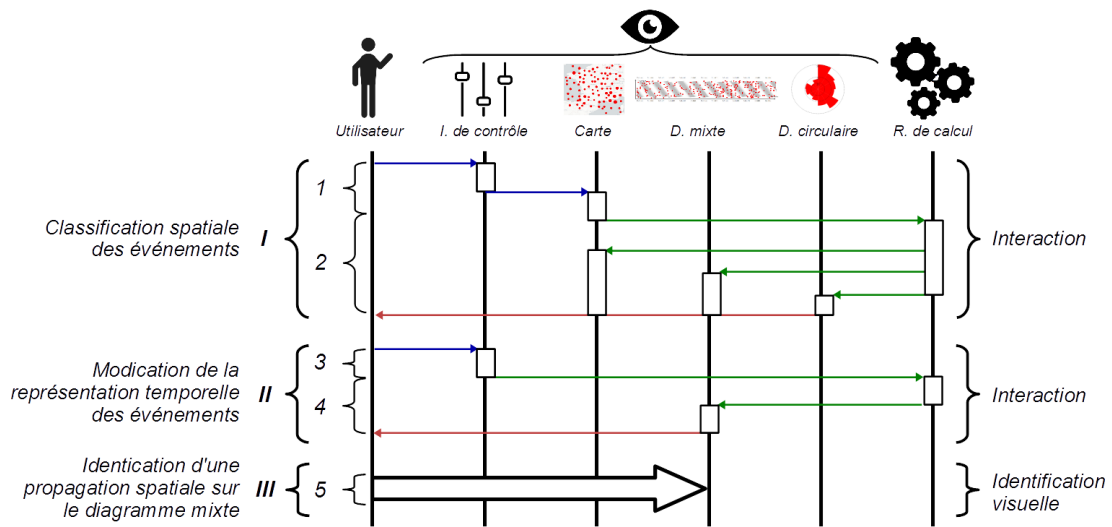


FIGURE 6.24 – Identification d'une propagation d'événements, selon une entrée spatiale

6.3.5 Analyse de récurrences cycliques spatialisées, selon une entrée spatiale

Cette analyse consiste à identifier des récurrences cycliques, localisées dans l'espace, au travers d'un découpage du territoire prédéfini. Un exemple peut être l'analyse de cycles dans l'apparition d'épisodes de fortes marées, en différents points définis d'une côte maritime. Le but de l'analyse est d'identifier si le phénomène d'apparition d'épisodes de fortes marées présente des cycles, si ces cycles sont spécifiques à certaines zones préalablement définies de la côte, ou si les épisodes de fortes marées réapparaissent selon des fréquences différentes dans ces différentes zones.

6.3.5.1 Procédé spécifique

L'identification d'un cycle au travers des diagrammes temporels, dans le cas d'une classification des événements selon leur lieu d'apparition, s'effectue selon le même procédé que celui décrit précédemment dans ce chapitre (Section 6.1.2).

Afin de guider l'utilisateur dans la recherche de cycle, l'environnement de géovisualisation effectue en arrière plan une transformée de Fourier sur les séries d'événements, afin de suggérer des pistes d'exploration à l'utilisateur. Dans le cas d'une classification des événements selon différentes zones spatiales, la transformée de Fourier peut être effectuée :

- de manière globale, sur l'ensemble de la série d'événements ;
- de manière individuelle, sur chaque série formée par les événements d'une même zone spatiale.

Dans ce second cas :

- Un signal temporel est créé pour chaque série formée par les événements d'une même zone.
- Une transformée de Fourier est effectuée pour chacun de ces signaux.
- Les fréquences les plus significatives pour chaque zone spatiale sont sélectionnées sur le spectre correspondant.
- Les périodes correspondantes sont projetées sur le curseur de modification de l'échelle cyclique sous la forme d'entités graphiques, prenant la couleur de la zone spatiale correspondante (Figure 6.25).
- L'utilisateur peut également explorer lui-même le spectre obtenu pour chaque zone.

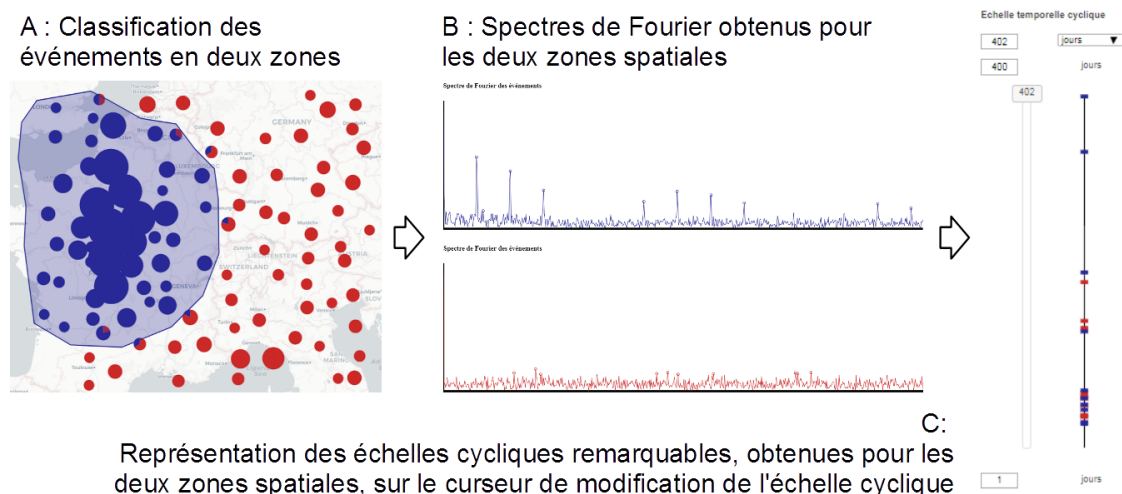


FIGURE 6.25 – Analyse de cycles par transformée de Fourier, selon des zones de classification spatiale

6.3.5.2 Scénario d'analyse

Cette analyse suit un scénario en dix étapes, décrites dans la Figure 6.26, et explicitées dans l'encadré suivant.

Identification d'une récurrence cyclique spatialisée, selon une entrée spatiale (*Partie 1*)

- I.** Classification spatiale des événements :
 1. Définition des zones de classification spatiales au travers de la carte, et choix des couleurs correspondantes.
 2. Classification des événements. Application de la couleur correspondante sur les entités graphiques dans la carte et les diagrammes temporels.
- II.** Analyse des événements par transformée de Fourier :
 3. Choix de l'utilisateur d'effectuer une analyse globale sur la totalité des événements, ou d'effectuer une analyse par classe d'événements.
 4. Analyse des événements par transformée de Fourier, et identification de potentiels comportements cycliques dans la série événementielle analysée. Suggestion des périodes correspondantes dans l'outil de modification de l'échelle temporelle cyclique.
- III.** Focalisation sur une ou plusieurs classes d'événements :
 5. Définition des zones spatiales à masquer au travers de la carte.
 6. Identification des entités graphiques à masquer dans la carte et les diagrammes temporels.
- IV.** Modification de la représentation temporelle des événements :
 7. Modification des critères de représentation des événements dans le diagramme temporel circulaire (Modification du nombre de sections du diagramme temporel circulaire).
 8. Calcul de la nouvelle représentation graphique. Modification du diagramme temporel circulaire.
 9. Modification des critères de représentation des événements dans le diagramme temporel mixte (événements *vs* agrégats; rayon des agrégats; étendue temporelle des agrégats).
 10. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans le diagramme temporel mixte.
- V.** Modification graduelle de l'échelle temporelle cyclique :
 11. Modification de la durée de l'échelle temporelle cyclique dans l'interface de contrôle.
 12. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans les diagrammes temporels.

Lors de la séquence V., l'échelle temporelle cyclique est modifiée jusqu'à l'identification :

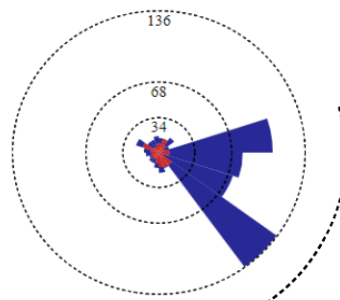
- d'un cluster spatio-temporel sur le diagramme circulaire (séquence VI) ;
- d'une récurrence cyclique spatialisée sur le diagramme temporel mixte (séquence VII).

L'échelle temporelle cyclique peut être modifiée graduellement :

- au voisinage des périodes suggérées par *GrAPHiST* à partir des résultats d'une transformée de Fourier sur les séries événements ;
- au voisinage d'une période identifiée par l'utilisateur lui-même sur le spectre obtenu par transformée de Fourier ;
- au voisinage d'une période correspondant à un harmonique des fréquences significatives du spectre obtenu ;
- au voisinage d'une période suspectée par l'utilisateur de correspondre à un cycle.

- VI.** Identification d'un cluster spatio-temporel sur le diagramme circulaire :

13. Identification d'une section angulaire du diagramme circulaire possédant une couleur majoritaire (Figure ci-dessous).

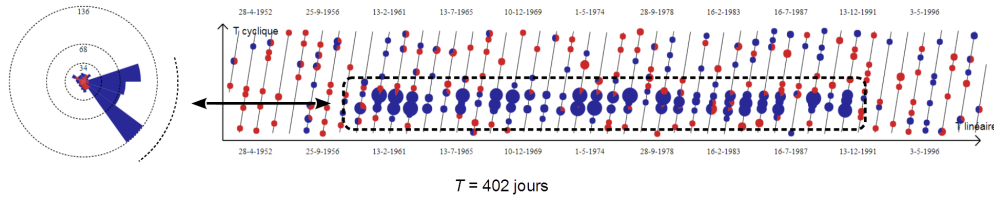


$T = 402$ jours

Identification d'une récurrence cyclique spatialisée, selon une entrée spatiale (*Partie 2*)

VII. Identification d'une récurrence cyclique spatialisée sur le diagramme temporel mixte :

- 14. Identification d'un alignement graphique sur le diagramme temporel mixte, formé par des entités de même couleur (Figure ci-dessous). Si un cluster spatio-temporel a été identifié (séquence VI), les événements formant l'alignement identifié ici doivent correspondre au cluster identifié précédemment (Figure de la page précédente).



VIII. Sélection des événements cycliques (Figure ci-dessous) :

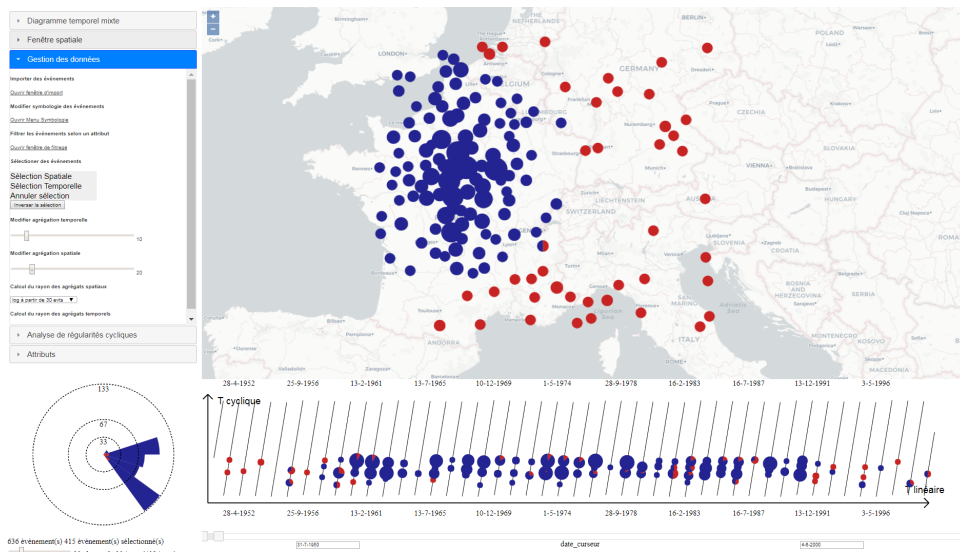
- 15. Sélection graphique, dans le diagramme temporel mixte, des événements identifiés dans la séquence VII.
- 16. Identification des événements à sélectionner. Mise en valeur des entités graphiques correspondantes dans la carte et les diagrammes temporels.

IX. Modification de la représentation spatiale des événements :

- 17. Modification des critères de représentation des événements dans la carte (événements vs agrégats ; rayon des agrégats ; étendue spatiale des agrégats).
- 18. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans la carte.

X. Identification de l'étendue spatiale de la récurrence cyclique spatialisée :

- 19. Identification, dans la carte, de l'étendue spatiale couverte par les événements sélectionnés (Figure ci-dessous).



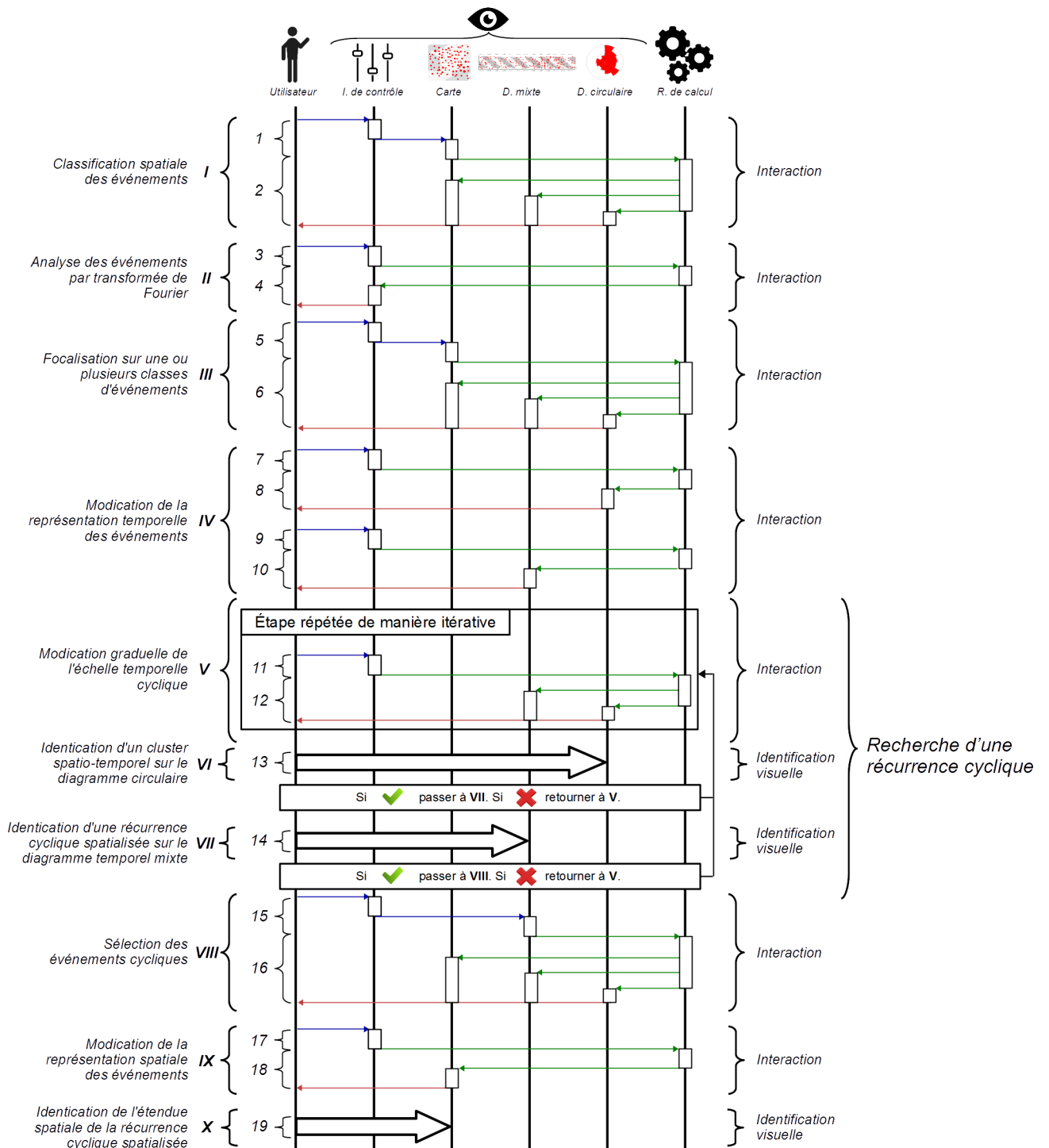


FIGURE 6.26 – Identification d'une récurrence cyclique spatialisée, selon une entrée spatiale

6.4 Analyse de patterns spatio-temporels, sans hypothèse *a priori* sur la distribution spatiale des événements

6.4.1 Procédé général

6.4.1.1 Classification temporelle et analyse au travers de la carte

À l'opposée de l'approche illustrée par *Time Wave* [Li, 2010] et *PerSE* [Swedberg and Peuquet, 2017], et à l'instar de celle illustrée par *Growth Ring Map* [Bak et al., 2009, Andrienko et al., 2011], l'analyse consiste ici à classer et à représenter les événements en fonction de leur moment d'apparition, puis à identifier directement dans la carte des structures dans la distribution spatio-temporelle des événements (Figure 6.27).

Cette approche utilise une représentation des événements en fonction de leur moment d'apparition relatif au temps linéaire, ou relatif à une échelle temporelle cyclique. Les événements sont représentés dans l'interface au moyen d'une variable visuelle : la valeur, dans le cas du temps linéaire, ou la couleur, dans le cas du temps cyclique.

6.4.1.2 L'utilisation des outils de sélection et de *brushing* pour affiner l'analyse

Dans cette approche, les outils de sélection forment également un outil supplémentaire pour l'analyse exploratoire, permettant d'améliorer la définition de l'étendue spatio-temporelle du pattern recherché.

Une fois qu'une structure est identifiée dans la carte, une sélection graphique des événements correspondants, au travers de la carte, permet de mettre en valeur leur composante temporelle dans les diagrammes. Il est alors possible de définir plus précisément l'étendue temporelle de la structure spatio-temporelle identifiée, au travers des diagrammes temporels.

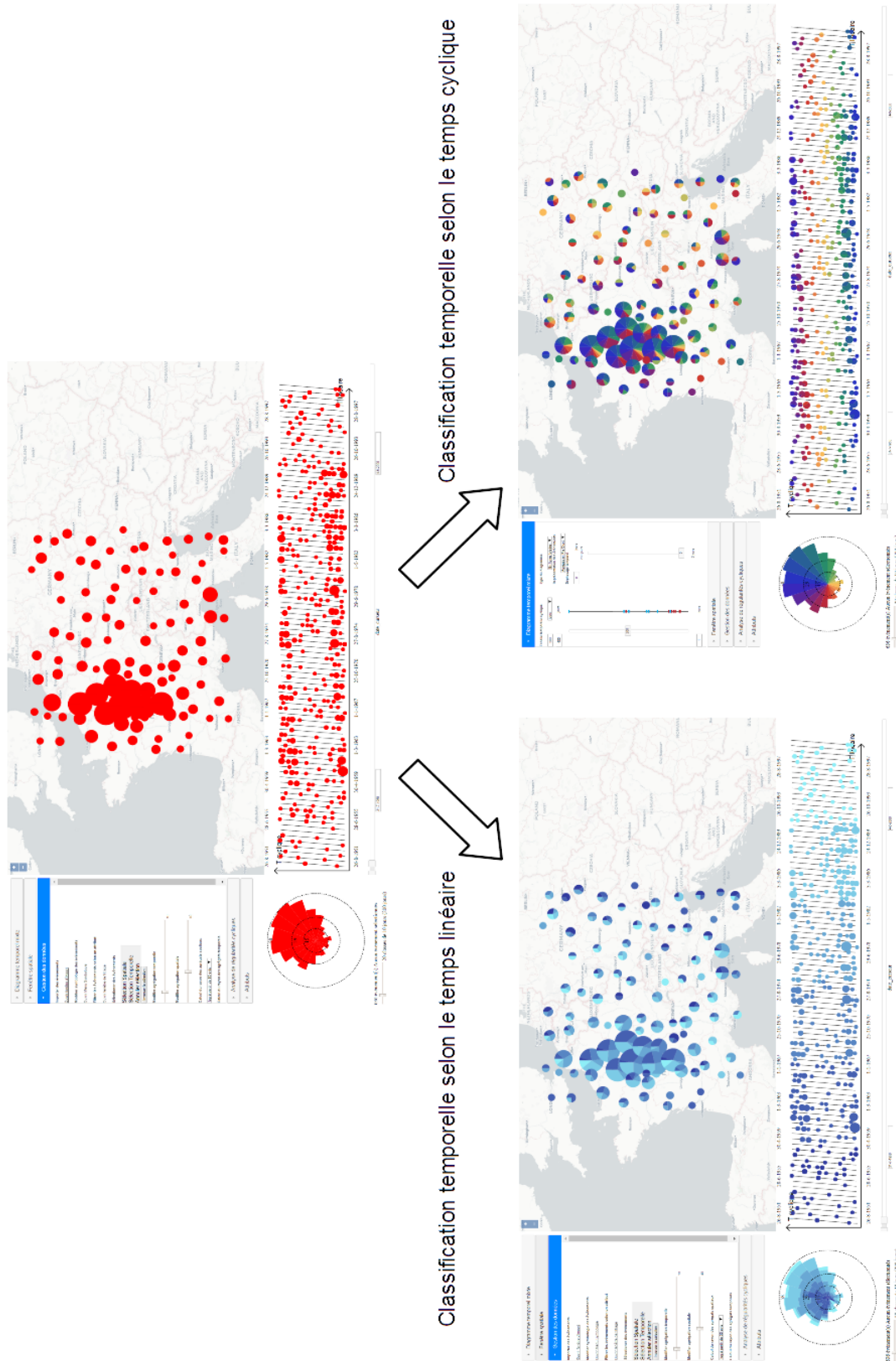


FIGURE 6.27 – Classification des événements selon leur moment d'apparition, relatif au temps linéaire ou relatif à une échelle cyclique

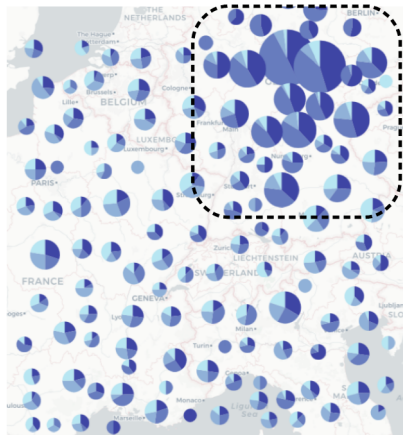
6.4.2 Analyse de clusters relatifs au temps linéaire, selon une entrée temporelle

Cette analyse consiste à identifier des concentrations d'événements, dans l'espace et dans le temps linéaire, au travers d'un découpage de l'étendue temporelle étudiée en différents intervalles temporels réguliers. Un exemple peut être l'analyse des accidents routiers sur une période de dix ans, en classant les événements selon des intervalles temporels d'une année. Le but de l'analyse est d'identifier si les accidents apparaissant durant les mêmes années se concentrent dans certaines zones de l'espace.

Cette analyse suit un scénario en six étapes, décrites dans la Figure 6.28, et explicitées dans l'encadré suivant.

Identification de clusters relatifs au temps linéaire, selon une entrée temporelle

- I.** Classification des événements selon le temps linéaire :
 1. Définition du nombre d'intervalles temporels de classification, et choix des couleurs utilisées.
 2. Classification des événements. Application de la valeur correspondante sur les entités graphiques dans la carte et les diagrammes temporels.
- II.** Modification de la représentation spatiale des événements :
 3. Modification des critères de représentation des événements dans la carte (événements *vs* agrégats ; rayon des agrégats ; étendue spatiale des agrégats).
 4. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans la carte.
- III.** Identification d'un cluster spatio-temporel sur la carte :
 5. Identification d'une concentration, sur la carte, d'entités ponctuelles de même couleur, ou d'entités *Pie Chart* présentant une valeur majoritaire ^a (Figure ci-dessous).



- IV.** Sélection des événements du cluster :
 6. Sélection graphique, dans la carte, des événements identifiés dans la séquence *III*.
 7. Identification des événements à sélectionner. Mise en valeur des entités graphiques correspondantes dans la carte et les diagrammes temporels.
- V.** Modification de la représentation temporelle des événements :
 8. Modification des critères de représentation des événements dans le diagramme temporel mixte (événements *vs* agrégats ; rayon des agrégats ; étendue temporelle des agrégats).
 9. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans le diagramme temporel mixte.
- VI.** Identification de l'étendue temporelle du cluster :
 10. Identification, dans le diagramme temporel mixte, de l'étendue temporelle linéaire couverte par les événements sélectionnés.

^a. Les figures présentées ont valeur d'exemple. La taille des *Pie Chart* dans la zone encadrée est un indice de la présence d'un pattern dans cette zone. Cependant, la carte a pour objectif de montrer un exemple de *Pie Chart* prenant une valeur majoritaire.

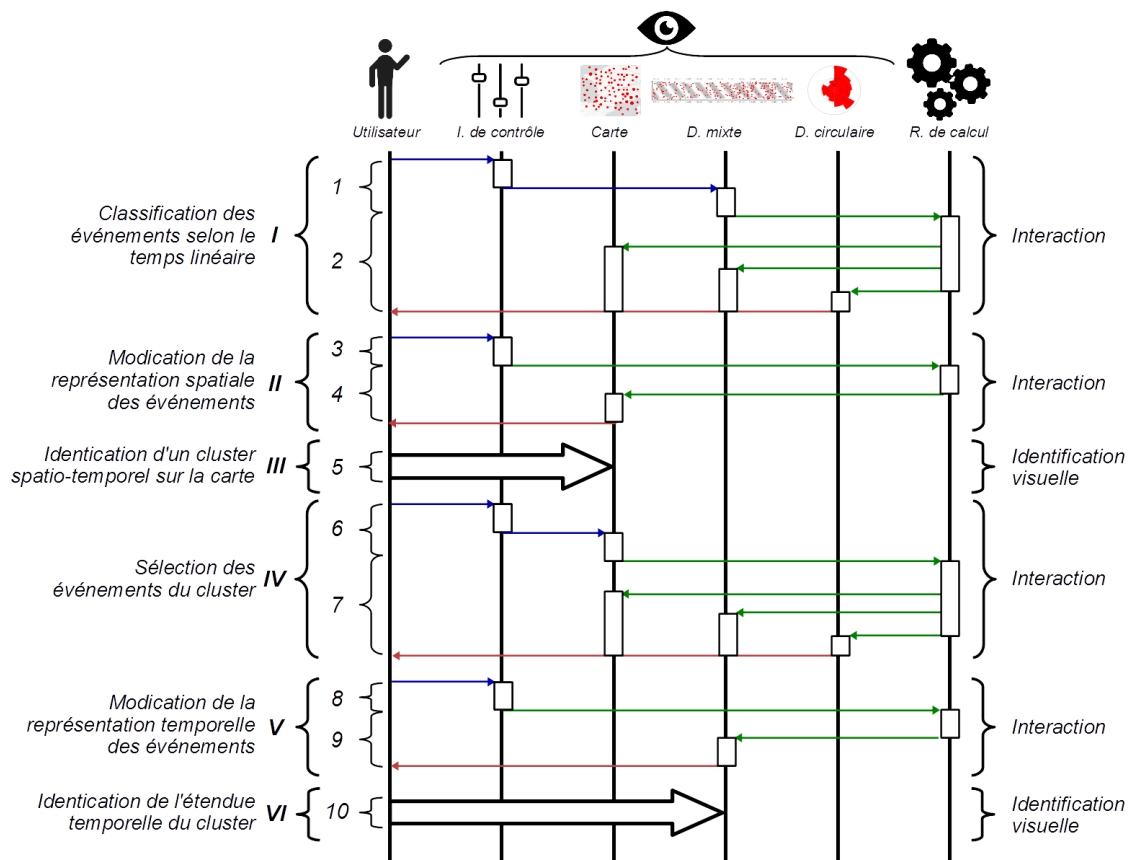


FIGURE 6.28 – Identification de clusters relatifs au temps linéaire, selon une entrée temporelle

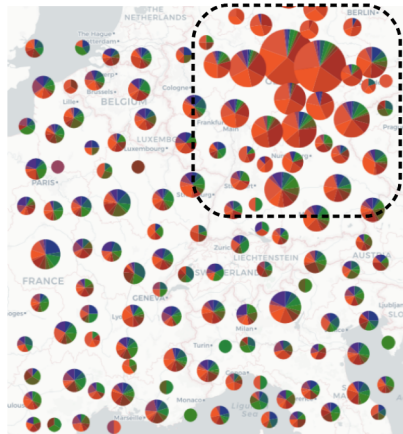
6.4.3 Analyse de clusters relatifs à une échelle temporelle cyclique, selon une entrée temporelle

Cette analyse consiste à identifier des concentrations d'événements, dans l'espace et dans certaines plages de temps relatives à une échelle cyclique, au travers d'une division de l'échelle cyclique choisie en différents intervalles temporels réguliers. Un exemple peut être l'analyse de l'apparition d'accidents routiers au travers d'une échelle cyclique d'une année divisée en douze mois. Le but de l'analyse est d'identifier si les accidents apparaissant durant les mêmes mois de l'année se concentrent dans certaines zones de l'espace.

Cette analyse suit un scénario en sept étapes, décrites dans la Figure 6.29, et explicitées dans l'encadré suivant.

Identification de clusters relatifs à une échelle cyclique, selon une entrée temporelle

- I.** Choix de l'échelle temporelle cyclique :
 1. Modification de la durée de l'échelle temporelle cyclique dans l'interface de contrôle.
 2. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans les diagrammes temporels.
- II.** Classification des événements selon le temps cyclique :
 3. Définition du nombre d'intervalles temporels de classification, du nombre de couleurs utilisées, et choix des couleurs.
 4. Classification des événements. Application de la couleur correspondante sur les entités graphiques dans la carte et les diagrammes temporels.
- III.** Modification de la représentation spatiale des événements :
 5. Modification des critères de représentation des événements dans la carte (événements *vs* agrégats ; rayon des agrégats ; étendue spatiale des agrégats).
 6. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans la carte.
- IV.** Identification d'un cluster spatio-temporel sur la carte :
 7. Identification d'une concentration, sur la carte, d'entités ponctuelles de même couleur, ou d'entités *Pie Chart* présentant une couleur majoritaire ^a (Figure ci-dessous).



- V.** Sélection des événements du cluster :
 8. Sélection graphique, dans la carte, des événements identifiés dans la séquence IV.
 9. Identification des événements à sélectionner. Mise en valeur des entités graphiques correspondantes dans la carte et les diagrammes temporels.
- VI.** Modification de la représentation temporelle des événements :
 10. Sélection graphique, dans la carte, des événements identifiés dans la séquence IV.
 11. Modification des critères de représentation des événements dans le diagramme temporel circulaire (Modification du nombre de sections du diagramme temporel circulaire).
- VII.** Identification de l'étendue temporelle du cluster :
 12. Identification, dans le diagramme temporel circulaire, de l'étendue temporelle cyclique couverte par les événements sélectionnés.

^a. Les figures présentées ont valeur d'exemple. La carte a pour objectif de montrer un exemple de *Pie Chart* prenant une couleur majoritaire.

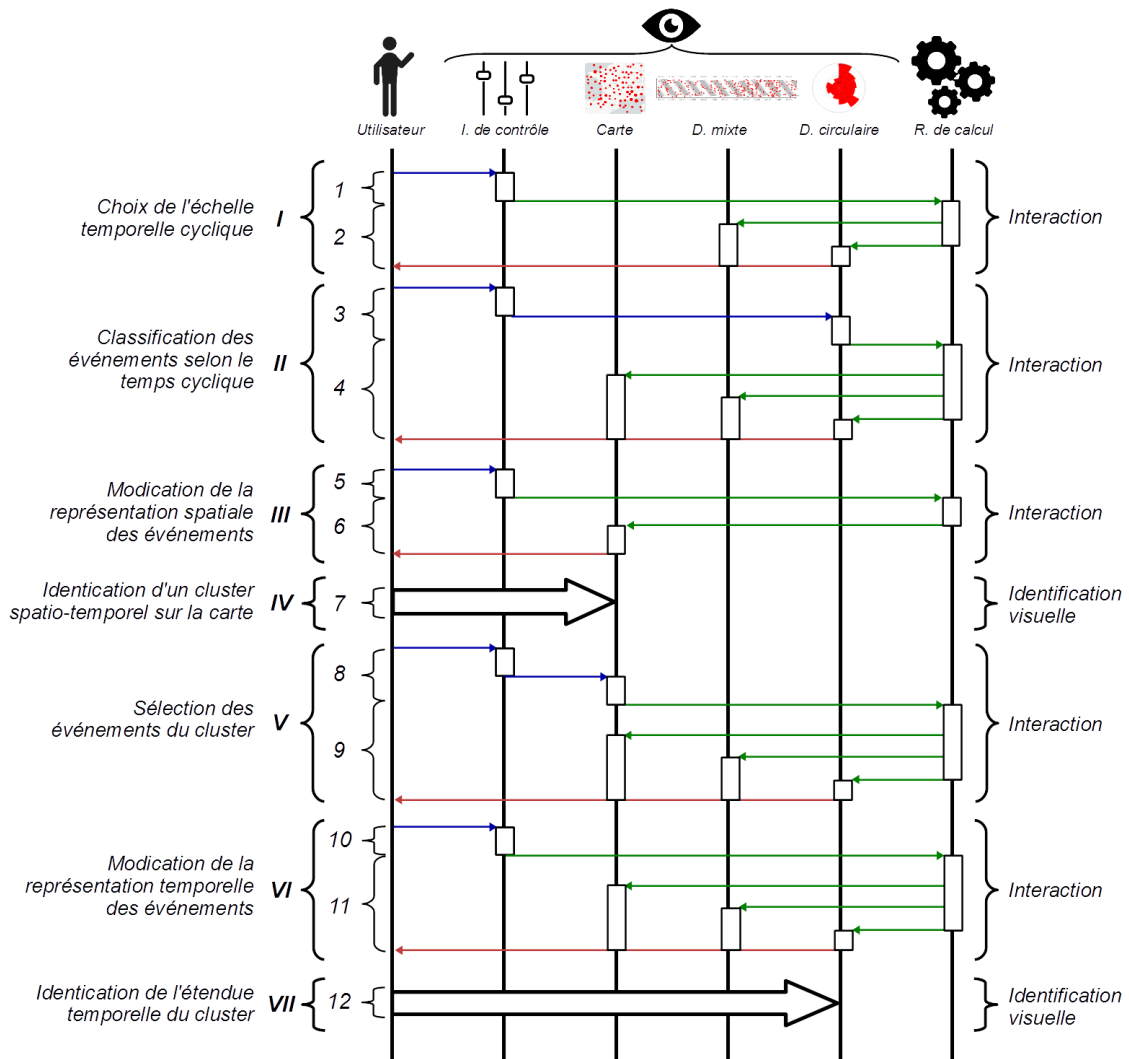


FIGURE 6.29 – Identification de clusters relatifs à une échelle cyclique, selon une entrée temporelle

6.4.4 Analyse d'une propagation spatiale d'événements au cours du temps, selon une entrée temporelle

Cette analyse consiste à identifier une propagation d'événements dans l'espace au cours du temps, au travers d'un découpage de l'étendue temporelle étudiée en différents intervalles temporels réguliers. Un exemple peut être l'analyse de la localisation de cas d'épidémie, apparaissant durant des intervalles temporels successifs d'une durée d'un mois. Le but de l'analyse est d'identifier si l'ordre d'apparition des cas d'épidémie dans le temps, correspond à leur localisation le long d'un axe directionnel dans l'espace.

Cette analyse suit un scénario en sept étapes, décrites dans la Figure 6.30, et explicitées dans l'encadré suivant.

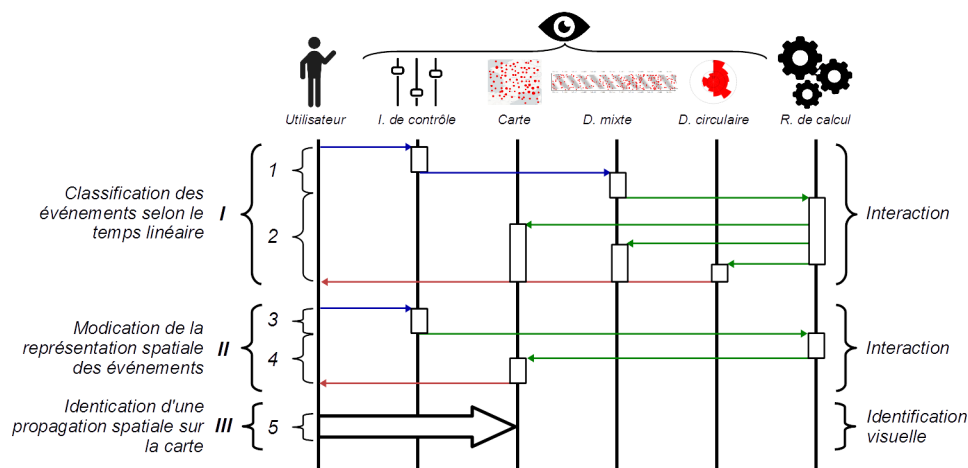
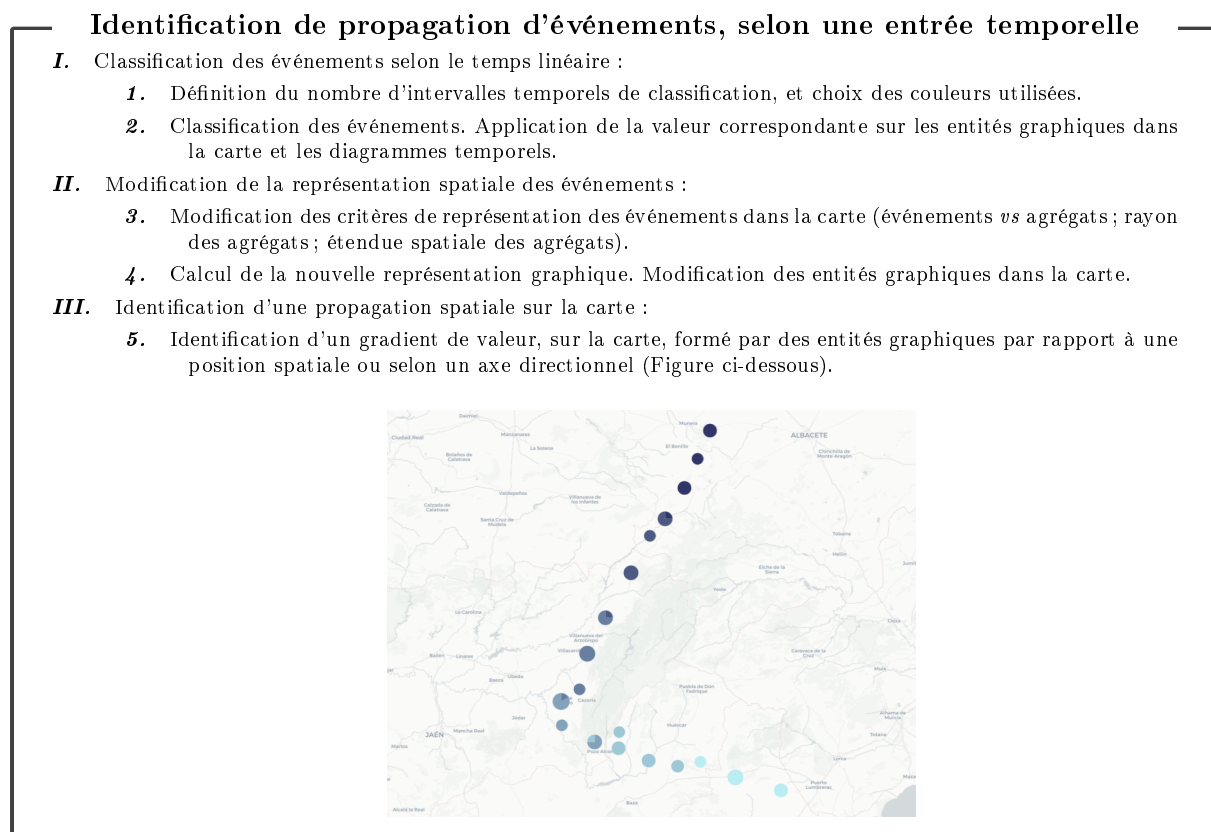


FIGURE 6.30 – Identification de propagation d'événements, selon une entrée temporelle

6.4.5 Analyse de récurrences cycliques spatialisées, selon une entrée temporelle

Cette analyse consiste à identifier des récurrences cycliques localisées dans l'espace. Un exemple peut être l'analyse de cycles dans l'apparition d'épisodes de fortes pluies, sans que les données soient rattachées à des points définis dans l'espace. Le but de l'analyse est d'identifier si le phénomène d'apparition d'épisodes de fortes pluies présente des cycles, si ces cycles sont spécifiques à certaines zones de l'espace, ou si les épisodes de fortes pluies réapparaissent selon des fréquences différentes dans différentes zones de l'espace.

L'analyse s'effectue directement sur la carte en utilisant une représentation des événements selon leur moment d'apparition relatif à une échelle cyclique. Pour classer et représenter les événements selon leur lieu d'apparition, l'échelle cyclique est divisée en intervalles temporels réguliers. Cependant, le nombre d'intervalles temporels utilisés pour classer les événements n'est pas fixe, et peut être modifié en cours d'analyse.

6.4.5.1 Procédé spécifique

Regroupement des événements proches dans l'espace et suggestion de pistes d'exploration. Afin de guider l'utilisateur dans la recherche de cycle, l'environnement de géovisualisation effectue en arrière-plan une transformée de Fourier sur les séries d'événements, afin de suggérer des pistes d'exploration à l'utilisateur. Dans le cas d'une classification des événements selon leur moment d'apparition, la transformée de Fourier peut être effectuée :

- de manière globale, sur l'ensemble de la série d'événements ;
- de manière individuelle, sur des séries formées par des groupes d'événements proches spatialement.

Dans ce second cas :

- Lors du déclenchement de l'analyse de Fourier, les événements sont regroupés selon leur proximité spatiale en utilisant une valeur seuil, modifiable par l'utilisateur. Le procédé est le même que celui utilisé pour créer les agrégats d'événements dans la représentation spatiale (Section 5.4.2). L'algorithme utilisé est celui de la librairie *JavaScript OpenLayers 3*. Cependant, la distance critique utilisée est plus importante.
- Une couleur est affectée, de manière automatique, à chaque regroupement spatial. L'enveloppe convexe de chaque regroupement est calculée et est représentée sur la carte (elle peut être masquée par la suite).
- Les regroupements d'événements créés forment des séries événementielles.
- Un signal temporel est calculé pour chaque série événementielle créée.
- Une transformée de Fourier est effectuée pour chacun de ces signaux.
- Les fréquences les plus significatives, pour chaque regroupement d'événements, sont sélectionnées sur le spectre correspondant.
- Les périodes correspondantes sont projetées sur le curseur de modification de l'échelle cyclique sous la forme d'entités graphiques, prenant la couleur affectée au regroupement d'événements correspondant.
- L'utilisateur peut également explorer lui-même le spectre obtenu pour chaque regroupement d'événements.

Recherche d'une récurrence cyclique spatialisée à l'aide de la carte. À l'inverse du procédé d'identification d'une récurrence cyclique sur un plan strictement temporel, l'identification de l'échelle temporelle cyclique, pour laquelle peut être observé des récurrences cycliques spatialisées dans les séries d'événements, est effectuée directement dans la fenêtre spatiale.

La recherche d'une récurrence spatialisée s'effectue en deux étapes :

- Dans un premier temps, l'utilisateur recherche, au travers de la carte, une échelle cyclique pour laquelle on observe la formation d'un cluster spatio-temporel, relatif au temps cyclique.
- Dans un second temps, les événements correspondant à ce cluster sont sélectionnés au travers de la carte, et mis en valeur dans les diagrammes temporels. L'observation d'un alignement graphique, formé par ces événements, dans le diagramme mixte permet de conclure à l'existence d'une récurrence cyclique spatialisée.

La formation d'un cluster dans la carte, au fur et à mesure que la durée de l'échelle cyclique s'approche de la période d'un cycle, est identifiable par la prise progressive, par un ensemble d'entités graphiques ou par des entités *Pie Chart*, d'une couleur majoritaire (Figure 6.31).

Le moment d'apparition, relatif à l'échelle cyclique, de l'ensemble des événements évolue lorsque la durée de l'échelle cyclique est modifiée (Section 6.1.2.3). Pour cette raison, les sections du diagramme circulaire, correspondant aux événements formant le cluster recherché, ne restent pas fixées dans la même zone du diagramme circulaire lors d'une modification de la durée de l'échelle cyclique (Figure 6.31). La couleur utilisée pour représenter les événements selon leur moment d'apparition, étant déterminée par les sections du diagramme circulaire, la couleur affectée aux événements formant le cluster dans la carte sera modifiée lors d'un changement de la durée de l'échelle cyclique.

Ainsi, si un ensemble d'entités graphiques, ou des entités *Pie Chart*, prennent une couleur majoritaire lorsque la durée de l'échelle cyclique approche la période d'un cycle, cette couleur majoritaire ne reste pas la même durant la modification graduelle de la durée de l'échelle cyclique (Figure 6.31).

La formation d'un cluster relatif au temps cyclique est ainsi visible dans la carte par une homogénéisation progressive de la couleur prise par un groupe d'entités graphiques, ou par l'accroissement de la portion d'entités *Pie Chart* occupée par une même couleur.

Lorsqu'un cluster spatio-temporel est identifié dans la carte, une sélection (pouvant être suivie d'un *brushing*) des entités graphiques permet de mettre en valeur la composante temporelle des événements correspondant dans les diagrammes temporels :

- Le diagramme temporel circulaire permet alors de définir l'intervalle temporel cyclique couvert par les événements du cluster.
- Le diagramme temporel mixte permet d'observer si ces événements forment un alignement parallèle à l'axe des abscisses. L'observation d'un tel alignement permet de conclure à l'existence d'une récurrence cyclique spatialisée.

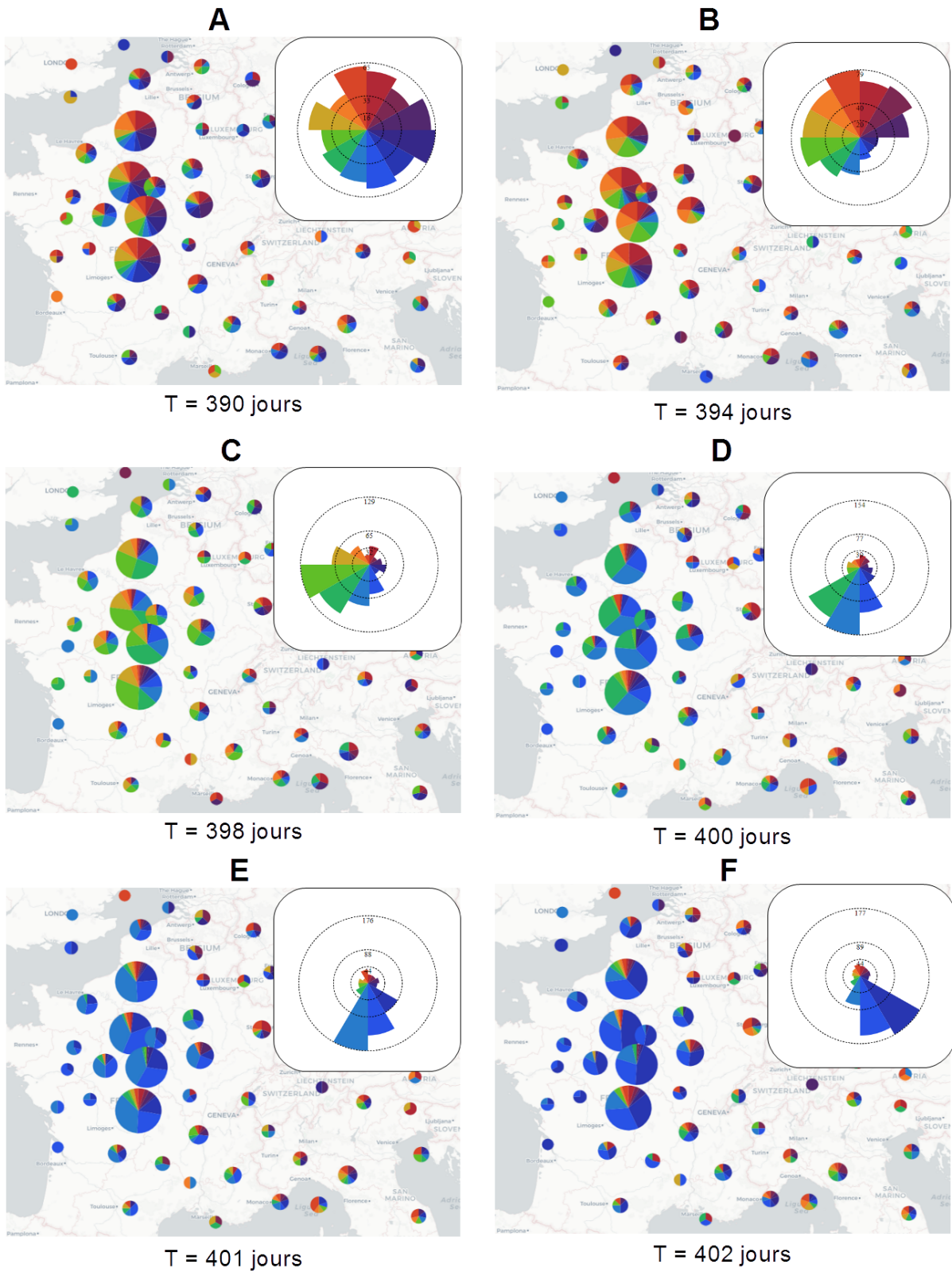


FIGURE 6.31 – Identification de la formation d'un cluster spatio-temporel dans la carte dans le cadre de la recherche d'une récurrence cyclisée spatialisée

6.4.5.2 Scénario d'analyse

Cette analyse suit un scénario en huit étapes, décrites dans la Figure 6.32, et explicitées dans l'encadré suivant.

Identification de récurrences cycliques spatialisées, selon une entrée temporelle (*Partie 1*)

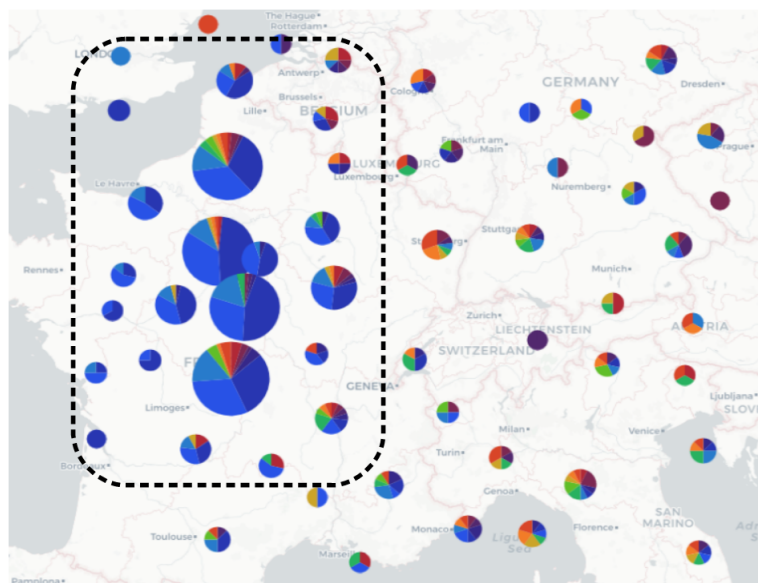
- I.** Classification des événements selon le temps cyclique :
 1. Définition du nombre d'intervalles temporels de classification, du nombre de couleurs utilisées, et choix des couleurs.
 2. Classification des événements. Application de la couleur correspondante sur les entités graphiques dans la carte et les diagrammes temporels.
- II.** Analyse des événements par transformée de Fourier :
 3. Choix de l'utilisateur d'effectuer une analyse globale sur la totalité des événements, ou d'effectuer une analyse par sous-ensemble d'événements groupés selon leur proximité spatiale.
 4. Analyse des événements par transformée de Fourier, et identification de potentiels comportements cycliques dans la série événementielle analysée. Suggestion des périodes correspondantes dans l'outil de modification de l'échelle temporelle cyclique.
- III.** Modification de la représentation spatiale des événements :
 5. Modification des critères de représentation des événements dans la carte (événements *vs* agrégats ; rayon des agrégats ; étendue spatiale des agrégats).
 6. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans la carte.
- IV.** Modification graduelle de l'échelle temporelle cyclique :
 7. Modification de la durée de l'échelle temporelle cyclique dans l'interface de contrôle.
 8. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans les diagrammes temporels, et dans la carte.

La séquence *IV.* est effectuée en boucle jusqu'à l'identification d'un cluster spatio-temporel dans la carte (séquence *V.*).

L'échelle temporelle cyclique peut être modifiée graduellement :

- au voisinage des périodes suggérées par *GrAPHiST* à partir des résultats d'une transformée de Fourier sur les séries événements ;
- au voisinage d'une période identifiée par l'utilisateur lui-même sur le spectre obtenu par transformée de Fourier ;
- au voisinage d'une période correspondant à un harmonique des fréquences significatives du spectre obtenu ;
- au voisinage d'une période suspectée par l'utilisateur de correspondre à un cycle.

- V.** Identification d'un cluster spatio-temporel sur la carte :
 9. Identification d'une concentration, sur la carte, d'entités ponctuelles de même couleur, ou prenant une couleur majoritaire (Figure ci-dessous).

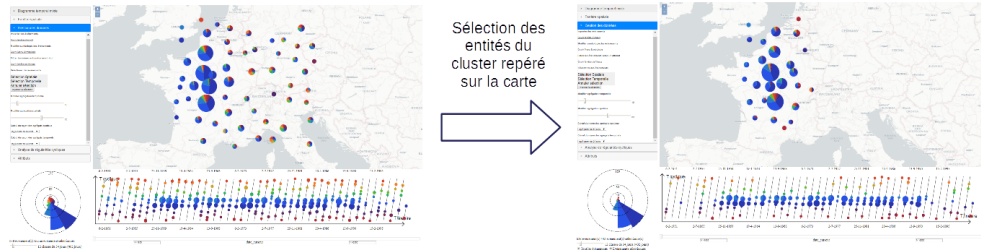


T = 402 jours

Identification de récurrences cycliques spatialisées, selon une entrée temporelle (*Partie 2*)

VI. Sélection des événements du cluster (Figure ci-dessous) :

- 10. Sélection graphique, dans la carte, des événements identifiés dans la séquence V.
- 11. Identification des événements à sélectionner. Mise en valeur des entités graphiques correspondantes dans la carte et les diagrammes temporels.

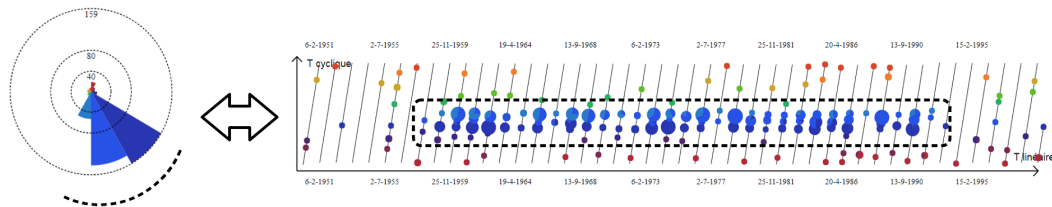


VII. Modification de la représentation temporelle des événements :

- 12. Modification des critères de représentation des événements dans le diagramme temporel circulaire (Modification du nombre de sections du diagramme temporel circulaire).
- 13. Calcul de la nouvelle représentation graphique. Modification du diagramme temporel circulaire et des entités graphiques dans la carte et le diagramme temporel mixte.
- 14. Modification des critères de représentation des événements dans le diagramme temporel mixte (événements *vs* agrégats; rayon des agrégats; étendue temporelle des agrégats).
- 15. Calcul de la nouvelle représentation graphique. Modification des entités graphiques dans le diagramme temporel mixte.

VIII. Identification d'une récurrence cyclique spatialisée sur le diagramme temporel mixte :

- 16. Identification d'un cluster temporel cyclique dans le diagramme temporel circulaire. Identification d'une section angulaire possédant un rayon plus important que les autres sections du diagramme (Figure ci-dessous).
- 17. Identification d'un alignement graphique sur le diagramme temporel mixte, formé par les entités du cluster identifié dans la sous-séquence 16 (Figure ci-dessous).



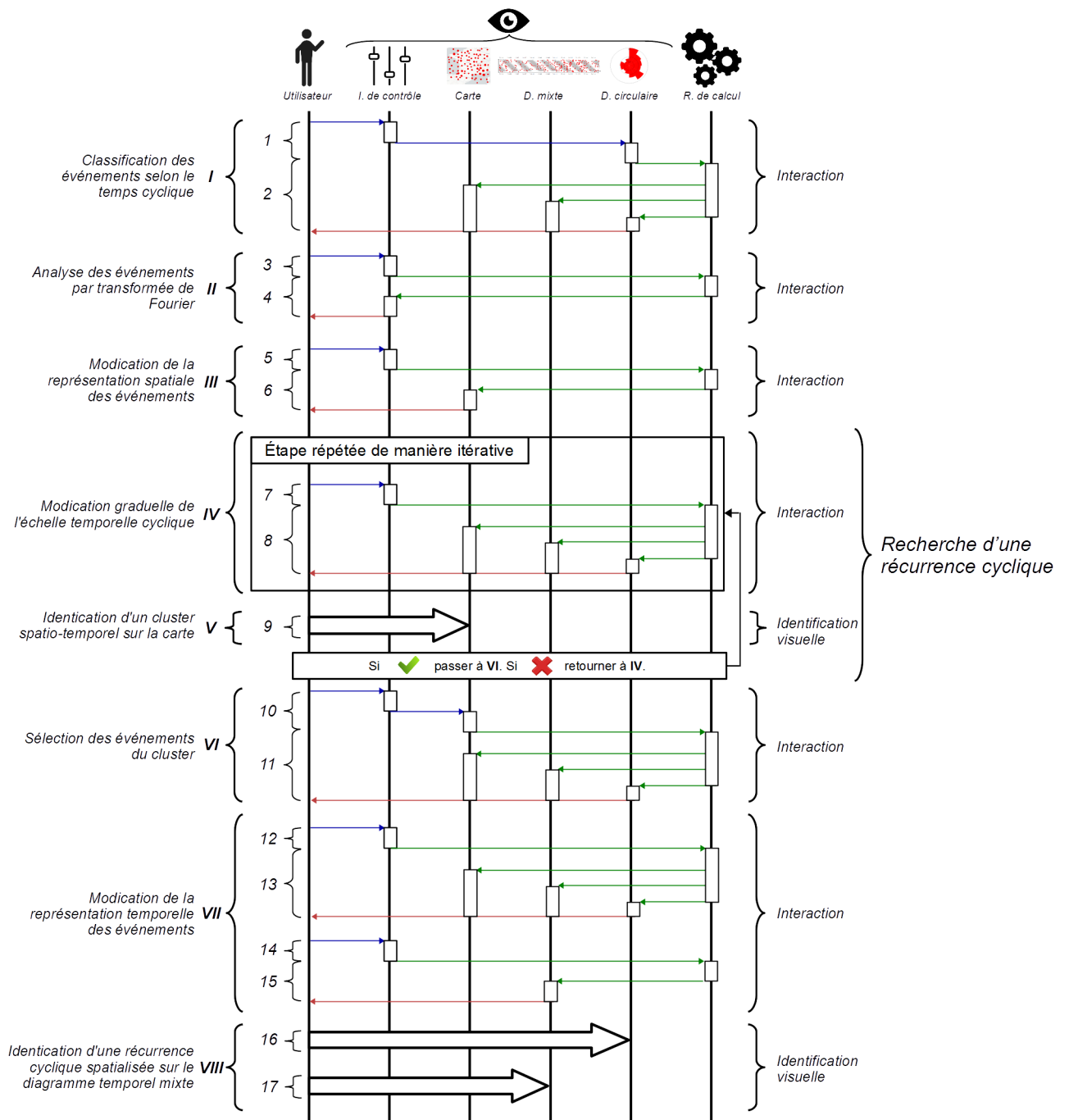


FIGURE 6.32 – Identification de récurrences cycliques spatialisées, selon une entrée temporelle

6.5 Synthèse

Ces scénarios d'analyse permettent de décrire les différentes étapes devant être effectuées par l'utilisateur, afin d'identifier les différents types de patterns auxquels nous nous intéressons, selon que l'analyse soit effectuée au travers d'un découpage prédéfini de l'espace, ou sans émettre d'hypothèse *a priori* sur la distribution spatiale des événements.

Conformément au principe de l'analyse exploratoire [Tukey, 1977, Andrienko and Andrienko, 2006], l'utilisateur occupe une place centrale dans chacun des scénarios :

- En fonction du phénomène étudié et des dynamiques que l'utilisateur souhaite analyser, l'utilisateur définit quel pattern il doit identifier et par quelle entrée (spatiale ou temporelle) il doit effectuer son analyse. Ces choix définissent le scénario d'analyse que l'utilisateur doit poursuivre.
- À chaque étape d'un scénario, l'utilisateur interprète, au travers des différentes composantes de l'interface de géovisualisation, sur quels objets d'intérêt l'analyse doit être approfondie. Il interagit ensuite avec l'environnement de géovisualisation, dans le but de mettre en évidence d'autres informations.
- En fonction du retour visuel de ces actions, l'utilisateur poursuit l'analyse exploratoire présentée dans le scénario.

L'analyse d'un jeu de données inconnues nécessite une phase d'exploration libre de la part de l'utilisateur, pour que celui-ci devienne *familier* avec ce jeu de données [Andrienko and Andrienko, 2006]. Cependant, une fois terminée cette phase de prise de connaissance de la donnée, l'utilisateur peut définir les dynamiques spatio-temporelles sur lesquelles doit porter son analyse, et par quelle entrée il peut l'effectuer.

La formalisation de l'utilisation de *GrAPHiST*, en différents scénarios d'analyse exploratoire, permet alors de définir comment l'utilisateur doit effectuer la suite de son analyse, selon les différents cas d'étude possibles. Cette formalisation permet de définir comment l'utilisateur doit interagir avec les différents composants de l'interface de *GrAPHiST*, dans le but d'identifier les dynamiques spatio-temporelles recherchées.

La question de la validation de notre approche se pose alors. Les méthodes de représentation graphique et interactive de *GrAPHiST*, ainsi que leurs modalités d'utilisation étant définies, il nous faut prouver la possibilité d'une identification de patterns dans les séries d'événements spatio-temporels, au moyen de *GrAPHiST*.

Conclusion

Nous présentons dans cette partie la structure et le fonctionnement de *GrAPHiST*, un environnement de visualisation intégrant les dimensions spatiale et temporelle, et offrant la possibilité d’une analyse exploratoire des dynamiques des phénomènes spatio-temporels.

L’environnement *GrAPHiST*

GrAPHiST est basé sur :

- l’utilisation d’un modèle de données générique des phénomènes spatio-temporels ;
- une interface de visualisation à multi-fenêtrage.

Modèle de données de *GrAPHiST*

Le modèle de données développé dans le cadre de *GrAPHiST* est conforme à l’approche de modélisation occurrente de la donnée spatio-temporelle présentée dans le chapitre 1 :

- L’élément central de ce modèle est l’*événement spatio-temporel*, entité limitée dans le temps et l’espace et possédant une identité propre.
- Chaque *événement* est composé de *sous-parties spatio-temporelles*, contiguës dans le temps, représentant les différents états de l’*événement* au cours du temps.
- Un *phénomène* est un regroupement d’*événements* relevant de la même problématique.

Ce modèle de données permet de représenter des événements ponctuels ou étendus dans le temps et l’espace. Dans le cadre de cette thèse, les données à représenter ont été limitées aux événements ponctuels.

Les représentations graphiques des événements dans les différentes fenêtres de l’outil

Les fenêtres de l’interface de *GrAPHiST* sont occupées par des représentations graphiques permettant d’identifier des structures dans la distribution spatiale ou temporelle des événements.

L’analyse d’un phénomène spatio-temporel s’effectue sur des séries d’événements dont la composante thématique est plus ou moins fixe. Pour cela, *GrAPHiST* ne comporte pas de fenêtre dédiée à la représentation visuelle de la dimension thématique. Cette dernière est prise en compte à travers la sélection des phénomènes à représenter.

La fenêtre spatiale est occupée par une carte d’inventaire, à l’instar de l’environnement *PerSE* [Swedberg and Peuquet, 2017]. Les événements peuvent y être représentés individuellement par une entité graphique, ou peuvent être représentés de manière agrégée.

La fenêtre temporelle est occupée par deux diagrammes synchronisés :

- Un diagramme temporel *mixte* pouvant se présenter :
 - sous la forme d’une courbe sinusoïdale, inspirée de *Time Coil* [Edsall and Peuquet, 1997] et *Time Wave* [Li, 2010] ;
 - sous la forme d’une courbe en dents de scie, inspirée des diagrammes *Tiles Map* [Mintz et al., 1997].

Les événements peuvent y être représentés individuellement par une entité graphique, ou peuvent être représentés de manière agrégée.

- Un diagramme temporel circulaire en secteur, inspiré du *Rose Charts* de Florence Nightingale [Nightingale, 1858]. Les événements y sont représentés de manière agrégée selon des intervalles temporels de durée régulière. L’échelle cyclique représentée dans ce diagramme est la même que dans le diagramme mixte.

La fenêtre temporelle est synchronisée à la fenêtre spatiale :

- les événements représentés dans une des fenêtres sont représentés dans l’autre fenêtre de l’interface ;
- une interaction de l’utilisateur au sein d’une fenêtre se répercute dans l’autre fenêtre.

Une troisième fenêtre est occupée par une interface de contrôle, permettant à l’utilisateur de modifier les critères de représentation des événements.

L’utilisation de la symbologie pour représenter le lien entre les fenêtres spatiale et temporelle

La mise en relation des composantes temporelle et spatiale des séries d’événements peut s’effectuer au moyen d’outils de sélection et de *brushing*. Cette mise en relation s’effectue également en introduisant une dimension supplémentaire dans une des fenêtres de *GrAPHiST*, au moyen de règles de symbologie, à l’instar de *Time Wave* [Li, 2010] et de *Growth Ring Map* [Bak et al., 2009, Andrienko et al., 2011]. Les événements peuvent être représentés :

- en fonction de leur lieu d’apparition dans la fenêtre temporelle ;
On parle alors d’une représentation selon **une entrée spatiale, orientée de la carte vers les diagrammes temporels**.
- en fonction de leur moment d’apparition relatif à une échelle linéaire dans la fenêtre spatiale ;
On parle alors d’une représentation selon **une entrée temporelle linéaire, orientée du diagramme temporel mixte vers la carte et le diagramme temporel circulaire**.
- en fonction de leur moment d’apparition relatif à une échelle cyclique dans la fenêtre spatiale.
On parle alors d’une représentation selon **une entrée temporelle cyclique, orientée du diagramme temporel circulaire vers la carte et le diagramme temporel mixte**.

Afin de représenter, au sein d’un agrégat d’événements, la proportion d’événements apparaissant dans la même zone spatiale ou durant le même intervalle temporel, les agrégats sont représentés sous la forme de *Pie Charts*. Cette représentation est utilisée dans la carte et dans le diagramme temporel mixte. Pour représenter ces proportions au sein du diagramme temporel mixte, dans le cas de représentation selon une entrée spatiale, ou selon une entrée temporelle linéaire, le diagramme circulaire représente, pour chaque section angulaire, chaque classe d’événements par une sous-section radiale.

L'analyse exploratoire au travers de *GrAPHiST*

L'analyse exploratoire au travers de notre environnement passe par une suite d'actions effectuée par l'utilisateur à travers les différentes fenêtres de l'interface. Durant ces actions, l'utilisateur modifie progressivement les critères de représentation des événements, jusqu'à rendre visible un motif graphique révélateur d'un pattern :

- L'analyse peut s'effectuer sur une seule des fenêtres de l'interface afin d'identifier une structure unidimensionnelle dans les distributions spatiale ou temporelle des événements.
- L'utilisateur peut projeter une composante additionnelle des événements dans une des fenêtres de l'environnement, au moyen des règles de symbologie, afin d'effectuer une analyse de patterns spatio-temporels.

L'analyse exploratoire de patterns unidimensionnels

GrAPHiST permet une analyse de clusters dans les distributions temporelle et spatiale des séries d'événements :

- des clusters spatiaux peuvent être identifiés dans la fenêtre spatiale ;
- des clusters temporels relatifs au temps linéaire peuvent être identifiés dans le diagramme temporel mixte, le long de l'axe des abscisses ;
- des clusters temporels relatifs au temps cyclique peuvent être identifiés :
 - dans le diagramme temporel mixte, le long de l'axe des ordonnées ;
 - dans le diagramme temporel circulaire.

L'utilisateur peut modifier la représentation des événements dans les fenêtres de l'interface, pour améliorer la définition des contours de ces structures.

L'utilisateur peut utiliser le diagramme temporel mixte pour rechercher des récurrences cycliques dans les séries événementielles, au moyen d'une modification dynamique et graduelle de l'échelle temporelle cyclique représentée, comme le préconisent les travaux relatifs à l'environnement *Spiral Graph* [Weber et al., 2001]. L'utilisateur modifie la durée T de l'échelle cyclique, jusqu'à observer la formation d'un motif graphique sur le diagramme mixte, lorsque T approche de la durée P de la période d'un cycle.

Afin de guider l'utilisateur dans sa recherche d'une récurrence cyclique, *GrAPHiST* effectue en arrière-plan une analyse quantitative par transformées de Fourier. Celles-ci permettent d'identifier les possibles récurrences cycliques présentes dans les séries d'événements, et de les renvoyer à l'utilisateur en guise de pistes d'exploration.

L'analyse exploratoire de patterns spatio-temporels

L'utilisateur peut combiner ces méthodes d'identification de patterns unidimensionnels, avec une représentation des événements selon une entrée spatiale ou temporelle, dans le but d'identifier des patterns spatio-temporels.

En utilisant une entrée spatiale, l'utilisateur peut analyser la distribution spatio-temporelle des événements à travers un découpage spécifique du territoire. Celui-ci peut être créé à la volée par l'utilisateur, ou correspondre à un découpage prédéfini du territoire. Cette seconde option permet de croiser l'analyse du phénomène spatio-temporel, avec la problématique de recherche ayant permis de créer le découpage du territoire utilisé : l'analyse peut être effectuée en utilisant un découpage du territoire selon différents niveaux de pollution, ou un découpage correspondant

à différents bassins versants de cours d'eau, etc.

En utilisant une entrée temporelle, l'utilisateur peut analyser la distribution spatio-temporelle des événements, sans effectuer d'hypothèse a priori sur leur distribution spatiale. L'utilisateur peut alors modifier le nombre d'intervalles temporels permettant de classer les événements selon leur moment d'apparition, afin d'identifier des patterns plus ou moins étendus dans le temps. Un nombre élevé d'intervalles temporels permet d'identifier des patterns moins étendus dans le temps, de déterminer plus précisément leurs contours, d'observer si ces patterns sont homogènes durant leur existence (Section 6.1.1.2). Un nombre élevé d'intervalles fait cependant intervenir plus de couleurs ou plus de valeurs pour représenter les événements selon leur moment d'apparition, rendant plus difficile la lecture de la carte.

Pour ces deux approches, les outils de sélection graphiques peuvent être utilisés afin d'affiner la définition des patterns identifiés.

Huit scénarios d'analyses exploratoires peuvent alors être formalisés, permettant l'identification des quatre dynamiques spatio-temporelles sur lesquelles nous nous focalisons, selon que l'analyse soit effectuée selon une approche spatiale (Section 6.3), ou temporelle (Section 6.4). Il est ainsi possible d'analyser la distribution spatio-temporelle des événements pour identifier :

- des clusters spatio-temporels relatifs au temps linéaire (Sections 6.3.2 et 6.4.2) ;
- des clusters spatio-temporels relatifs au temps cyclique (Sections 6.3.3 et 6.4.3) ;
- des propagations d'événements au cours du temps (Sections 6.3.4 et 6.4.4) ;
- des récurrences cycliques spatialisées d'événements (Sections 6.3.5 et 6.4.5).

Limites et travaux à effectuer

Les limites posées par les choix de représentation

Le développement de *GrAPHiST* ayant été empirique, le choix de limiter notre représentation aux événements ponctuels dans le temps et l'espace a été effectué au cours de l'implémentation. Cela a néanmoins influé sur nos choix de représentation graphique par la suite.

Le choix des règles de symbologie combiné à une représentation des événements proches par des agrégats, prenant la forme de *Pie Charts*, se justifie ainsi pour des entités spatio-temporelles ponctuelles dans le temps et l'espace. Cependant, ce choix s'avère problématique pour l'analyse d'événements étendus dans le temps et l'espace.

Cas des événements étendus dans l'espace

Dans le cas d'événements étendus dans l'espace, par exemple, la représentation de leur étendue spatiale dans la carte empêche la représentation d'événements spatialement proches sous la forme d'agrégats représentés par une seule entité ponctuelle. D'autres environnements de géovisualisation utilisent alors la transparence pour représenter des entités spatiales suffisamment proches pour être superposés. L'utilisation de la transparence rend cependant difficile l'utilisation de règles de symbologie pour représenter le temps dans la carte.

Il convient alors d'intégrer d'autres procédés de représentation si l'on veut étendre notre approche à l'analyse d'événements non ponctuels.

Cas des événements étendus dans le temps

Un autre cas de figure n'étant pas pris en compte par notre proposition est le cas d'événements étendus dans le temps. La question se pose alors de la représentation d'entités duratives

dans les diagrammes temporels.

Si des événements étendus dans le temps peuvent être représentés par des entités graphiques linéaires, suivant la courbe du diagramme temporel mixte, comme dans la Figure 5.7, le cas d'événements proches temporellement pose les mêmes problèmes de superposition que les entités étendues dans l'espace.

Le cas du diagramme temporel circulaire pose notamment la question suivante : un événement s'étendant sur un intervalle temporel représenté par trois sections du diagramme est-il comptabilisé dans ces trois sections, ou dans la seule section représentant son moment d'apparition.

Enfin, un dernier cas de figure non exploré est le cas d'événements possédant plusieurs sous-parties temporelles. La question se pose la représentation de ces différentes sous-parties dans la carte et dans les diagrammes temporels.

La nécessité d'une validation par l'expérimentation

Si les propositions effectuées, en termes de représentations graphiques, d'interactions, ou d'utilisation, permettent en théorie de fournir la possibilité d'une analyse exploratoire des dynamiques spatio-temporelles, il est nécessaire, pour prouver ce postulat, d'effectuer une expérimentation sur des jeux de données réels.

GrAPHiST doit ainsi être appliquée à l'exploration de jeux de données dans le but d'y retrouver les différents patterns cités. Une identification d'un exemple des différentes dynamiques que nous voulons analyser pourra permettre de valider la possibilité d'une analyse des dynamiques des phénomènes spatio-temporels au moyen d'une analyse exploratoire au travers de notre environnement de géovisualisation. Différents cas d'application ont ainsi été réalisés, et sont présentés dans les chapitres suivants (Chapitre 7, Chapitre 8 et Chapitre 9).

Quatrième partie

Cas d'étude

Introduction

Dans cette partie, nous présentons une application de l'analyse exploratoire, proposée par *GrAPHiST*, à différents cas d'études, relatifs à des jeux de données réels ou simulés.

Objectif

L'objectif des analyses présentées dans cette partie est de valider l'hypothèse, selon laquelle les propositions émises au cours des chapitres précédents (en termes de modélisation, de représentations graphiques, et de possibilités d'interaction dans le chapitre 5, et en termes de scénarios d'utilisation dans le chapitre 6) rendent possible une analyse exploratoire des phénomènes spatio-temporels, en mesure d'identifier les dynamiques décrites dans le chapitre 2.

Pour atteindre cet objectif, nous effectuons au moyen de l'interface de géovisualisation que nous avons développée, une série d'analyses exploratoires sur différents jeux de données. Lors de ces analyses, nous tentons d'identifier selon les procédés décrits dans le chapitre 6 des patterns présents dans les jeux de données utilisés.

Justification de l'approche

La validation d'une représentation cartographique passe généralement par des expérimentations auprès d'utilisateurs.

La méthode expérimentale

Dans le domaine de la psychologie, une expérimentation est définie par Grawitz (cité dans [Catanas, 2002, Saint-Marc, 2017]) comme *une observation provoquée portant sur une situation créée et contrôlée par le chercheur qui a pour but de valider ou d'invalider une ou des hypothèses issues d'un système théorique*. Une hypothèse est une *affirmation prédictive* de la forme suivante : *la variable X devrait produire l'effet Y*.

Une expérimentation fait intervenir plusieurs objets [Nicolas et al., 2014, Saint-Marc, 2017] :

- les critères permettant de valider l'hypothèse, appelés *facteurs unitaires*, mesurables lors de l'expérimentation au travers de *variables dépendantes* (appelées ainsi, car elles sont dépendantes du phénomène observé) ;
- les différentes modalités, contrôlées par l'expérimentateur, permettant de faire varier la situation de test. Ces modalités sont appelées les *facteurs principaux*, et sont mesurables par des variables *indépendantes* ;
- les autres sources de variation de la situation de test, pouvant influencer les résultats, mais n'étant pas destinées à être testées, sont appelées les *facteurs secondaires*.

Une expérimentation consiste à tester l'hypothèse, à travers l'observation des *facteurs unitaires* :

- en faisant varier l’environnement de test selon les modalités décrites par les *facteurs principaux* : l’expérimentateur fait varier la valeur des *variables indépendantes* et observe l’effet de cette évolution sur les *variables dépendantes* ;
- en contrôlant la variabilité des *facteurs secondaires* afin d’éviter l’introduction d’erreurs dans l’interprétation des résultats : l’expérimentation s’effectue donc en milieu contrôlé [Hall, 1998].

L’expérimentation auprès d’utilisateurs dans le domaine de la cartographie

Dans le domaine de la cartographie et de la visualisation, les *tests-utilisateurs* visent le plus souvent à évaluer la lisibilité des représentations cartographiques et l’efficacité de la carte pour la transmission d’informations [Saint-Marc, 2017]. Dans le cas d’une interface de visualisation, une représentation graphique est considérée comme efficace si elle est plus rapide à interpréter, permet de tirer plus de conclusions, et conduit à faire moins d’erreurs d’interprétation qu’une autre représentation [Saint-Marc, 2017].

L’expérimentation a pour but de valider l’hypothèse selon laquelle la représentation cartographique proposée A est plus efficace qu’une autre représentation cartographique B pour analyser une information. Au cours de cette expérimentation :

- les *facteurs unitaires* sont les critères permettant de décrire si une représentation cartographique est efficace : rapidité d’interprétation, conclusions données sur les résultats, nombres d’erreurs d’interprétation, etc ;
- les *facteurs principaux* sont les différentes représentations cartographiques à tester : ici les représentations A et B ;
- les *facteurs secondaires* regroupent l’ensemble des paramètres devant être contrôlés par l’utilisateur pour ne pas fausser les résultats.

Le test d’une carte ou d’une interface de géovisualisation implique ainsi de mesurer les *variables dépendantes* de l’expérimentation, au moyen d’indicateurs témoignant de la facilité du raisonnement de l’utilisateur [Saint-Marc, 2017]. Ces mesures sont généralement effectuées lors d’expérimentations où des participants doivent effectuer une série de tâches au moyen de la représentation cartographique. Celles-ci sont effectuées lors d’expérimentation de laboratoire, dans laquelle des participants sont convoqués en un lieu défini aux conditions environnementales contrôlées, ou lors d’expérimentation web, dans laquelle l’expérimentation est effectuée au travers d’un questionnaire diffusé sur Internet [Kinkeldey et al., 2014, Saint-Marc, 2017]. D’autres modes d’expérimentations existent, ne demandant pas d’accomplissement de tâches par les participants, et s’effectuent sous la forme d’entretiens avec l’expérimentateur, ou sous la forme de groupe de discussion autour de l’outil.

Lors de ces expérimentations, deux types de tâches sont proposées [Kinkeldey et al., 2014, Saint-Marc, 2017] :

- les évaluations objectives, pour lesquelles on peut mesurer, objectivement, l’exactitude des réponses (retrouver une valeur parmi un ensemble, identifier une entité) ;
- les évaluations subjectives, qui regroupent des tâches d’auto-évaluation du ressenti des participants (évaluer la confiance dans une réponse, information sur les difficultés rencontrées). Ces évaluations subjectives peuvent être évaluées qualitativement (analyse du contenu de textes) ou quantitativement (statistiques). Ces évaluations subjectives sont généralement mises en relation avec les mesures d’efficacité de la lecture de la carte [Saint-Marc, 2017].

La plupart des expérimentations portant sur la lisibilité des représentations cartographiques mesurent *la vitesse ou la précision des sujets à accomplir des tâches dans un ensemble graphique ou visuel* [Castner, 1983, Chesneau, 2006, Saint-Marc, 2017]. Les deux variables dépendantes généralement utilisées pour mesurer l'efficacité de la lecture d'une carte sont le temps de réponse et le taux de bonnes réponses [Kinkeldey et al., 2014, Saint-Marc, 2017]. Il est généralement demandé aux participants d'évaluer leur utilisation de l'outil, leur confiance dans les résultats trouvés, et l'intuitivité de l'interface [Saint-Marc, 2017].

En parallèle à ces expérimentations, des méthodes sont élaborées pour étudier le mode de lecture et de raisonnement de l'utilisateur au travers d'une interface. Parmi celles-ci, nous pouvons citer :

- l'oculométrie cognitive [Irina Fabrikant et al., 2008, Henderson and Hollingworth, 1998, Saint-Marc, 2017], permettant d'enregistrer les points de fixation des yeux du participant sur la carte ou dans l'interface de géovisualisation ;
- l'enregistrement de traces d'utilisation logicielles [Choquet and Iksal, 2007, Jansen, 2008, Saint-Marc, 2017], qui permet l'enregistrement des séquences d'interactions réalisées par le participant au sein de l'interface.

L'efficacité d'une représentation cartographique est ainsi testée en combinant les résultats de mesures qualitatives, issus de l'oculométrie ou de l'enregistrement de traces, aux résultats de mesures quantitatives [Saint-Marc, 2017].

Notre approche expérimentale

Avant d'effectuer ce type d'expérimentations, permettant de comparer l'efficacité de notre proposition par rapport à d'autres méthodes d'analyses existantes, nous devons valider l'hypothèse selon laquelle il est possible d'analyser les dynamiques des phénomènes spatio-temporels au moyen de notre approche. Nous devons vérifier si l'analyse exploratoire d'un jeu de données spatio-temporelles, au moyen *GrAPHiST*, permet l'identification de ces dynamiques.

Dans un premier temps, l'objectif de notre approche expérimentale est de déterminer si une identification visuelle des patterns, décrits dans le chapitre 2, au cours d'une analyse exploratoire de données spatio-temporelles événementielle, est rendue possible grâce à notre interface de géovisualisation. Le but n'est pas d'étudier l'efficacité de notre approche en termes de facilité d'utilisation, mais de valider si celle-ci est fonctionnelle.

Avant de comparer l'efficacité de notre proposition avec celles d'autres méthodes d'analyse, nous devons également valider le caractère généraliste de notre approche. Beaucoup de méthodes d'analyse des phénomènes spatio-temporels sont développées de manière *ad hoc* [Andrienko and Andrienko, 2006]. Effectuer une étude de l'efficacité de notre approche, par rapport à d'autres méthodes, nécessite alors d'effectuer une expérimentation dans un domaine d'étude défini. Il convient alors de montrer que notre approche d'analyse exploratoire peut être appliquée dans le cadre de différentes problématiques. Notre proposition pourra alors être comparée aux méthodes d'analyses existantes, utilisées pour chacune de ces problématiques.

Dans un second temps, l'objectif de nos expérimentations est donc de valider l'hypothèse selon laquelle il est possible d'appliquer notre environnement de géovisualisation à l'analyse exploratoire de phénomènes spatio-temporels de différentes natures.

Les expérimentations que nous décrivons dans cette partie ne font donc pas appel à des tests-utilisateurs, ou à l'utilisation d'appareils d'oculométrie ou d'enregistrement de traces logicielles.

Dans un premier temps, l'expérimentation consiste en l'application de notre proposition à une série de jeux de données, dans lesquels des patterns sont présents. Lors de cette expérimentation, nous tentons de retrouver ces patterns au travers d'une analyse exploratoire. L'hypothèse que nous émettons alors est que si notre interface de géovisualisation permet de retrouver des patterns dans ces jeux de données, des patterns équivalents pourront être identifiés dans n'importe quel jeu de données présentant la même structure.

Dans un second temps, nous appliquons notre proposition d'analyse exploratoire à différents phénomènes, et nous présentons les opportunités d'analyse offerte par *GrAPHiST* dans chacun de ces jeux de données.

Procédé général

Dans les trois chapitres de cette partie, nous effectuons une analyse exploratoire de données spatio-temporelles, centrée sur l'identification de différents patterns.

Reproduction d'un récit historique de la guerre de Sécession

Dans un premier chapitre (7), nous analysons des données historiques relatives aux lieux et dates de batailles de la guerre de Sécession, ayant eu lieu sur le territoire des États-Unis d'Amérique au milieu du *XIXe* siècle. L'objectif de l'analyse effectuée dans ce chapitre est d'analyser la distribution spatio-temporelle des événements afin de reproduire un récit du conflit.

Cette analyse passe par l'identification de deux types de patterns spatio-temporels :

- des clusters spatio-temporels relatifs au temps linéaire, correspondant aux principales zones de combat au cours de la guerre ;
- des propagations spatiales des événements au cours du temps, correspondant au déplacement de la ligne de front au cours de l'avancement du conflit.

Le récit de la guerre de Sécession est connu. Des cartes représentant les différentes phases du conflit existent. Cette connaissance du récit de la guerre et les cartes à disposition nous servent de références historiques que nous devons confronter aux résultats de notre analyse exploratoire. Notre objectif n'est cependant pas de comparer ces cartes à notre outil de géovisualisation en termes d'efficacité pour retranscrire le récit du conflit.

Identification de cycles spatialisés dans des séries d'événements

Dans un second chapitre (8), nous analysons une série de jeux de données simulées, contenant chacun une série d'événements cycliques localisés dans l'espace, mais dont la période du cycle et la zone d'apparition sont définies aléatoirement, et une série d'événements de lieu et de moment d'apparition aléatoires. L'objectif de l'analyse est de retrouver, pour chaque jeu de données, la période et l'étendue temporelle du cycle, ainsi que la zone d'apparition des événements correspondants.

Cette analyse passe donc par l'identification de récurrences cycliques spatialisées, mais également par l'identification de clusters spatio-temporels relatifs au temps cyclique, celle-ci pouvant être une étape dans l'identification visuelle de cycles.

Les jeux de données sont créés de manière procédurale, et les paramètres déterminant le cycle et la zone d'apparition des événements de la première série sont stockés dans un fichier annexe, de manière à ce que ces paramètres puissent être accessibles à posteriori, mais inconnus lors de l'expérimentation. Ces paramètres sont révélés à l'utilisateur en fin d'expérimentation, et servent à évaluer la précision des résultats trouvés lors de l'analyse.

Identification de clusters spatio-temporels

Dans le dernier chapitre (9) de cette partie, nous effectuons une série d'analyses exploratoires dans le but d'identifier des clusters spatio-temporels, relatifs au temps linéaire ou cyclique, présents ou non dans différents jeux de données réelles. Ces jeux de données concernent :

- un recensement des délits et crimes sur le territoire de la ville de Chicago, de 2001 à 2017 ;
- un recensement des accidents routiers autour de la ville de Grenoble, de 2005 à 2015 ;
- des inondations survenues en France métropolitaine, de 1800 à 2015.

Contrairement aux deux premiers chapitres de cette partie, nous ne disposons pas de connaissances a priori sur d'éventuels patterns spatio-temporels présents dans les données. Les expérimentations décrites dans ce chapitre n'ont pas pour objectif de valider les fonctionnalités de notre interface de géovisualisation, en retrouvant des patterns connus au moyen de notre outil.

Les résultats de ces analyses exploratoires sur des données réelles, correspondant à différentes problématiques, ont cependant pour but de présenter les possibilités offertes par nos propositions dans l'analyse exploratoire de phénomènes spatio-temporels, et de montrer le caractère généraliste de notre approche.

Chapitre	Données utilisées	Résultats attendus
<i>VII</i>	Données relatives aux champs de bataille de la Guerre de Sécession, États-Unis, 1861-1865	Identification de propagations spatiales reflétant le récit historique du conflit
<i>VIII</i>	Données simulées présentant des cycles de réapparition de périodes inconnues	Identification de récurrences cycliques spatialisées
<i>IX</i>	Données criminologiques, Chicago, 2001-2017 Données d'accidents routiers, région de Grenoble, 2005-2015 Données d'inondations, France métropolitaine, 1800-2015	Identification de clusters spatio-temporels, dans le temps linéaire et dans le temps cyclique

Chapitre 7

Analyse de données historiques sur la guerre de Sécession

Sommaire

7.1	Objectif	276
7.1.1	Identifier des patterns spatio-temporels relatifs au temps linéaire dans les séries événementielles	276
7.1.2	Recréer le récit d'un phénomène à partir de données historiques	276
7.1.3	Le choix des données du cas d'étude	276
7.2	Données	278
7.2.1	Nature des données	278
7.2.2	Implémentation dans notre modèle de données	279
7.3	Procédé	279
7.3.1	Principe	279
7.3.2	Choix des épisodes à analyser, et objectifs de ces analyses	280
7.3.3	Chronologie des épisodes analysés	280
7.3.4	Scénarios d'analyse	284
7.4	Résultats	285
7.4.1	Analyse des différentes phases du conflit	285
7.4.2	Analyse des engagements du théâtre de l'Est	290
7.5	Synthèse	293

7.1 Objectif

7.1.1 Identifier des patterns spatio-temporels relatifs au temps linéaire dans les séries événementielles

L'objectif de l'expérimentation décrite dans ce chapitre est de valider l'hypothèse selon laquelle notre interface de visualisation permet une analyse exploratoire, rendant possible l'identification de patterns relatifs au temps linéaire dans des séries événementielles. Ces patterns concernent :

- la présence de clusters spatio-temporels selon le temps linéaire.
- la présence d'une propagation spatiale d'événements au cours du temps.

7.1.2 Recréer le récit d'un phénomène à partir de données historiques

Dans les chapitres 1 et 5, nous décrivons le phénomène comme une méta-entité spatio-temporelle, limitée dans le temps et l'espace, formée d'événements regroupés sur un critère thématique. Le phénomène se rapproche alors d'un groupement thématique, rassemblant un ensemble d'événements semblables, apparaissant durant l'étendue temporelle et dans l'emprise spatiale étudiées.

Un phénomène *Inondation* peut regrouper des événements représentant des cas d'inondation particuliers, sur la durée et l'emprise spatiale du phénomène, sans que ces événements soient liés entre eux autrement que par leur similarité thématique. Le phénomène modélisé ne représente pas un ensemble distinct et tangible, identifiable dans la réalité.

À l'inverse, un phénomène *Inondations relatives au Cyclone X* (X étant le nom du cyclone) peut regrouper différents événements représentant des cas d'inondation particuliers, tous relatifs au cyclone X . Ces événements peuvent alors être appréhendés comme des épisodes du cyclone, lui-même représenté au travers du phénomène *Inondations relatives au Cyclone X*.

Les données de séries événementielles relatives à un épisode historique limité dans le temps et l'espace (tel qu'un cyclone) peuvent alors être interprétées comme la manifestation des différentes phases de cet épisode historique, représenté par un phénomène.

Un des enjeux de la visualisation de patterns spatio-temporels relatifs au temps linéaire dans ces données devient alors de recréer l'historique des différentes phases de l'épisode historique représenté. La visualisation de l'évolution de la distribution spatiale des événements au cours du temps permet de créer du récit à partir de données historiques, permettant une meilleure compréhension des phénomènes représentés.

La visualisation de clusters spatio-temporels relatifs au temps linéaire permet d'identifier les différentes phases des phénomènes analysées dans le temps et l'espace. La visualisation de la propagation spatiale d'événements au cours du temps permet, quant à elle, d'identifier l'enchaînement de ces phases, et la progression du phénomène au cours du temps.

7.1.3 Le choix des données du cas d'étude

Dans le cadre de cette expérimentation, nous avons donc choisi de nous intéresser à des données relatives à un épisode historique tangible, identifiable dans la réalité, et étendu dans le temps et l'espace. Ces données doivent pouvoir être représentées sous la forme d'événements, assimilables à une localisation ponctuelle dans le temps et l'espace.

Plusieurs cas d'étude peuvent être utilisés dans le cadre de cette expérimentation. Les exemples suivants peuvent être cités :

- l'analyse spatio-temporelle d'une épidémie, représentée sous la forme de données événementielles recensant la localisation et le moment d'apparition des différents cas d'infection ;
L'analyse doit permettre d'identifier les différentes phases de la diffusion de l'épidémie, ainsi que les axes de propagation.
- l'analyse des représentations d'un artiste musical au cours d'une tournée ;
- l'analyse de différents relevés de position GPS d'un véhicule au cours du temps.

Nous avons finalement opté pour l'analyse spatio-temporelle de données événementielles représentant les lieux et dates des batailles de la guerre de Sécession, survenue de 1861 à 1865 aux États-Unis d'Amérique. L'outil devra permettre de reproduire un récit du conflit, et de ces différentes phases, au travers de la visualisation de l'évolution du lieu des batailles au cours du temps.

Le choix de ce cas d'étude se justifie pour plusieurs raisons :

- Ce jeu de données répond aux besoins de l'étude sur le format de données à utiliser. L'ensemble des événements représentés s'étend sur une longue période et sur un territoire étendu, mais chaque bataille de la guerre de Sécession a eu lieu sur une zone relativement restreinte et sur une durée relativement courte, permettant de considérer cette bataille comme un événement ponctuel à l'échelle de l'intervalle temporel et de la zone d'étude (que ce soit à l'échelle des *USA* ou à l'échelle d'un état).
- Ces données permettaient une analyse du déroulement du conflit selon plusieurs échelles temporelles et spatiales. L'analyse pouvait être effectuée à une échelle spatiale globale sur toute la durée du conflit, ou sur la durée d'une campagne militaire dans un théâtre d'opérations en particulier.
- Le conflit représenté par le jeu de données a été en grande partie une guerre de mouvement, si bien qu'à l'échelle d'une campagne militaire, les armées ne restaient que peu immobiles. Ce conflit a également vu l'invasion progressive du territoire du camp vaincu par le camp vainqueur, si bien que la ligne d'affrontement s'est déplacée progressivement dans le territoire du camp vaincu au cours de la guerre.
Ainsi, la représentation des engagements du conflit, que ce soit à l'échelle globale du conflit ou à l'échelle d'une campagne militaire, peut donner à voir des propagations spatiales des événements au cours du temps.
- Enfin, les différents épisodes de ce conflit ayant été particulièrement étudiés par les historiens. L'histoire du conflit est relativement bien connue, et les sources d'informations disponibles sont nombreuses. Plusieurs documents cartographiques ont également été produits et permettent une visualisation des différents épisodes du conflit dans l'espace et le temps (Figure 7.1).
Ces derniers peuvent représenter l'évolution globale du conflit, ainsi que le déroulement de chaque campagne militaire. Ces cartes, et les nombreuses sources sur l'histoire du conflit sont autant de référence pour évaluer le récit produit au travers de l'outil.

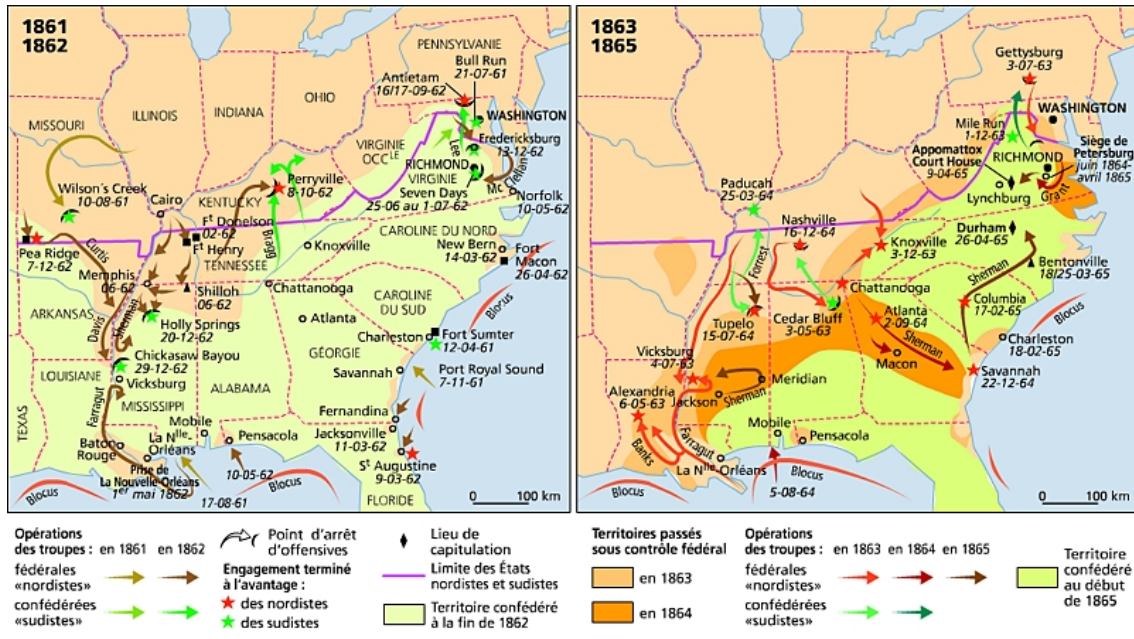


FIGURE 7.1 – Guerre de Sécession ou guerre civile américaine [Larousse, 2016]

7.2 Données

7.2.1 Nature des données

Les données dont nous disposons sont issues d'un projet de carte interactive destinée à servir de support pour un projet intitulé *Jamestown to Appomattox : Mapping US History using GIS, Volume 1*. Les données ont été compilées, à partir de données du *National Park Service*, par Robert Stewart, Jeremy Newcomb, et Chris Bunin, dans le cadre de *The Virginia Experiment, a Teaching American History Grant* [Robert Stewart, 2007].

Les données se présentent sous la forme d'une série d'entités ponctuelles, chaque entité correspondant à une bataille de la guerre de Sécession. Le lieu de chaque engagement est ainsi résumé à un point dans l'espace, et est représenté par l'emplacement de l'entité ponctuelle correspondante sur la carte. Pour chaque bataille nous disposons d'une série d'information concernant :

- les dates de début et fin de l'engagement ;
- le nom de la bataille ;
- la campagne militaire dans laquelle la bataille prend part ;
- les commandants des deux forces en présence ;
- l'effectif des troupes des deux forces en présence ;
- le vainqueur de l'engagement ;
- le nombre de victimes pour les deux camps.

Bataille	Campagne	Lieu	Date début	Date fin	Com. USA	Com. CSA	Troupes USA	Troupes CSA	Vainqueur	Victimes USA	Victimes CSA
1st Bull Run	Manassas	38.8147, -77.5227	21-07-1861	21-07-1861	Irvin McDowell	P.G.T. Beauregard	35732	34000	CSA	2708	1982
Gettysburg	Gettysburg	39.811, -77.225	01-07-1863	03-07-1863	George G. Meade	Robert E. Lee	104256	34000	USA	23049	28000

Certaines batailles de la guerre de Sécession se sont étalées sur une durée de plusieurs jours (plus de dix jours pour certaines). Nous avons cependant décidé de les traiter comme ponctuelle dans le temps, en ne considérant que la date du début de l'engagement. Chaque bataille est alors traitée comme une entité spatio-temporelle ponctuelle dans le temps et l'espace.

L'ensemble des batailles ont lieu du *12 avril 1861* au *9 mai 1865*, et sont majoritairement localisées dans le sud-est des États-Unis.

7.2.2 Implémentation dans notre modèle de données

Les données brutes sont disponibles sous la forme d'un tableau. Nous utilisons un script *SQL* pour entrer ces événements dans notre base de données, en classant les informations de chaque événement selon la structure décrite par le *Pyramid Model*. Nous créons ainsi pour chaque entrée du tableau une entité événement, composée d'une unique sous-partie temporelle, et possédant neuf arguments thématiques (Figure 7.2).

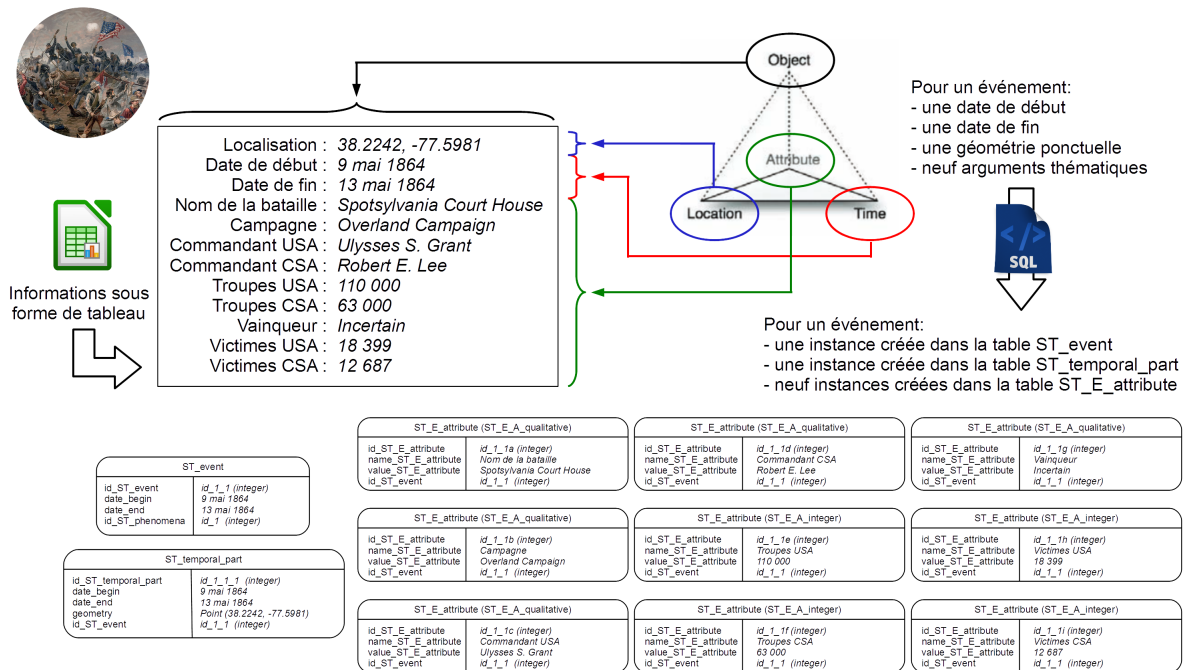


FIGURE 7.2 – Implémentation des données relatives à la guerre de Sécession dans notre modèle

7.3 Procédé

7.3.1 Principe

Ce conflit, étendu sur 4 ans, a principalement été une guerre de mouvement, comprenant différentes phases. Au cours de ces différentes phases, les armées des deux camps ont effectué des manœuvres, offensives ou défensives, au cours de différentes campagnes dont le but était l'invasion du territoire de l'adversaire. Ces campagnes se sont déroulées sur des étendues spatiales et temporelles limitées, et ont été le cadre de déplacements d'armées dans l'espace au cours du temps. Ces déplacements ont été ponctués d'affrontements, qui sont représentés au travers des données dont nous disposons.

L'analyse de la distribution spatio-temporelle des batailles de la guerre de la Sécession doit permettre d'identifier le déroulement spatio-temporel des différentes campagnes et phases du

conflit. Un récit du conflit doit alors être produit au travers de ce que nous voyons dans l'interface de visualisation. Puis nous comparons ce récit avec ceux déjà produits par les historiens.

7.3.2 Choix des épisodes à analyser, et objectifs de ces analyses

Nous nous intéressons, en premier lieu, à une analyse globale du conflit afin d'en déterminer les différentes grandes phases. La visualisation de l'évolution temporelle de la distribution spatiale des batailles aura pour but d'identifier le déroulement de la conquête du territoire des états sécessionnistes du sud par les armées du nord, au travers du déplacement des zones de conflits au cours du temps.

Dans un second temps, nous effectuons une analyse centrée sur un théâtre d'opérations particulier pendant une période plus restreinte. Nous avons choisi d'étudier le théâtre d'opérations de l'Est, entre mai 1864 et avril 1865, période qui a vu la progression des armées du Nord au travers de l'état de Virginie, sudiste, et la poursuite des armées confédérées. La visualisation de la distribution spatio-temporelle des batailles aura pour but de retranscrire le déroulement de la campagne, ses différentes phases, et la progression des armées nordistes au cours de l'offensive.

7.3.3 Chronologie des épisodes analysés

Dans les paragraphes suivants, nous décrivons brièvement le déroulement global du conflit, puis un déroulement plus détaillé des opérations dans le théâtre d'opérations de l'Est, entre mai 1864 et avril 1865. Cette chronologie est basée sur les articles francophones et anglophones de *Wikipédia* [Wikipedia, 2018a] sur la guerre de Sécession. Cette chronologie sera ensuite comparée au récit produit par l'analyse des données au travers de l'interface de visualisation.

7.3.3.1 Chronologie de la guerre de Sécession

La guerre de Sécession, ou guerre civile américaine (*The Civil War* aux États-Unis) se déroule du 12 avril 1861 au 9 avril 1865, et voit s'affronter les États-Unis d'Amérique (l'*Union*, désignée comme le *Nord* ou *USA*), à onze états sécessionnistes du sud des États-Unis, rassemblés au sein des États Confédérés d'Amérique (la *Confédération*, désignée comme le *Sud* ou *CSA*). Le conflit s'est achevé par une victoire de l'*Union* et par la réintégration des états sécessionnistes.

Hormis quelques théâtres d'affrontement à l'ouest du Mississippi, la majorité des combats eurent lieu au sud-est des États-Unis, communément appelé le *Sud* des États-Unis (Figure 7.3).



FIGURE 7.3 – Batailles de la guerre de Sécession selon le nombre de victimes [The Washington Post, 2011]

Les combats eurent notamment lieu :

- entre Washington D.C. (capitale de l'Union) et Richmond (capitale de la Confédération), dans l'état de Virginie ;
Ce théâtre d'affrontements fut appelé le *théâtre de l'Est*.
- le long de la ligne des Appalaches, dans la région des états du Kentucky et du Tennessee ;
Ce théâtre d'affrontements fut appelé le *théâtre de l'Ouest*.
- le long du fleuve Mississippi.

La stratégie de la Confédération fut défensive. Aussi les différentes phases du conflit furent surtout à l'initiative du Nord, qui envahit progressivement le territoire du Sud durant les quatre années du conflit. Une chronologie brève du conflit peut alors être décrite ainsi en 5 périodes selon les années du conflit (Figure 7.4) :

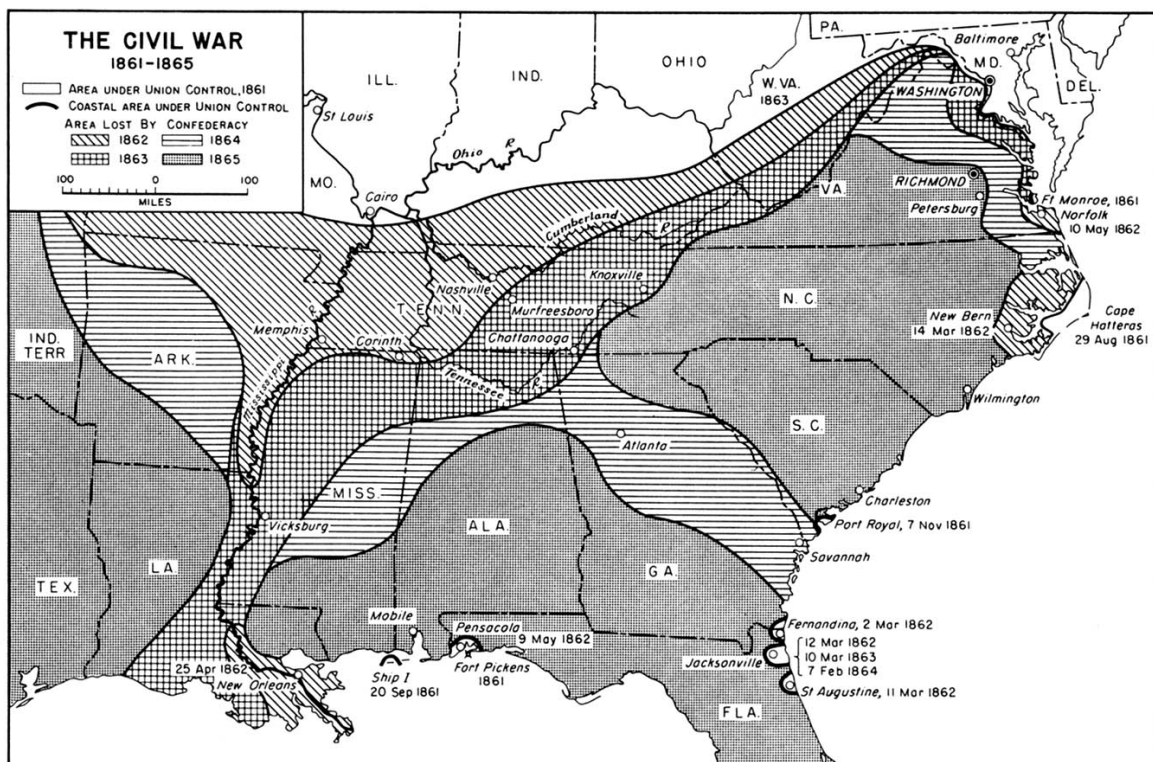


FIGURE 7.4 – Territoires perdus par la Confédération lors de la guerre de Sécession [Office of the Chief of Military History, 1989]

- En 1861, le Nord s'assure le contrôle des frontières avec les états sécessionnistes, et établit un blocus naval pour empêcher la Confédération de commercer avec l'extérieur.
- En 1862, le Nord s'assure le contrôle des contreforts des Appalaches, et pénètre au Tennessee. Sur le théâtre de l'Est, suite à des campagnes infructueuses visant à capturer Richmond le front reste stable. Le Nord progresse le long du Mississippi vers le sud et la forteresse de Vicksburg. Simultanément, le Nord débarque à l'embouchure du Mississippi et capture la Nouvelle-Orléans.
- En 1863, le Nord prend Vicksburg, puis Port Hudson près de Bâton Rouge, achevant la prise de contrôle du Mississippi. L'Union progresse dans le théâtre de l'Ouest et prend Chattanooga. Dans le théâtre de l'Est, le front reste stable, et le Sud échoue dans sa dernière tentative d'invasion des territoires nordistes.

- En 1864, le Nord prend Atlanta, capitale de l'état de Géorgie, sur le théâtre de l'Ouest. Le général *Unioniste* William Tecumseh Sherman effectue une *Marche vers la Mer*, afin de prendre le port de Savannah, et de couper le territoire de la *Confédération* en deux. Dans le théâtre de l'Est, le Nord progresse vers le sud au cours de l'*Overland Campaign*, et commence les sièges de Richmond et de Petersburg.
- En 1865, l'armée nordiste de Sherman remonte vers le nord au cours de la campagne des Carolines. Dans le théâtre de l'Est, le Nord prend Richmond et Petersburg, et poursuit les armées sudistes vers l'ouest, où elles se rendent.

7.3.3.2 Déroulement des actions du théâtre nord-est, de mai 1864 à avril 1865

Contexte. À partir de mai 1864, l'armée nordiste, dirigée par Ulysses Grant, déclenche une offensive dans le théâtre de l'Est (Figure 7.5). Cette offensive vise, non pas la capture de Richmond, capitale de la *Confédération*, mais la destruction ou la reddition de l'armée confédérée de Virginie du Nord, dirigée par Robert Lee.

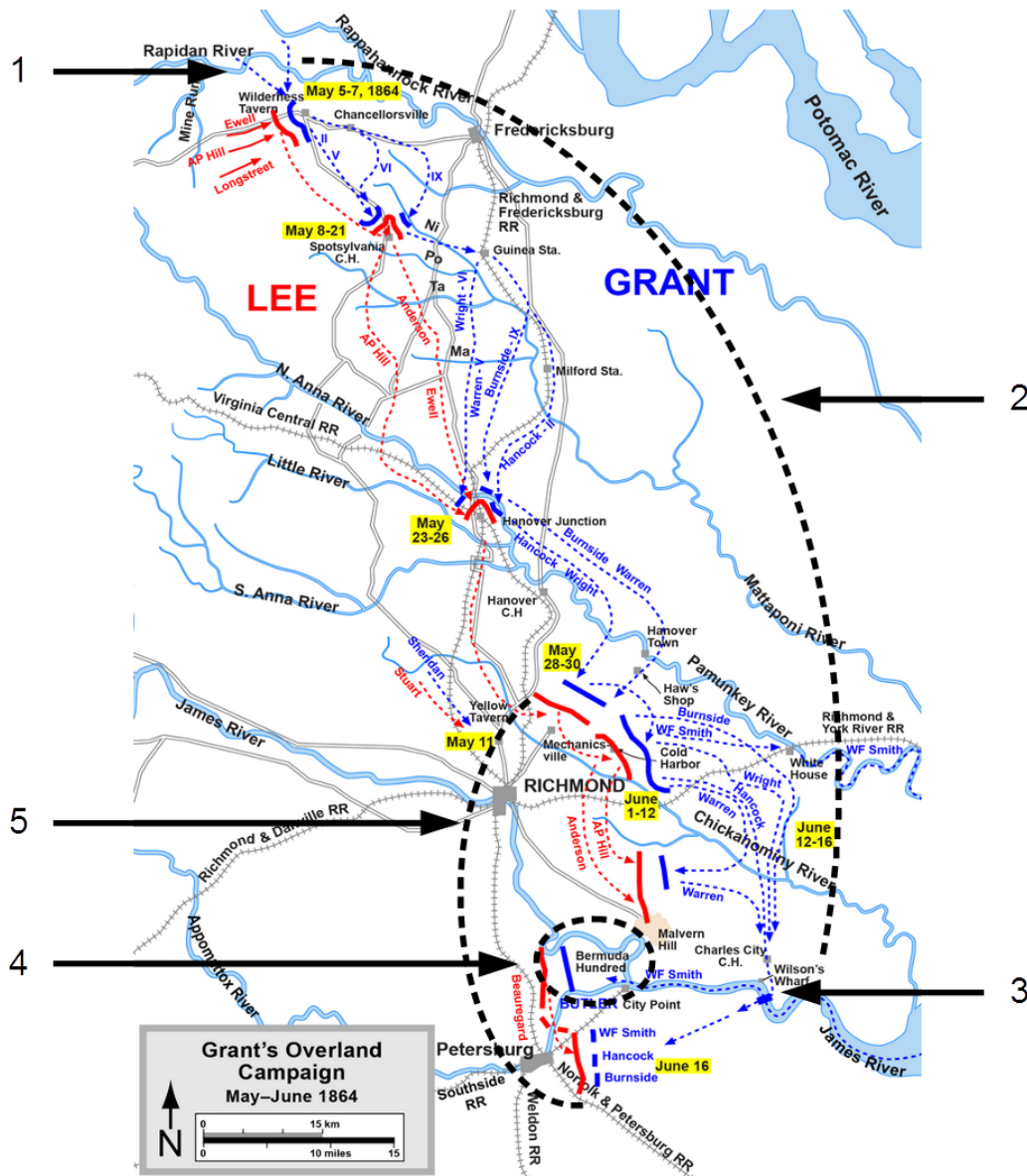


FIGURE 7.5 – Campagne de l'Overland, de mai à juin 1864 [Jespersen, 2012b]

L'Overland Campaign. Lors de la première partie de l'offensive, appelée campagne de l'*Overland* ou *Grant's Overland Campaign*, le Nord tente de vaincre l'armée confédérée en rase-campagne.

Le Nord progresse en territoire sudiste, et tente plusieurs fois de prendre de flanc l'armée confédérée. Lee, bien qu'infligeant de nombreuses défaites tactiques à l'armée nordiste au cours de l'avance de celle-ci, recule progressivement vers le sud de la Virginie, afin de ne pas subir de défaite face à la supériorité numérique des troupes nordistes. Durant cette période, le Nord progresse vers le sud depuis la *Rapidan River* (Repère 1 de la Figure 7.5), et contourne Richmond par l'est (Repère 2 de la Figure 7.5).

Lors de la dernière partie de la campagne de l'*Overland*, Grant change de stratégie et fait traverser la *James River* à son armée afin de menacer Petersburg (Repère 3 de la Figure 7.5). Cette ville, noeud ferroviaire et fluvial vital pour la défense de la capitale confédérée, ne peut être abandonnée par Lee. Ce dernier doit se résoudre à s'enfermer dans Petersburg et à en tenir le siège, en juin 1864.

Campagne de Bermuda Hundred. Au moment où Grant traverse la *Rapidan River*, l'*Union* débarque des troupes entre Richmond et Petersburg, depuis la *James River* (Repère 4 de la Figure 7.5). Ce débarquement vise à couper la voie de chemin de fer entre les deux villes. Ce débarquement échoue, et les troupes nordistes sont bloquées sur leur lieu de débarquement. Cette manoeuvre porte le nom de campagne de *Bermuda Hundred*.

Siège de Petersburg. La seconde partie de l'offensive, appelée campagne de Richmond-Petersburg ou siège de Petersburg, consiste en une série de batailles durant laquelle l'armée nordiste tente de prendre d'assaut la ville de Petersburg (Repère 5 de la Figure 7.5). Ce siège dure 9 mois, du 15 juin 1864 au 25 mars 1865, et prend la forme d'une guerre de position (ou guerre de tranchées).

Campagne d'Appomattox. Au bout de 9 mois de siège, les forces nordistes sont sur le point de prendre la ville. Lee décide d'évacuer Petersburg et Richmond, et se déplace vers l'ouest, le long de l'*Appomattox River*, dans le but de rejoindre d'autres armées confédérées en Caroline du Nord.

La campagne d'Appomattox qui s'ensuit est la dernière phase de l'offensive, dans laquelle Grant poursuit Lee au cours de sa retraite vers l'ouest (Figure 7.6). Cette poursuite est ponctuée d'affrontements, jusqu'à la bataille finale d'*Appomattox Court House*, le 9 avril, où Grant reçoit la reddition de Lee. Cette date est considérée comme marquant la fin de la guerre de Sécession.

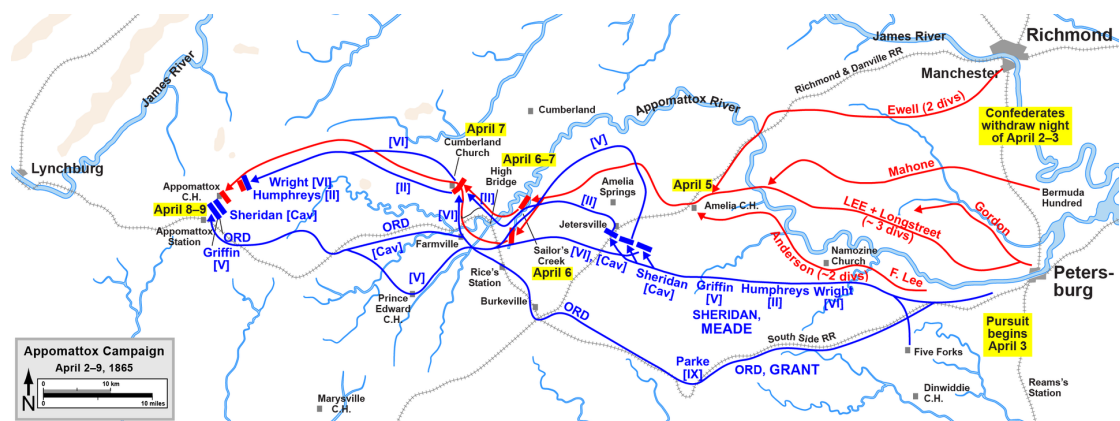


FIGURE 7.6 – Campagne d'Appomattox, du 2 au 9 avril 1865 [Jespersen, 2012a]

7.3.4 Scénarios d'analyse

7.3.4.1 Analyse globale des différentes phases du conflit

L'étendue temporelle étudiée est fixée entre le *1er janvier 1861* et le *31 décembre 1865*. L'échelle temporelle cyclique est fixée à 365 jours. L'emprise spatiale est choisie de manière à englober les états de l'Arkansas et de la Louisiane à l'ouest, la ville de la Nouvelle-Orléans au sud, et la ville de New York au nord-est.

Les événements sont représentés en fonction de leur moment d'apparition relatif au temps linéaire, selon cinq classes, correspondant aux cinq années du conflit de 1861 à 1865. Les événements sont représentés en utilisant la variable visuelle de la valeur, de manière à ce que les événements les plus anciens soient représentés par une couleur sombre, et les événements les plus récents par une couleur claire.

Les événements sont analysés en étant, dans un premier temps, représentés de manière non-agrégée, puis en étant représentés de manière agrégée.

L'analyse s'effectue visuellement, au moyen de la carte de la fenêtre spatiale. Elle doit permettre d'identifier des propagations spatiales d'événements au cours du temps, ainsi que des clusters spatio-temporels selon le temps linéaire. Une suite de clusters temporels correspondant à des intervalles temporels linéaires successifs, disposés le long d'un axe sur la carte, peut être interprété comme une propagation spatiale au cours du temps.

À partir des observations, un récit des différentes phases de l'invasion des territoires sudistes par le Nord doit être produit, puis comparée à des cartes et des chronologies du conflit déjà produites.

7.3.4.2 Analyse du théâtre de l'Est, de l'*Overland Campaign* à la bataille d'Appomattox

Trois analyses sont envisagées :

- La première doit s'effectuer sur l'ensemble du théâtre de l'Est, sur toute la période de temps allant du début de l'*Overland Campaign* à la bataille d'Appomattox. L'analyse doit donc porter sur la campagne de l'*Overland*, la campagne de *Bermuda Hundred*, le siège de Petersburg et la campagne d'*Appomattox*.
- La seconde analyse est centrée sur le déroulement de l'*Overland Campaign* et de la campagne de *Bermuda Hundred*.
- La troisième analyse est centrée sur le déroulement de la campagne d'*Appomattox*.

Pour ces trois analyses, l'emprise spatiale est choisie de manière à englober les villes de Fredericksburg au nord, West Point à l'est, Petersburg au sud, et Lynchburg à l'ouest.

Les événements sont représentés sur la carte de manière agrégée. Les événements sont représentés selon leur moment d'apparition relatif au temps linéaire. Les événements sont représentés en utilisant la variable visuelle de la valeur de manière à représenter les batailles les plus anciennes par une couleur sombre, et les batailles les plus récentes par une couleur claire.

L'analyse s'effectue visuellement, au moyen de la carte de la fenêtre spatiale. L'analyse doit permettre d'identifier des propagations spatiales d'événements au cours du temps, ainsi que des clusters spatio-temporels selon le temps linéaire. Les récits produits au travers de l'identification

des patterns doivent ensuite être comparés aux cartes et chronologies du conflit produites par les historiens.

Première analyse. L'étendue temporelle est réduite à la période entre le *1er mai 1864* et le *30 avril 1865*. L'échelle temporelle cyclique est fixée à 30 jours. Les événements sont représentés selon douze classes de valeurs de saturation, représentant les douze mois entre mai 1864 et avril 1865. L'observation de la carte doit permettre d'établir un récit des différentes phases de l'offensive nordiste dans le théâtre de l'Est du *1er mai 1864* au *30 avril 1865*.

Seconde analyse. L'étendue temporelle étudiée est fixée entre le *1er mai 1864* et le *31 juin 1864*. L'échelle temporelle cyclique est fixée à 7 jours. Les événements sont représentés selon neuf classes de valeurs de saturation, représentant chacune environ une semaine du *1er mai 1864* et le *31 juin 1864*. L'observation de la carte doit permettre de visualiser la progression des troupes nordistes en Virginie au cours de la campagne de l'*Overland*.

Troisième analyse. L'étendue temporelle étudiée est fixée entre le *2 avril 1865* et le *9 avril 1865*. L'échelle temporelle cyclique est fixée à sept jours. L'emprise spatiale reste inchangée par rapport à la première analyse. Les événements sont représentés selon sept classes de valeurs de saturation, représentant chacune une journée du *2 avril 1865* et le *9 avril 1865*. L'observation de la carte doit permettre de visualiser la poursuite des troupes confédérées par Grant au cours de la campagne d'*Appomattox*.

7.4 Résultats

7.4.1 Analyse des différentes phases du conflit

La Figure 7.7 montre une capture d'écran de l'interface de géovisualisation une fois saisis les paramètres décrits précédemment, et une fois les événements représentés selon leur moment d'apparition relatif au temps linéaire.

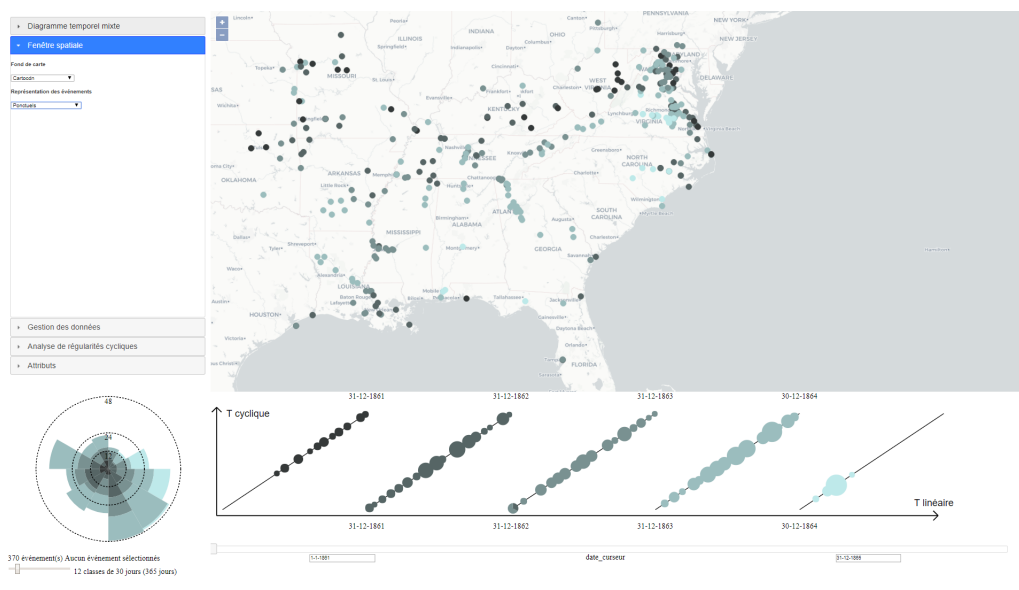


FIGURE 7.7 – Visualisation des batailles de la guerre de Sécession

7.4.1.1 Distribution des événements selon une échelle cyclique annuelle

Le diagramme temporel circulaire montre une plus faible proportion d'événements durant l'hiver. Les armées de l'époque prenaient en effet durant cette période leurs quartiers d'hiver, la majorité des campagnes étant effectuées à partir du printemps quand les conditions climatiques étaient plus clémentes (Figure 7.8).

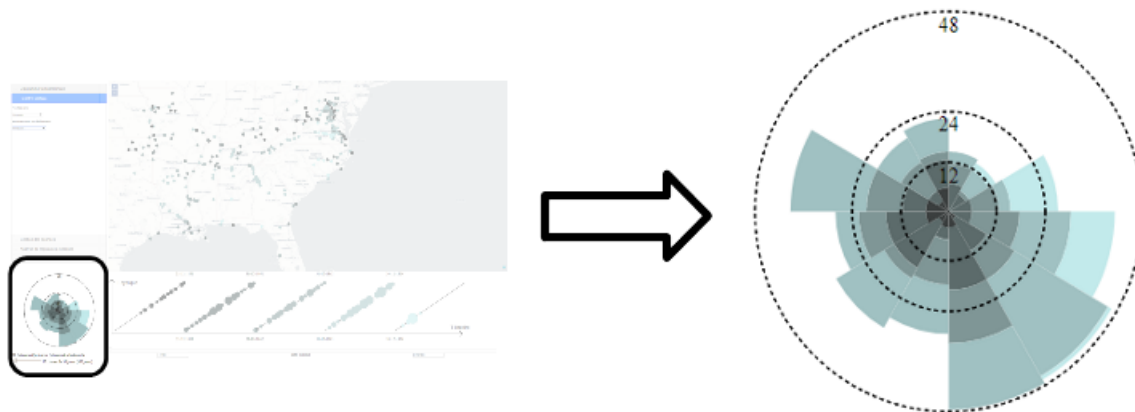


FIGURE 7.8 – Distribution temporelle des combats de la guerre de Sécession selon une échelle cyclique d'une année

7.4.1.2 Identification des différentes phases du conflit dans la fenêtre spatiale, à l'aide d'une représentation non-agrégée

Le grand nombre d'entités ponctuelles sur la carte gêne la lecture, et rend difficile la visualisation de l'évolution du territoire occupé par le Nord durant le conflit. Cela est notamment le cas pour le théâtre de l'Est (Repère 1 dans la Figure 7.9). Dans cette zone, les engagements ont eu lieu sur un territoire relativement restreint, et plusieurs campagnes successives ont eu lieu dans les mêmes régions.

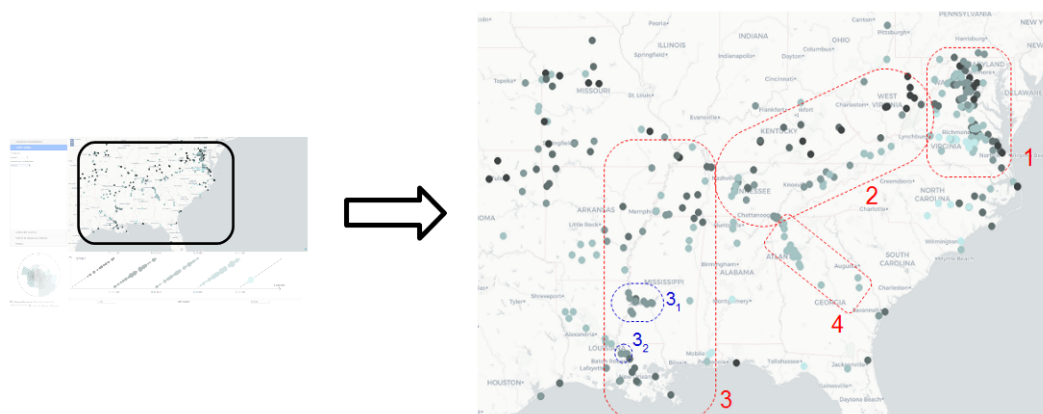


FIGURE 7.9 – Analyse des différentes phases de la guerre de Sécession, avec une représentation non-agrégée

Dans le cadre du théâtre de l'Ouest, les différentes phases de l'évolution du territoire conquis par le Nord, présentées dans la Figure 7.4, sont un peu plus visibles :

- Au niveau de la région des Appalaches, on observe un déplacement global vers le sud des lieux de conflits au cours du temps (Repère 2 dans la Figure 7.9). Ce déplacement se

poursuit au niveau d'Atlanta, où l'on observe un rapide déplacement des engagements vers l'océan, correspondant à la *Marche vers la Mer* de Sherman (Repère 4 dans la Figure 7.9).

- Au niveau du Mississippi (Repère 3 dans la Figure 7.9), les événements les plus anciens sont localisés au nord du fleuve ainsi qu'à la Nouvelle-Orléans, et correspondent aux premières avancées des troupes nordistes en 1862 dans la Figure 7.4. Les événements les plus récents correspondant à la prise de Vicksburg et de Port Hudson sont visibles sur les zones 3₁ et 3₂ de la Figure 7.9.

7.4.1.3 Identification des différents épisodes du conflit dans le théâtre de l'Ouest, à l'aide d'une représentation non-agrégée

Si la retranscription des différentes phases du conflit, présentées dans la Figure 7.4 semble assez malaisée avec une représentation non-agrégée des combats, celle-ci permet cependant de retranscrire plus précisément le déroulement des différentes campagnes dans un théâtre d'opérations.

Dans la Figure 7.10, nous comparons une carte présentant le récit des opérations du théâtre occidental de la guerre de Sécession, avec la représentation obtenue avec notre interface de visualisation. Nous observons qu'il est possible de retrouver le déroulement des différentes campagnes ayant eu lieu sur le théâtre de l'Ouest, présentées sur la carte de référence. Par ordre chronologique, on observe :

- la prise de La Nouvelle-Orléans à l'embouchure du Mississippi (3₁) ;
- l'avancée de l'armée de l'*Union* au nord du Mississippi (3₂ et 3₃) ;
- l'invasion du Kentucky par l'armée confédérée (3₄), suivie de sa poursuite par l'armée nordiste au Tennessee (3₅), puis à Chattanooga (3₆) ;
- parallèlement, la prise de Vicksburg (3₇) et de Port Hudson (3₈) par l'*Union* ;
- la prise d'Atlanta (3₉) et la *Marche vers la Mer* (3₁₀) de Sherman ;
- la bataille de la baie de Mobile (3₁₁) ;
- la campagne des Carolines (3₁₂).

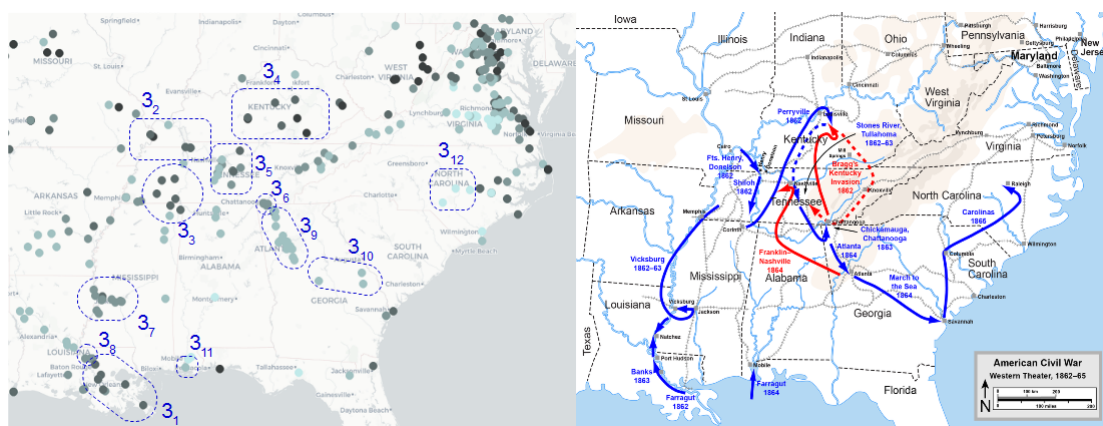


FIGURE 7.10 – Analyse du déroulement des campagnes sur le théâtre de l'Ouest [Jespersen, 2012c]

7.4.1.4 Identification des différentes phases du conflit dans la fenêtre spatiale, à l'aide d'une représentation agrégée

La Figure 7.11 montre une capture d'écran de l'interface de géovisualisation une fois les événements représentés de manière agrégée.

Cette représentation agrégée offre moins de détails que la représentation non-agrégée sur le lieu des différents engagements. Aussi le déroulement précis de chaque campagne est plus difficile à visualiser que dans la Figure 7.10.

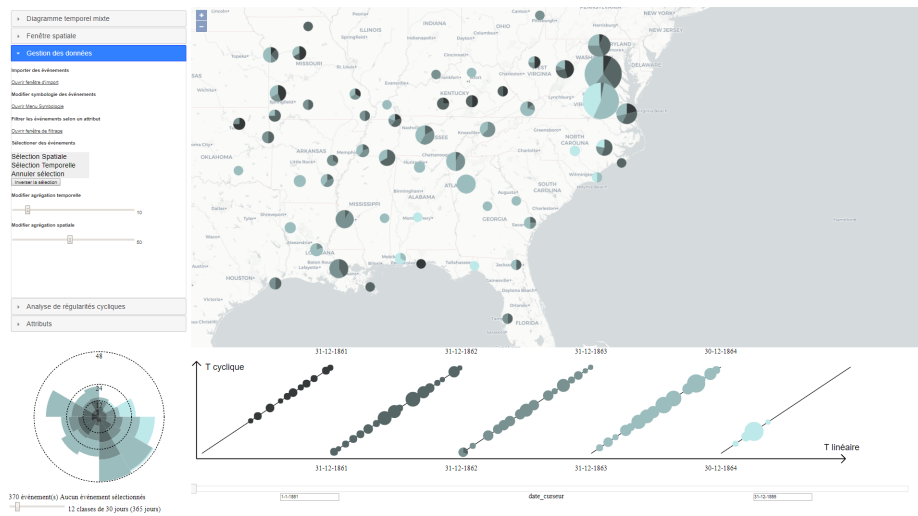


FIGURE 7.11 – Visualisation des batailles de la guerre de Sécession (représentation agrégée)

Cependant, en réduisant le nombre d'entités visibles sur la carte, tout en gardant visible la proportion des engagements selon les différentes années du conflit, elle permet une meilleure identification de l'avancée de l'Union dans les territoires du Sud (présentée dans la Figure 7.4).

Au niveau de la ligne des Appalaches (Repère 2 Figure 7.12), on identifie l'avancée de l'Union au travers d'un dégradé de la valeur de saturation du nord vers le sud, qui se poursuit avec des valeurs de saturation faible au niveau d'Atlanta et de la Géorgie (Repère 4 Figure 7.12).

Sur le théâtre d'opérations du Mississippi (Repère 3 Figure 7.12), on observe les valeurs les plus sombres au nord du fleuve et à son embouchure. Les valeurs les plus claires sont observables au niveau de Vicksburg (Repère 3₁), et également au niveau de l'embouchure du fleuve (Repère 3₂), Port Hudson étant situé non loin de la Nouvelle-Orléans.

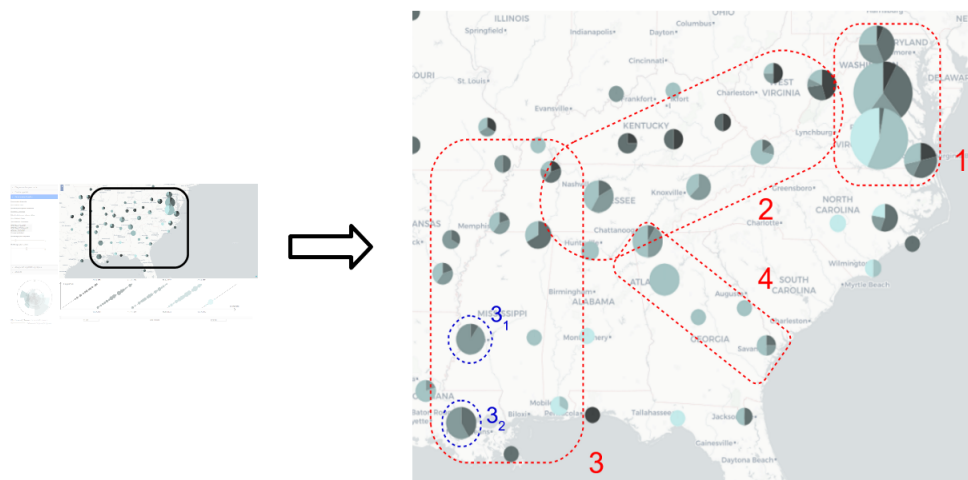


FIGURE 7.12 – Analyse des différentes phases de la guerre de Sécession, avec une représentation agrégée

7.4.1.5 Comparaison des représentations agrégées et non-agrégée des événements dans le théâtre de l'Est

C'est particulièrement au niveau du théâtre de l'Est (Repère 1 de la Figure 7.12) que la représentation agrégée des événements offre une meilleure possibilité d'analyse que la représentation non-agrégée.

Ce théâtre d'opérations a été le principal lieu d'affrontement durant le conflit. Cette région peu étendue spatialement a été le lieu de plusieurs campagnes militaires infructueuses, à l'initiative des deux camps, dans le but de capturer la capitale adverse. Les offensives successives de chacun des deux camps ont provoqué le déplacement de la ligne de front dans les deux directions de l'axe nord-sud, reflétant les avancées et retraits des deux camps lors de ces campagnes, tout au long du conflit. Ce n'est qu'à partir de 1864 que la ligne de front se déplace définitivement vers le sud.

Ainsi, bien que l'avancée progressive de l'*Union* vers le sud se soit étalée sur trois années, certains engagements du début de la guerre se sont déroulés plus au sud que d'autres engagements survenus plus tard dans le conflit. Combiné au grand nombre d'événements localisés dans une zone étroite de la carte, ceci rend plus difficile l'identification de l'évolution globale du territoire contrôlé par l'*Union* dans ce théâtre d'opérations. La visualisation d'une propagation globale des événements vers le sud est "entravée" par l'apparition d'événements récents apparaissant dans le nord de la zone d'étude.

Une représentation non-agrégée des événements produit ainsi un amas de points superposés, de différentes valeurs, ne permettant pas l'observation d'un gradient de valeur selon un axe directionnel. Cette représentation rend difficile l'identification d'une évolution globale du territoire occupé par l'*Union* au cours de la guerre (Repère C de la Figure 7.13).

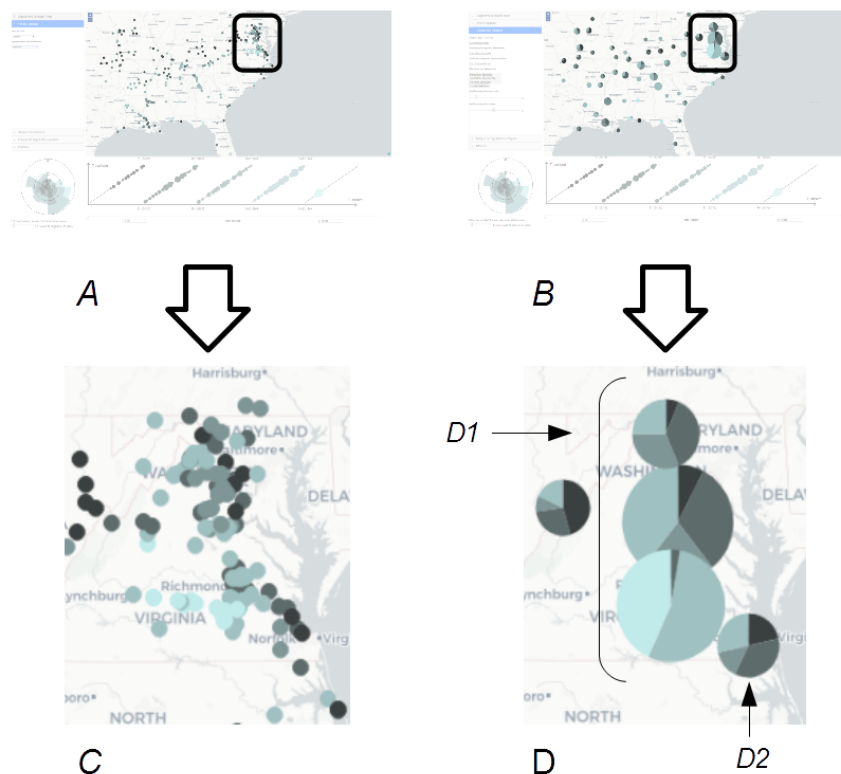


FIGURE 7.13 – Comparaison représentations non-agrégée et agrégée

À l'inverse une représentation agrégée permet de limiter le nombre d'entités graphiques visibles à l'écran, en regroupant les événements spatialement proches. Les événements du théâtre d'opérations étant localisés le long d'un axe nord-sud, les agrégats formés regroupent ces événements selon différentes plages de distance le long de cet axe. La représentation des agrégats sous la forme de *Pie Chart* permet alors de représenter à quel moment du conflit ont eu la majorité des affrontements survenant dans chaque plage de distance de l'axe nord-sud.

La représentation agrégée permet alors d'identifier une propagation globale des événements du nord vers le sud, témoignant d'un déplacement progressif de la ligne de front vers le sud au cours du conflit (Repère *D* Figure 7.13). Plus l'entité graphique est située vers le sud, plus la proportion de couleurs claires, représentant les combats les plus récents, est importante (Repère *D1* Figure 7.13). Ce gradient de valeur, du nord vers le sud, peut être interprété comme une propagation spatiale des événements au cours de l'intervalle temporel couvert par le conflit.

On observe également un cluster spatio-temporel formé par des événements apparaissant à une date antérieure aux événements situés plus au nord-ouest (Repère *D2* Figure 7.13). L'emplacement de ce cluster rassemblant des événements majoritairement plus anciens semble contredire l'hypothèse selon laquelle plus les batailles ont eu lieu vers le sud, plus elles sont anciennes.

La raison de l'existence de ce cluster est que les batailles concernées correspondent à une tentative de l'*Union*, appelée *Campagne de la Péninsule*, de contourner les défenses confédérées par un débarquement dans la péninsule de Virginie, située plus au sud que la ligne de front principale. Cette campagne s'est déroulée de *mars 1862* à *juillet 1862*. Les batailles correspondantes sont alors représentées par une couleur plus sombre que les batailles ayant eu lieu plus au nord-ouest qui, quant à elles, se sont déroulées en *1864*.

7.4.2 Analyse des engagements du théâtre de l'Est

7.4.2.1 Première analyse : engagements du théâtre de l'Est de l'*Overland Campaign* à la bataille d'Appomattox

La Figure 7.14 montre une capture d'écran de l'interface de géovisualisation une fois l'emprise spatiale réduite à la zone du théâtre de l'Est et l'emprise temporelle réduite à la période allant du *1er mai 1864* au *30 avril 1865*.

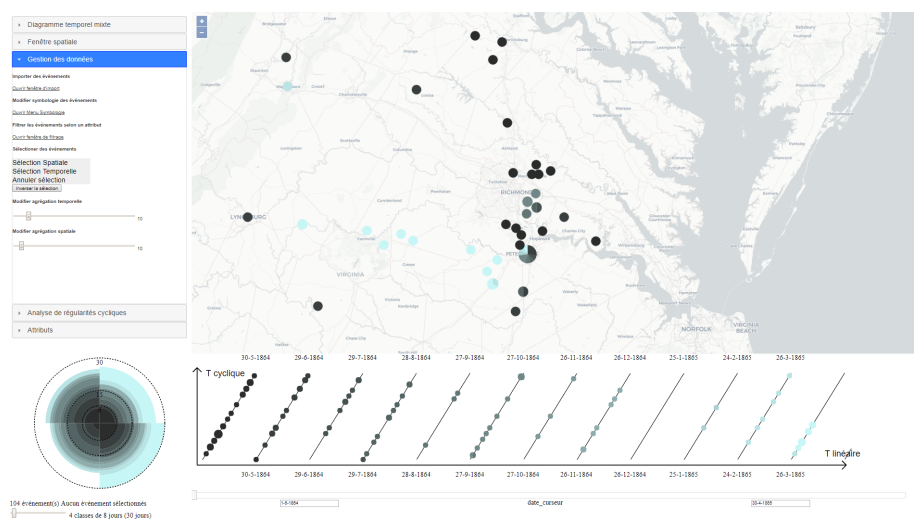


FIGURE 7.14 – Visualisation des engagements du théâtre de l'Est, du *1er mai 1864* au *30 avril 1865*

Les événements sont représentés de manière agrégée selon leur moment d'apparition relatif au temps linéaire, selon douze intervalles temporels d'un mois. Une représentation agrégée est choisie afin que les entités représentant les engagements du siège de Petersburg, apparaissant dans une zone spatiale restreinte, ne se superposent pas. La distance critique déterminant l'agrégation d'événements spatialement proche est réduite par rapport aux analyses précédentes.

Les entités spatiales les plus sombres correspondent à la première phase de l'offensive nordiste, et peuvent être regroupées en deux clusters correspondant à :

- l'avancée des troupes de l'*Union* au cours de l'*Overland Campaign*, s'étendant de la *Rapidan River* à la *James River* selon un axe nord-nord-ouest/sud-sud-est (Repère 1 de la Figure 7.15) ;
- le débarquement des troupes nordistes au cours de la campagne de *Bermuda Hundred*, localisé sur une zone restreinte proche de la *James River* (Repère 2 de la Figure 7.15).

Deux clusters regroupant des entités de couleur moins sombre sont localisés dans des zones spatiales très restreintes au nord et au sud de la *James River*. Ces clusters correspondent à la fixation des combats autour de Richmond (Repère 3 de la Figure 7.15) et Petersburg (Repère 4 de la Figure 7.15) au cours du siège de ces villes.

Enfin, on observe à l'est de Petersburg un cluster regroupant des entités de couleur claire, regroupée le long d'un axe ouest-nord-ouest/est-sud-est. Ce cluster correspond aux engagements de la campagne d'Appomatox (Repère 5 de la Figure 7.15).



FIGURE 7.15 – Analyse des engagements du théâtre de l'Est, du 1^{er} mai 1864 au 30 avril 1865

7.4.2.2 Seconde analyse : engagements du théâtre de l'Est du 1^{er} mai 1864 au 1^{er} juillet 1864

Dans la Figure 7.16, nous comparons une carte présentant le récit de l'*Overland Campaign*, avec la représentation obtenue avec notre interface des événements allant du 1^{er} mai 1864 au 1^{er}

juillet 1864. Les événements sont représentés selon neuf intervalles temporels d'une semaine.

On observe trois éléments graphiques pouvant être mis en comparaison avec les éléments de la carte de référence :

- un dégradé de valeur des entités du nord vers le sud selon un axe nord-nord-ouest/sud-sud-est, représentant l'avancée des troupes nordistes au cours de l'*Overland Campaign* (Repères 1 et 4 de la Figure 7.16) ;
- un cluster spatio-temporel regroupant des entités de couleur sombre entre les villes de Richmond et de Petersburg, correspondant aux engagements de la campagne de *Bermuda Hundred*, survenant au même moment que débute l'*Overland Campaign* (Repères 2 et 5 de la Figure 7.16) ;
- un cluster spatio-temporel regroupant des entités de couleur claire au sud de la *James River*, correspondant aux derniers affrontements de l'*Overland Campaign* (Repères 3 et 6 de la Figure 7.16) ;

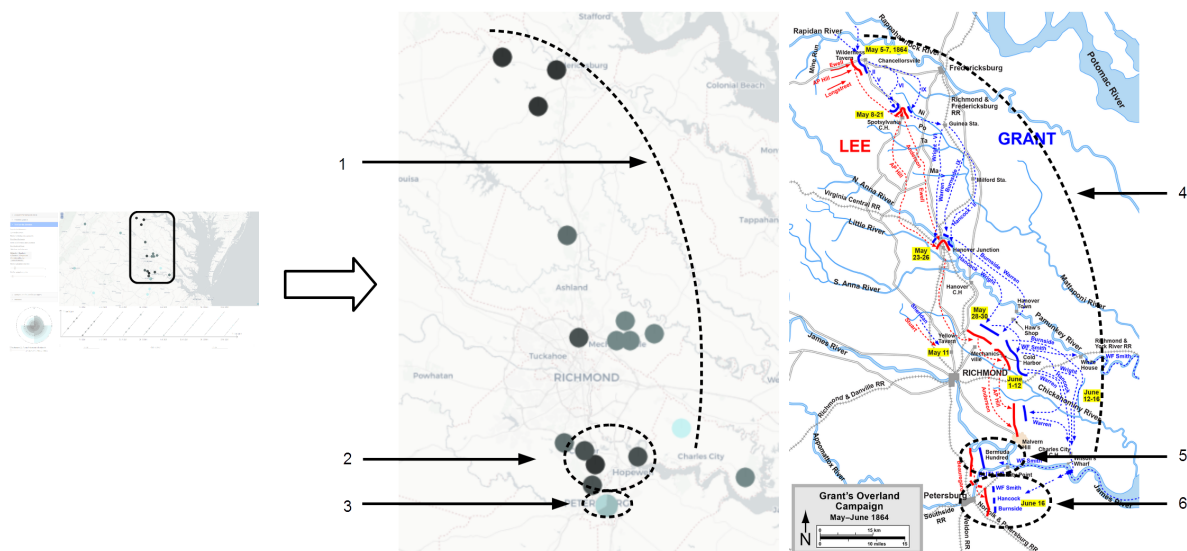


FIGURE 7.16 – Visualisation des engagements du 1er mai 1864 au 1er juillet 1864, et comparaison avec un récit existant de l'*Overland Campaign* [Jespersen, 2012b]

7.4.2.3 Troisième analyse : engagements du théâtre de l'Est du 2 avril 1865 au 9 avril 1865

Dans la Figure 7.17, nous comparons une carte présentant le récit de la campagne d'*Appomattox*, avec la représentation obtenue avec notre interface des événements allant du 2 avril 1865 au 9 avril 1865. Les événements sont représentés selon sept intervalles temporels d'une journée.

On observe un dégradé de la valeur des entités de l'est vers l'ouest. Ce dégradé correspond aux engagements ponctuant la poursuite des armées confédérées par l'*Union*, depuis l'évacuation de Petersburg jusqu'à la bataille d'*Appomattox Court House*, comme le représente la carte de référence dans la Figure 7.17.

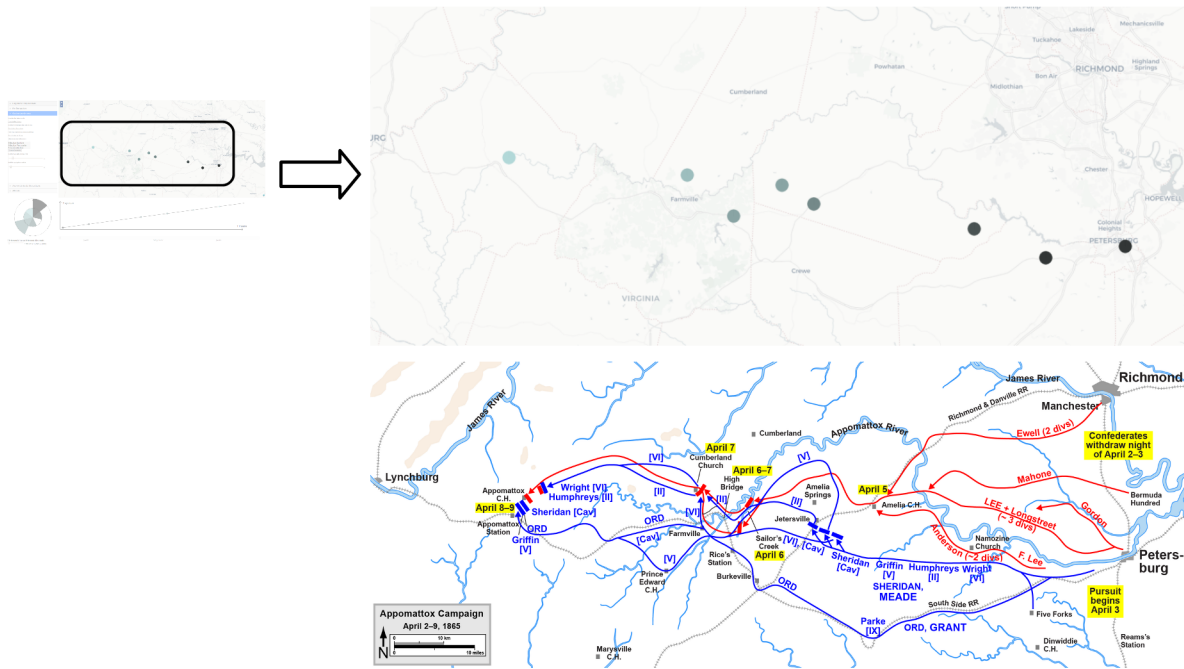


FIGURE 7.17 – Visualisation des engagements du 2 avril 1865 au 9 avril 1865, et comparaison avec un récit existant de la campagne d'Appomattox [Jespersen, 2012c]

7.5 Synthèse

L'analyse des batailles de la guerre de Sécession au moyen de *GrAPHiST* a permis la création de récits de différentes phases du conflit, au travers de l'identification visuelle de clusters spatio-temporels relatifs au temps linéaire et de l'identification visuelle de propagation spatiale d'événements au cours du temps.

L'analyse de la distribution spatio-temporelle des différentes zones d'engagement et de leur déplacement au cours du temps a notamment permis une description :

- de l'évolution globale du conflit et de l'avance progressive des troupes de l'Union en territoire confédéré (Figure 7.12) ;
- des différentes phases du théâtre occidental (Figure 7.10) ;
- des différentes campagnes de l'offensive nordiste de 1864-1865 dans le théâtre oriental (Figure 7.15) ;
- du déplacement des troupes au cours de l'*Overland Campaign* (Figure 7.16) et de la campagne d'*Appomattox* (Figure 7.17).

Les utilisations comparées des représentations agrégée et non-agrégée des événements, au niveau du *théâtre d'opérations de l'est*, ont montré l'avantage d'une représentation des agrégats d'événements sous la forme de *Pie Chart* pour identifier les structures principales dans la distribution spatio-temporelle de larges séries d'événements (Section 7.4.1.5). La Figure 7.13 montre que la représentation agrégée permet d'éviter la surcharge d'entités graphiques et leur superposition sur la carte, permettant de mettre en valeur l'évolution globale du territoire contrôlé par l'Union au cours du conflit.

L'analyse exploratoire des données relatives aux batailles de la guerre de Sécession, au travers de notre outil, dans le but d'identifier des patterns spatio-temporels relatifs au temps linéaire,

permet ainsi la reconstitution du récit de cet épisode historique. Nous émettons l'hypothèse, qu'une analyse exploratoire similaire sur d'autres données permet de créer le récit d'autres épisodes historiques, et donc de créer de l'information sur ces épisodes.

Les possibilités d'application peuvent être multiples. L'analyse de données relatives à des cas d'épidémie peut permettre de reproduire le récit de la propagation d'une épidémie et d'effectuer de possibles hypothèses sur ses facteurs. L'analyse de données relatives à une culture archéologique peut permettre de reproduire le récit de sa diffusion au cours de l'Histoire.

Chapitre 8

Identification de récurrences cycliques spatialisées dans les données événementielles

Sommaire

8.1	Objectif	296
8.1.1	Identifier des patterns spatio-temporels relatifs au temps cyclique dans les séries événementielles	296
8.1.2	Le choix des données du cas d'étude	298
8.2	Données	299
8.2.1	Nature des données	299
8.2.2	Construction des données	299
8.2.3	Implémentation dans notre modèle de données	303
8.3	Procédé	304
8.3.1	Principe	304
8.3.2	Les différents jeux de données	305
8.3.3	Scénario d'analyse	306
8.4	Résultats	307
8.4.1	Analyse des jeux de données du groupe <i>A</i>	307
8.4.2	Analyse des jeux de données du groupe <i>B</i>	316
8.5	Limites et synthèse	323
8.5.1	Distribution temporelle des événements cycliques dans les jeux de données	323
8.5.2	Conclusion	324

8.1 Objectif

8.1.1 Identifier des patterns spatio-temporels relatifs au temps cyclique dans les séries événementielles

L'objectif de l'expérimentation décrite dans ce chapitre est de valider l'hypothèse selon laquelle notre interface de visualisation permet une analyse exploratoire, rendant possibles la recherche et l'identification de patterns relatifs au temps cyclique dans des séries événementielles. Ces patterns concernent :

- la présence de clusters spatio-temporels selon le temps cyclique ;
- la présence de récurrences cycliques spatialisées.

8.1.1.1 Identification visuelle de patterns relatifs au temps cyclique

Tel qu'il a été conçu et construit, *GrAPHiST* permet de visualiser le caractère cyclique d'une série d'événements lorsque la période correspondante (ou celle correspondant à un harmonique du cycle) est représentée par l'échelle cyclique du diagramme temporel mixte. Un cluster ou une récurrence cyclique peuvent alors être identifiés au sein des représentations graphiques de l'interface de géovisualisation, comme nous le présentons dans le chapitre 6.

Une concentration des événements dans le temps cyclique (cluster temporel relatif au temps cyclique) est identifiable par :

- la présence d'une ou de plusieurs sections angulaires présentant un rayon plus important que la moyenne dans le diagramme circulaire ;
- la concentration d'entités graphiques le long de l'axe des ordonnées dans le diagramme mixte en dents de scie, ou dans les mêmes sections de sinusöïde dans le diagramme mixte sinusöïdal.

Dans le cas d'une représentation des événements selon leur moment d'apparition relatif au temps cyclique, cette concentration (spatio-)temporelle (cluster spatio-temporel relatif au temps cyclique) peut être identifiée sur la carte par une concentration spatiale d'entités de même couleur.

Cette concentration temporelle dans le temps cyclique peut alors former la partie visible, dans le diagramme temporel circulaire ou dans la carte, d'une récurrence cyclique régulière, ayant pour période la durée de l'échelle cyclique représentée. Une fois identifiée une concentration des événements dans le temps cyclique, une récurrence cyclique de ces événements peut être identifiée par un alignement des entités graphiques correspondantes dans le diagramme temporel mixte, parallèle à l'axe des abscisses.

8.1.1.2 Enjeux de l'expérimentation

Identification de cycles de période inconnue. L'analyse de patterns spatio-temporels cycliques dans les données événementielles peut s'effectuer en fixant à l'avance la durée de l'échelle cyclique représentée dans le diagramme temporel mixte. La recherche de structures spatio-temporelles cycliques est effectuée par après, pour une série d'échelles cycliques choisie arbitrairement par l'utilisateur.

Cette approche, utilisée dans *Time Wave* [Li, 2010] et *PerSE* [Swedberg and Peuquet, 2017], peut être justifiée pour l'analyse de phénomènes au travers d'échelles calendaires, dont on soupçonne une correspondance avec la distribution temporelle des événements correspondants. L'existence de patterns dans la distribution spatio-temporelle d'événements relatifs à des phénomènes climatiques peut ainsi être analysée au travers d'une échelle cyclique d'une année, reflétant le

cycle des saisons.

Cette approche convient notamment à l'analyse de clusters spatio-temporels, où le but est moins d'identifier si un phénomène possède un caractère cyclique, que d'analyser la distribution spatiale d'événements considérés comme cycliques a priori, et la relation entre la zone d'apparition des événements et leur moment d'apparition au sein de l'échelle cyclique représentée.

Un des enjeux de cette expérimentation est au contraire d'évaluer la possibilité d'identifier un pattern relatif à une échelle cyclique, donc la durée serait inconnue.

Pour permettre une analyse exploratoire de ces patterns, la représentation visuelle ne doit pas uniquement permettre d'identifier le caractère cyclique d'une série d'événements, lorsque la période correspondante est représentée. La représentation visuelle doit permettre à l'utilisateur d'évaluer, lorsqu'il modifie l'échelle cyclique représentée dans le diagramme mixte, si cette dernière s'approche ou s'éloigne d'une période pour laquelle un cycle est observable. Dans ces conditions, l'utilisateur peut identifier un cycle en modifiant l'échelle cyclique, pas à pas, vers la durée de la période correspondante. L'expérimentation décrite dans ce chapitre doit permettre d'évaluer si cette analyse exploratoire est possible.

Comme nous l'avons montré dans le chapitre 6 l'identification de clusters spatio-temporels relatifs au temps cyclique est une des étapes dans la recherche de récurrences cycliques au moyen de notre interface. Aussi, la recherche de cycles de période inconnue sert ici de cadre à l'identification des deux types de patterns.

Identification de patterns cycliques dans des données réelles. Un autre enjeu de l'expérimentation de valider si l'analyse exploratoire offerte par notre outil est en mesure de rechercher et d'identifier des patterns relatifs au temps cyclique au sein de données réelles.

Les événements présentant un caractère cyclique au sein de ces jeux de données sont alors susceptibles :

- de posséder une dispersion spatiale plus ou moins élevée, et d'apparaître dans des zones spatiales plus ou moins étendues ;
- de posséder une dispersion temporelle plus ou moins élevée, et d'apparaître durant des intervalles temporels cycliques plus ou moins étendus.

Ces jeux de données réels peuvent également comporter une proportion plus ou moins élevée d'événements aléatoires.

L'expérimentation doit donc permettre d'évaluer si notre interface de visualisation permet l'identification de patterns cycliques au sein de tels jeux de données.

8.1.1.3 Identification de récurrences cycliques spatialisées

Afin de valider notre hypothèse, nous décidons d'effectuer une série d'analyses sur différents jeux de données. Ces analyses ont pour but d'identifier, pour chaque jeu de données, une récurrence cyclique spatialisée.

Afin d'évaluer la possibilité d'utiliser notre interface sur des données réelles, les jeux de données doivent présenter des caractéristiques différentes. Celles-ci concernent :

- la dispersion temporelle des événements cycliques, soit la durée de l'intervalle temporel cyclique durant laquelle ces événements apparaissent (un événement apparaît tous les x jours, plus ou moins k jours) ;

- la dispersion spatiale des événements cycliques, soit l'étendue de la zone spatiale dans laquelle ces événements apparaissent ;
- le bruit, soit la proportion d'événements de composante temporelle et spatiale aléatoire dans le jeu de données.

8.1.2 Le choix des données du cas d'étude

Notre but est de rechercher des cycles localisés dans les séries événementielles. Il nous faut donc utiliser des séries d'événements susceptibles de présenter de telles structures spatio-temporelles. Certains jeux de données réels sont connus pour présenter des caractéristiques cycliques, comme c'est le cas pour des données liées à des phénomènes météorologiques. Les phénomènes de crue, par exemple, présentent des cycles annuels, décennaux, ou centennaux.

Or, la connaissance de l'existence de ces cycles et de leur période nous influence dans l'analyse de telles séries événementielles. Afin d'évaluer la possibilité d'identifier des cycles de période inconnue, au moyen de l'analyse exploratoire offerte par notre interface, nous devons utiliser des jeux de données nous empêchant d'être influencés par une connaissance a priori ou une suspicion des cycles recherchés.

L'expérimentation décrite dans ce chapitre nécessite alors l'utilisation de jeux de données présentant des cycles, mais dont la période de réapparition est inconnue.

Comme évoqué précédemment ces jeux de données doivent également présenter différentes valeurs de dispersion temporelle et spatiale de leurs événements cycliques, ainsi qu'une proportion plus ou moins importante d'événements aléatoires.

Cette expérimentation est effectuée dans le cadre de l'évaluation fonctionnelle de l'outil. L'objectif n'est pas d'analyser un phénomène réel dans le but d'en découvrir un caractère cyclique inconnu, mais de montrer que de telles caractéristiques seraient observables au moyen de l'outil. Le caractère réel des jeux de données à utiliser implique surtout la présence d'événements cycliques présentant une dispersion spatiale et une dispersion temporelle plus ou moins élevées dans les jeux de données à tester, ainsi qu'une proportion plus ou moins importante d'événements aléatoires. Si ces conditions sont réunies, le fait que les données fassent effectivement référence à un phénomène réel n'est pas une condition nécessaire à l'expérimentation.

Ainsi, nous utilisons dans le cadre de cette expérimentation des séries événementielles fictives, que nous créons de manière procédurale. Ces jeux de données comportent une série d'événements cyclique, de période inconnue et définie aléatoirement lors de la création de chaque jeu de donnée. Afin d'imiter des jeux de données réels, les événements de chaque série d'événements cyclique possèdent une dispersion temporelle et une dispersion spatiale plus ou moins grande. Chaque jeu de données possède également une proportion d'événements aléatoire. Ces trois critères peuvent être paramétrés pour la création de chaque jeu de données.

Cette approche nous permet d'avoir à disposition une série de jeux de données, dont nous pouvons être sûrs qu'ils présentent un caractère cyclique, mais dont nous ignorons la période, car celle-ci est choisie aléatoirement. Cette approche nous permet également de créer des jeux de données correspondant à différentes valeurs de dispersion spatiale et temporelle, et à différentes proportions d'événements aléatoires dans le jeu de données.

La durée de la période d'une série cyclique est sauvegardée lors de la création d'un jeu de données, mais n'est pas transmise directement à l'expérimentateur. Le but de l'expérimentation est alors d'identifier cette série cyclique au moyen de l'outil, et une fois l'expérimentation

effectuée, de comparer les résultats trouvés avec les informations stockées en base.

Nous émettons alors l'hypothèse que si des cycles sont identifiables dans ces jeux de données fictifs, imitant le réel, des cycles équivalents seront identifiables dans des jeux de données réels.

8.2 Données

8.2.1 Nature des données

Les jeux de données fictifs créés dans le cadre de cette expérimentation regroupent des événements ponctuels dans le temps et l'espace. Les événements créés ne possèdent pas d'attributs, la dimension thématique n'étant pas prise en compte dans l'analyse. Les composantes temporelle et spatiale des événements de chaque jeu de données sont créées de manière procédurale. Les événements de chaque jeu de données sont considérés comme formant un phénomène distinct.

Chaque jeu de données est composé de deux séries d'événements :

- une série d'événements apparaissant de manière cyclique, dans une même zone de l'espace ;
- une série d'événements apparaissant aléatoirement dans le temps et l'espace.

Les cycles présents dans les jeux de données se caractérisent par la réapparition d'un ou de plusieurs événements, dans des intervalles temporels réguliers. La période du cycle, son nombre d'instances, la date où le cycle débute et le nombre d'événements apparaissant lors d'une instance du cycle sont définis aléatoirement.

Pour chaque jeu de données créé, différentes informations sont stockées en base et cachées à l'utilisateur. Ces informations concernent :

- la période de la série événementielle cyclique ;
- l'emprise temporelle de la série événementielle cyclique ;
- l'emprise spatiale de la série événementielle cyclique.

Chaque jeu de données correspond une valeur particulière prise par trois paramètres intervenant dans la création du jeu de données. Ces paramètres sont :

- la dispersion temporelle des événements cycliques, ou la durée des intervalles temporels réguliers où apparaissent ces événements ;
- la dispersion spatiale des événements cycliques, ou la taille de la zone spatiale où apparaissent ces événements ;
- la proportion d'événements aléatoires dans le jeu de données.

8.2.2 Construction des données

8.2.2.1 Délimitation de l'étendue d'étude

Les événements que nous simulons apparaissent sur un territoire et durant un intervalle de temps limités.

L'étendue spatiale choisie pour l'expérimentation correspond à une zone comprise entre les parallèles 41^e nord et 50^e nord, et les méridiens 5^e ouest et 9^e est.

La carte de notre fenêtre spatiale utilise une projection *WGS 84/Pseudo-Mercator*, standard *de facto* des représentations cartographiques sur le *Web*. Dans cette projection, les méridiens et

les parallèles forment un quadrillage régulier.

Notre zone spatiale d'étude, limitée par des méridiens et des parallèles, correspond à une zone trapézoïdale dans la réalité, mais est représentée par un espace rectangulaire dans notre représentation cartographique (Figure 8.1).



FIGURE 8.1 – Emprise de notre zone d'étude dans la fenêtre spatiale

L'étendue temporelle choisie pour l'expérimentation est comprise entre le *1er janvier 1900*, et le *1er janvier 2000*.

Les événements traités dans cette expérimentation étant des événements fictifs, la nature du territoire dans lequel ils apparaissent, et l'époque durant laquelle ils apparaissent, importent peu. Un événement peut apparaître en *2001* ou en *2020*, dans une zone de montagne ou en pleine mer, sans que cela ait une importance pour l'analyse que nous effectuons ici.

Les dates calendaires représentées dans le diagramme et le fond de carte utilisés dans la fenêtre spatiale ont pour unique tâche d'offrir des points de repère pour l'analyse visuelle des événements dans l'interface.

8.2.2.2 Construction des séries d'événements cycliques

Composante spatiale des événements cycliques. Les événements cycliques apparaissent à l'intérieur d'une zone spatiale définie par :

- un centre C correspondant à un point ayant pour coordonnées géographiques $longitude_C$ et $latitude_C$;
- la dispersion spatiale $disp_s$ de la série d'événements, définissant la taille de la zone.

Méridiens et parallèles formant un quadrillage dans une projection en *WGS 84/Pseudo-Mercator*, nous exprimons, dans le cadre de nos analyses, la valeur $disp_s$ en degrés, et non en distance métrique. La longitude des événements créés est ainsi comprise entre $longitude_C - disp_s$ et $longitude_C + disp_s$. La latitude des événements créés est comprise entre $latitude_C - disp_s$ et $latitude_C + disp_s$.

Afin d'obtenir une distribution des événements fictifs centrée autour du point C , nous créons les événements dans un cercle de centre C et de rayon $disp_s$ (Figure 8.2).

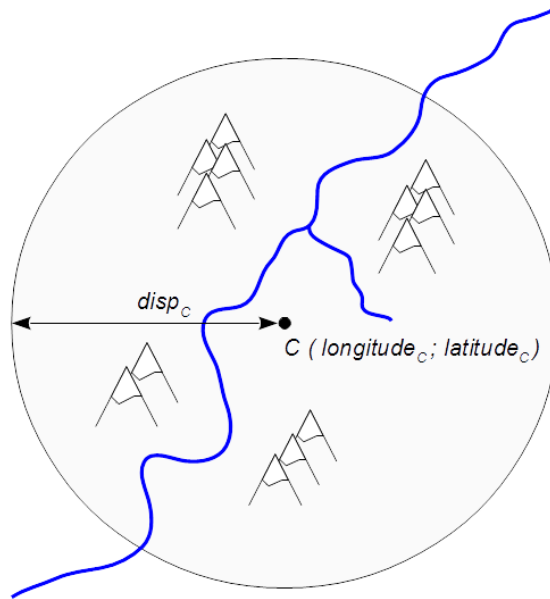


FIGURE 8.2 – Zone d'apparition des événements fictifs

Soient $random_r$ et $random_\theta$ des variables aléatoires choisies entre 0 et 1 pour chaque événement, les coordonnées géographiques $longitude_{evt}$ et $latitude_{evt}$ d'un événement d'une série cyclique sont déterminées par la formule (8.1).

$$\begin{cases} longitude_{evt} = longitude_C + disp_s * random_r * \cos(2 * \pi * random_\theta) \\ latitude_{evt} = latitude_C + disp_s * random_r * \sin(2 * \pi * random_\theta) \end{cases} \quad (8.1)$$

Le quadrillage formé sur la carte par les méridiens et les parallèles n'est cependant pas régulier :

- La distance graphique entre deux méridiens sur carte est constante sur l'ensemble de la carte.
- La distance graphique entre deux parallèles sur carte est égale à celle entre deux méridiens au niveau de l'équateur, puis augmente quand on se rapproche des pôles.

Au niveau de la zone d'étude, la distance graphique entre deux parallèles sur la carte est près de deux fois plus grande que celle entre deux méridiens.

Ainsi, la zone d'apparition des événements ne forme pas un cercle sur la carte, mais une ellipse. La zone décrite ne correspond pas à une zone de forme circulaire dans le monde réel. Elle ne forme une ellipse que dans notre représentation spatiale projetée en *WGS 84/Pseudo-Mercator*.

Composante temporelle des événements cycliques. Les événements cycliques apparaissent durant différents intervalles temporels, réapparaissant régulièrement selon une certaine période (Figure 8.3). Ces intervalles temporels correspondent aux différentes instances, I_n , du cycle observé par la série d'événements. Ces instances apparaissent cycliquement selon une période cyclique P , le nombre total d'instances étant égal à N .

Une série d'événements commence à une date initiale T_0 , date où apparaît la première instance I_0 de la série. Chaque instance I_n de la série d'événements cyclique correspond à une date T_n

calculée à partir de T_0 en fonction du numéro de l'instance en question, n , et de la durée de la période P .

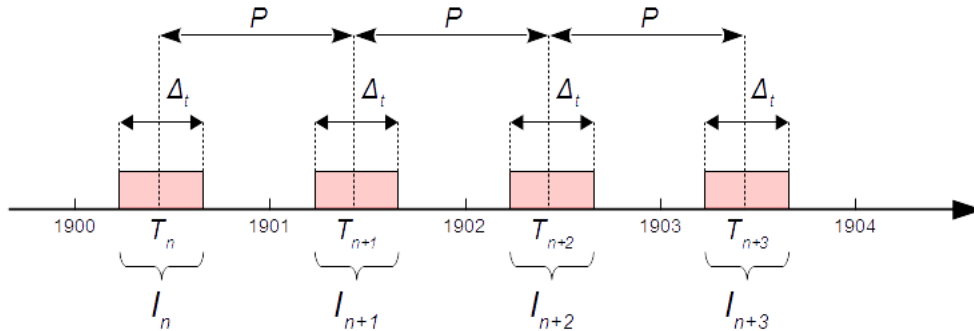


FIGURE 8.3 – Intervalles d'apparition des événements fictifs

Chaque instance I_n correspond à un intervalle temporel f_n centré autour de la date centrale T_n . L'étendue Δ_t de cet intervalle est définie en fonction de P , de manière à ce que $\Delta_t = disp_t * P$, avec $disp_t$ correspondant à la dispersion temporelle des événements. Ce critère peut varier entre 0% et 50%.

Pour chaque instance I_n , un nombre variable d'événement k_n apparait dans l'intervalle f_n . Pour chaque instance, ce nombre est choisi aléatoirement entre deux valeurs k_{min} et k_{max} .

Soit $random_{evt}$ une variable aléatoire choisie entre -1 et 1 pour chaque événement, la date d'apparition t_{evt} des événements d'une série cyclique est déterminée par la formule (8.2).

$$t_{evt} = T_0 + (P * n + random_{evt} * (1/2) * P * disp_t) \quad (8.2)$$

Constantes intervenant dans la création des séries d'événements cycliques. Les termes $longitude_C$, $latitude_C$, $disp_s$, P , N , T_0 , k_{min} , k_{max} et $disp_t$ désignent des constantes, fixées pour la création d'une série d'événements cycliques. Pour chaque série :

- $longitude_C$ est choisi aléatoirement entre -5 et 9 , longitudes des extrêmes de l'étendue spatiale d'étude ;
- $latitude_C$ est choisi aléatoirement entre 41 et 50 , latitudes des extrêmes de l'étendue spatiale d'étude ;
- T_0 est choisi aléatoirement entre le *1er janvier 1900* et le *1er janvier 2000*, soit dans l'intervalle formé par l'étendue temporelle d'étude ;
- P est définie aléatoirement dans un intervalle allant de *180 jours* (environ six mois) à *2005 jours* (environ cinq ans et six mois), et s'exprime en nombre de jours ;
- N est choisi aléatoirement entre 10 et 100 ;
- k_{min} est fixé à 1 ;
- k_{max} est fixé à 10 .

Les valeurs de $longitude_C$, $latitude_C$, T_0 et P sont choisies aléatoirement pour chaque série d'événements cycliques, à l'intérieur d'un intervalle fixe et commun à tous les jeux de données.

Les valeurs de $disp_s$ et $disp_t$ sont choisies aléatoirement pour chaque série d'événements cycliques, à l'intérieur d'un intervalle dont les bornes sont définies manuellement pour chaque jeu de données.

8.2.2.3 Construction des séries d'événements aléatoires

Composante spatiale des événements aléatoires. Les coordonnées géographiques $longitude_{al}$ et $latitude_{al}$ des événements aléatoires sont choisies aléatoirement dans l'intervalle formé par les longitudes et latitudes extrêmes de la zone spatiale d'étude : -5 et 9 pour la longitude, 41 et 50 pour la latitude.

Composante temporelle des événements aléatoires. La date d'apparition t_{al} des événements aléatoires est choisie aléatoirement dans l'intervalle formée par les bornes de l'étendue temporelle d'étude, soit entre le 1er janvier 1900, et le 1er janvier 2000.

8.2.3 Implémentation dans notre modèle de données

Les séries d'événements aléatoires ainsi que les séries d'événements cycliques sont créées et intégrées dans notre base de données au moyen de scripts SQL, en classant les informations de chaque événement selon la structure décrite par le Pyramid Model. Nous créons pour chaque cas une entité événement, composée d'une unique sous-partie temporelle (Figure 8.4).

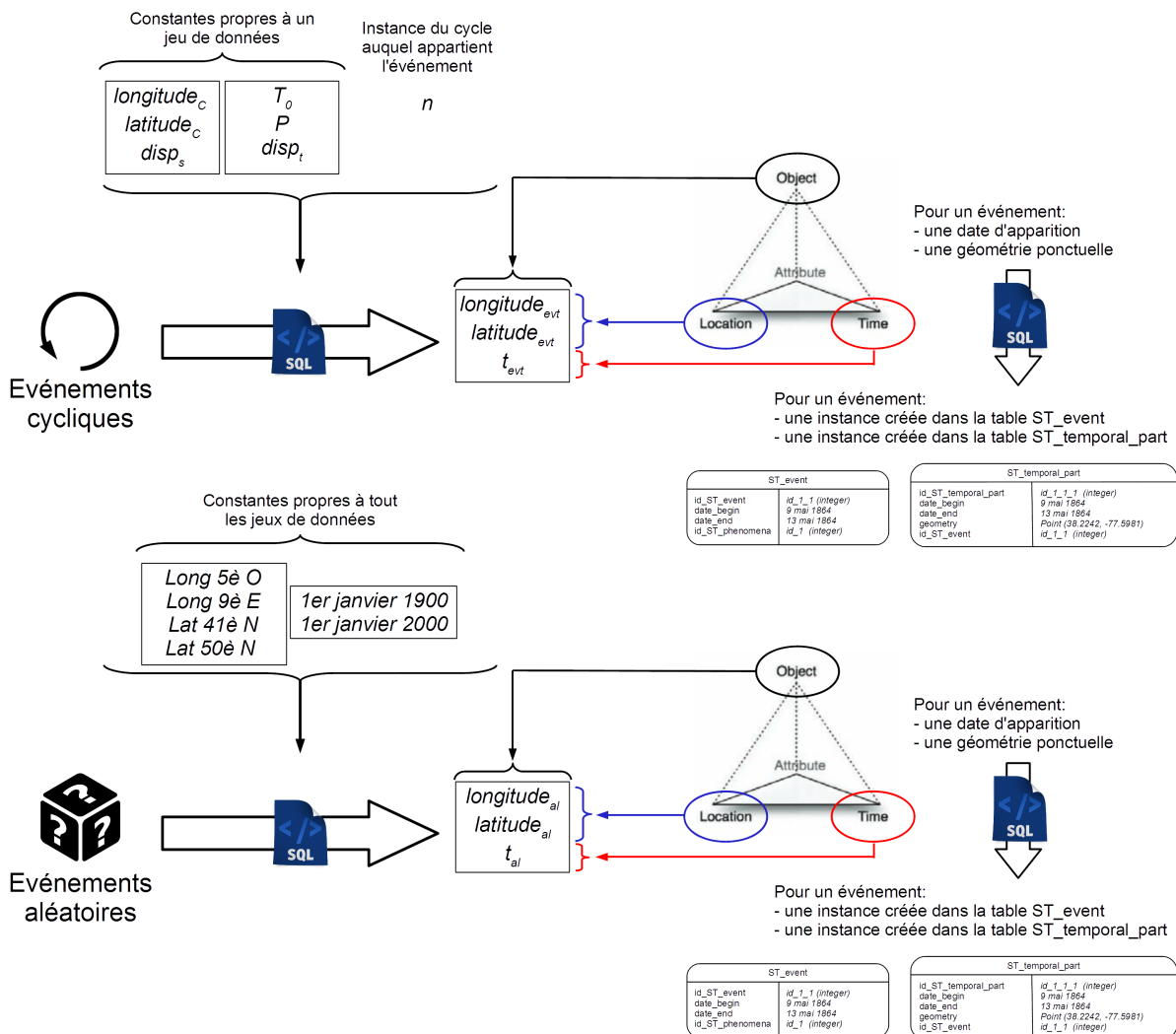


FIGURE 8.4 – Implémentation des données fictives dans notre modèle

Lors de la création d'un jeu de données, les informations relatives à la série d'événements cycliques créée sont stockées en base dans une table dédiée. Ces informations concernent :

- les coordonnées $longitude_C$ et $latitude_C$ du centre C de la zone d'apparition des événements cycliques ;
- leur dispersion spatiale $disp_s$;
- leur dispersion temporelle $disp_t$;
- la période P du cycle ;
- la date T_0 de la première instance du cycle ;
- la date T_N de la dernière instance du cycle.

8.3 Procédé

8.3.1 Principe

Chaque jeu de données fait l'objet d'une analyse exploratoire, dans le but d'identifier la série d'événements cycliques créée, et les caractéristiques de cette série (Figure 8.5). Ces dernières sont :

- l'emprise spatiale des événements cycliques, soient les coordonnées $longitude_C$ et $latitude_C$ du centre de la zone d'apparition des événements cycliques, ainsi que leur dispersion spatiale $disp_s$;
- la période P du cycle ;
- l'emprise temporelle des événements cycliques, soient les dates T_0 et T_N ;
- leur dispersion temporelle $disp_t$.

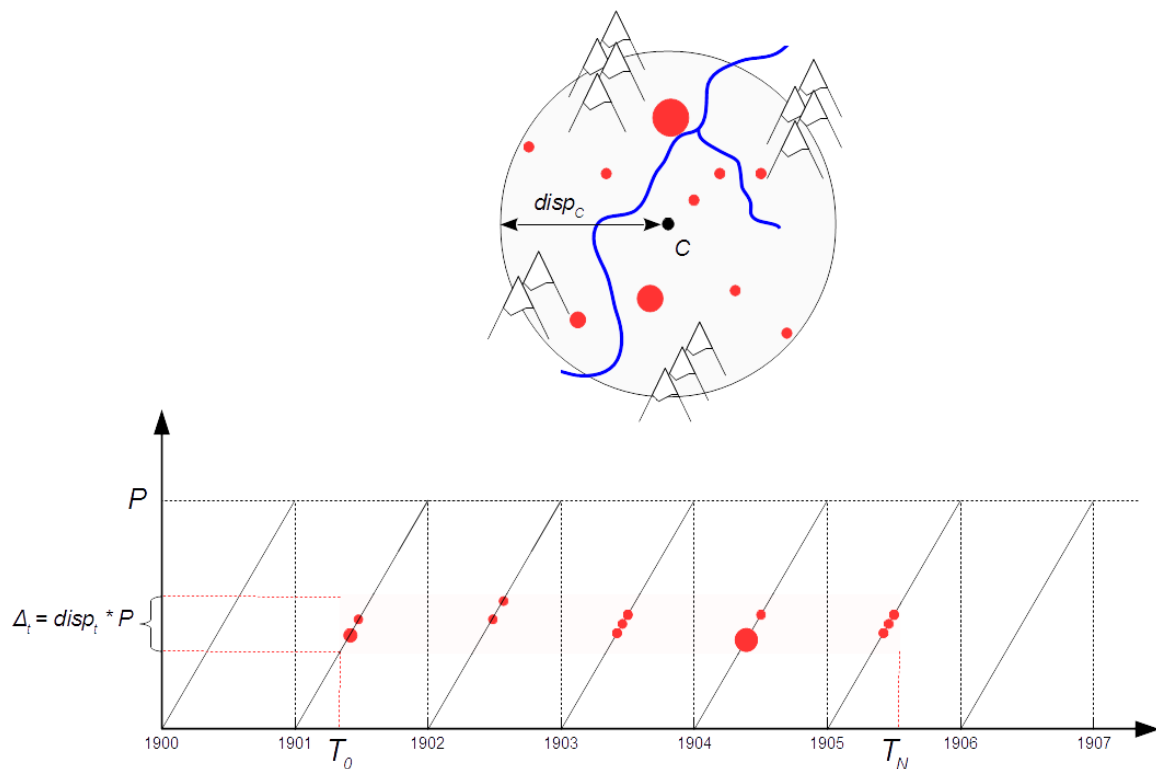


FIGURE 8.5 – Identification des caractéristiques d'une série d'événements cycliques

Une fois l'expérimentation terminée, une comparaison est effectuée entre les résultats trouvés par l'utilisateur grâce à l'interface de visualisation, et les données stockées en base.

8.3.2 Les différents jeux de données

Nous effectuons une analyse exploratoire sur huit jeux de données, correspondant à différentes valeurs de dispersion temporelle et spatiale des événements cycliques, et à différentes proportions d'événements aléatoires.

Les jeux de données sont divisés en deux groupes que nous nommerons A et B :

- les jeux de données du groupe A possèdent quatre événements aléatoires pour un événement cyclique ;
- les jeux de données du groupe B possèdent neuf événements aléatoires pour un événement cyclique.

Les événements des séries cycliques possèdent une dispersion temporelle choisie aléatoirement :

- entre 8% (intervalle temporel d'un mois pour une période d'une année) et 25% (intervalle temporel de trois mois pour une période d'une année) pour les jeux A_1 , A_3 , B_5 , et B_7 ;
- entre 25% (intervalle temporel de trois mois pour une période d'une année) et 50% (intervalle temporel de six mois pour une période d'une année) pour les jeux A_2 , A_4 , B_6 , et B_8 .

Les événements des séries cycliques possèdent une dispersion spatiale choisie aléatoirement :

- entre 2.5 et 4 degrés pour les jeux A_1 , A_2 , B_5 , et B_6 ;
- entre 1.5 et 2.5 degrés pour les jeux A_3 , A_4 , B_7 , et B_8 .

Le nombre d'événements contenu dans chaque jeu de données est aléatoire. Celui-ci dépend :

- du nombre d'événements cycliques créés, dépendant :
 - du nombre d'instances du cycle, défini aléatoirement pour chaque jeu de données ;
 - du nombre d'événements apparaissant lors d'une instance du cycle, défini aléatoirement pour chaque instance.
- de la proportion d'événements aléatoires dans le jeu de données.

Jeu	Prop. d'evts aléatoires	$disp_t$	$disp_s$	Nbre d'evts
A_1	80%	[8%,25%]	[2.5 degrés,4 degrés]	285
A_2	80%	[25%,50%]	[2.5 degrés,4 degrés]	490
A_3	80%	[8%,25%]	[1.5 degrés,2.5 degrés]	507
A_4	80%	[25%,50%]	[1.5 degrés,2.5 degrés]	290
B_5	90%	[8%,25%]	[2.5 degrés,4 degrés]	610
B_6	90%	[25%,50%]	[2.5 degrés,4 degrés]	1010
B_7	90%	[8%,25%]	[1.5 degrés,2.5 degrés]	570
B_8	90%	[25%,50%]	[1.5 degrés,2.5 degrés]	880

8.3.3 Scénario d'analyse

8.3.3.1 Identification des événements cycliques

La première étape de l'analyse que nous effectuons sur chaque jeu de données consiste à identifier la série d'événements cyclique.

La zone d'apparition des événements cyclique étant inconnue, nous utilisons une représentation des événements selon leur moment d'apparition relatif à l'échelle cyclique. Pour limiter le nombre d'entités visuelles présentes dans l'interface, nous utilisons pour chaque analyse une représentation agrégée des événements. L'identification des événements cycliques s'effectue alors en suivant le scénario d'analyse décrit dans le chapitre 6.

Une fois les événements représentés selon leur moment d'apparition relatif au temps cyclique, l'utilisateur modifie graduellement l'échelle cyclique représentée dans le diagramme mixte jusqu'à identifier un cluster spatio-temporel cyclique sur la carte. L'utilisateur peut utiliser, pour s'orienter dans sa recherche, les périodes retournées par la transformée de Fourier effectuée par l'interface sur les données événementielles.

Une fois qu'un cluster est identifié sur la carte, l'utilisateur sélectionne graphiquement les entités du cluster dans la carte, afin de mettre en valeur les événements correspondants dans les diagrammes temporels. L'utilisateur utilise alors ces diagrammes pour identifier l'alignement d'entités graphiques correspondant à la série d'événements cycliques.

La période P du cycle est identifiée à la fin de cette étape, ainsi que les dates T_0 et T_N .

8.3.3.2 Redéfinition des paramètres de représentation des événements

Une fois les événements cycliques identifiés sur le diagramme temporel, la deuxième étape de notre analyse consiste à modifier les paramètres de représentation des événements. Les événements sont toujours représentés selon leur moment d'apparition relatif à l'échelle cyclique. Cette manoeuvre a pour but que les événements apparaissant durant l'intervalle temporel cyclique, occupé par la série identifiée, soient représentés de manière distincte des autres événements du jeu de donnée.

Le nombre de sections du diagramme temporel circulaire est augmenté de manière à déterminer les limites du cluster temporel formé par les événements cycliques. La dispersion temporelle $disp_t$ des événements est alors identifiée.

Les couleurs et le nombre de classes utilisées pour représenter les événements sont modifiés, afin que l'intervalle temporel occupé par les événements cycliques soit mis en valeur.

La sélection d'événements effectuée précédemment est désactivée. L'utilisateur utilise la nouvelle représentation pour estimer plus précisément les limites spatiales de la récurrence cyclique dans la carte.

Les paramètres $longitude_C$, $latitude_C$, et $disp_s$ sont identifiés lors de cette étape en suivant le procédé suivant :

- a. L'utilisateur estime les bornes Nord, Sud, Est, Ouest de la zone d'apparition des événements.
- b. L'utilisateur identifie les longitudes et latitudes correspondant à ces extrémités.
- c. Les coordonnées du point C sont calculées à partir de ces latitudes et longitudes exprimées.

- d. La valeur de $disp_s$ est estimée à partir de l'écart entre $longitude_C$ et la longitude des bornes Est et Ouest, ou à partir de l'écart entre $latitude_C$ et la latitude des bornes Nord et Sud. L'écart est le même en nombre de degrés malgré la forme ellipsoïdale de la zone d'apparition des événements.

L'utilisateur peut également effectuer une nouvelle sélection des événements dans la carte pour réévaluer les valeurs de T_0 et T_N .

8.3.3.3 Comparaison des résultats trouvés avec les informations de référence

Une fois l'analyse terminée, l'utilisateur récupère les paramètres stockés en base correspondant au jeu de données testé, et les compare aux résultats trouvés.

8.4 Résultats

Nous présentons, dans cette section, les résultats obtenus lors de nos analyses pour les différents jeux de données créés, et leur comparaison avec les paramètres stockés en base.

Pour les jeux de données A_1 et B_8 , nous présentons en détail les étapes suivies pour identifier les caractéristiques des séries d'événements cycliques. Pour les autres jeux de données, nous présentons les résultats obtenus, ainsi que les illustrations de ces résultats dans l'interface de visualisation.

8.4.1 Analyse des jeux de données du groupe A

8.4.1.1 Jeu A_1

La Figure 8.6 présente la distribution spatiale et temporelle du jeu de données A_1 .

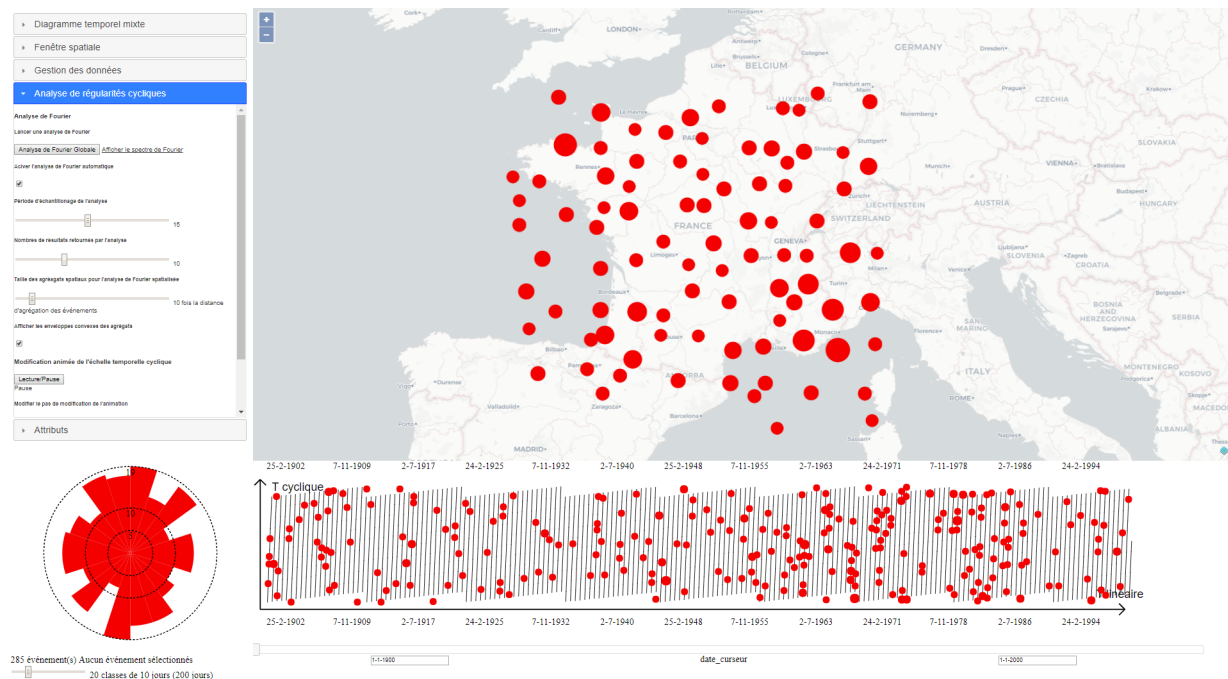


FIGURE 8.6 – Distribution spatio-temporelle du jeu A_1

Lors de l'import, une analyse par transformée de Fourier est effectuée automatiquement sur les données. Plusieurs propositions d'échelles cycliques sont renvoyées (Figure 8.7).

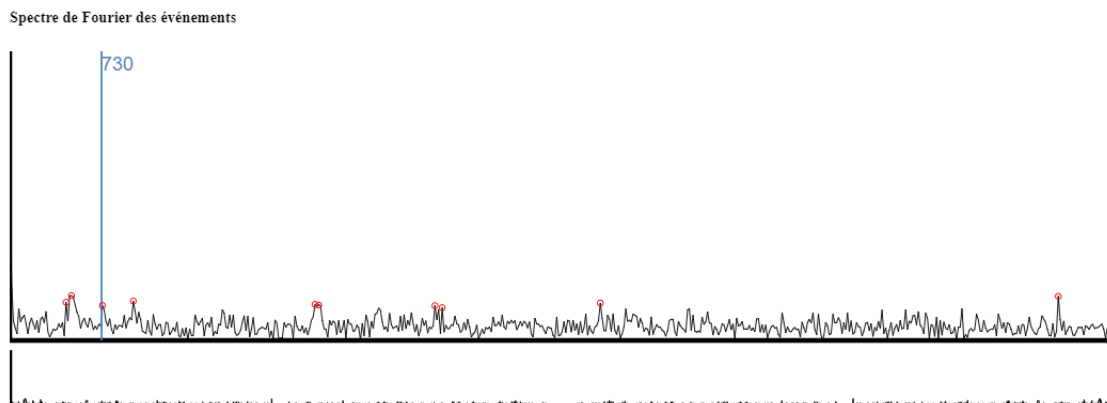


FIGURE 8.7 – Spectre de Fourier du jeu A_1

Une des fréquences significatives renvoyées par transformée de Fourier est égale à 1074 jours. Après avoir affecté cette valeur à l'échelle temporelle cyclique T , nous représentons les événements selon leur moment d'apparition relatif au temps cyclique (Repères A et B de la Figure 8.8). Nous utilisons six couleurs de base et six classes temporelles.

Nous observons sur la carte un cluster au sud-est de la zone étudiée, la couleur bleue y étant majoritaire. Les entités de ce cluster sont sélectionnées au travers de la carte. Les entités non sélectionnées sont masquées (Repères B et C de la Figure 8.8).

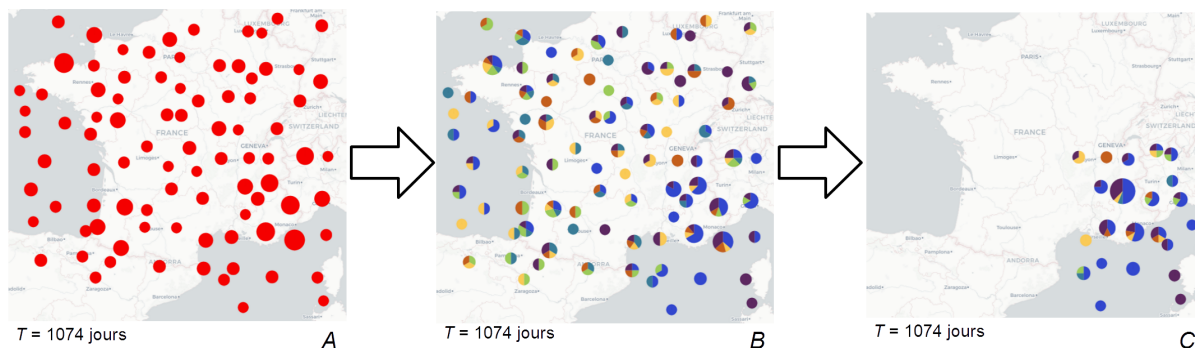


FIGURE 8.8 – Identification d'un cluster spatio-temporel dans le jeu A_1

La composante temporelle des événements sélectionnés est représentée sur les diagrammes temporels (Figure 8.9). On observe une surreprésentation de la première et de la sixième section du diagramme circulaire. Dans le diagramme mixte, on observe deux alignements graphiques, situés sur les deux extrêmes verticaux du diagramme.

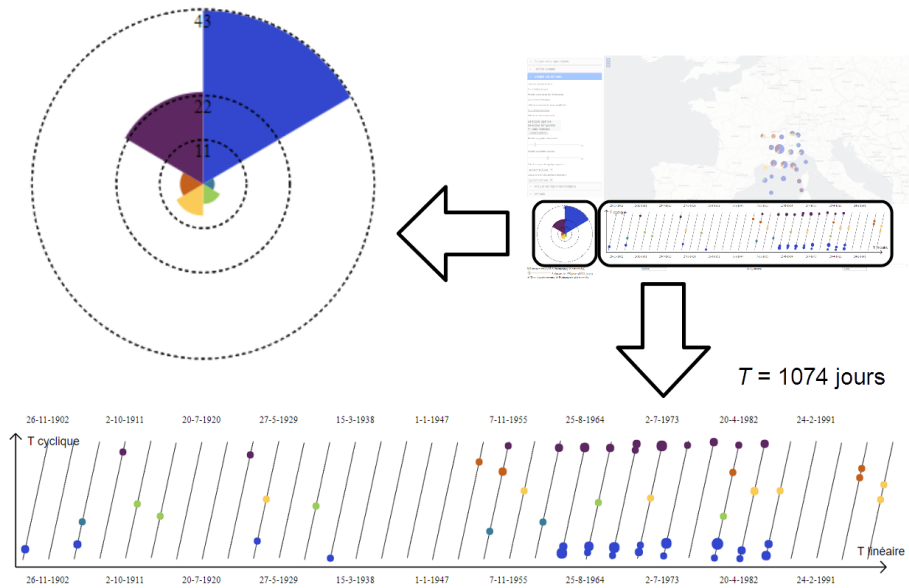


FIGURE 8.9 – Visualisation de la composante temporelle du cluster spatio-temporel dans le jeu A_1

Le déphasage du diagramme temporel mixte est modifié afin d'obtenir un seul alignement d'entités graphiques sur le diagramme mixte (Figure 8.10). L'alignement graphique obtenu est légèrement ascendant. Dans le diagramme temporel circulaire, les troisième et quatrième sections sont surreprésentées.

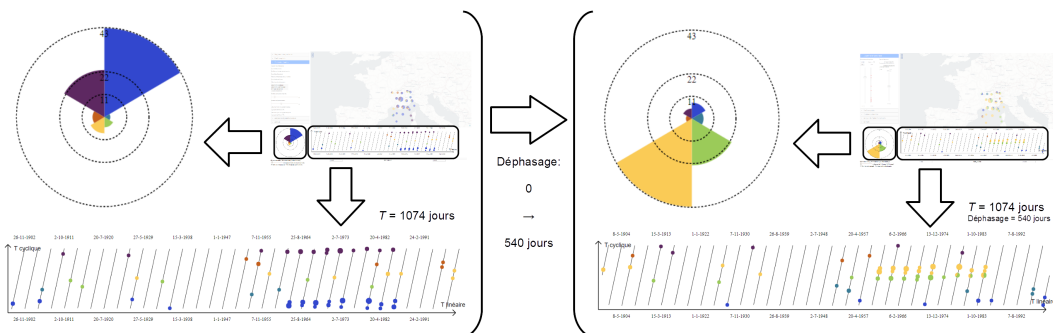


FIGURE 8.10 – Modification du déphasage du diagramme temporel mixte

L'échelle cyclique T est progressivement augmentée jusqu'à 1079 jours (Figure 8.11). L'alignement devient parallèle à l'axe des abscisses du diagramme mixte.

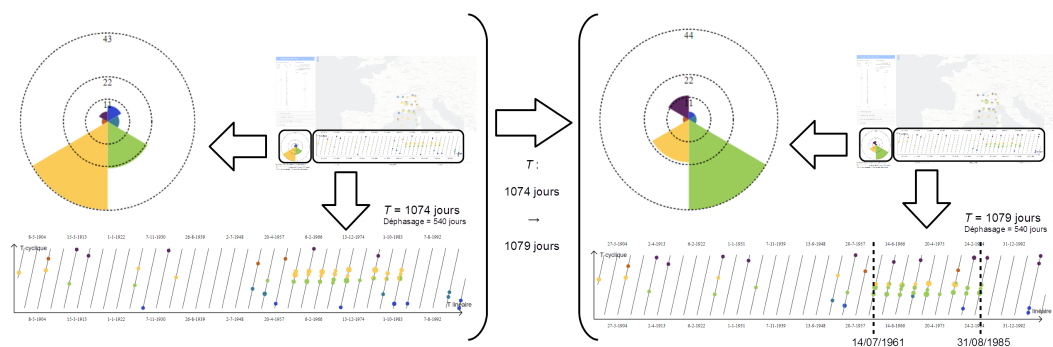


FIGURE 8.11 – Identification de la période de la série événementielle cyclique dans le jeu A_1

Les événements correspondants apparaissent du 14/07/1961 au 31/08/1985 (Figure 8.12). Sur le diagramme circulaire, les troisième et quatrième sections sont surreprésentées.

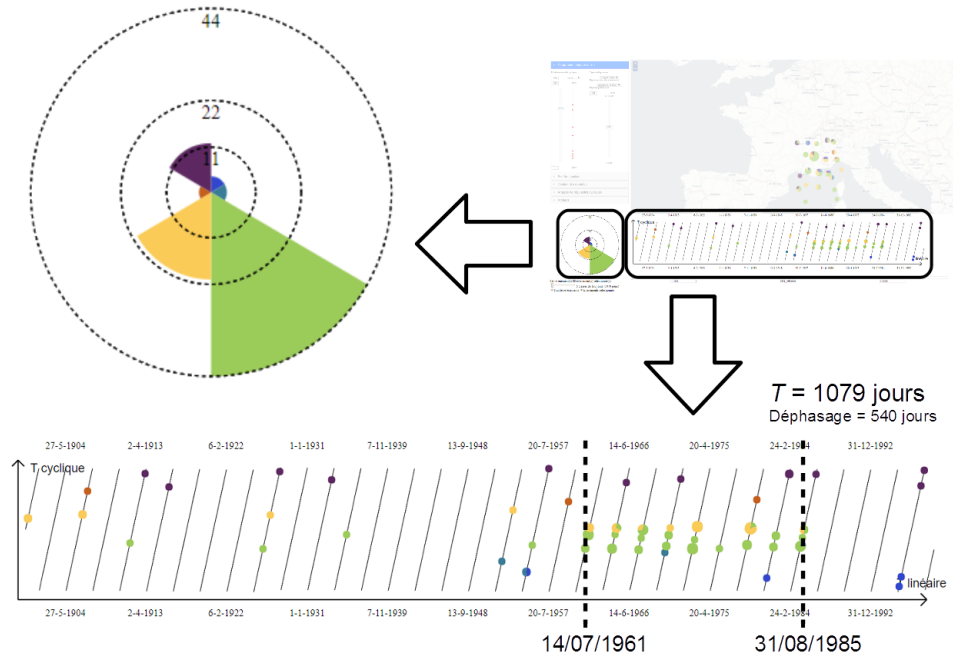


FIGURE 8.12 – Identification de l’emprise temporelle de la série événementielle cyclique dans le jeu A_1

Nous portons le nombre de sections du diagramme circulaire à 60 sections (Repères 1 et 2 de la Figure 8.13). Nous observons que la forte proportion d’événements est représentée par douze sections du diagramme circulaire, ce qui correspond à 20% de l’échelle cyclique. La dispersion temporelle des événements cycliques est donc de 20%.

Nous modifions le déphasage du diagramme temporel mixte pour placer la première des douze sections identifiées au point d’origine du diagramme temporel circulaire (Repères 2 et 3 de la Figure 8.13).

La forte proportion d’événements s’étendant sur 20% de l’échelle cyclique, nous portons le nombre de sections du diagramme circulaire à cinq, afin que cette forte proportion d’événements soit représentée par une seule section du diagramme (Repères 3 et 4 de la Figure 8.13). Ces événements sont alors représentés par une même couleur.

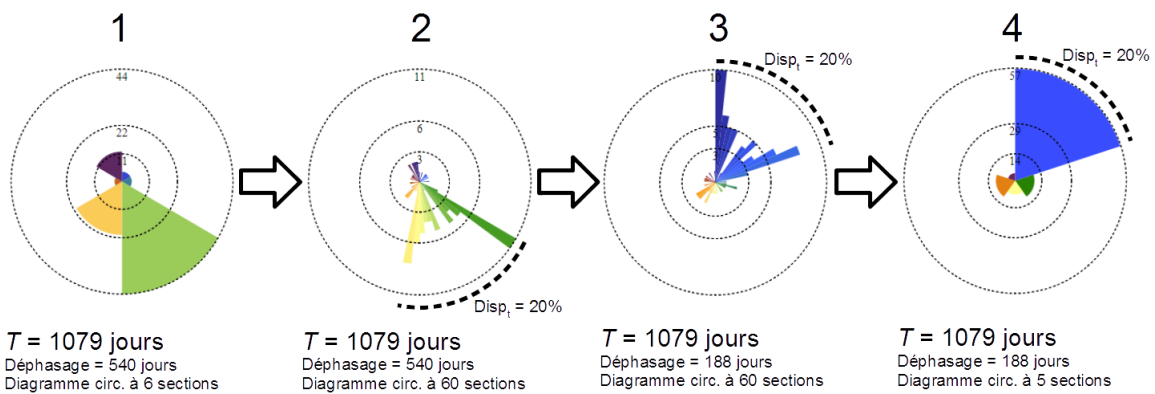


FIGURE 8.13 – Modification des paramètres de représentation des événements du jeu A_1

Les événements du cluster identifiés auparavant sont désélectionnés. On obtient alors la représentation donnée dans la Figure 8.14.

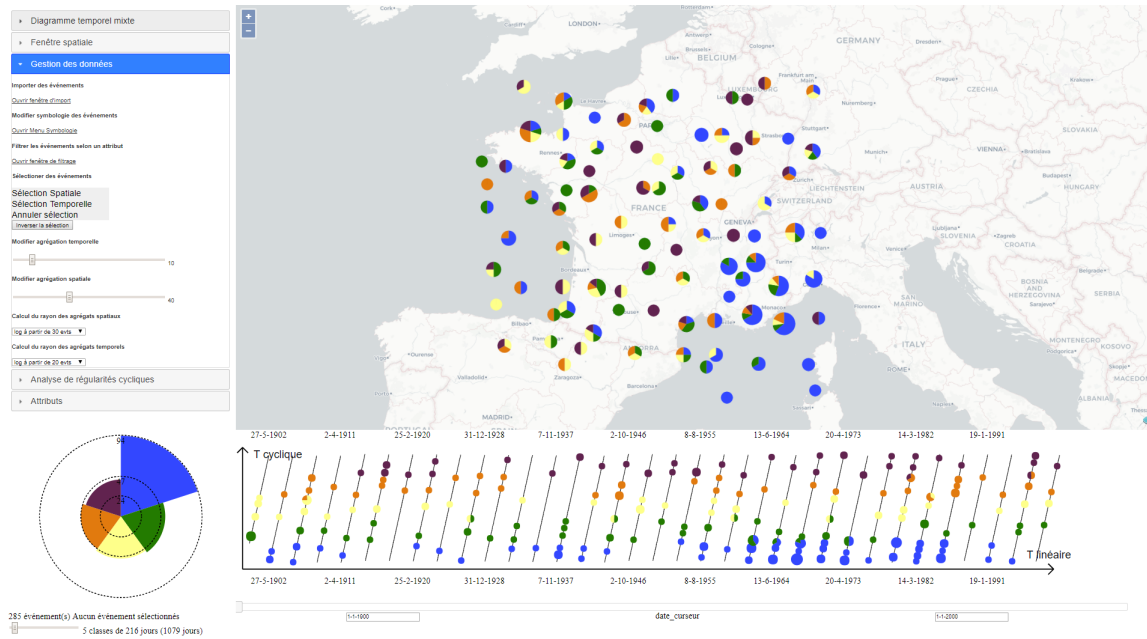


FIGURE 8.14 – Visualisation d’une récurrence spatialisée dans le jeu A_1

Les événements cycliques sont localisés sur le quart sud-est de la zone d’étude, entre 4.0 et 9.0 de longitude, et entre 41.1 et 46.3 de latitude. Nous comparons ces résultats aux données stockées en base dans le tableau ci-dessous. La Figure 8.15 présente cette comparaison sous forme graphique.

	Résultats trouvés	Informations stockées en base
$longitude_C$	6.5	6.4
$latitude_C$	43.7	44.3
Dispersion spatiale $disp_s$	2.5 degrés	3.4 degrés
T_0	14/07/1961	17/07/1961
T_N	31/08/1985	20/09/1985
Période P	1079 jours	1077 jours
Dispersion temporelle $disp_t$	20.0%	20.5%

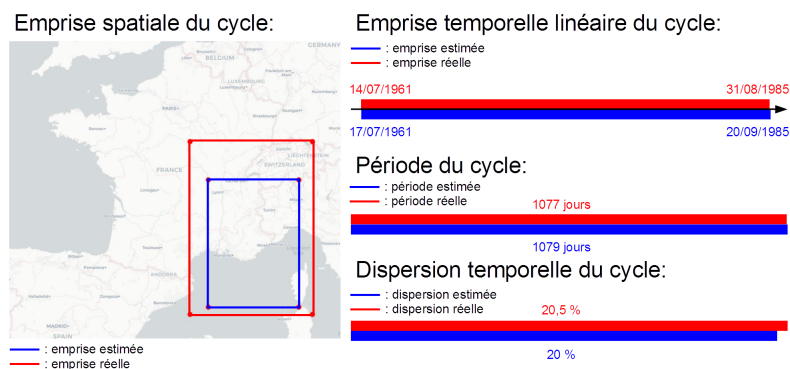


FIGURE 8.15 – Comparaison des résultats obtenus avec les données de référence pour le jeu A_1

8.4.1.2 Jeu A_2

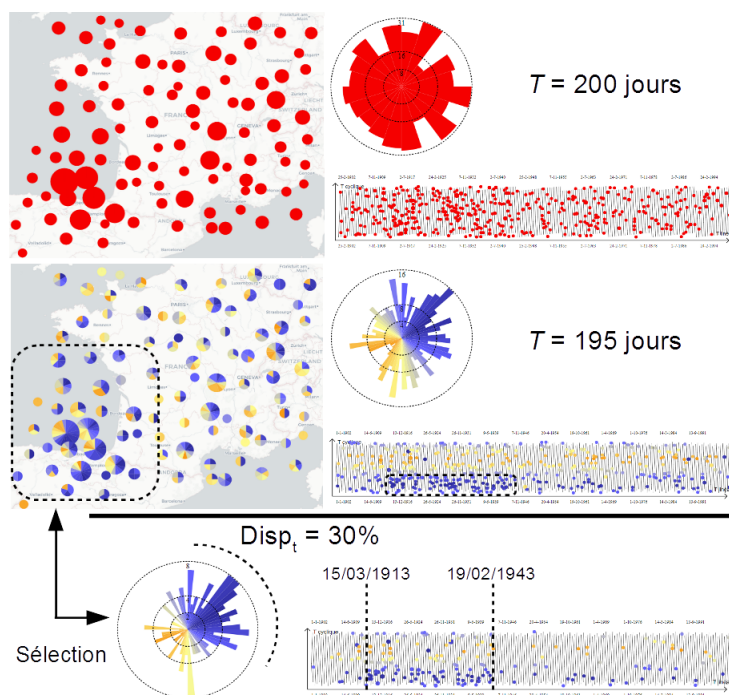


FIGURE 8.16 – Analyse du jeu A_2

La Figure 8.16 correspond à la représentation des événements du jeu A_2 au travers des fenêtres temporelle et spatiale de l'interface, en début et en fin d'analyse. Nous comparons les résultats obtenus avec les données stockées en base dans le tableau suivant. La Figure 8.17 présente cette comparaison sous forme graphique.

	Résultats trouvés	Informations stockées en base
$longitude_C$	-2.0	-3.4
$latitude_C$	44.5	43.9
Dispersion spatiale $disp_s$	3.0 degrés	3.1 degrés
T_0	15/3/1913	30/09/1909
T_N	19/02/1943	08/03/1943
Période P	195 jours	581 jours
Dispersion temporelle $disp_t$	30.0%	43.0%

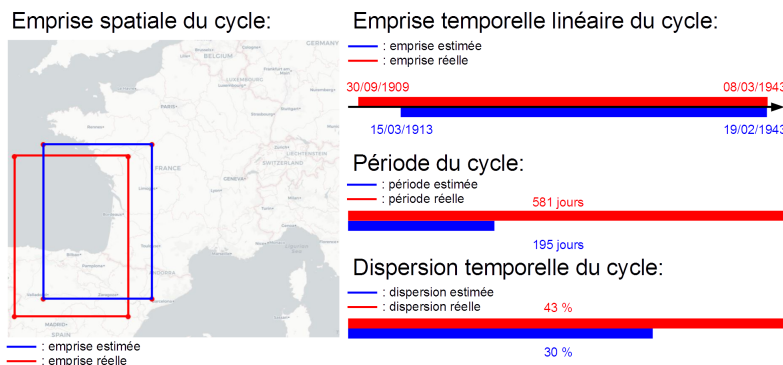


FIGURE 8.17 – Comparaison des résultats obtenus avec les données de référence pour le jeu A_2

La série cyclique du jeu de données correspond à celle que nous avons identifiée dans la Figure 8.16. Cependant la période que nous avons identifiée correspond à un harmonique de rang trois

de la réelle période de la série cyclique. La durée de 581 jours inscrite en base est une valeur proche du triple de la période de 185 jours identifiée visuellement dans la Figure 8.16.

Une représentation des événements selon une échelle cyclique de 581 jours permet d'identifier la même récurrence cyclique localisée (Figure 8.18). Cette représentation permet alors d'identifier une dispersion temporelle de 43% et une étendue temporelle allant du 30/09/1909 au 08/03/1943.

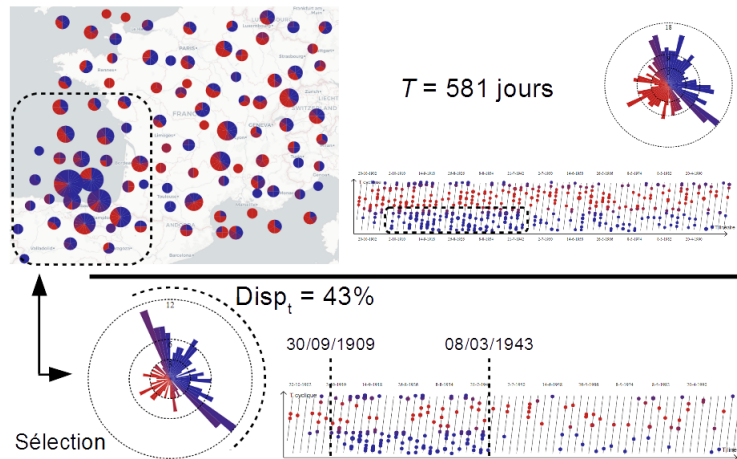


FIGURE 8.18 – Événements du jeu A_2 représentés selon une échelle cyclique de 581 jours

Parmi les échelles cycliques suggérées d'après les résultats de la transformée de Fourier, une échelle cyclique de 570 jours était proposée. Une représentation des événements du jeu A_2 selon cette échelle cyclique (Figure 8.19) permettait d'identifier une récurrence cyclique localisée. Nous émettons l'hypothèse que l'identification de la période de 581 jours aurait pu être possible en prenant, comme point de départ de l'analyse, l'échelle de 570 jours suggérée par l'interface.

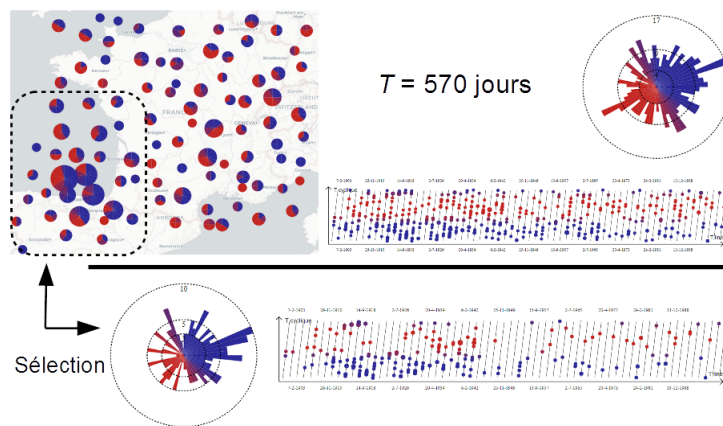


FIGURE 8.19 – Événements du jeu A_2 représentés selon une échelle cyclique de 570 jours

8.4.1.3 Jeu A_3

La Figure 8.20 correspond à la représentation des événements du jeu A_3 au travers des fenêtres temporelle et spatiale de l'interface, en début et en fin d'analyse.

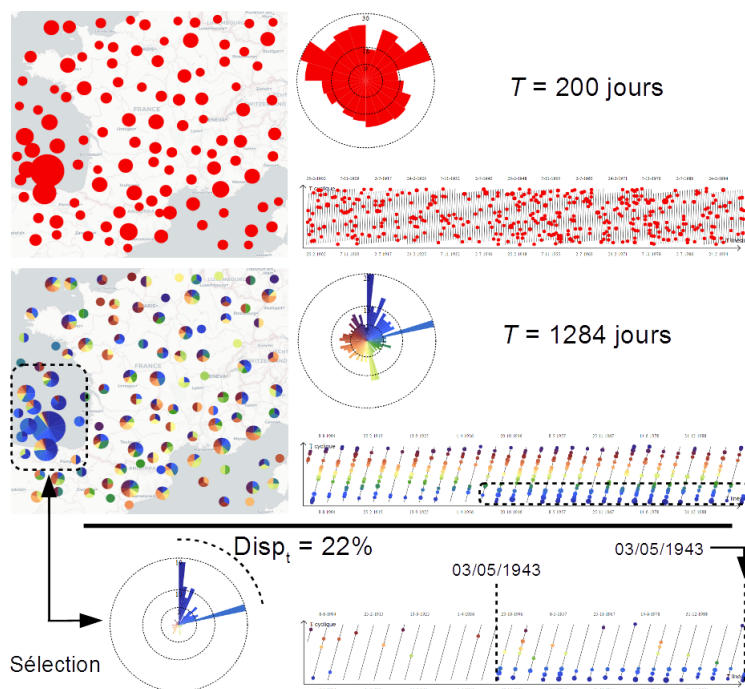


FIGURE 8.20 – Analyse du jeu A_3

Nous comparons les résultats obtenus avec les données stockées en base dans le tableau suivant. La Figure 8.21 présente cette comparaison sous forme graphique.

	Résultats trouvés	Informations stockées en base
$longitude_C$	-2.8	-3.9
$latitude_C$	44.7	44.2
Dispersion spatiale $disp_s$	2.0 degrés	1.9 degrés
T_0	03/05/1943	13/05/1943
T_N	07/12/1999	28/12/1999
Période P	1284 jours	1283 jours
Dispersion temporelle $disp_t$	22.0%	20.8%

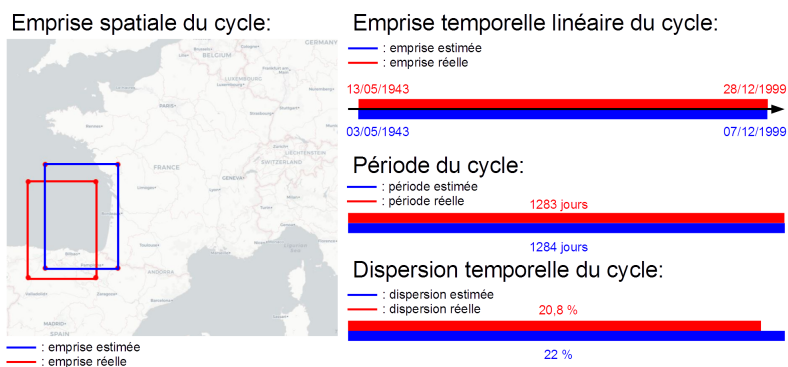


FIGURE 8.21 – Comparaison des résultats obtenus avec les données de référence pour le jeu A_3

8.4.1.4 Jeu A_4

La Figure 8.22 correspond à la représentation des événements du jeu A_4 au travers des fenêtres temporelle et spatiale de l'interface, en début et en fin d'analyse.

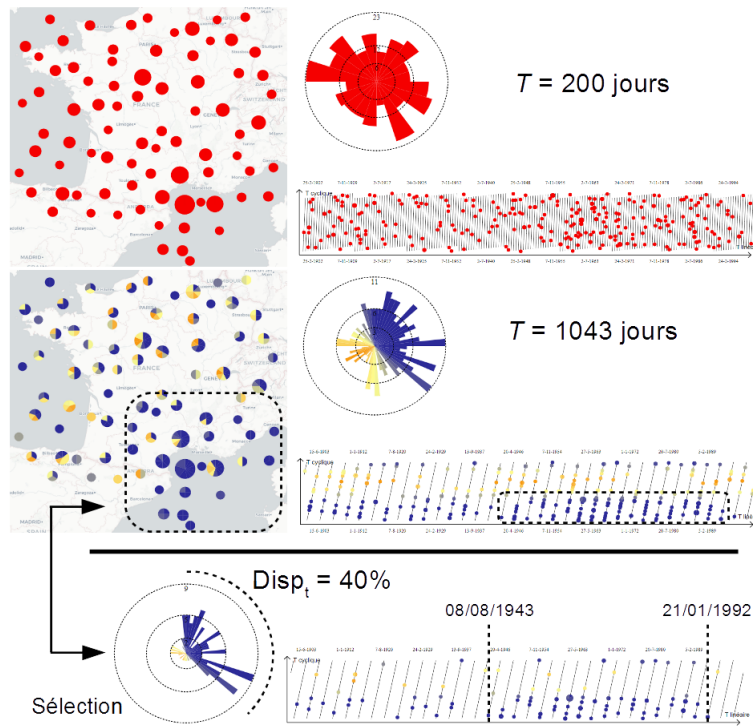


FIGURE 8.22 – Analyse du jeu A_4

Nous comparons les résultats obtenus avec les données stockées en base dans le tableau suivant. La Figure 8.23 présente cette comparaison sous forme graphique.

	Résultats trouvés	Informations stockées en base
$longitude_C$	6.0	3.8
$latitude_C$	41.9	42.5
Dispersion spatiale $disp_s$	2.9 degrés	2.5 degrés
T_0	08/08/1943	21/05/1946
T_N	21/01/1992	17/01/1992
Période P	1043 jours	1037 jours
Dispersion temporelle $disp_t$	40.0%	37.0%

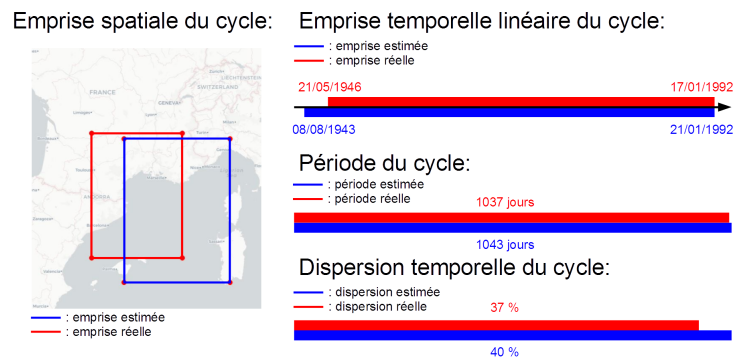


FIGURE 8.23 – Comparaison des résultats obtenus avec les données de référence pour le jeu A_4

8.4.2 Analyse des jeux de données du groupe B

8.4.2.1 Jeu B_5

La Figure 8.24 correspond à la représentation des événements du jeu B_5 au travers des fenêtres temporelle et spatiale de l'interface, en début et en fin d'analyse. Nous comparons les résultats obtenus avec les données stockées en base dans le tableau suivant. La Figure 8.25 présente cette comparaison sous forme graphique.

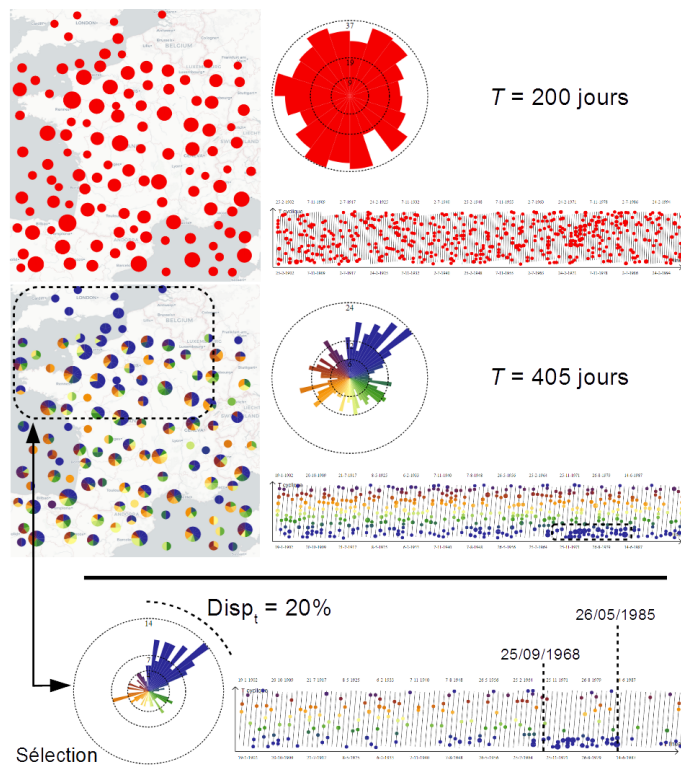


FIGURE 8.24 – Analyse du jeu B_5

	Résultats trouvés	Informations stockées en base
$longitude_C$	0.2	0.8
$latitude_C$	49.0	49.5
Dispersion spatiale $disp_s$	3.5 degrés	4.0 degrés
T_0	25/09/1968	17/05/1966
T_N	26/05/1985	06/06/1985
Période P	405 jours	406 jours
Dispersion temporelle $disp_t$	20.0%	15.0%

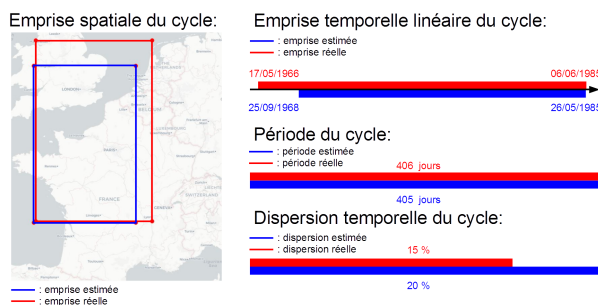


FIGURE 8.25 – Comparaison des résultats obtenus avec les données de référence pour le jeu B_5

8.4.2.2 Jeu B_6

La Figure 8.26 correspond à la représentation des événements du jeu B_6 au travers des fenêtres temporelle et spatiale de l'interface, en début et en fin d'analyse. Nous comparons les résultats obtenus avec les données stockées en base dans le tableau suivant. La Figure 8.27 présente cette comparaison sous forme graphique.

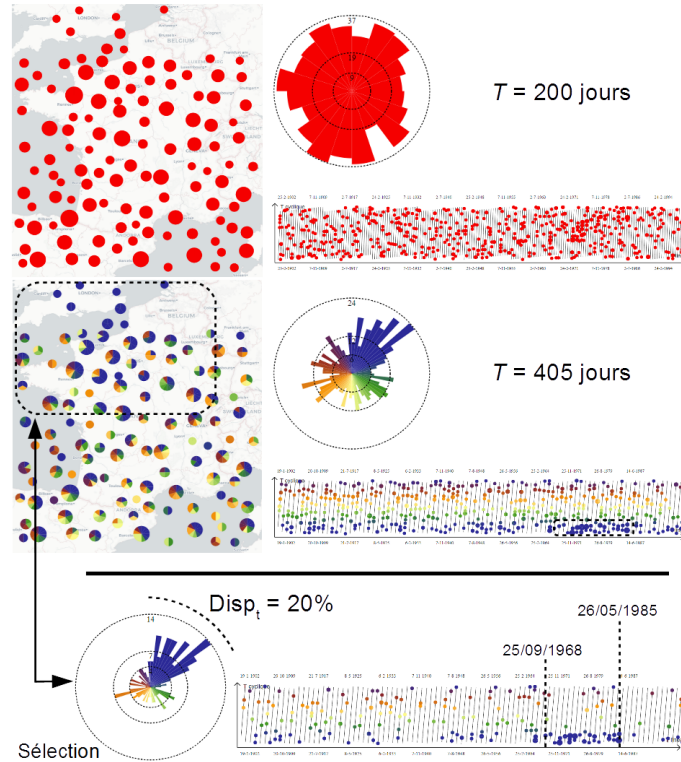


FIGURE 8.26 – Analyse du jeu B_6

	Résultats trouvés	Informations stockées en base
$longitude_C$	6.3	6.5
$latitude_C$	47.6,	47.0
Dispersion spatiale $disp_s$	2.6 degrés	2.8 degrés
T_0	03/05/1922	21/10/1933
T_N	08/04/1999	12/09/1990
Période P	1028 jours	1024 jours
Dispersion temporelle $disp_t$	45.0%	33.0%

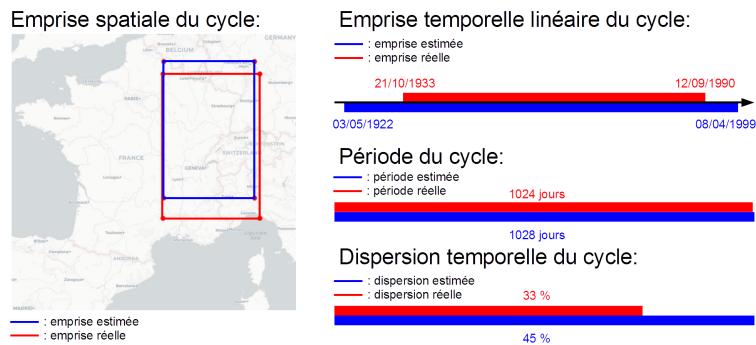


FIGURE 8.27 – Comparaison des résultats obtenus avec les données de référence pour le jeu B_6

8.4.2.3 Jeu B_7

La Figure 8.28 correspond à la représentation des événements du jeu B_7 au travers des fenêtres temporelle et spatiale de l'interface, en début et en fin d'analyse. Nous comparons les résultats obtenus avec les données stockées en base dans le tableau suivant. La Figure 8.29 présente cette comparaison sous forme graphique.

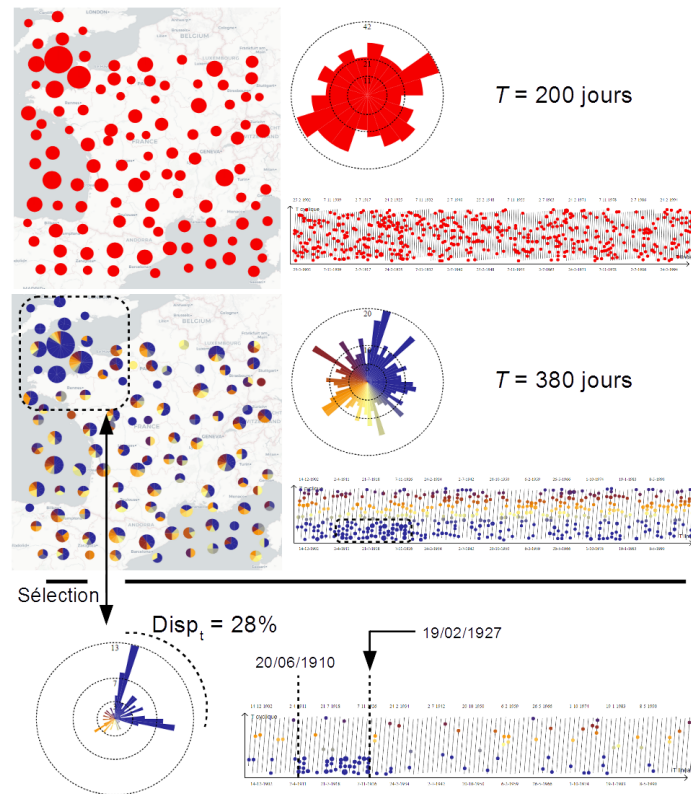


FIGURE 8.28 – Analyse du jeu B_7

	Résultats trouvés	Informations stockées en base
$longitude_C$	-2.2	-3.0
$latitude_C$	49.4,	49.8
Dispersion spatiale $disp_s$	2.5 degrés	1.8 degrés
T_0	20/06/1910	30/04/1911
T_N	19/02/1927	09/03/1927
Période P	380 jours	380 jours
Dispersion temporelle $disp_t$	28.0%	25.0%

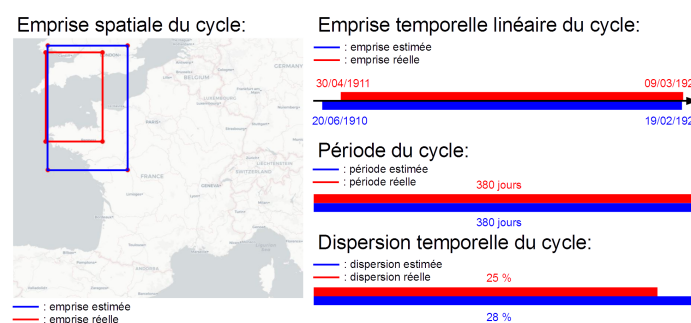


FIGURE 8.29 – Comparaison des résultats obtenus avec les données de référence pour le jeu B_7

8.4.2.4 Jeu B_8

La Figure 8.30 présente la distribution spatiale et temporelle du jeu de données B_8 .

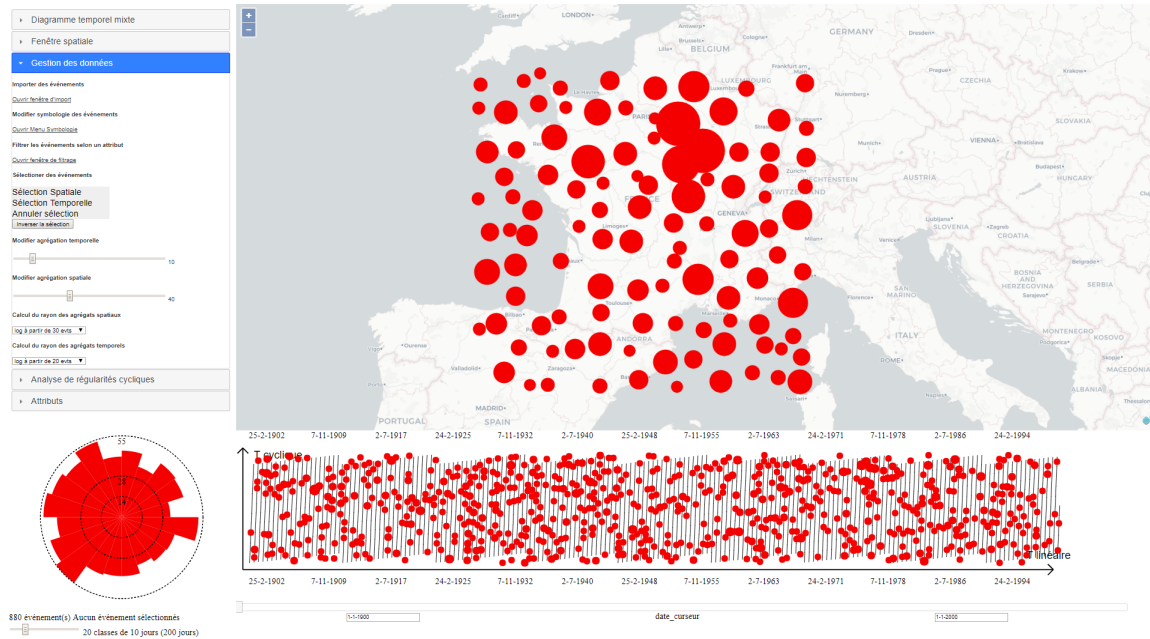


FIGURE 8.30 – Distribution spatio-temporelle du jeu B_8

Nous représentons les événements selon leur moment d'apparition relatif au temps cyclique (Figure 8.31). Nous utilisons six couleurs de base et six classes temporelles. Lors de cette opération, les événements sont regroupés selon leur proximité spatiale en différentes zones. La distance critique, permettant de créer ces groupes d'événements proches, est fixée à six fois la distance critique définissant les agrégats d'événements représentés sur la carte.

L'outil regroupe les événements en six groupes (Figure 8.31), et une transformée de Fourier est effectuée pour chacune des zones. Une des fréquences significatives identifiées sur le spectre de Fourier obtenu pour le groupe d'événements *rouge* correspond à une échelle temporelle cyclique de 1304 jours.

Une fois cette valeur affectée à l'échelle cyclique du diagramme temporel mixte, nous observons sur la carte un cluster au nord-est de la zone étudiée, les couleurs orange, jaunes et vertes y étant majoritaires. Les entités de ce cluster sont sélectionnées au travers de la carte. Les entités non sélectionnées sont masquées (Figure 8.31).

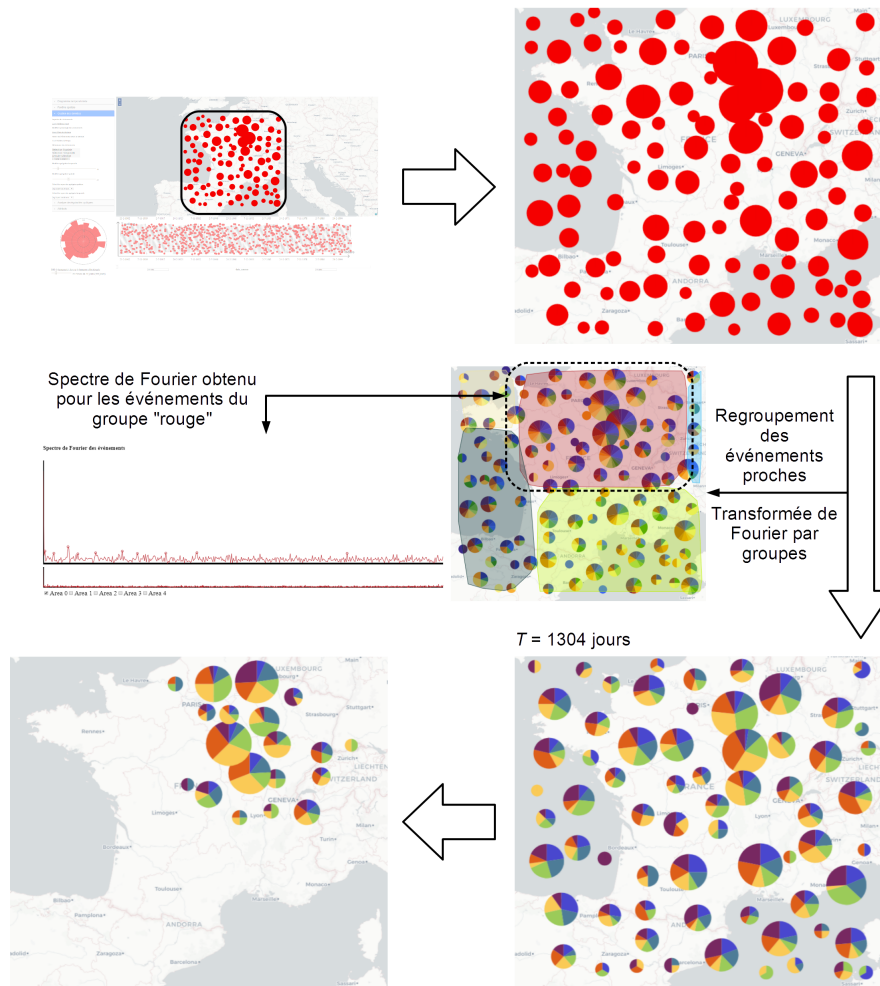


FIGURE 8.31 – Identification d'un cluster spatio-temporel dans le jeu B_8

La composante temporelle des événements sélectionnés est représentée sur les diagrammes temporels (Figure 8.32). On observe une surreprésentation des troisième, quatrième et cinquième sections du diagramme circulaire. Dans le diagramme mixte, on observe un alignement, légèrement descendant, des événements correspondants.

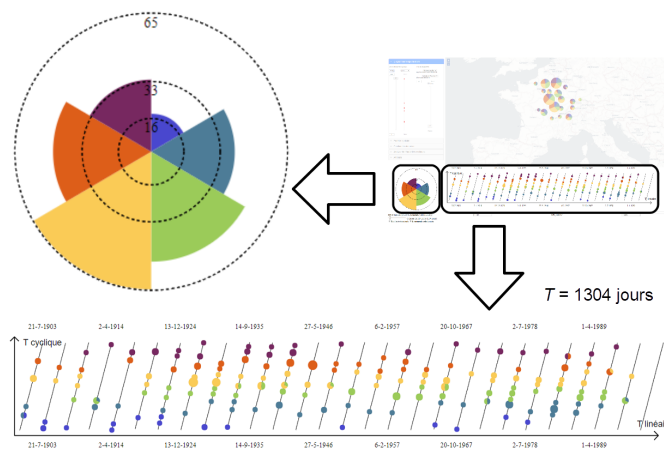


FIGURE 8.32 – Visualisation de la composante temporelle du cluster spatio-temporel dans le jeu B_8

L'échelle cyclique T est progressivement diminuée jusqu'à 1299 jours (Figure 8.33). L'aligne-

ment devient parallèle à l'axe des abscisses du diagramme mixte.

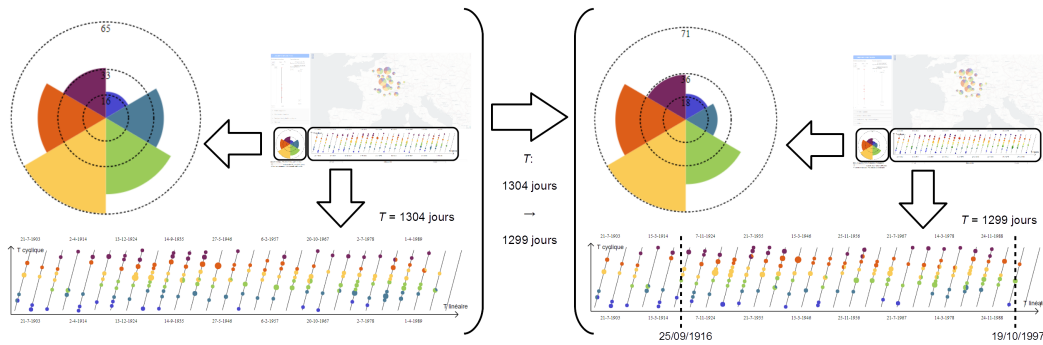


FIGURE 8.33 – Identification de la période de la série événementielle cyclique dans le jeu B_8

Les événements correspondants apparaissent du 25/09/1916 au 19/10/1997 (Figure 8.34). Sur le diagramme circulaire, les troisième, quatrième et cinquième sections sont toujours surreprésentées, mais le rayon des première, deuxième et sixième sections a diminué.

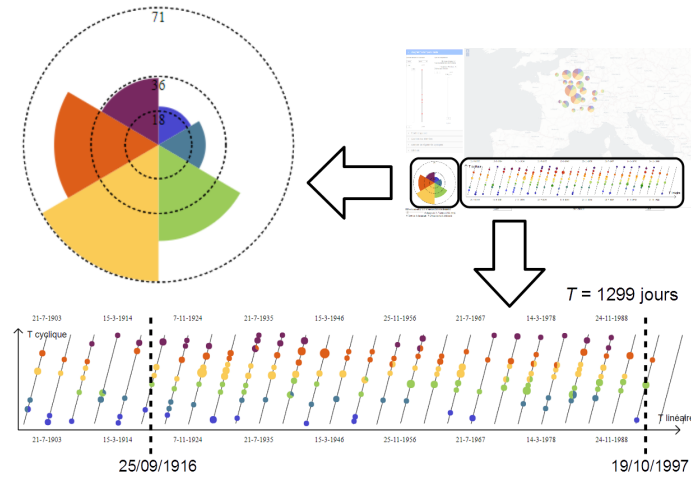


FIGURE 8.34 – Identification de l'emprise temporelle de la série événementielle cyclique dans le jeu B_8

Nous portons le nombre de sections du diagramme circulaire à 60 sections (Repères 1 et 2 de la Figure 8.35). Nous observons que la forte proportion d'événements est représentée par vingt-quatre sections du diagramme circulaire, ce qui correspond à 40% de l'échelle cyclique. La dispersion temporelle des événements cycliques est donc de 40%.

Nous modifions le déphasage du diagramme temporel mixte pour placer la première des vingt-quatre sections identifiées au point d'origine du diagramme temporel circulaire (Repères 2 et 3 de la Figure 8.35).

La forte proportion d'événements s'étendant sur 40% de l'échelle cyclique, nous portons le nombre de sections du diagramme circulaire à cinq, afin que cette forte proportion d'événements soit représentée par les deux premières sections du diagramme (Repères 3 et 4 de la Figure 8.35). Nous modifions les couleurs utilisées afin de représenter les événements correspondant à ces deux sections par une même couleur, et de représenter le reste des événements par une couleur distincte.

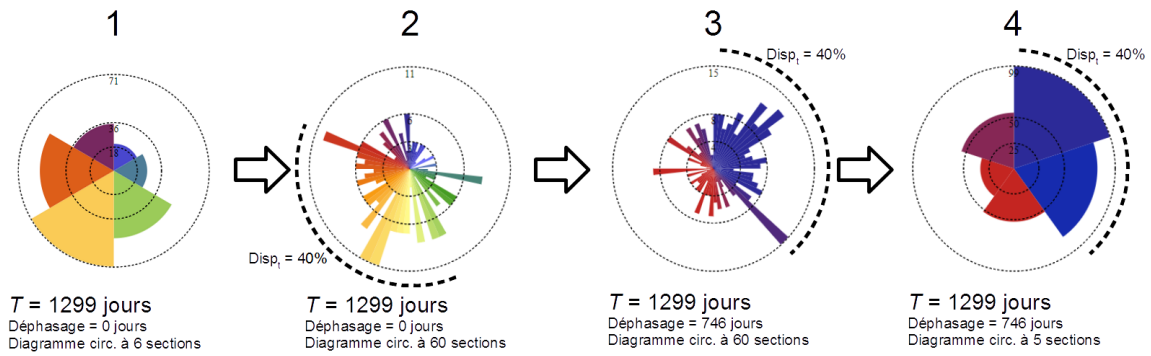


FIGURE 8.35 – Modification des paramètres de représentation des événements du jeu B_8

Les événements du cluster identifiés auparavant sont désélectionnés. On obtient alors la représentation donnée dans la Figure 8.36.

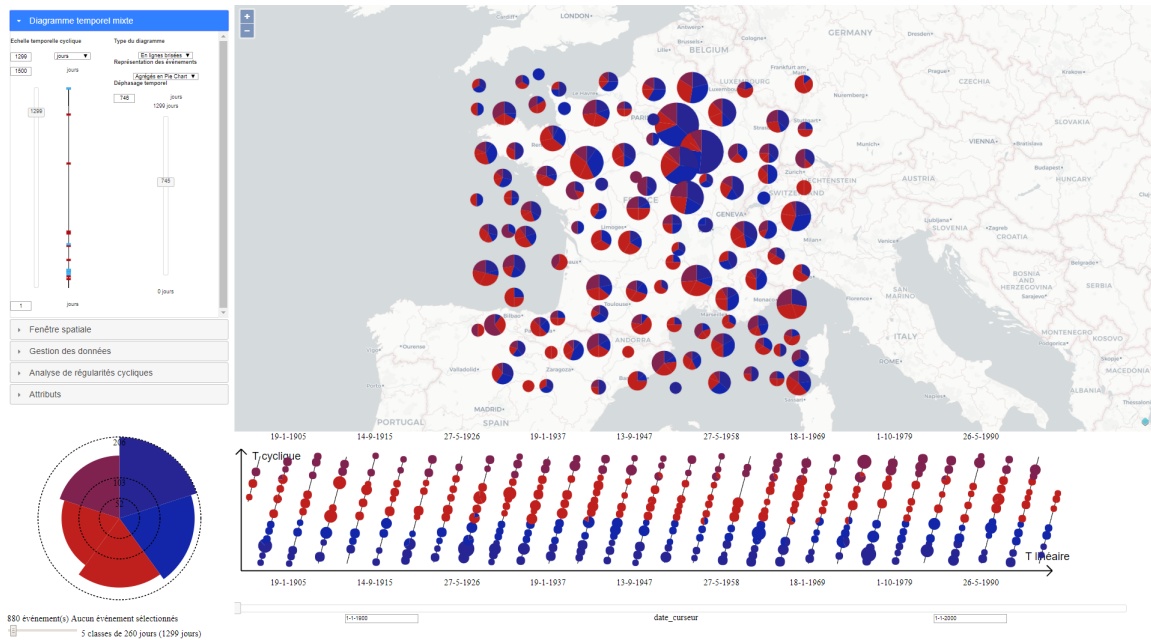


FIGURE 8.36 – Visualisation d’une récurrence spatialisée dans le jeu B_8

Les événements cycliques sont localisés au nord-est de la zone d’étude, entre 2.1 et 5.8 de longitude, et entre 46.6 et 50.3 de latitude. Dans le tableau suivant, nous comparons les résultats obtenus aux données stockées en base. La Figure 8.37 présente cette comparaison sous forme graphique.

	Résultats trouvés	Informations stockées en base
$longitude_C$	3.9	3.8
$latitude_C$	48.4,	48.1
Dispersion spatiale $disp_s$	1,8 degrés	1.8 degrés
T_0	25/09/1916	01/13/1920
T_N	19/10/1997	21/10/1990
Période P	1299 jours	1278 jours
Dispersion temporelle $disp_t$	40.0%	37.0%

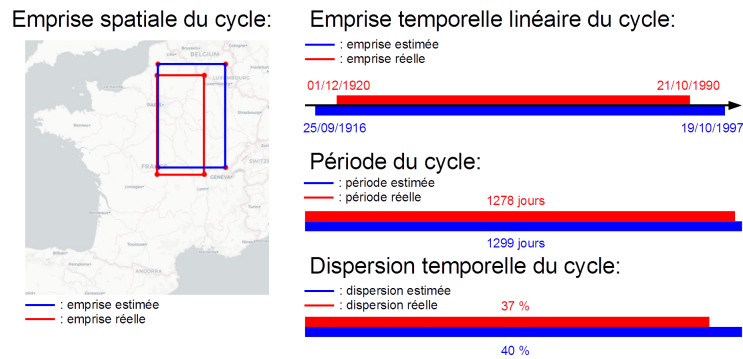


FIGURE 8.37 – Comparaison des résultats obtenus avec les données de référence pour le jeu B_8

8.5 Limites et synthèse

8.5.1 Distribution temporelle des événements cycliques dans les jeux de données

Lors de l'expérimentation, nous avons découvert que pour la majorité des jeux de données, la distribution temporelle des événements à l'intérieur de leur intervalle d'apparition montrait une surreprésentation d'événements ayant lieu au début et à la fin de chaque intervalle. Nous avons pris conscience par après que cette distribution était due à la manière dont nous avons implémenté la formule déterminant la composante temporelle de nos événements cycliques.

Dans la formule $t_{evt} = T_0 + (P * n + random_{evt} * (1/2) * P * disp_t)$, le terme $random_{evt}$ définissant une variable aléatoire choisie entre -1 et 1 , a été implémenté au moyen d'une fonction sinusoïdale. Ceci a engendré une surreprésentation des valeurs -1 et 1 pour le terme $random_{evt}$, ce qui a abouti à une concentration des événements en début et fin des intervalles temporels cycliques.

La conséquence de cette distribution est que les clusters, correspondant aux événements cycliques, ne présentent pas une concentration homogène des événements sur leur étendue temporelle. Ce caractère non homogène est notamment visible dans les diagrammes temporels circulaires, en particulier pour les jeux de données présentant une forte dispersion temporelle. Les séries d'événements gardent cependant leur caractère cyclique.

Le caractère non homogène des clusters, présentant des pics du nombre d'événements au début et à la fin de chaque intervalle, peut faciliter la définition de l'emprise temporelle du cluster dans le temps cyclique, en permettant de mieux identifier ses bornes.

En revanche, cette distribution temporelle des événements dans les clusters peut rendre plus difficile l'identification de ces derniers au travers de la carte. Parmi les couleurs représentant les événements des clusters à identifier, celles correspondant aux bornes du cluster sont celles présentant la plus grande différence entre elles. L'utilisateur peut ainsi avoir plus de difficulté à observer la prise d'une couleur majoritaire dans les *Pie Chart* de la carte.

Or, l'objectif principal de cette expérimentation est d'identifier l'existence d'un cycle dans les séries d'événements, en effectuant dans un premier temps une identification de clusters au travers de la carte. L'atteinte de cet objectif n'est pas rendue plus aisée avec l'existence de ce biais dans la création des jeux de données. Malgré celui-ci, on peut donc considérer l'expérimentation comme valable.

Ce caractère homogène n'est pas nécessaire pour effectuer notre expérimentation, notre objectif principal étant d'identifier l'existence d'un cycle, en effectuant dans un premier temps une identification d'un cluster au travers de la carte.

8.5.2 Conclusion

La série d'analyses exploratoires effectuées a permis d'identifier les récurrences cycliques spatialisées présentes dans chaque jeu de données.

Pour chaque jeu de données, la période identifiée est très proche de la période réelle, dont la durée est stockée en base. L'écart entre la valeur estimée et la valeur réelle est, pour chaque jeu de données hormis le jeu A_2 , inférieur à 2% de la valeur réelle.

Dans le cas du jeu A_2 , la série événementielle cyclique est identifiée, mais la période estimée correspond à un harmonique de rang 3 du cycle identifié. La durée de la période estimée est ainsi trois fois inférieure à la durée de la période réelle.

Cette erreur aurait pu être évitée si, lors de l'analyse, l'absence d'entités graphiques sur plusieurs lignes brisées du diagramme temporel mixte avait été identifiée. Parmi les échelles cycliques suggérées par *GrAPHiST* lors de l'analyse, une échelle cyclique de 570 jours était proposée. La sélection de cette échelle, proche de la durée de la période réelle (581 jours), comme point de départ de l'analyse, aurait pu permettre une estimation assez précise de la durée de la période du cycle. Nous émettons l'hypothèse que si ce type d'erreur d'interprétation peut être récurrent, il peut facilement être corrigé par l'utilisateur en cours d'analyse.

Le tableau suivant présente, pour chaque jeu de données, l'écart entre le cycle estimé lors de l'analyse et le cycle réel, selon les différents critères identifiés :

- la colonne *Tx imprécision de la période estimée* correspond au taux d'imprécision de la durée estimée de la période du cycle, par rapport à la durée réelle de la période ;
- la colonne *Précision de l'emprise spatiale estimée* correspond à la précision de l'emprise spatiale estimée du cycle, par rapport à l'emprise spatiale réelle. Ce critère est défini de manière qualitative ;
- la colonne *Tx imprécision de l'emprise temporelle linéaire estimée* correspond au taux d'imprécision de l'emprise temporelle linéaire estimée du cycle, par rapport à l'emprise temporelle linéaire réelle ;
- la colonne *disp_t estim. vs disp_t réelle* correspond à la différence, exprimée en pourcentage, entre la dispersion temporelle estimée des événements cycliques, et leur dispersion temporelle réelle.

<i>Jeu</i>	<i>Tx imprécision de la période estimée</i>	<i>Précision de l'emprise spatiale estimée</i>	<i>Tx imprécision de l'emprise temporelle linéaire estimée</i>	<i>disp_t estim. vs disp_t réelle</i>	Nbre d'evts
A_1	1.9%	imprécise	0.2%	0.5%	285
A_2	X	assez imprécise	11.7%	15.0%	490
A_3	0.1%	relativement précise	0.1%	1.2%	507
A_4	0.6%	très imprécise	6.8%	3.0%	290
B_5	0.2%	relativement précise	14.3%	6.0%	610
B_6	0.4%	relativement précise	26.0%	12.0%	1010
B_7	0%	très imprécise	4.9%	3.0%	570
B_8	1.6%	assez imprécise	13.8%	3.0%	880

L'analyse a également permis de localiser la zone spatiale d'apparition des événements cycliques dans chaque jeu de données :

- Concernant les jeux de données présentant une dispersion spatiale moins élevée (jeux A_3 , A_4 , B_7 , B_8) :
 - l'emprise spatiale estimée pour le jeu A_3 est de même taille que l'emprise réelle, et légèrement décalée vers le sud-ouest. L'estimation de l'emprise spatiale est relativement précise ;
 - l'emprise spatiale estimée pour le jeu A_4 est plus petite que l'emprise réelle, et décalée vers le nord-ouest. L'estimation de l'emprise spatiale est très imprécise ;
 - l'emprise spatiale estimée pour le jeu B_7 est plus petite que l'emprise réelle, et décalée vers le nord-ouest. L'estimation de l'emprise spatiale est très imprécise ;
 - l'emprise spatiale estimée pour le jeu B_8 est plus petite que l'emprise réelle, et décalée vers l'ouest. L'estimation de l'emprise spatiale est assez imprécise.

L'emprise estimée est généralement plus petite que l'emprise réelle. L'emprise estimée est généralement imprécise, sauf pour le cas du jeu A_3 où l'emprise estimée est plus proche de l'emprise réelle.

- Pour les jeux de données présentant une dispersion spatiale plus élevée (jeux A_1 , A_2 , B_5 , B_6) :
 - l'emprise spatiale estimée pour le jeu A_1 est plus grande que l'emprise réelle, et légèrement décalée vers le nord. L'estimation de l'emprise spatiale est imprécise ;
 - l'emprise spatiale estimée pour le jeu A_2 est de même taille que l'emprise réelle, et légèrement décalée vers le sud-ouest. L'estimation de l'emprise spatiale est assez imprécise ;
 - l'emprise spatiale estimée pour le jeu B_5 est plus grande que l'emprise réelle, et légèrement décalée vers le nord-est. L'estimation de l'emprise spatiale est relativement précise ;
 - l'emprise spatiale estimée pour le jeu B_6 est plus grande que l'emprise réelle, et légèrement décalée vers le sud. L'estimation de l'emprise spatiale est relativement précise.

L'emprise estimée est généralement plus grande que l'emprise réelle. L'emprise estimée reste peu précise, mais est paradoxalement plus précise que pour les jeux de données présentant une dispersion spatiale moindre, hormis pour le cas du jeu A_1 .

L'analyse permet ainsi d'identifier globalement la localisation des événements cycliques. La définition des contours de cette zone manque toutefois de précision.

L'analyse a permis d'identifier l'emprise temporelle linéaire couverte par le cycle, pour chaque jeu de données :

- Pour les jeux de données présentant 80% d'événements aléatoires (jeux A), l'emprise temporelle linéaire est estimée avec un taux d'imprécision allant de 0.2% (jeu A_1), à 11.7% (jeu A_3), par rapport à la valeur réelle.
- Pour les jeux de données présentant 90% d'événements aléatoires (jeux B), l'emprise temporelle linéaire est estimée avec un taux d'imprécision allant de 4.9% (jeu B_7), à 26.0% (jeu B_6), par rapport à la valeur réelle.

L'estimation de l'emprise temporelle linéaire couverte par le cycle manque ainsi de précision, notamment quand la proportion d'événements aléatoires devient plus importante.

L'analyse des récurrences a également permis de définir l'étendue de l'intervalle temporel, durant lequel apparaissent les événements cycliques :

- Pour les jeux de données présentant une dispersion temporelle allant de 8% à 25% (jeux A_1, A_3, B_5, B_7), la différence, exprimée en pourcentage, entre les dispersions temporelles estimées et réelles varie entre 0.5% à 6%.
- Pour les jeux de données présentant une dispersion temporelle allant de 25% à 50% (jeux A_2, A_4, B_6, B_8), la différence, exprimée en pourcentage, entre les dispersions temporelles estimées et réelles varie entre 3% à 15%.

L'estimation de l'emprise temporelle cyclique couverte par le cycle manque ainsi de précision, notamment lorsque les événements apparaissent dans un intervalle temporel plus large.

En comparant la précision de nos résultats avec le nombre d'événements contenu dans chaque jeu de données, il apparaît que :

- l'estimation de l'emprise spatiale des événements récurrents est plus précise pour les jeux de données contenant le plus grand nombre d'événements (Repère A de la Figure 8.38) ;
- l'estimation de l'emprise temporelle linéaire des événements récurrents est plus précise pour les jeux de données contenant le plus petit nombre d'événements (Repère B de la Figure 8.38) ;
- l'estimation de l'emprise temporelle cyclique des événements récurrents est plus précise pour les jeux de données contenant le plus petit nombre d'événements (Repère C de la Figure 8.38).

Cependant, le nombre de jeux de données que nous analysons, et donc le nombre de résultats que nous possédons, n'est pas suffisant pour conclure à l'existence de corrélations entre la précision de ces résultats et le nombre d'événements contenus dans le jeu de données.

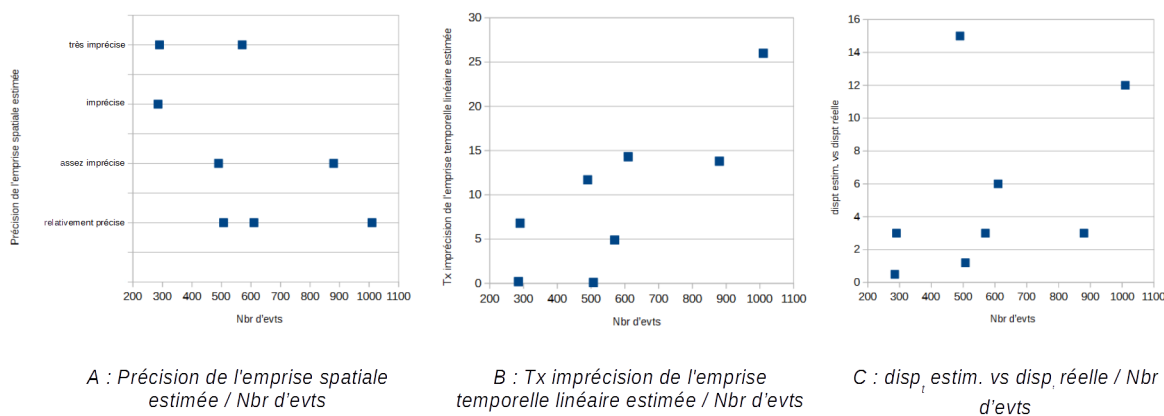


FIGURE 8.38 – Précision des résultats obtenus par rapport au nombre d'événements

L'interface de visualisation manque pour le moment de précision dans l'estimation des emprises, spatiale et temporelle, couvertes par des événements cycliques. Cette expérimentation a cependant montré qu'une analyse exploratoire, effectuée au moyen de *GrAPHiST*, permet de révéler l'existence d'un cycle localisé, d'identifier sa période, et d'estimer sa position globale dans l'espace et dans le temps au sein d'un jeu de données majoritairement aléatoires.

L'identification visuelle de clusters et de récurrences spatialisées dans de tels jeux de données étant possible au moyen de notre interface, nous effectuons l'hypothèse que notre approche peut être utilisée afin d'identifier des structures analogues dans des jeux de données réels.

Une possibilité d'application de la recherche de clusters pourrait être l'analyse de la distribution spatio-temporelle d'événements, liés à des phénomènes météorologiques, selon plusieurs échelles calendaires. De la même manière, ces événements pourraient faire l'objet d'une recherche de récurrences spatialisées, dans le but d'identifier une réapparition cyclique du phénomène, qui présenterait différentes périodes dans différentes régions de l'espace.

Chapitre 9

Analyse de clusters spatio-temporels dans des jeux de données réels

Sommaire

9.1	Principe et objectifs	330
9.1.1	Application de notre approche sur des données réelles	330
9.1.2	Choix des données	330
9.1.3	Structures recherchées	330
9.2	Données	331
9.2.1	Données d'accidents routiers	331
9.2.2	Données d'inondations	332
9.2.3	Données criminologiques	334
9.3	Procédé général	335
9.3.1	Analyse selon une entrée temporelle	336
9.3.2	Analyse selon une entrée spatiale	337
9.4	Analyse des données d'accidents routiers	337
9.4.1	Scénario d'analyse	337
9.4.2	Résultats	339
9.4.3	Synthèse	342
9.5	Analyse des données d'inondations	342
9.5.1	Scénario d'analyse	342
9.5.2	Résultats	344
9.5.3	Synthèse	349
9.6	Analyse des données criminologiques	350
9.6.1	Scénarios d'analyse	350
9.6.2	Résultats	355
9.6.3	Synthèse	366
9.7	Synthèse générale	367

9.1 Principe et objectifs

9.1.1 Application de notre approche sur des données réelles

Dans les chapitres précédents, nous montrions que notre interface de géovisualisation pouvait être utilisée afin d'identifier des patterns présents dans des jeux de données spatio-temporelles sous format événementiel. Ces patterns étaient relatifs au temps linéaire (Chapitre 7), ou relatifs au temps cyclique (Chapitre 8).

Dans ce chapitre, nous appliquons notre approche d'analyse exploratoire à différents jeux de données réelles, et nous tentons d'identifier des structures dans leur distribution spatio-temporelle, afin d'illustrer les possibilités d'utilisation de notre interface.

9.1.2 Choix des données

Afin d'illustrer le caractère générique de notre approche, nous effectuons ces analyses sur une série de jeux de données liés à différentes problématiques métiers. Ces jeux de données concernent des accidents routiers, des données d'inondations et des données de criminologie :

- Le premier jeu de données que nous analysons rassemble des données relatives à des accidents routiers sur le territoire de la France métropolitaine, de 2005 à 2015.
- Le second jeu de données concerne des épisodes d'inondation sur le territoire de la France métropolitaine depuis la fin du XVIII^e siècle à nos jours.
- Le troisième jeu de données rassemble des données de criminologie relatives à des délits et crimes ayant eu lieu sur le territoire de la ville de Chicago de 2001 à 2017.

9.1.3 Structures recherchées

Lors des analyses, nous cherchons à identifier des clusters spatio-temporels relatifs au temps cyclique ou au temps linéaire. Nous nous intéressons donc à identifier des zones spatiales où la distribution temporelle des événements se distingue des autres zones de la carte, ou bien des intervalles temporels, linéaires ou cycliques, où la distribution spatiale des événements correspondants est particulière.

9.1.3.1 Analyse selon le temps linéaire

Au travers d'une analyse de clusters selon le temps linéaire, nous étudions l'évolution de ces phénomènes au cours du temps et selon l'espace, au travers de zones spatiales définies à l'avance ou non.

Le but de cette analyse est d'identifier les lieux et moments d'apparition de possibles pics d'intensité de ces phénomènes, ou d'identifier une évolution globale (augmentation, diminution) de l'intensité de ces phénomènes selon leur localisation dans l'espace.

9.1.3.2 Analyse selon le temps cyclique

Au cours d'une analyse de clusters selon le temps cyclique, nous étudions la distribution spatio-temporelle des phénomènes au travers d'échelles calendaires, comme la journée ou l'année. L'analyse peut s'effectuer au travers de zones spatiales définies à l'avance, en étudiant la distribution temporelle des événements dans chacune de ces zones.

Le but de cette analyse est d'identifier si des concentrations des événements dans l'espace correspondent à une concentration des mêmes événements durant certaines périodes de la journée ou de l'année. L'analyse peut également permettre d'identifier à quelle période de la journée ou

de l'année ces événements ont majoritairement lieu, et si cette période est la même pour chaque portion de l'espace.

9.2 Données

9.2.1 Données d'accidents routiers

9.2.1.1 Description des données

Origine Les données dont nous disposons proviennent de la plateforme de diffusion de données publiques de l'État français, *Data.gouv.fr* [État français, 2017c], et sont produites par l'*Observatoire national interministériel de la sécurité routière (Onisr)* [État français, 2017b].

Format des données Les données relatives aux accidents de la route se présentent sous la forme de plusieurs fichiers :

- un fichier relatif à l'accident en lui-même, contenant la localisation ponctuelle de l'accident, son moment d'apparition enregistré avec une granularité temporelle de l'ordre de la minute, et des informations thématiques portant notamment sur les conditions de météorologie et de luminosité de l'accident ;
- un fichier relatif au lieu de l'accident, comportant des informations thématiques sur l'état de la route, le type de voirie, etc ;
- un fichier relatif à chaque véhicule impliqué dans l'accident, comportant des informations thématiques sur chaque véhicule ;
- un fichier relatif à chaque personne impliquée dans l'accident, comportant des informations thématiques sur chaque personne.

Limites des données utilisées Lors d'une première visualisation des données disponibles sur le portail *Data.gouv.fr*, nous avons réalisé que de nombreuses données ne possédaient pas de coordonnées valides, les rendant impossibles à représenter sur une carte. Ceci peut être dû à des erreurs de manipulation lors de la création des données, ou au fait que la localisation de l'événement n'est pas renseignée pour certains accidents. Ces données n'ont pas pu être utilisées lors de notre analyse.

Les données présentent également des manques, les données relatives à certains départements ou certaines communes n'étant pas renseignées pour certaines périodes. Par exemple, les données relatives à l'année 2009 n'étaient pas disponibles pour de nombreux départements.

9.2.1.2 Implémentation dans notre modèle de données

Les données provenant des quatre fichiers importés depuis le portail *data.gouv.fr* sont compilées au moyen d'un script *SQL*, afin de créer une entité spatio-temporelle unique pour chaque accident routier (Figure 9.1). Les données relatives aux différents usagers et aux différents véhicules impactés par un accident sont résumées par des variables renseignant le nombre d'usagers blessés ou décédés lors de l'accident, et le nombre de véhicules impactés.

Les informations de chaque entité sont classées selon la structure décrite par le *Pyramid Model*, et entrées dans la base au moyen d'un script *SQL*. Nous créons pour chaque entité du tableau un événement, composé d'une sous-partie temporelle, et possédant huit arguments thématiques.

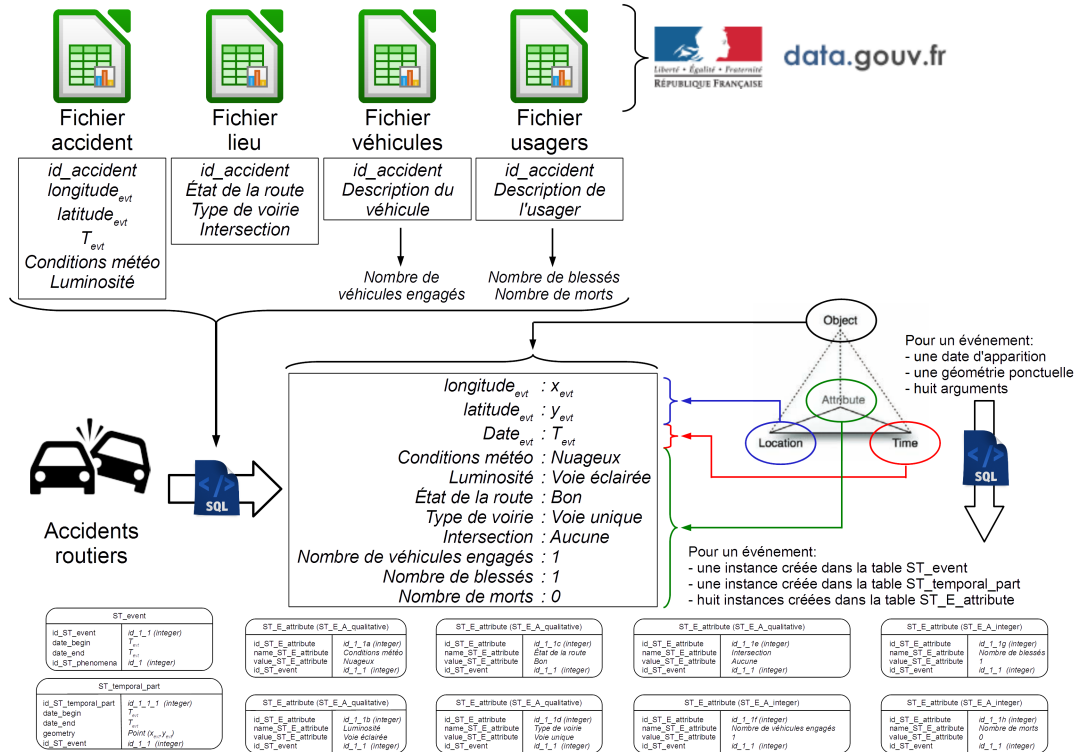


FIGURE 9.1 – Implémentation des données relatives aux accidents routiers dans notre modèle

9.2.2 Données d'inondations

9.2.2.1 Description des données

Origine Les données dont nous disposons proviennent de la *Base de Données Historiques sur les Inondations (BDHI)* [État français, 2017a], produite par le *Ministère de l'Écologie, du Développement durable et de l'Énergie*, en collaboration avec le laboratoire *Irstea*, l'équipe *Steamer* du *Laboratoire d'informatique de Grenoble*, et *Acthys Diffusion*.

Cette base de données renseigne les données relatives aux inondations survenues sur le territoire de la France métropolitaine depuis la fin du *XVIIIe* siècle.

Format des données Les données se présentent sous un format événementiel.

Pour chaque cas d'inondation, la base de données *BDHI* renseigne l'ensemble des communes touchées, les dates de début et fin de l'inondation, ainsi qu'une suite d'informations thématiques relatives à l'inondation. Chaque événement recensé dans la base de données représente l'impact d'une inondation sur une commune.

Ainsi pour une inondation, dénommée x , débutant à T_1 et finissant à T_E et impactant n communes, la *BDHI* comporte n événements (un pour chaque commune), partageant les dates de début et de fin T_1 et T_E .

Si une commune est touchée par plusieurs inondations au cours du temps, un événement est créé pour chaque inondation touchant cette commune.

Chaque événement possède :

- le code *INSEE* de la commune touchée ;
- les dates de début et de fin de l'inondation ;

- plusieurs informations concernant la dénomination de l'inondation, le nombre de victimes, le type de l'inondation (crue pluviale, forte marée, rupture de digue ou de barrage ...), etc. Ces données sont communes à tous les événements traitant de la même inondation.

Limites des données utilisées Pour les besoins de l'expérimentation, nous avons adapté les données de la *BDHI* à notre format de représentation d'événements ponctuels.

Nous avons ainsi considéré les événements comme ponctuels dans le temps en ne prenant en compte que la date du début de l'inondation, et comme ponctuels dans l'espace en résumant leur localisation au barycentre du territoire de la commune touchée.

Chaque cas d'inondation est donc représenté par un ensemble d'événements apparaissant au même moment, partageant la même composante thématique, chacun représentant une commune touchée. La composante spatiale de l'inondation forme alors un nuage de point sur la carte.

Les événements correspondant à différentes inondations touchant les mêmes communes sont alors superposés dans la carte, et les événements d'une même inondation touchant différentes communes sont superposés dans le diagramme temporel mixte.

Les événements traités ici représentent le fait qu'une commune soit touchée par une inondation. Ainsi, si le débordement d'une rivière touche quatre communes, ce phénomène sera retranscrit par quatre événements. L'objectif de cette analyse n'est pas d'améliorer nos connaissances sur les phénomènes d'inondation en France métropolitaine, mais d'évaluer l'intérêt de l'outil pour la recherche de clusters spatio-temporels dans des données d'inondation.

9.2.2.2 Implémentation dans notre modèle de données

Les données importées depuis la *BDHI* ne possèdent pas de géométrie, cependant chaque événement possède le code *INSEE* de la commune correspondante.

Afin de créer la composante spatiale de l'événement relatif à l'impact d'une commune par une inondation, nous utilisons un fichier *shapefile* contenant la géométrie des communes françaises, que nous relient aux données de la *BDHI* grâce au code *INSEE*. Nous calculons pour chaque commune les coordonnées du barycentre du territoire correspondant, et nous considérons ces coordonnées géographiques comme celles de l'événement. Ces opérations sont effectuées au moyen du logiciel *QGIS* (Figure 9.2).

Les informations de chaque entité sont classées selon la structure décrite par le *Pyramid Model*, et entrées dans la base au moyen d'un script *SQL*. Nous créons pour chaque entité du tableau un événement, composé d'une sous-partie temporelle, et possédant trois arguments thématiques (Figure 9.2).

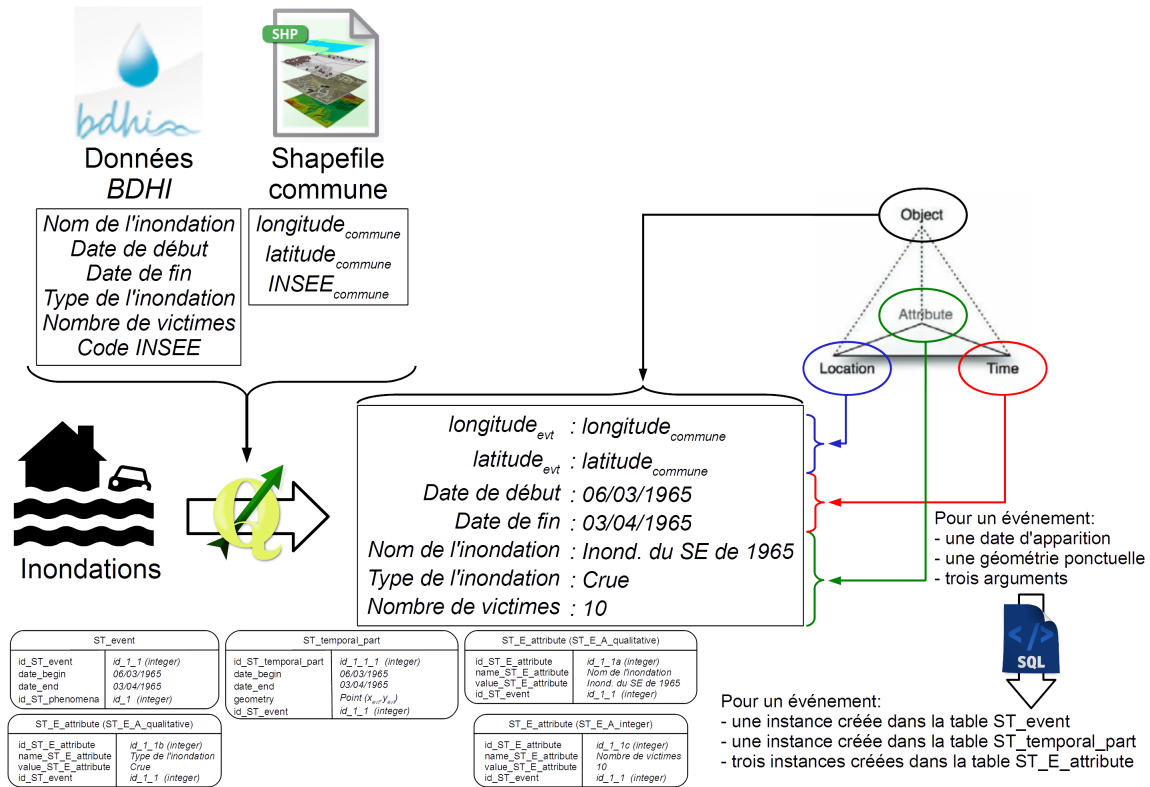


FIGURE 9.2 – Implémentation des données relatives aux inondations dans notre modèle

9.2.3 Données criminologiques

9.2.3.1 Description des données

Origine Les données dont nous disposons proviennent du portail de données en ligne de la ville de Chicago, USA (*Chicago Data Portal* [City of Chicago, 2017]). Chaque donnée correspond à un délit ou crime rapporté par la police de Chicago (*Chicago Police Department*) depuis 2001. Cette base de données est régulièrement mise à jour.

Format des données Les données de criminologie se présentent sous un format événementiel. Chaque événement correspond à :

- des coordonnées géographiques, le lieu de l'événement étant résumé à un point dans l'espace ;
- un moment d'apparition, la granularité utilisée étant la minute ;
- une série d'informations portant sur le type de délit ou de crime commis, la description du lieu, les circonstances de l'événement, etc.

Limites des données utilisées Lors d'une première visualisation des données, effectuée après leur import depuis le portail de la ville de Chicago, nous avons réalisé que de nombreuses données ne possédaient pas de coordonnées valides. Ces données n'ont pas pu être utilisées pour l'analyse. Nous émettons cependant l'hypothèse que, malgré ce manque de données, les cas d'étude présentés ici sont pertinents pour montrer les opportunités d'analyse offertes par *GrAPHiST*.

9.2.3.2 Implémentation dans notre modèle de données

Les données brutes provenant du *Chicago Data Portal* sont disponibles sous la forme d'un tableau. Nous utilisons un script *SQL* pour introduire ces événements dans notre base de don-

nées, en classant les informations de chaque événement selon la structure décrite par le *Pyramid Model*. Nous créons pour chaque entrée du tableau une entité événement, composée d'une unique sous-partie temporelle, et possédant quatre arguments thématiques (Figure 9.3).

Au vu du nombre important de données, nous focalisons notre analyse sur certains types de délit/crime : les homicides, les *crimes violents* regroupant les vols avec violence (*robbery*) et les cambriolages (*burglary*). Les événements correspondants sont regroupés en effectuant une sélection des données brutes selon leurs attributs thématiques (Figure 9.3).

Pour chaque type de délit/crime, un phénomène distinct est créé pour regrouper les événements correspondants. Notre base de données contient ainsi un phénomène *Homicides*, un phénomène *Cambriolages*, et un phénomène *Crimes violents*.

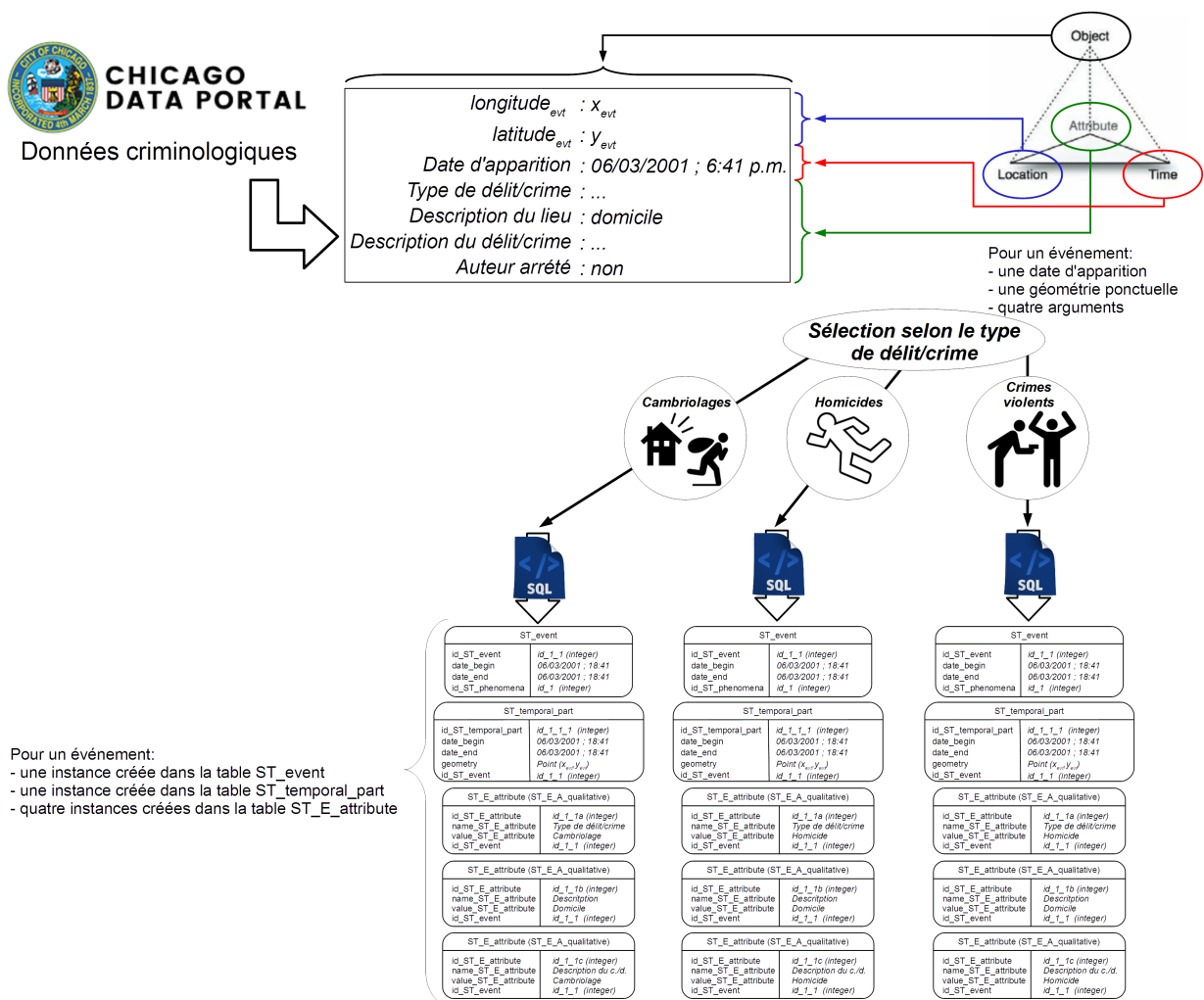


FIGURE 9.3 – Implémentation des données de criminologie dans notre modèle

9.3 Procédé général

Lors des analyses décrites dans ce chapitre, nous cherchons à identifier des clusters spatio-temporels, relatifs au temps cyclique ou au temps linéaire, dans les jeux de données présentés précédemment. Cette analyse peut s'effectuer selon une entrée temporelle, ou selon une entrée

spatiale au travers de zones spatiales définies à l'avance.

Lors de ces analyses, nous ne suspectons pas une distribution spatio-temporelle particulière des événements a priori. Aussi, nous effectuons une première analyse des jeux de données selon une entrée temporelle, en représentant les événements en fonction de leur moment d'apparition, relatif au temps linéaire ou à une échelle temporelle cyclique.

Pour compléter la première analyse, pour l'approfondir en partant des premiers résultats trouvés, ou pour prendre un autre point de vue sur les données si la première analyse n'a pas produit de résultats concluants, nous effectuons une seconde analyse selon une entrée spatiale.

Pour cela nous classons les événements selon leur appartenance à différentes zones spatiales, et nous les représentons en fonction de leur lieu d'apparition. Pour chaque jeu de données, nous classons les événements selon un découpage territorial plus ou moins relié à la problématique étudiée.

9.3.1 Analyse selon une entrée temporelle

9.3.1.1 Analyse selon le temps linéaire

L'analyse consiste à représenter les événements selon leur moment d'apparition linéaire en utilisant la variable visuelle de la valeur, et à identifier sur la carte une concentration spatiale d'entités présentant une valeur similaire.

Le nombre d'intervalles temporels linéaires utilisés pour classer les événements a une incidence sur la forme de la structure spatio-temporelle que l'on recherche.

L'utilisation de nombreux intervalles temporels, représentés par de nombreuses valeurs variant peu d'un intervalle à l'autre, permet d'identifier des pics d'intensité du phénomène, localisés dans l'espace et peu étendus dans le temps.

En revanche, l'utilisation d'un nombre limité d'intervalles, représentés par un nombre limité de valeurs variant de manière plus marquée d'un intervalle à l'autre, peut permettre d'identifier des variations globales du phénomène, localisées dans l'espace, mais étendues dans le temps.

Prenons l'exemple d'événements apparaissant sur une étendue temporelle totale de dix ans :

- L'utilisateur peut choisir d'utiliser dix intervalles d'un an. La représentation obtenue peut permettre d'identifier, dans une zone de la carte, un pic d'intensité du phénomène durant une durée d'un an. Cette structure est visible par un regroupement spatial d'entités ayant pour valeur majoritaire celle représentant l'année correspondante.
- L'utilisateur peut également choisir d'utiliser deux intervalles de cinq ans. Cette représentation obtenue peut permettre d'identifier, dans une zone de la carte, l'augmentation, la diminution, ou la stagnation de l'intensité du phénomène observé au cours du temps. Pour cela, l'utilisateur doit identifier, dans la zone spatiale correspondante, la proportion des valeurs représentant les deux périodes de cinq ans parmi la population d'événements concernés.

9.3.1.2 Analyse selon une échelle temporelle cyclique

L'analyse consiste à représenter les événements selon leur moment d'apparition cyclique en utilisant la variable visuelle de la couleur, et à identifier sur la carte une concentration spatiale

d'entités présentant une couleur similaire.

Le nombre d'intervalles temporels cycliques utilisés pour classer les événements, et les couleurs utilisées pour représenter ces derniers ont également une incidence sur la recherche de structures spatio-temporelles.

L'utilisation de nombreux intervalles et de nombreuses couleurs permet d'identifier des pics d'intensité cyclique du phénomène, dont l'intervalle temporel d'apparition est plus restreint.

En revanche, le nombre d'intervalles et de couleurs utilisés, peut être déterminé de manière à reproduire la structure temporelle d'échelles calendaires, afin d'analyser la distribution spatio-temporelle des événements au travers de ces échelles.

9.3.2 Analyse selon une entrée spatiale

L'analyse consiste à représenter les événements selon l'appartenance de leur lieu d'apparition à différentes zones spatiales en utilisant la variable visuelle de la couleur, et à identifier une structure temporelle formée par des événements de couleur similaire sur les diagrammes temporels.

Les zones utilisées pour classer les événements sont créés à l'avance au moyen d'outils *SIG* dédiés, et sont importées dans un second temps dans l'interface. Ces zones correspondent à un découpage du territoire en lien avec le phénomène analysé.

L'analyse selon le temps linéaire s'effectue en observant la disposition des événements de couleur similaire le long de l'axe des abscisses du diagramme temporel mixte.

L'analyse selon le temps cyclique s'effectue :

- en observant la disposition des événements de couleur similaire le long de l'axe des ordonnées du diagramme temporel mixte ;
- en observant la distribution temporelle cyclique des événements de chaque zone dans le diagramme temporel circulaire. Le nombre de sections du diagramme peut être modifié. Un nombre important de sections permet d'identifier un pic d'intensité du phénomène apparaissant dans un intervalle temporel restreint. Le nombre de sections peut également être choisi afin de reproduire la structure d'une échelle temporelle calendaire.

L'utilisateur peut également masquer les événements de certaines zones de classification spatiale, afin de focaliser l'analyse sur un groupe d'événements en particulier.

9.4 Analyse des données d'accidents routiers

9.4.1 Scénario d'analyse

Au vu du nombre important de données, nous nous focalisons sur les accidents survenus autour de la ville de Grenoble. L'analyse s'effectue sur une zone centrée sur Grenoble, englobant les massifs de la Chartreuse (au nord), du Vercors (au sud-ouest) et de Belledonne (à l'est). L'emprise temporelle de l'étude s'étend du *1er janvier 2005* au *1er janvier 2016*.

Les événements sont représentés sur la carte et dans le diagramme temporel mixte sous un format agrégé.

Les manques présents dans les données disponibles sur le portail *Data.gouv.fr* nous empêchent d'effectuer une analyse des événements selon le temps linéaire sur l'étendue temporelle choisie. Par exemple, nous ne possédons pas de données sur l'année 2009, et nous ne possédons pas de données concernant des accidents survenus sur la commune de Grenoble avant l'année 2010.

Nous nous limitons donc pour ce jeu de données à une analyse selon le temps cyclique. Nous émettons l'hypothèse que les événements d'accidents peuvent présenter des distributions spatiales particulières selon la saison de l'année observée. Les accidents routiers pourraient être plus fréquents en hiver qu'en été, et se dérouler majoritairement dans les zones de montagne lors de la saison hivernale.

Les données sont utilisées pour une analyse selon une échelle cyclique d'un an, afin d'étudier la distribution spatio-temporelle des accidents routiers selon les quatre saisons de l'année.

L'échelle temporelle cyclique est fixée à 365 jours. L'analyse s'effectue dans un premier temps selon une approche temporelle, en utilisant une classification des événements selon leur moment d'apparition cyclique, en classant les événements au moyen de quatre intervalles de trois mois. Les événements sont ensuite représentés grâce à quatre couleurs différentes selon leur appartenance à ces intervalles temporels.

La phase du diagramme temporel mixte est modifiée de manière à ce que la première instance de l'échelle cyclique de 365 jours débute un 21 décembre. Les quatre intervalles temporels utilisés pour classer les événements correspondent ainsi aux quatre saisons de l'année (Figure 9.4).

L'étendue temporelle d'étude étant d'une durée de dix ans, le décalage entre les dates représentant le début de la première et de la dernière instance de l'échelle cyclique sera inférieur à trois jours. Les intervalles temporels permettant de classer les événements ayant une étendue de trois mois, nous émettons l'hypothèse que ce décalage n'aura pas d'incidence significative sur les résultats de l'analyse.

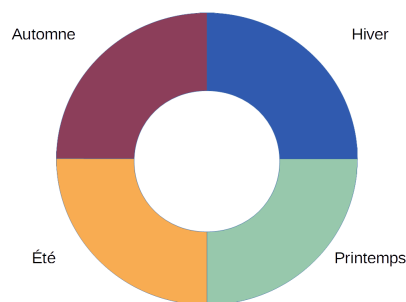


FIGURE 9.4 – Division de l'échelle temporelle de 365 jours en quatre sections

L'analyse selon une entrée temporelle est ensuite complétée par une analyse selon une entrée spatiale.

L'analyse s'effectue sur le diagramme temporel circulaire une fois les événements classés selon leur lieu d'apparition. Les zones utilisées pour la classification spatiale sont au nombre de trois, et sont créées de manière à correspondre :

- au territoire occupé par la commune de Grenoble ;
- aux territoires des massifs de Chartreuse, du Vercors, et de Belledonne ;
- aux zones, hors Grenoble, situées dans les vallées de l'Isère et du Drac.

L'analyse de la composante temporelle des événements est dans un premier temps effectuée pour l'ensemble des zones, puis pour chaque zone en particulier. Pour effectuer ce découpage, nous avons émis l'hypothèse que le nombre d'accidents routiers peut être différent selon le territoire étudié (zone de montagne ou zone de vallée ; zone urbaine ou zone rurale).

Pour créer ces zones, nous avons utilisé un fichier vectoriel contenant la géométrie des communes couvrant le territoire de la zone étudiée. Nous avons ensuite fusionné les géométries des communes de manière à créer la géométrie des trois zones citées précédemment. La Figure 9.5 montre une copie d'écran de la représentation, au travers de notre interface, des trois zones créées.

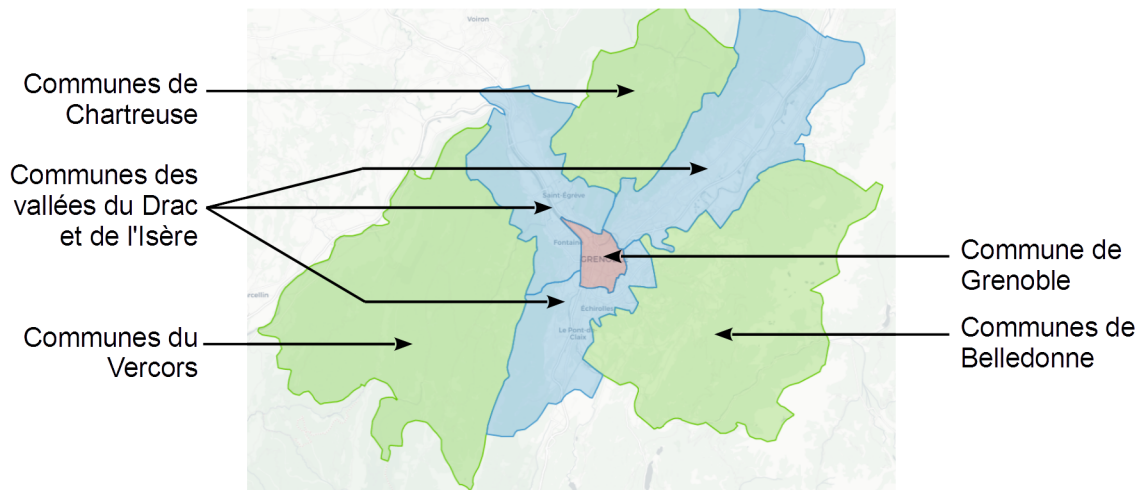


FIGURE 9.5 – Zones créées pour la classification spatiale des événements d'accidents routiers

9.4.2 Résultats

La Figure 9.6 montre une capture d'écran de l'interface de géovisualisation une fois saisis les paramètres décrits précédemment.

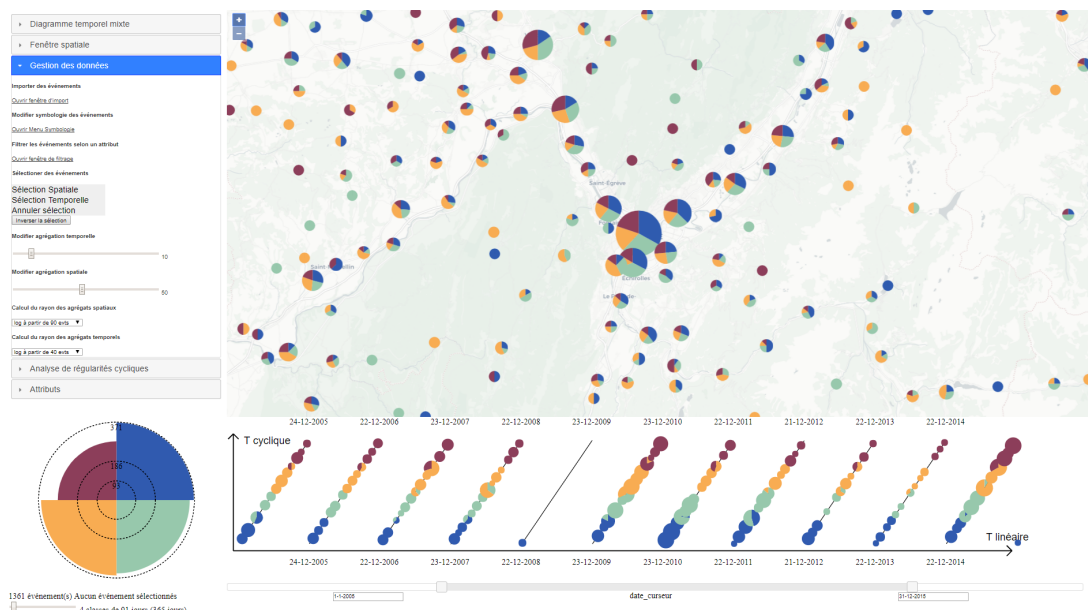


FIGURE 9.6 – Visualisation des accidents routiers dans la région de Grenoble, du 1er janvier 2005 au 1er janvier 2016

Dans cette carte, les événements apparaissant en dehors des zones définies dans la Figure 9.5 sont représentés, à l'instar des événements apparaissant au nord-ouest du Vercors. Ils ne sont cependant pas pris en compte dans l'analyse.

Le diagramme temporel cyclique montre une distribution temporelle des événements relativement constante durant l'hiver, le printemps et l'été (le nombre d'accidents apparaissant en hiver étant relativement plus important). L'automne présente un nombre d'événements moins important que les trois autres saisons (Figure 9.7).

On observe que la majorité des événements apparaissent au niveau de la métropole de Grenoble (Figure 9.7). La proportion d'événements apparaissant durant l'hiver et le printemps y est plus forte que dans le reste de la région étudiée (Figure 9.7). En revanche, la proportion d'événements apparaissant durant l'automne, et surtout durant l'été, y est plus faible que dans le reste de la région étudiée.

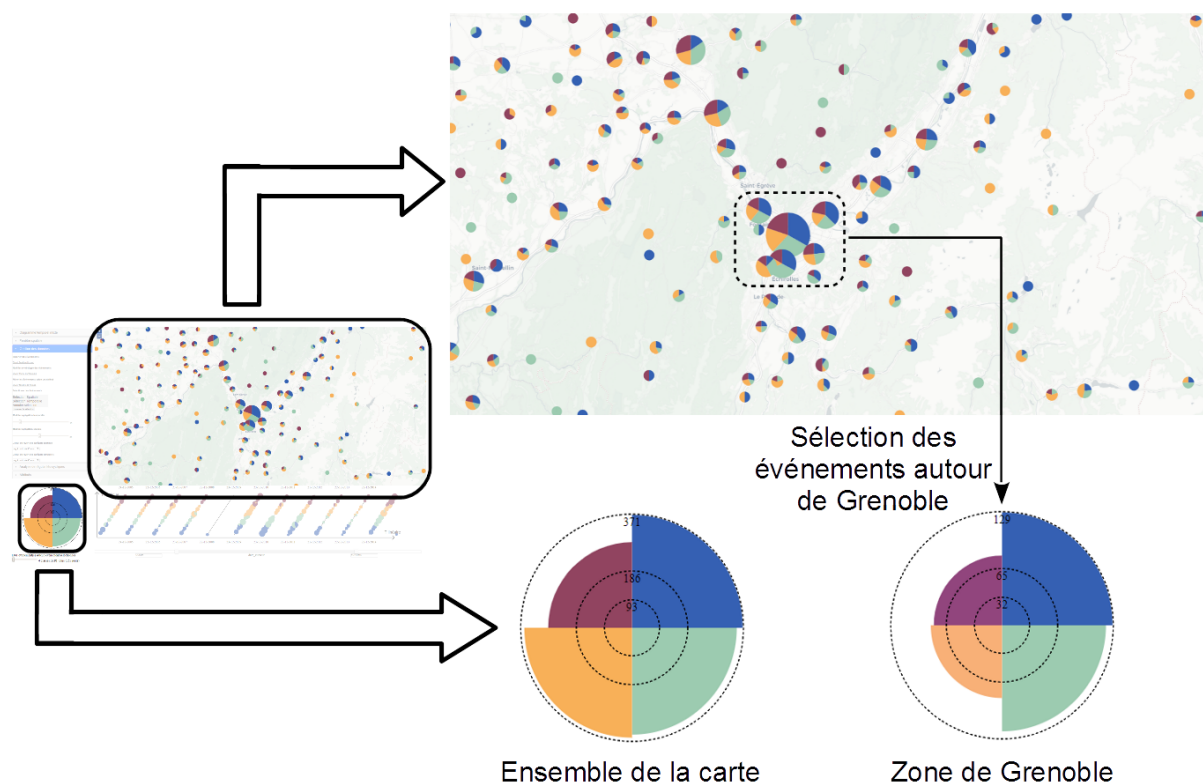


FIGURE 9.7 – Visualisation des accidents routiers autour de la ville de Grenoble

Pour approfondir cette analyse, et comparer la distribution temporelle des événements apparaissant dans le territoire de Grenoble, avec la distribution des événements survenant dans le reste de la zone d'étude, nous représentons les événements en fonction de leur lieu d'apparition.

La Figure 9.8 montre une capture d'écran de l'interface de géovisualisation une fois les événements classés selon le découpage territorial présenté dans la Figure 9.5. Les événements apparaissant en dehors des trois zones de classification sont masqués.

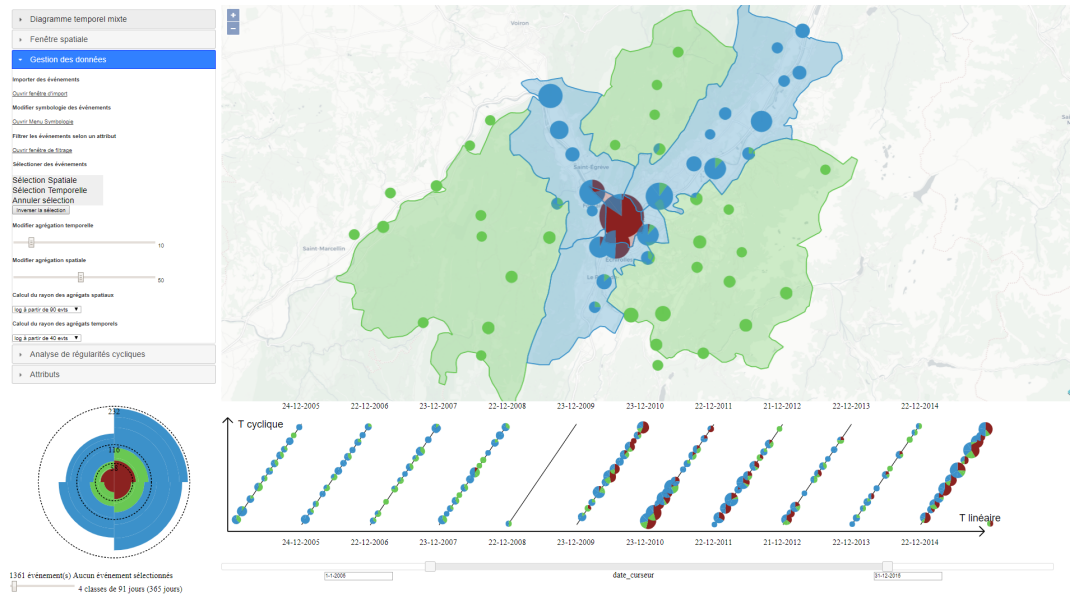


FIGURE 9.8 – Représentation des accidents routiers autour de la ville de Grenoble, selon leur lieu d’apparition

Le diagramme temporel circulaire montre une plus grande proportion d’apparitions d’événements durant l’hiver, puis durant le printemps. L’automne est la saison présentant le plus petit nombre d’apparitions d’événements (Figure 9.9). Nous comparons ensuite la distribution temporelle cyclique des événements de chaque zone (Figure 9.9).

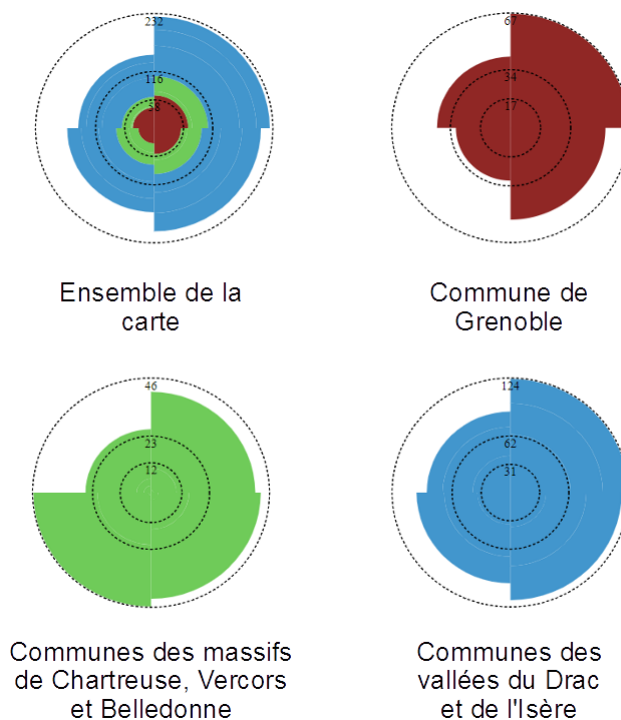


FIGURE 9.9 – Distribution des accidents routiers par zone

On distingue que la proportion d’accidents survenant l’hiver est beaucoup plus importante sur la commune de Grenoble que dans les autres territoires de l’étendue spatiale étudiée, tandis que la proportion d’accidents survenant l’été y est beaucoup moins importante.

On observe également que les événements survenant dans les massifs montagneux apparaissent majoritairement au printemps, et surtout l'été. Cette majorité d'événements apparaissant durant l'été ne se retrouve ni dans les autres zones, ni en prenant en compte l'ensemble de l'étendue spatiale étudiée.

9.4.3 Synthèse

Une première analyse, effectuée selon une entrée temporelle, permet d'identifier que la zone rassemblant la commune de Grenoble et ses alentours présente :

- la plus forte concentration spatiale d'accidents routiers de la région étudiée ;
- une proportion plus élevée d'accidents survenant durant l'hiver et le printemps ;
- une proportion plus faible d'accidents survenant durant l'automne, et surtout durant l'été.

Une seconde analyse, effectuée selon une entrée spatiale, permet d'identifier que :

- les accidents survenant sur l'ensemble de la région étudiée apparaissent majoritairement en hiver, et les accidents survenant en automne sont minoritaires ;
- les accidents survenant sur la commune de Grenoble apparaissent majoritairement en hiver, et les accidents survenant en été sont minoritaires ;
- les accidents survenant dans les communes des massifs de Chartreuse, du Vercors et de Belledonne apparaissent majoritairement en été, et les accidents survenant en automne sont minoritaires ;
- les accidents survenant dans les communes des vallées du Drac et de l'Isère présentent la même distribution temporelle cyclique que l'ensemble de la région étudiée.

On observe ainsi deux zones, la zone de Grenoble et les massifs de Chartreuse, du Vercors et de Belledonne, présentant des distributions temporelles cycliques distinctes, qui se différencient de l'ensemble de la zone étudiée.

9.5 Analyse des données d'inondations

9.5.1 Scénario d'analyse

Les événements sont analysés sur l'entièreté du territoire de la France métropolitaine, et sur une étendue temporelle allant du *1er janvier 1900* au *1er janvier 2016*.

La visualisation des événements sur la carte s'effectue dans un premier temps avec une représentation agrégée des événements, puis avec une représentation non-agrégée. Les événements sont représentés de manière agrégée dans le diagramme temporel mixte.

Nous émettons l'hypothèse que les phénomènes d'inondations sont reliés au cycle des saisons. Nous effectuons pour ce jeu de données une analyse selon une échelle cyclique d'un an, afin d'étudier la distribution spatio-temporelle des inondations selon les différents mois de l'année.

L'analyse s'effectue dans un premier temps selon une approche temporelle, en représentant les événements au travers d'une échelle cyclique d'une année, et en classant les événements selon leur moment d'apparition cyclique au moyen de douze intervalles d'un mois. Nous utilisons six couleurs primaires pour la représentation des événements (Figure 9.10).

Nous utilisons auparavant une échelle de 365 jours pour représenter une échelle cyclique d'une année. La durée exacte d'une année astronomique est cependant égale à 365,242 jours.

Sur une étendue temporelle de dix ans, le décalage produit entre les dates représentant le début de la première et de la dernière instance de l'échelle cyclique est inférieur à trois jours, et peut être négligé si la durée des intervalles utilisés pour classer les événements est de trois mois.

Sur une étendue temporelle de 115 ans, ce décalage est proche d'un mois, ce qui n'est plus négligeable, surtout si la durée des intervalles de classification des événements est également d'un mois.

Pour représenter les événements au travers d'une échelle cyclique d'une durée plus proche de l'année astronomique, nous modifions la granularité de notre représentation temporelle, et nous utilisons une échelle de 525 948 minutes, soit environ 365,242 jours. En utilisant cette échelle, le décalage produit entre les dates représentant le début de la première et de la dernière instance de l'échelle cyclique devient inférieur à une demi-journée.

La phase du diagramme temporel mixte est modifiée afin que le début de chaque instance de la période cyclique coïncide avec le premier janvier. Les douze classes temporelles créées correspondent au découpage temporel représenté dans la Figure 9.10.

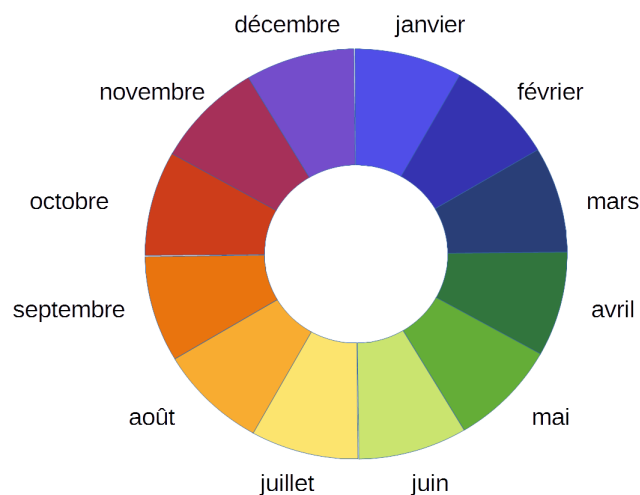


FIGURE 9.10 – Division de l'échelle temporelle d'un an en douze sections avec six couleurs primaires

L'analyse selon une entrée temporelle est ensuite complétée par une analyse selon une entrée spatiale.

L'analyse s'effectue sur le diagramme temporel circulaire une fois les événements classés selon leur lieu d'apparition. Les zones utilisées correspondent aux territoires des bassins versants des principaux cours d'eau de France métropolitaine (fleuves et rivières).

Afin de réduire le nombre de classes, nous avons fusionné certains bassins versants, en rattachant au bassin versant d'un cours d'eau le bassin versant de ses affluents.

La Figure 9.11 montre une copie d'écran de la représentation des zones obtenues, au travers de notre interface.

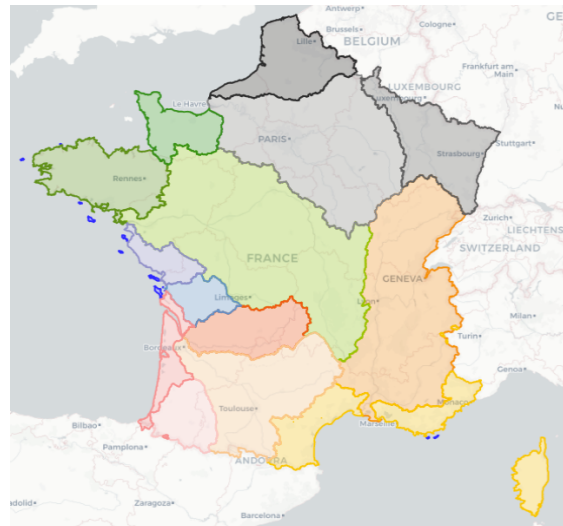


FIGURE 9.11 – Zones créées à partir des bassins versants pour la classification spatiale des événements d’inondation

9.5.2 Résultats

La Figure 9.12 montre une capture d’écran de l’interface de géovisualisation une fois saisi les paramètres décrits précédemment. La représentation des événements y est non-agrégée.

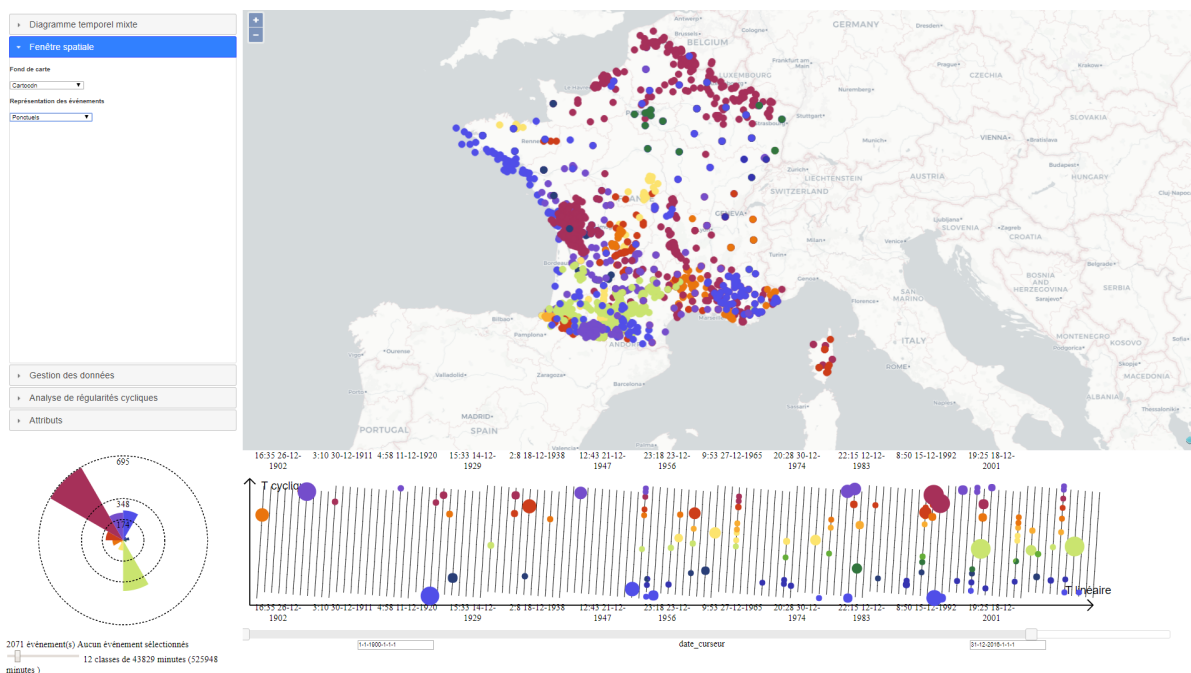


FIGURE 9.12 – Visualisation des cas d’inondations sur le territoire de la France métropolitaine, du 1^{er} janvier 1900 au 1^{er} janvier 2016

Le diagramme temporel circulaire montre que les événements apparaissent en majorité durant le mois de novembre, puis durant le mois de juin. On observe une moindre proportion d’appari-tions d’événements, quoique toujours forts, durant les mois décembre et janvier. Les événements apparaissant durant le reste de l’année sont peu représentés, et apparaissent majoritairement en octobre, puis en septembre (Figure 9.13).

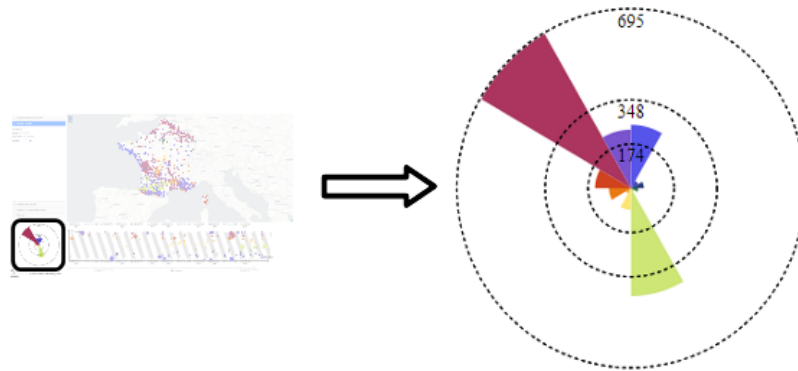


FIGURE 9.13 – Distribution temporelle des inondations durant l'année

La Figure 9.14 montre la représentation cartographique obtenue en choisissant une représentation agrégée (Repère 1) et non-agrégée (Repère 2) des événements.

La composante spatiale des événements correspondant au barycentre des communes touchées, les événements touchant une même commune se superposent sur la carte. Une représentation non-agrégée des événements peut alors engendrer une perte d'information, car dans ce cas de superposition seul le dernier événement touchant la commune est visible.

À l'inverse, une représentation agrégée permet de visualiser le nombre d'événements apparaissant au même endroit dans l'espace, et permet de visualiser la proportion de ces événements en fonction de leur moment d'apparition.

Cependant, la représentation non-agrégée permet ici une bonne identification du contour des zones touchées. Les entités ponctuelles relatives à une même inondation forment des grappes d'entités ponctuelles de même couleur, bien visibles sur la carte, laissant deviner la géométrie des cours d'eau en crue, à l'origine des inondations (Repère 1 de la Figure 9.14).

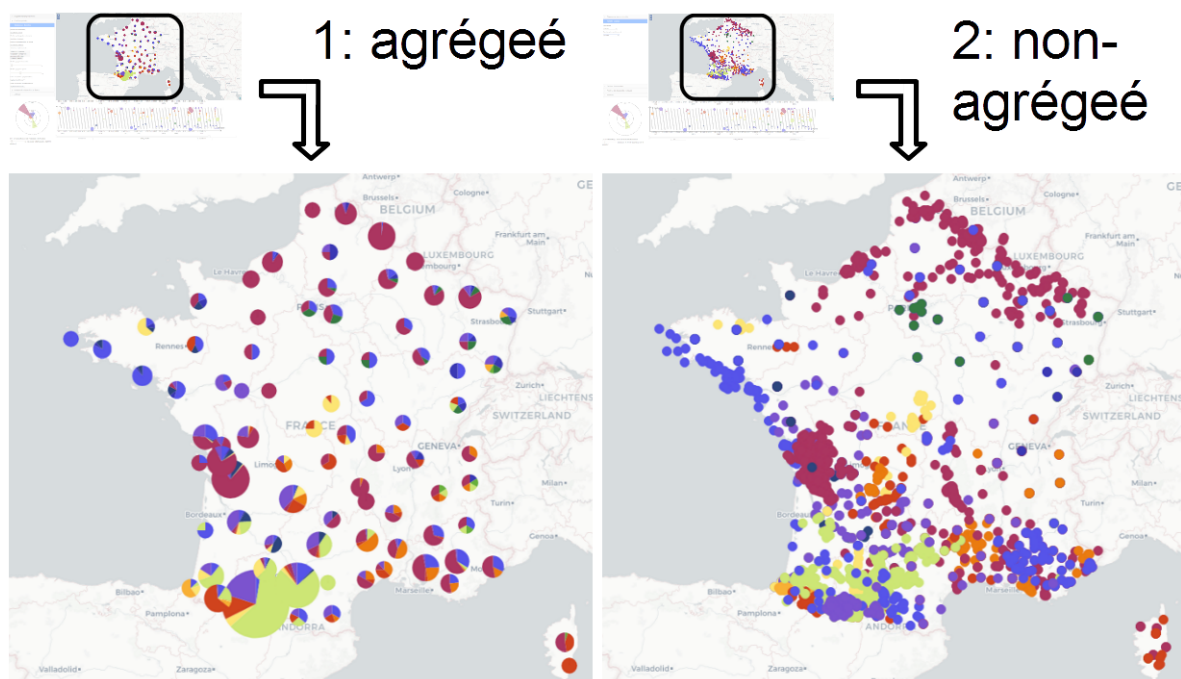


FIGURE 9.14 – Représentation des inondations selon une échelle cyclique d'un an

On retrouve dans ces deux cartes une concentration spatiale des événements au niveau des Pyrénées, et de la Charente.

Plusieurs clusters spatio-temporels sont également visibles dans la Figure 9.15 :

- la forte proportion d'événements apparaissant durant le mois de novembre se retrouve sur plusieurs zones de la carte, notamment au nord et au nord-est (Repère 6), au sud-est (Repère 9), en Auvergne (Repère 11), au niveau de la Charente (Repère 1) ;
- les événements apparaissant au sud-est (Repère 9) apparaissent majoritairement en novembre, mais une minorité apparaît en janvier ;
- les événements de la côte bretonne (Repère 5) et de la côte vendéenne (Repères 1 et 5) apparaissent majoritairement en janvier ;
- les événements des Pyrénées apparaissent majoritairement en juin (Repères 8 et 10), et une forte minorité apparaît en décembre, octobre et janvier ;
- les événements de la zone du Périgord (Repère 2) apparaissent majoritairement en décembre, mais une forte minorité apparaît en octobre et en septembre ;
- les événements du Pays basque (Repère 7) montrent une forte proportion d'événements apparaissant en août, une minorité apparaît en juillet ;
- les événements apparaissant en Corse (Repère 4) apparaissent majoritairement en octobre, une minorité apparaît en novembre ;
- au niveau du Gard, on observe de fortes proportions d'événements apparaissant en septembre (Repère 3).

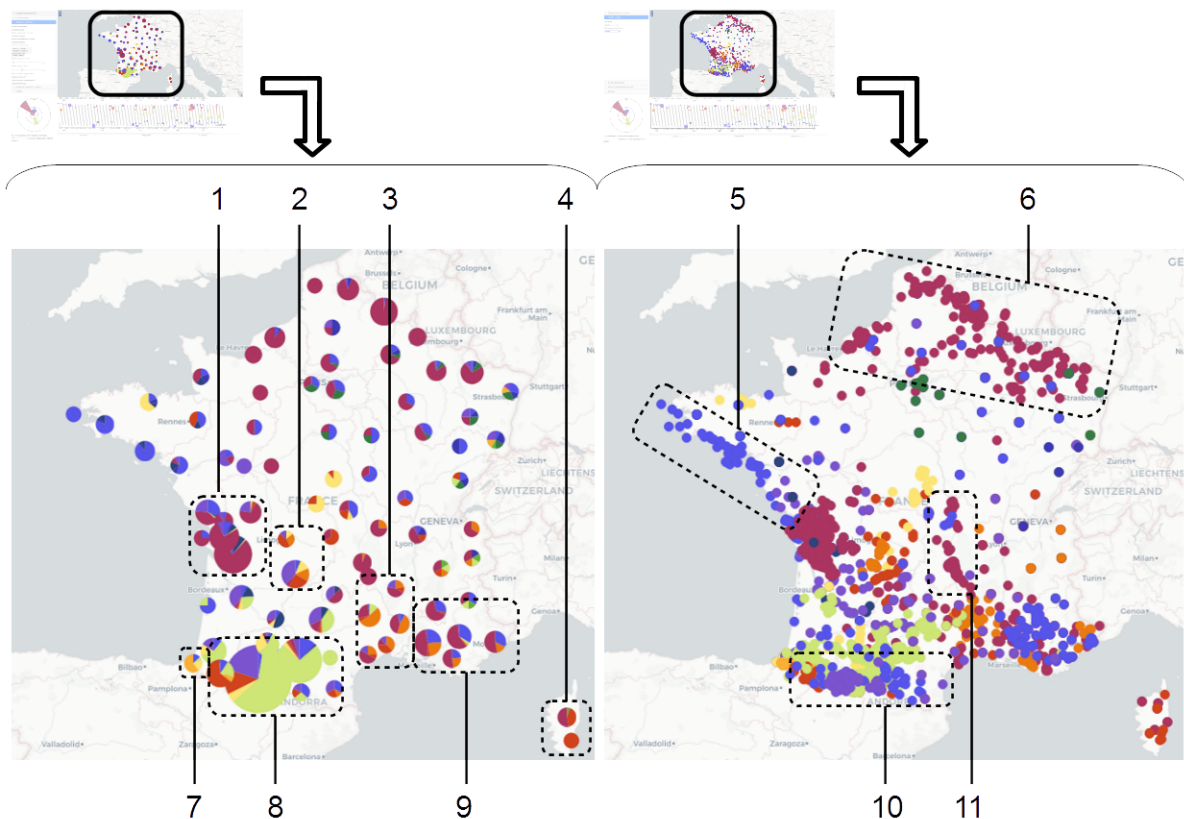


FIGURE 9.15 – Analyse des inondations selon une échelle cyclique d'un an

À la lecture des cartes de la Figure 9.15, la zone présentant le plus de diversité dans le moment d'apparition des événements au cours de l'année est celle située au sud du Périgord, entre la côte

atlantique et le Rhône. La Figure 9.16 montre le rendu visuel d'une analyse restreinte sur cette zone.

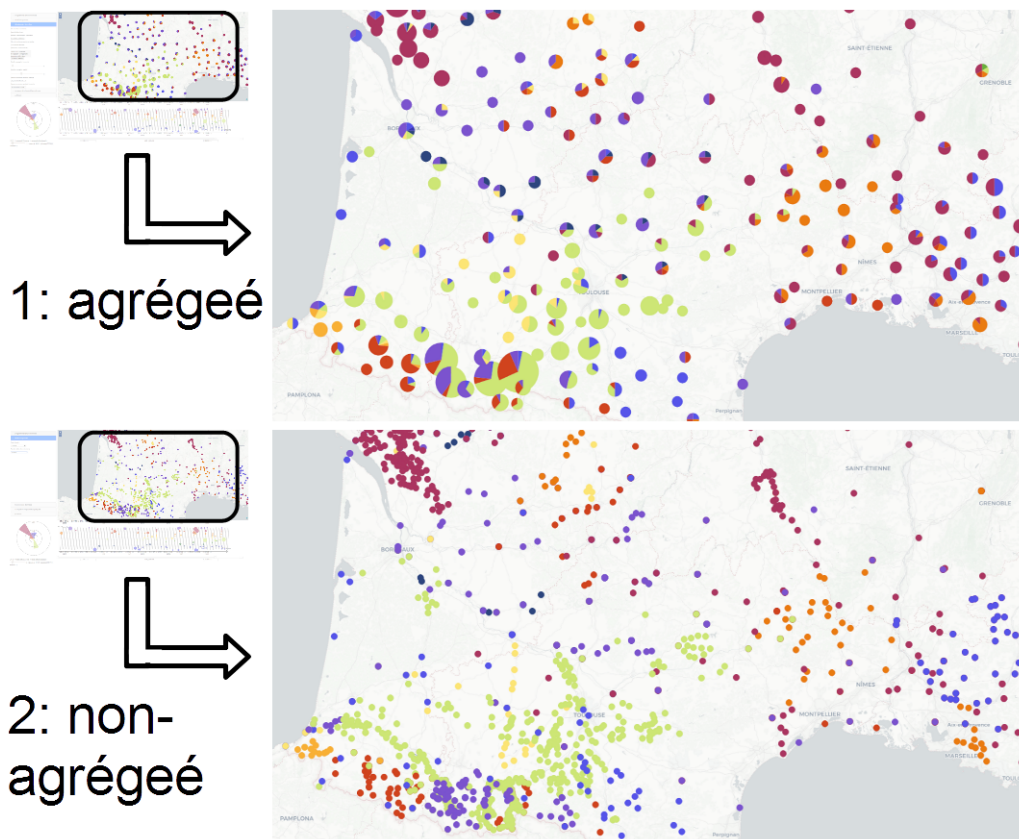


FIGURE 9.16 – Représentation des inondations du sud-ouest de la France selon une échelle cyclique d'un an

On retrouve dans la Figure 9.17 certains clusters mis en évidence dans la Figure 9.15. On observe notamment une concentration d'événements survenant :

- en novembre au niveau des Charentes (Repère 1) ;
- dans les mois de décembre, d'octobre et de septembre dans le Périgord (Repère 4) ;
- en novembre en Auvergne (Repère 8) ;
- en juillet et août au Pays basque (Repère 2) ;
- en septembre dans le Gard (Repère 6).

On repère également un cluster formé dans la zone de Marseille (Repère 10) par des événements apparaissant majoritairement en septembre.

La Figure 9.17 permet également une meilleure identification des clusters au niveau des Pyrénées :

- les événements apparaissant en juin sont situés dans le bassin de la Garonne, de sa source jusqu'à Bordeaux, ainsi qu'au nord des Pyrénées dans les vallées de l'Adour et du Gave de Pau (Repères 5 et 9) ;
- les événements apparaissant en octobre (Repère 3), et en décembre (Repère 7) sont situés dans les montagnes, au sud du bassin de l'Adour.

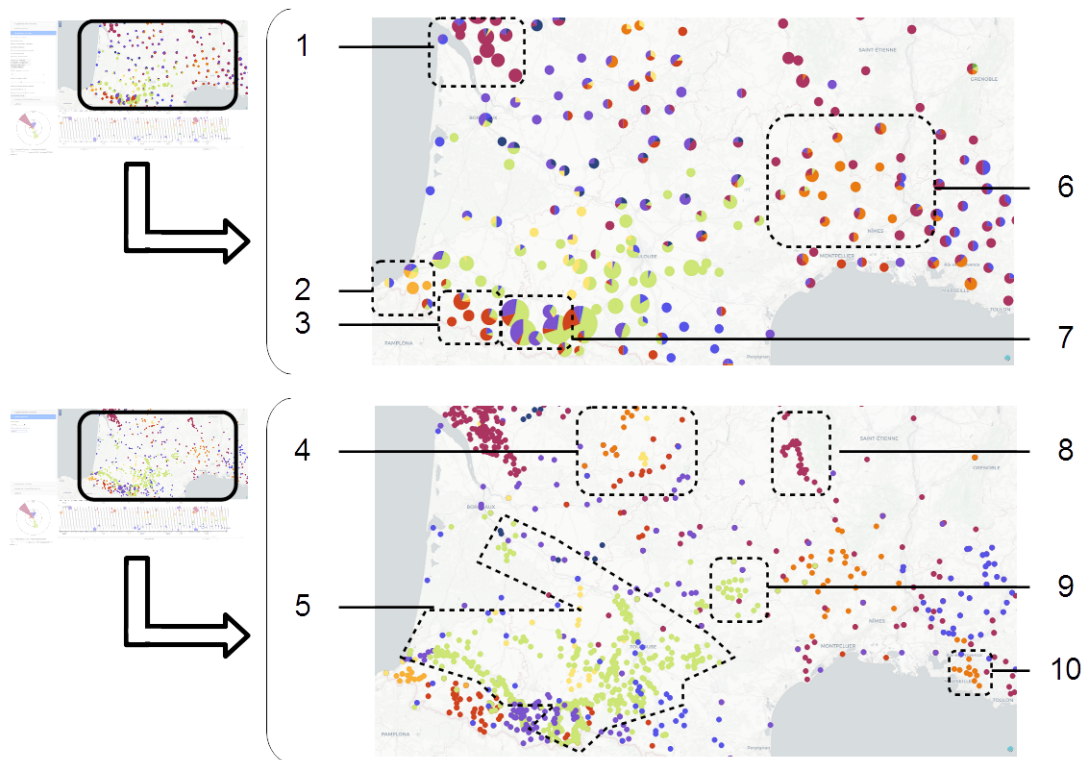


FIGURE 9.17 – Analyse des inondations du sud-ouest de la France selon une échelle cyclique d’un an

Afin de vérifier si les clusters identifiés lors de l’analyse depuis une approche temporelle correspondent à des bassins versants, nous effectuons une seconde analyse selon une entrée spatiale, en classant les événements selon l’appartenance de leurs lieux d’apparition aux zones représentées dans la Figure 9.11. Nous focalisons notre analyse sur les bassins versants de la Charente, de la Dordogne, de la Garonne, de la côte sud-ouest, et de l’Adour, représentés dans la Figure 9.18.

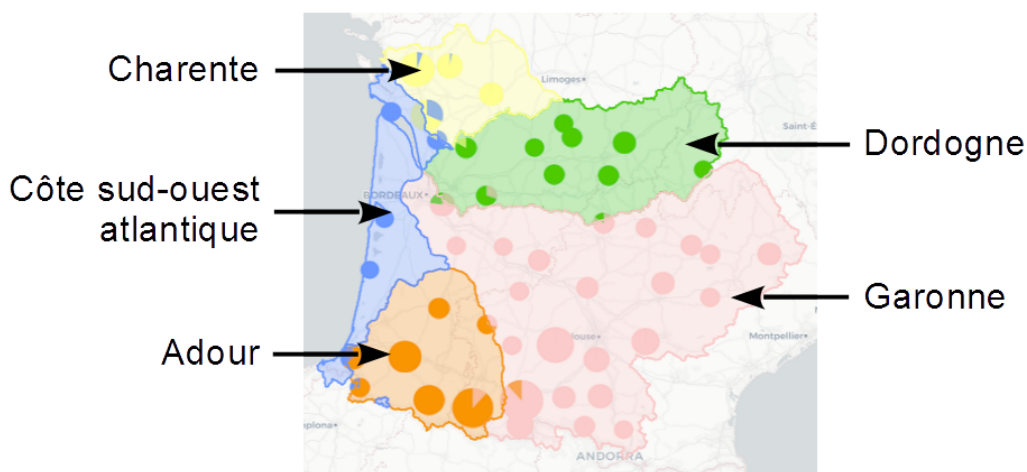


FIGURE 9.18 – Bassins versants du sud-ouest

Pour l’ensemble de la zone d’étude (Repère 1 de la Figure 9.19), les événements apparaissent majoritairement en juin, comme cela avait été identifié dans la Figure 9.15, puis dans une moindre mesure en novembre puis en décembre. Les événements apparaissent majoritairement dans le bassin de la Garonne, puis, dans une moindre mesure, majoritairement dans les bassins versants de la Charente et de l’Adour.

Nous comparons ensuite la distribution temporelle cyclique des événements pour chaque bassin versant (Repères 2 à 6 de la Figure 9.19).

On observe au travers de cette figure que trois zones présentent des distributions temporelles cycliques bien distinctes :

- Les événements apparaissant dans les bassins versants de la Garonne et de l'Adour (Repères 2 et 3 de la Figure 9.19) apparaissent majoritairement en juin. À l'intérieur de ces zones, le bassin de l'Adour présente une plus forte proportion d'événements apparaissant durant l'automne et l'hiver.
- Les événements apparaissant majoritairement en novembre sont situés dans les bassins situés proches de la côte, correspondant aux bassins versants des fleuves côtiers et au bassin versant de la Charente (Repères 4 et 6 de la Figure 9.19).
- Enfin, le bassin versant présentant la plus forte proportion d'événements apparaissant durant l'hiver, malgré de fortes minorités apparaissant en octobre et en novembre, est celui de la Dordogne (Repère 5 de la Figure 9.19).

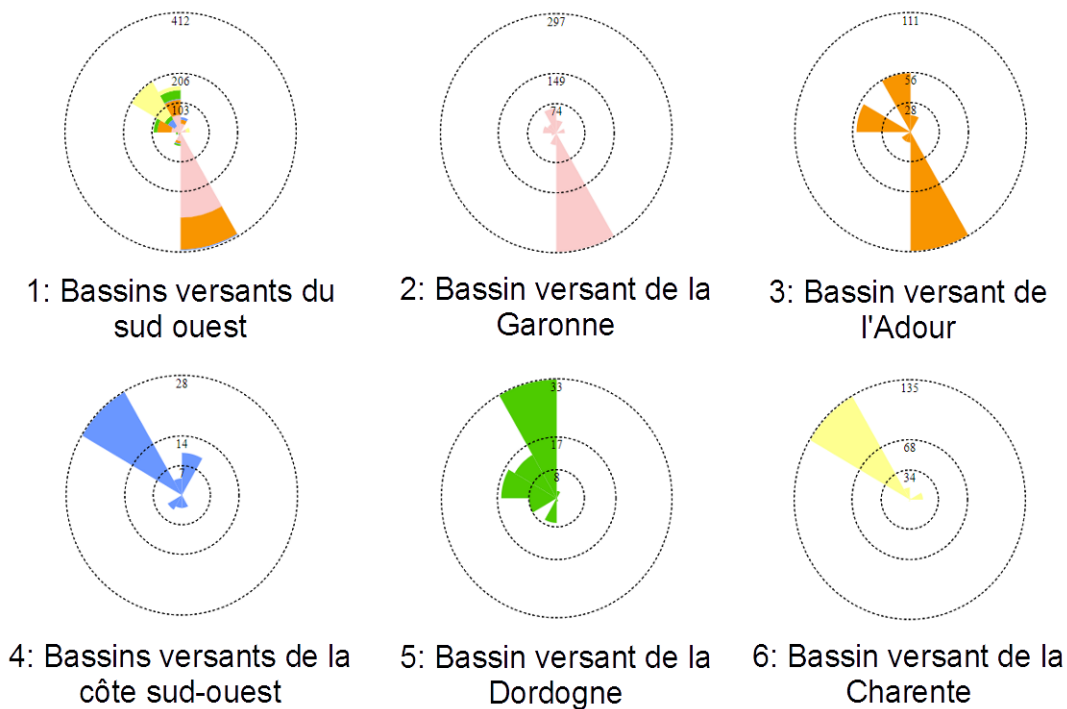


FIGURE 9.19 – Analyse de la distribution temporelle cyclique des événements dans les bassins versants du sud-ouest

On peut donc émettre l'hypothèse que les clusters repérés dans la Figure 9.17 correspondent à ces trois ensembles de bassins versants.

9.5.3 Synthèse

Une première analyse, effectuée selon une entrée temporelle, permet d'identifier que :

- les inondations apparaissent majoritairement durant les mois de novembre, puis de juin, puis dans une moindre mesure durant les mois de décembre et janvier ;
- les inondations, survenant durant les mois de novembre, décembre et janvier, apparaissent dans différentes zones de la carte (nord-est, Charentes, sud-est, et Auvergne pour le mois

de novembre, côtes bretonnes et vendéennes, sud-est, Pyrénées pour les mois de décembre et janvier) ;

- les inondations apparaissant en juin sont localisées dans le sud-ouest ;
- la région sud-ouest présente le plus de diversité dans la distribution temporelle des événements.

Une seconde analyse, effectuée selon une entrée temporelle et centrée sur la zone du sud-ouest, permet d'identifier que les inondations survenant :

- en novembre apparaissent majoritairement dans les Charentes ;
- en juin apparaissent dans le bassin de la Garonne ainsi qu'au nord des Pyrénées, dans les vallées de l'Adour et du Gave de Pau ;
- dans le Pays basque, apparaissent en juillet et en août ;
- dans l'ouest des Pyrénées, du Pays basque aux sources de la Garonne, apparaissent en octobre et en décembre ;
- dans le Périgord, apparaissent dans les mois de décembre, d'octobre et de septembre.

Une troisième analyse, effectuée selon une entrée spatiale et centrée sur les bassins versants du sud-ouest, permet d'identifier que :

- sur l'ensemble des bassins versants, les inondations apparaissent majoritairement en juin, puis en novembre et en décembre ;
- les inondations apparaissent majoritairement dans le bassin versant de la Garonne ;
- les inondations survenant dans les bassins versants de la Garonne et de l'Adour apparaissent majoritairement en juin ;
- les inondations survenant dans les bassins versants de la Charente et des fleuves côtiers de la côte sud-ouest apparaissent majoritairement en novembre ;
- les inondations survenant dans le bassin versant de la Dordogne apparaissent majoritairement en décembre.

On observe ainsi, dans la région sud-ouest de la France métropolitaine, deux grandes zones où les inondations présentent une distribution cyclique distincte :

- dans les bassins versants des fleuves prenant leur source dans les Pyrénées, la Garonne et l'Adour, les inondations apparaissent majoritairement au début de l'été, en juin ;
- dans les bassins versants des fleuves côtiers ou des fleuves prenant leur source dans le Massif central, la Dordogne et la Charente, les inondations apparaissent majoritairement en novembre et en décembre, à la fin de l'automne et au début de l'hiver.

9.6 Analyse des données criminologiques

9.6.1 Scénarios d'analyse

Les analyses s'effectuent sur l'intégralité de la ville de Chicago. Au vu du nombre important de données, les événements sont représentés sur la carte et dans le diagramme temporel mixte sous un format agrégé.

Les analyses s'effectuent dans un premier temps selon une approche temporelle en représentant les événements selon leur moment d'apparition. Le scénario d'analyse suivi pour chaque jeu de données est décrit par la suite.

L'analyse peut être ensuite complétée par une analyse selon une entrée spatiale. Pour cela, l'analyse est effectuée sur les diagrammes temporels au moyen d'une représentation des événements selon leur lieu d'apparition. Pour cela nous classons les événements en utilisant des zones polygonales correspondant au territoire des *Chicago's Sides*, un regroupement administratif utilisé par la ville de Chicago (Figure 9.20).

L'analyse de la composante temporelle des événements est effectuée dans un premier temps pour l'ensemble des zones, puis pour chaque *Side* en particulier.

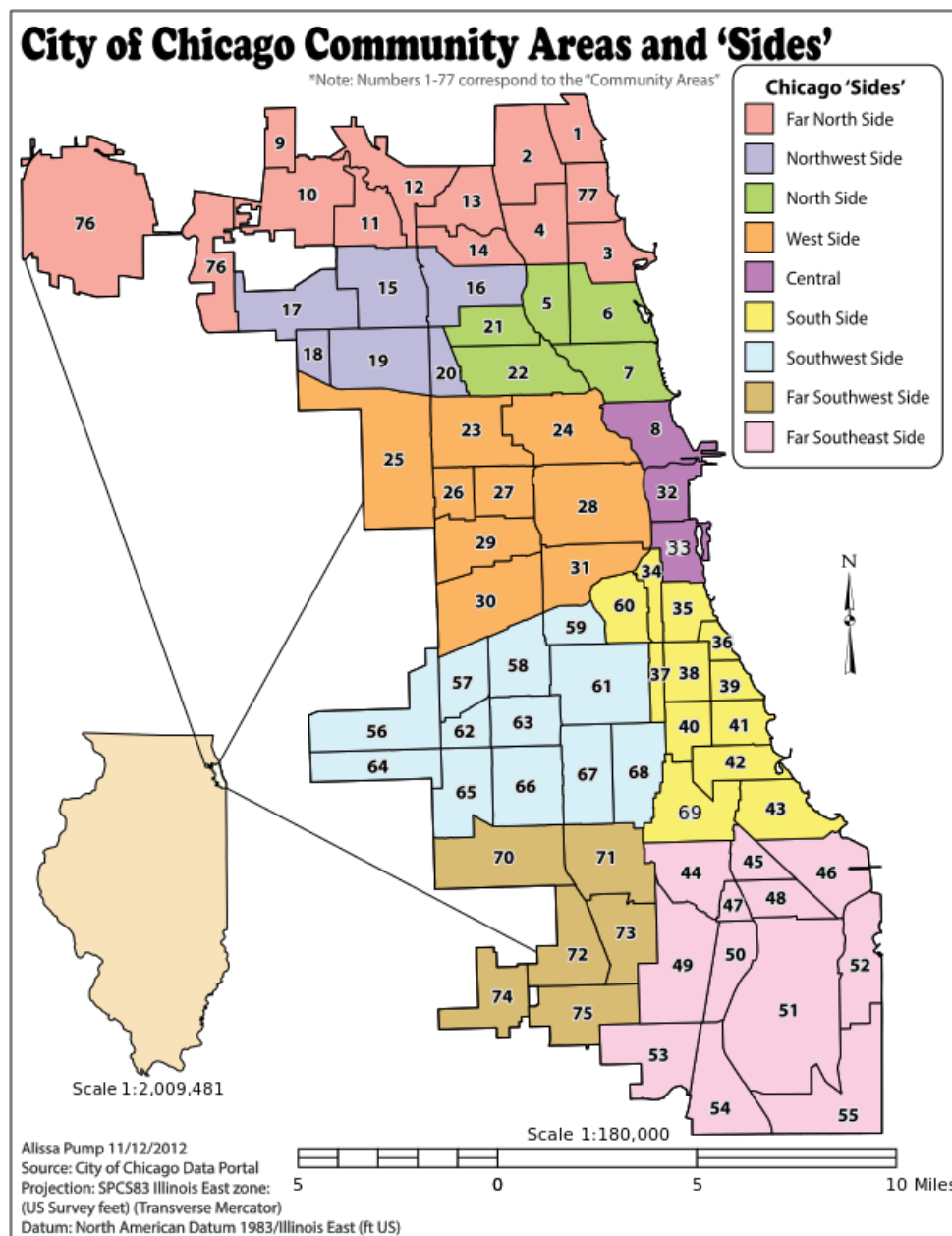


FIGURE 9.20 – Division administrative de la ville de Chicago [City of Chicago, 2017]

9.6.1.1 Cambriolages

Nous émettons l'hypothèse que les événements de cambriolages peuvent être plus fréquents à certains moments de la journée, et que les zones majoritairement touchées peuvent être différentes

selon l'heure de la journée. Les cambriolages peuvent par exemple avoir lieu plus fréquemment la nuit dans les zones d'activité, ou peuvent toucher les zones résidentielles majoritairement durant la journée.

Pour vérifier cette hypothèse, nous effectuons une analyse selon une échelle cyclique d'une journée. L'emprise temporelle de l'étude s'étend du *1er janvier 2017* au *1er juillet 2017*. L'analyse sur les données de cambriolage est effectuée dans un premier temps selon une entrée temporelle.

La granularité temporelle choisie pour la représentation temporelle est l'heure. Nous représentons les événements selon leur moment d'apparition dans une échelle cyclique de 24 heures grâce à six couleurs primaires et en utilisant huit classes temporelles de trois heures chacune.

Le déphasage des diagrammes temporels est fixé de manière à ce que le début des instances de l'échelle cyclique commence à midi. Les huit classes temporelles correspondent au découpage temporel représenté dans la Figure 9.21.

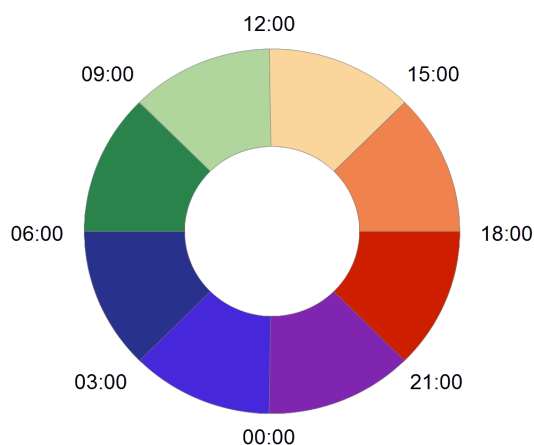


FIGURE 9.21 – Division de l'échelle temporelle de 24h en huit sections avec six couleurs primaires

L'analyse peut ensuite être complétée par une approche spatiale en utilisant le diagramme circulaire, au moyen d'une classification spatiale des événements par *Side* (Figure 9.20).

9.6.1.2 Homicides

L'objectif de cette analyse est d'étudier l'évolution du nombre d'homicides sur une période longue, selon leur lieu d'apparition sur le territoire de la ville de Chicago. L'analyse effectuée sur ce jeu de données est ainsi relative au temps linéaire.

L'emprise temporelle de l'étude s'étend du *1er janvier 2001* au *31 décembre 2015*. L'échelle temporelle cyclique est fixée à 365 jours, et le nombre de sections dans le diagramme temporel cyclique est fixé à 12 sections d'un mois.

L'analyse s'effectue dans un premier temps selon une entrée temporelle. Les événements sont représentés selon leur moment d'apparition dans le temps linéaire, en classant les événements selon trois intervalles temporels linéaires de cinq ans.

L'analyse peut ensuite être complétée par une approche spatiale en utilisant le diagramme temporel mixte, au moyen d'une classification spatiale des événements par *Side* (Figure 9.20).

9.6.1.3 Crimes violents

Objectif du cas d'étude En 2013, l'*Institution for Social and Policy Studies* de l'Université de Yale a produit une étude sur l'évolution du nombre de délits et crimes sur le territoire de la ville de Chicago entre 1965 et 2013 [Papachristos, 2013].

Les résultats présentés par cette étude comportent entre autres deux représentations cartographiques, présentant la distribution spatiale des *crimes violents*, regroupant les vols avec violence (*robbery*) et les homicides, et leur évolution dans le temps selon la *community area* où le crime avait eu lieu. La *community area* étant une subdivision administrative de la ville de Chicago (un *Side* est formé de plusieurs *community areas*).

La première représentation cartographique est une carte des *crimes violents* survenus entre le 1er janvier et le 30 novembre 2011 par *community area* (Repère 1 de la Figure 9.22). La représentation de la distribution spatiale des événements, dans cette carte, diffère de celle proposée par *GrAPHiST*. La carte produite par l'étude de l'Université de Yale représente le nombre de *crimes violents* apparaissant dans chaque *community areas*, rapporté au nombre d'habitants. Notre représentation montre, quant à elle, des données brutes, sans les représenter relativement au nombre d'habitants.

La seconde représentation est une carte de l'évolution du nombre de crimes violents par *community area* entre 2011 et 2013 (Repère 2 de la Figure 9.22). Cette carte a été obtenue en calculant, pour chaque *community area*, la variation entre le nombre de crimes violents survenus entre le 1er janvier et le 30 novembre 2011, et le nombre de crimes violents survenus entre le 1er janvier et le 30 novembre 2013.

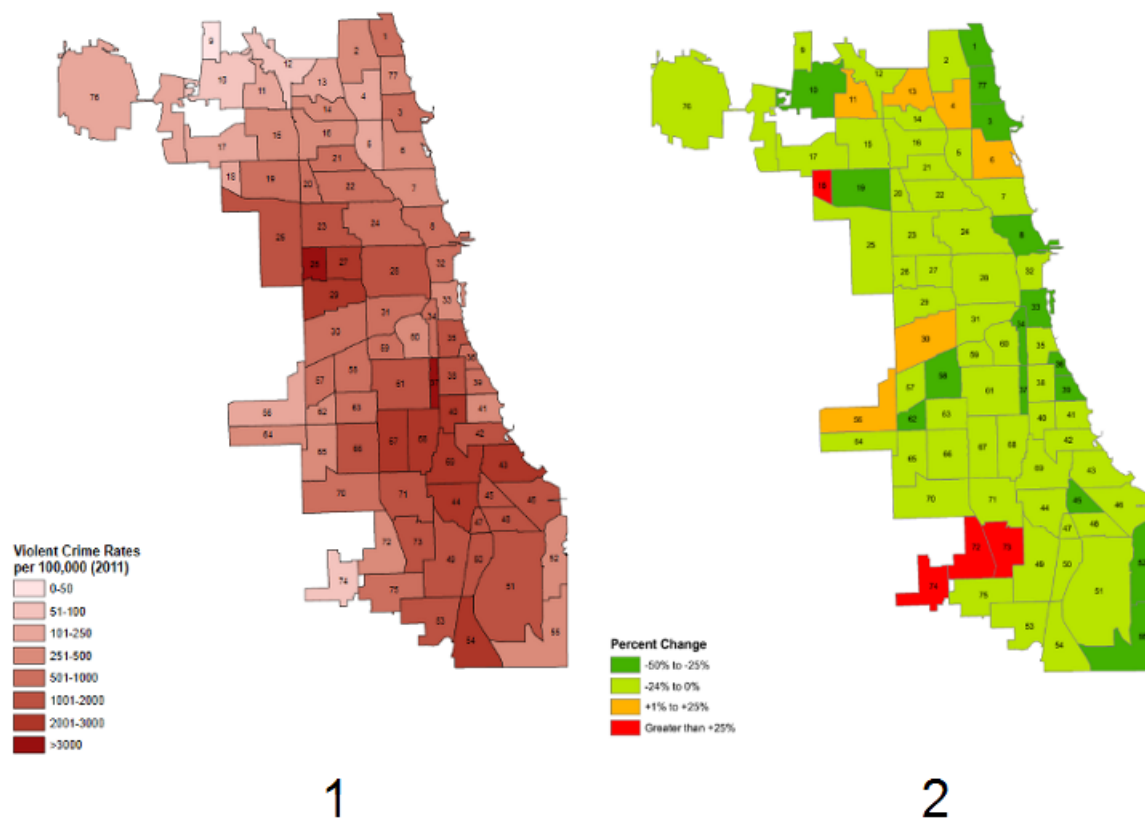


FIGURE 9.22 – Cartes de la répartition des crimes violents à Chicago en 2011 (1) et de leur évolution entre 2011 et 2013 (2) [Papachristos, 2013]

Pour ce cas d'étude, nous effectuons une analyse de la distribution spatio-temporelle des événements de *crimes violents*, sur le territoire de la ville de Chicago, entre 2011 et 2013.

Du fait de la perte de données lors de l'import d'événements ne possédant pas de coordonnées valides depuis le *Chicago Data Portal*, nous ne pouvons pas comparer nos résultats avec ceux produits par l'étude de l'Université de *Yale*. En revanche, nous pouvons comparer la représentation cartographique que nous produisons lors de notre analyse avec les représentations cartographiques produites par l'étude de l'Université de *Yale*.

Notre analyse suit alors un double objectif :

- D'une part, le but de l'analyse est de visualiser la distribution spatiale des *crimes violents* survenus entre le *1er janvier* et le *30 novembre* 2011, et de comparer la représentation cartographique obtenue avec la carte produite par l'étude de l'Université de *Yale* (Repère 1 de la Figure 9.22).
- D'autre part, le but de l'analyse est de visualiser l'évolution, entre 2011 et 2013, du nombre de *crimes violents* selon le lieu d'apparition de ces crimes, et de comparer la représentation cartographique obtenue avec la carte produite par l'étude de l'Université de *Yale* (Repère 2 de la Figure 9.22).

Procédés mis en place L'emprise temporelle est fixée du *1er janvier 2011* au *31 décembre 2013*, et l'emprise spatiale choisie couvre l'entièreté du territoire de la ville de Chicago. Les événements sont représentés de manière agrégée sur la carte et dans le diagramme temporel mixte.

Afin de pouvoir identifier l'évolution temporelle de la distribution spatiale des événements sur la carte, nous effectuons une analyse de la distribution spatio-temporelle des événements, relative au temps linéaire, selon une approche temporelle.

Les événements sont représentés selon leur moment d'apparition relatif au temps linéaire, en les classant selon deux intervalles temporels d'un an et demi. Une valeur forte est associée aux événements les plus anciens, et une valeur faible est associée aux événements les plus récents. Les événements apparaissant durant l'année 2011 sont ainsi représentés par une couleur sombre, et les événements apparaissant durant l'année 2013 sont représentés par une couleur claire.

Les événements survenant du *1er décembre 2011* au *31 janvier 2012*, et les événements apparaissant du *1er décembre 2013* au *31 décembre 2013* sont sélectionnés puis masqués. Une fois ces événements masqués, la couleur sombre représente les événements apparaissant entre le *1er janvier* et le *30 novembre* 2011, et la couleur claire représente les événements apparaissant entre le *1er janvier* et le *30 novembre* 2013.

Les événements étant représentés de manière agrégée sur la carte, la taille des agrégats permet d'identifier les zones spatiales présentant les plus fortes ou les plus faibles concentrations de *crimes violents*, comme dans la carte présentée dans la Figure 9.22 (Repère 1). Les événements de l'année 2013 peuvent être masqués afin de visualiser la distribution spatiale des événements pour la seule année 2011.

La couleur majoritaire d'un agrégat dans la carte permet alors de définir si, dans la zone spatiale correspondant à l'étendue de l'agrégat, les *crimes violents* sont en augmentation, en diminution, ou en stagnation, de la même manière que la carte présentée dans la Figure 9.22 (Repère 2) :

- si la couleur majoritaire d'un agrégat est la couleur sombre, le nombre de *crimes violents* dans la zone spatiale correspondante diminue entre 2011 et 2013, car la proportion

d'événements anciens est plus élevée.

- si la couleur majoritaire d'un agrégat est la couleur claire, le nombre de *crimes violents* dans la zone spatiale correspondante augmente entre 2011 et 2013, car la proportion d'événements récents est plus élevée.
- si les deux couleurs possèdent la même proportion dans un agrégat, le nombre de *crimes violents* dans la zone spatiale correspondante n'évolue pas entre 2011 et 2013, car le nombre d'événements récents est égal au nombre d'événements anciens.

La proportion des deux couleurs dans un agrégat permet également de quantifier la variation du nombre de *crimes violents* entre 2011 et 2013. Une occupation des trois quarts du *pie chart* par la couleur sombre correspond à une diminution de 50% du nombre de *crimes violents* dans la zone correspondante, car les événements récents sont deux fois moins nombreux que les événements anciens.

9.6.2 Résultats

9.6.2.1 Cambriolages

La Figure 9.23 montre une capture d'écran de l'interface de géovisualisation une fois saisis les paramètres décrits précédemment.

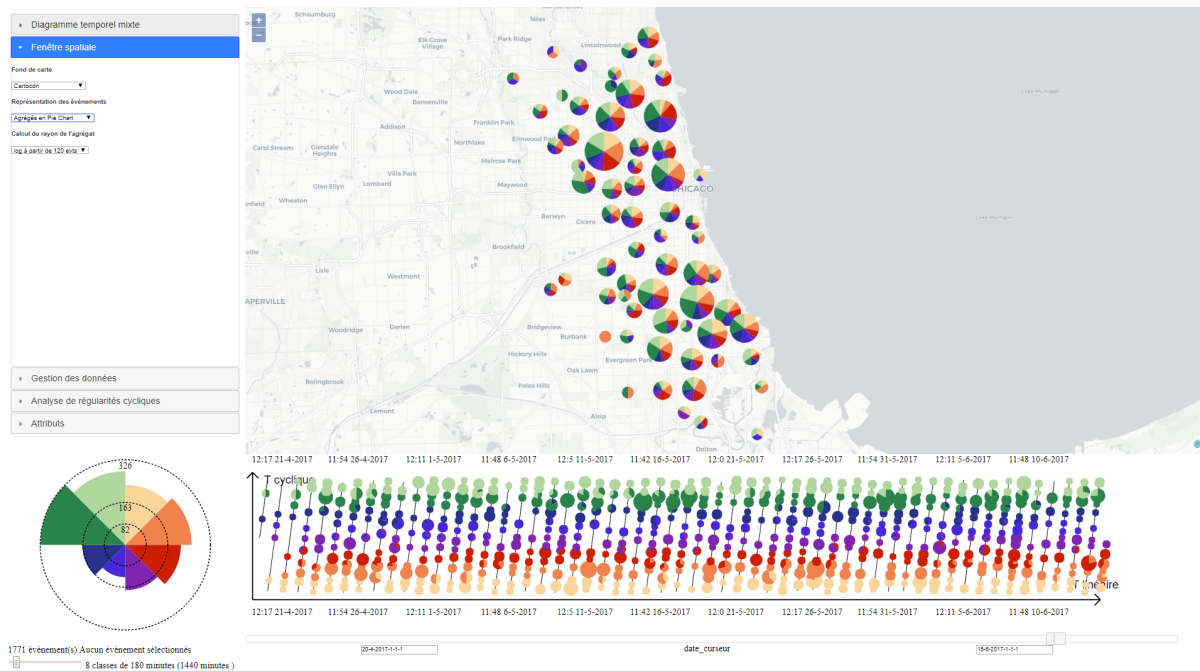


FIGURE 9.23 – Visualisation des événements de cambriolages du 1er janvier 2017 au 1er juillet 2017

La distribution temporelle des événements dans le diagramme temporel cyclique montre que les cambriolages apparaissent le plus souvent en journée, les plages horaires allant de 21h à 6h étant celles présentant le plus petit nombre d'apparitions d'événements (Figure 9.24). Les plages horaires du matin, entre 6h et 12h, sont celles présentant le plus grand nombre d'événements.

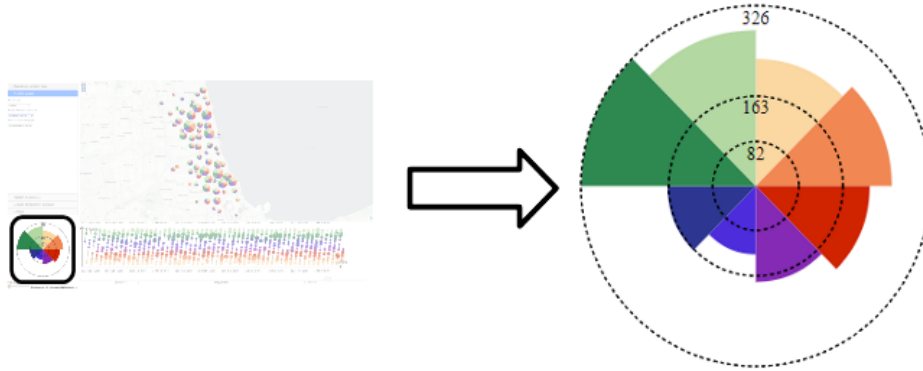


FIGURE 9.24 – Distribution temporelle des événements de cambriolages

On observe sur la carte certaines zones où le nombre de cambriolages survenus entre 21h et 6h est plus important que dans l'ensemble de la carte (Repères 1, 2, 3 et 6 de la Figure 9.25).

À l'inverse, certaines zones présentent une très faible proportion de cambriolages survenant la nuit, et une proportion d'événements survenant entre 6h et 12h plus forte que dans le reste de la carte. Le matin dans ces plages horaires est plus faible que la moyenne (Repères 4 et 5 de la Figure 9.25).

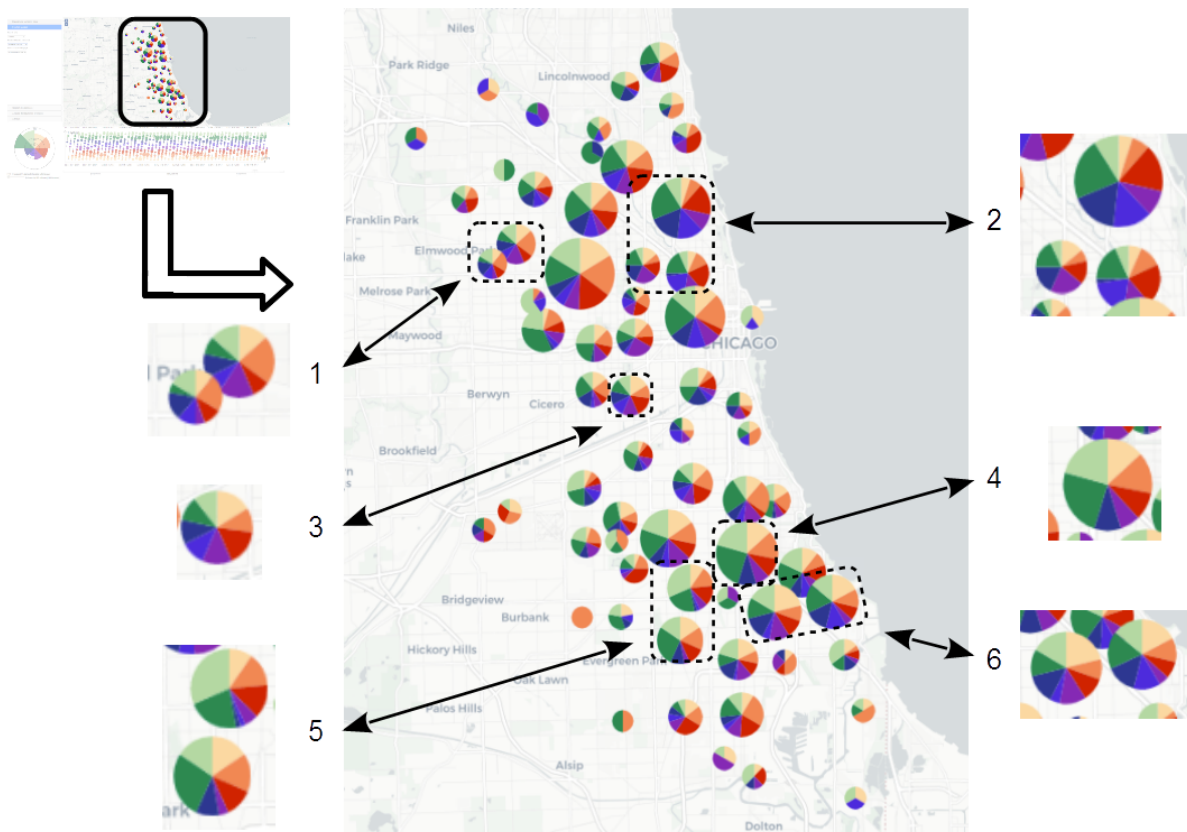


FIGURE 9.25 – Distribution spatio-temporelle des événements de cambriolages

Nous effectuons une sélection des événements apparaissant dans les zones correspondant aux Repères 2 et 5 de la Figure 9.25), et nous comparons leur distribution temporelle relative au temps cyclique avec la distribution temporelle de l'ensemble des événements de la zone d'étude. La zone correspondant au Repère 2 est située dans le *North Side*, tandis que la zone correspon-

nant au Repère 5 est située à la frontière entre le *Far Southwest Side* et le *Southeast Side* (Figure 9.20).

La distribution temporelle des événements de la zone correspondant au Repère 2 présente les caractéristiques suivantes (Figure 9.26) :

- Les événements apparaissent majoritairement entre $18h$ et $9h$, soit en début de journée, en début de soirée et la nuit.
- Les pics d'intensité du phénomène ont lieu en début de journée, et en début de soirée, dans les plages horaires $[6h-9h]$ et $[18h-21h]$. La proportion d'événements apparaissant en début de soirée est supérieure à la moyenne de l'ensemble de la zone d'étude.
- Un nombre d'événements, équivalent à la somme des événements survenus durant ces deux pics d'activité, apparaît en pleine nuit entre $21h$ et $6h$. La proportion d'événements apparaissant durant ces plages horaires est supérieure à la moyenne de l'ensemble de la zone d'étude.
- Les événements ayant lieu en pleine journée, entre $9h$ et $18h$ sont très minoritaires. La proportion d'événements apparaissant en pleine journée est très inférieure à la moyenne de l'ensemble de la zone d'étude.

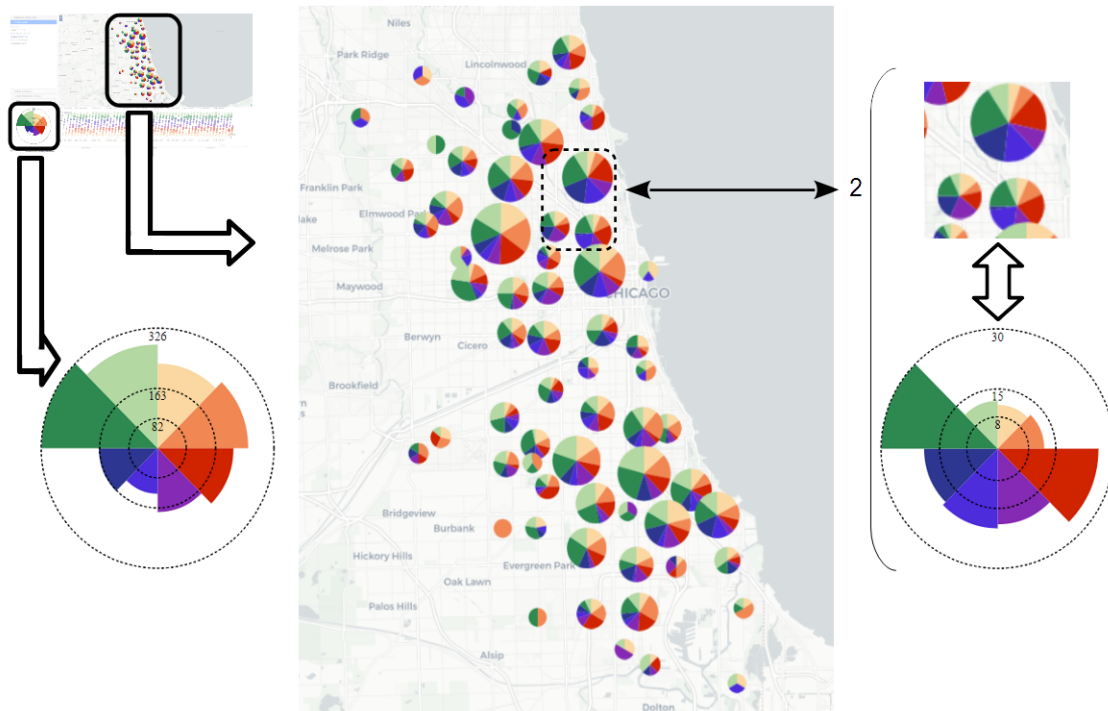


FIGURE 9.26 – Distribution temporelle des événements de la zone correspondant au Repère 2

La distribution temporelle des événements de la zone correspondant au Repère 5 présente les caractéristiques suivantes (Figure 9.27) :

- Les événements apparaissent majoritairement entre $6h$ et $12h$, à l'instar de la moyenne de l'ensemble de la zone d'étude. Cependant la proportion d'événements apparaissant dans ces plages horaires est supérieure à la moyenne.
- La proportion d'événements apparaissant de $12h$ à $21h$ est sensiblement la même que pour l'ensemble de la zone d'étude.
- Les événements ayant lieu la nuit, entre $18h$ et $6h$ sont très minoritaires. La proportion d'événements apparaissant en pleine journée est très inférieure à la moyenne de l'ensemble de la zone d'étude.

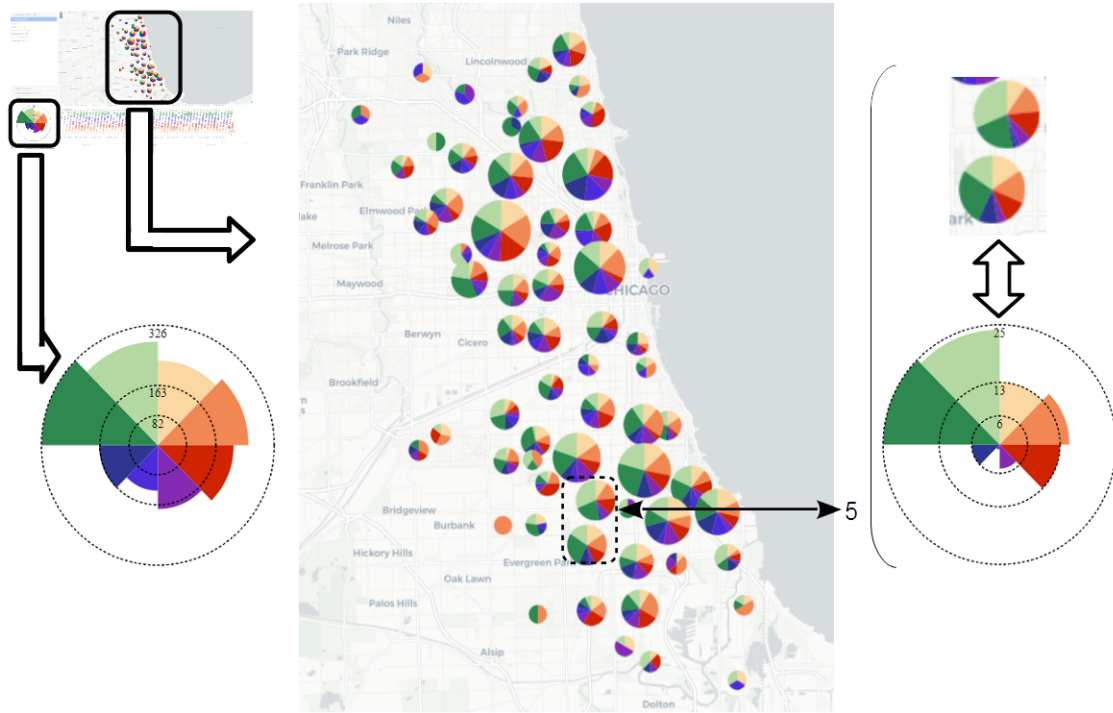


FIGURE 9.27 – Distribution temporelle des événements de la zone correspondant au Repère 5

Pour compléter cette analyse, afin d'identifier s'il existe une différence significative dans la distribution temporelle des cambriolages selon les différents quartiers de la ville, nous effectuons une seconde analyse selon une approche spatiale (Figure 9.28). Nous classons pour cela les événements en utilisant le découpage administratif présenté dans la Figure 9.20.

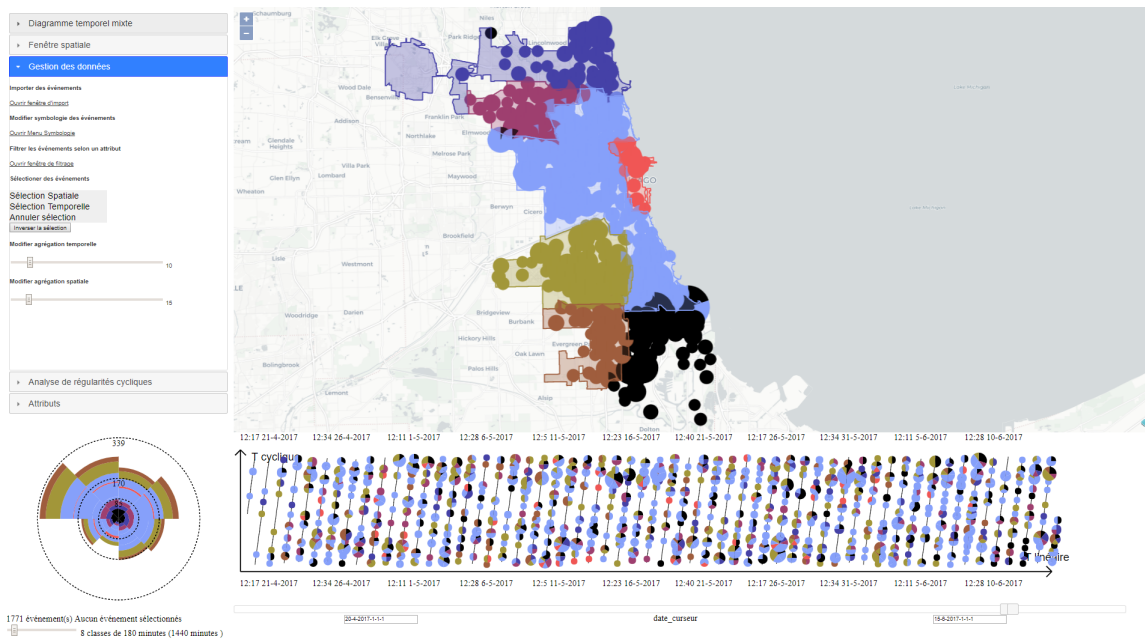


FIGURE 9.28 – Représentation des événements de cambriolages par *Side*

Nous tentons notamment d'analyser si la distribution temporelle des événements du *North Side* (Repère 2 de la Figure 9.25) correspond aux résultats que nous avons trouvés lors de notre première analyse, et si la distribution temporelle de ce quartier diffère de l'ensemble de la ville.

Nous tentons également d’analyser si une différence significative existe entre la distribution temporelle des événements apparaissant dans le *Center*, comprenant des quartiers d’affaires comme le *Loop* [Wikipedia, 2018b], et la distribution temporelle des événements apparaissant dans les quartiers entourant le centre-ville (*North Side*, *West Side*, *South Side*), comprenant d’importants quartiers résidentiels [Wikipedia, 2018b].

La Figure 9.28 présente une capture d’écran de notre outil une fois les événements représentés selon leurs lieux d’apparition. Les événements survenant dans le *West Side*, le *South Side*, et le *North Side* se sont vus attribuer la même couleur, dans le but de comparer la distribution temporelle de l’ensemble des événements de ces trois quartiers, avec la distribution temporelle des événements survenant dans le centre-ville.

Dans la Figure 9.29, nous comparons la distribution temporelle cyclique des événements zone par zone. Les quartiers du *West Side*, du *South Side*, et du *North Side* ont également été traités comme un seul ensemble.

On retrouve dans le diagramme représentant la distribution temporelle des événements du *North Side* (Repère 1 de la Figure 9.29) une distribution temporelle similaire aux résultats que nous avons trouvés lors de notre première analyse :

- Les événements apparaissent majoritairement entre 18h et 9h, soit en début de journée, en début de soirée, et la nuit.
- Le pic d’intensité du phénomène a lieu en début de journée, entre 6h et 9h. La proportion d’événements apparaissant en début de soirée entre 18h et 21h est inférieure à ce que nous avons identifié lors de notre première analyse.
- Un nombre d’événements, équivalent au nombre d’événements apparaissant entre 6h et 9h, apparait en pleine nuit entre 21h et 6h.
- Les événements ayant lieu en pleine journée, entre 9h et 18h sont minoritaires.



FIGURE 9.29 – Comparaison des distributions temporelles cycliques des cambriolages par zones

Ce diagramme montre également que le *North Side* présente une plus grande proportion d’événements nocturnes, et une plus faible proportion d’événements apparaissant en pleine journée, que les autres *Sides* de Chicago (en dehors du centre-ville). Dans la plupart des autres zones spatiales (Figure 9.29), la majorité des événements a lieu la journée, avec :

- un pic d'intensité entre 6h et 9h pour le *West Side*, le *South Side*, le *Northwest Side* et le *Southwest Side* (Repères 2, 3, 7 et 9 de la Figure 9.29) ;
- deux pics d'intensité entre 6h et 9h et entre 15h et 28h pour le *Far Southwest Side* (Repère 6 de la Figure 9.29) ;
- deux pics d'intensité entre 6h et 9h et entre 18h et 21h pour le *Far North Side* (Repère 8 de la Figure 9.29) ;
- un pic d'intensité entre 9h et 15h pour le *Southeast Side* (Repère 10 de la Figure 9.29).

À l'inverse du *North Side*, la majorité des *Sides* en dehors du *Center* présente ainsi une distribution temporelle des cambriolages proche de celle de l'ensemble de la ville de Chicago (Repère 11 de la Figure 9.29), dans le sens où la proportion d'événements survenant la journée est plus forte que la proportion des événements nocturnes.

À l'inverse de l'ensemble [*North Side ; West Side ; South Side*] (Repère 4 de la Figure 9.29) qui présente une forte proportion d'événements étalée entre 9h et 18h, avec un pic d'intensité entre 6h et 9h, la distribution temporelle des événements du centre-ville (Repère 5 de la Figure 9.29) montre une forte proportion d'événements arrivant, soit en pleine nuit entre 0h et 3h, soit au milieu de la journée entre 12h et 15h.

9.6.2.2 Homicides

La Figure 9.30 montre une capture d'écran de l'interface de géovisualisation une fois saisis les paramètres décrits précédemment.

On observe, pour la plupart des zones de la ville, une légère diminution du nombre d'homicides durant chaque quinquennat : la proportion de la partie sombre est la plus importante pour la plupart des *pie chart*, et la proportion de la partie claire est la plus claire.

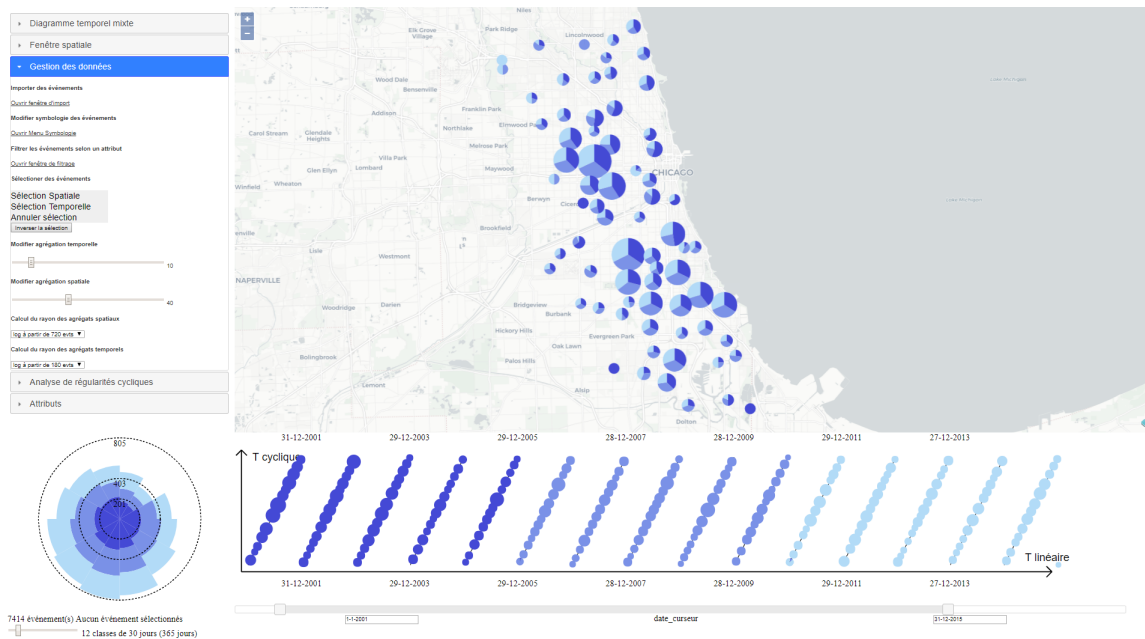


FIGURE 9.30 – Visualisation des données relatives aux homicides du 1er janvier 2001 au 31 décembre 2015

Cette diminution reste cependant faible dans certaines zones, comme dans le *West Side* (Repère 1 de la Figure 9.31), les proportions des événements de chaque intervalle temporel étant très proches.

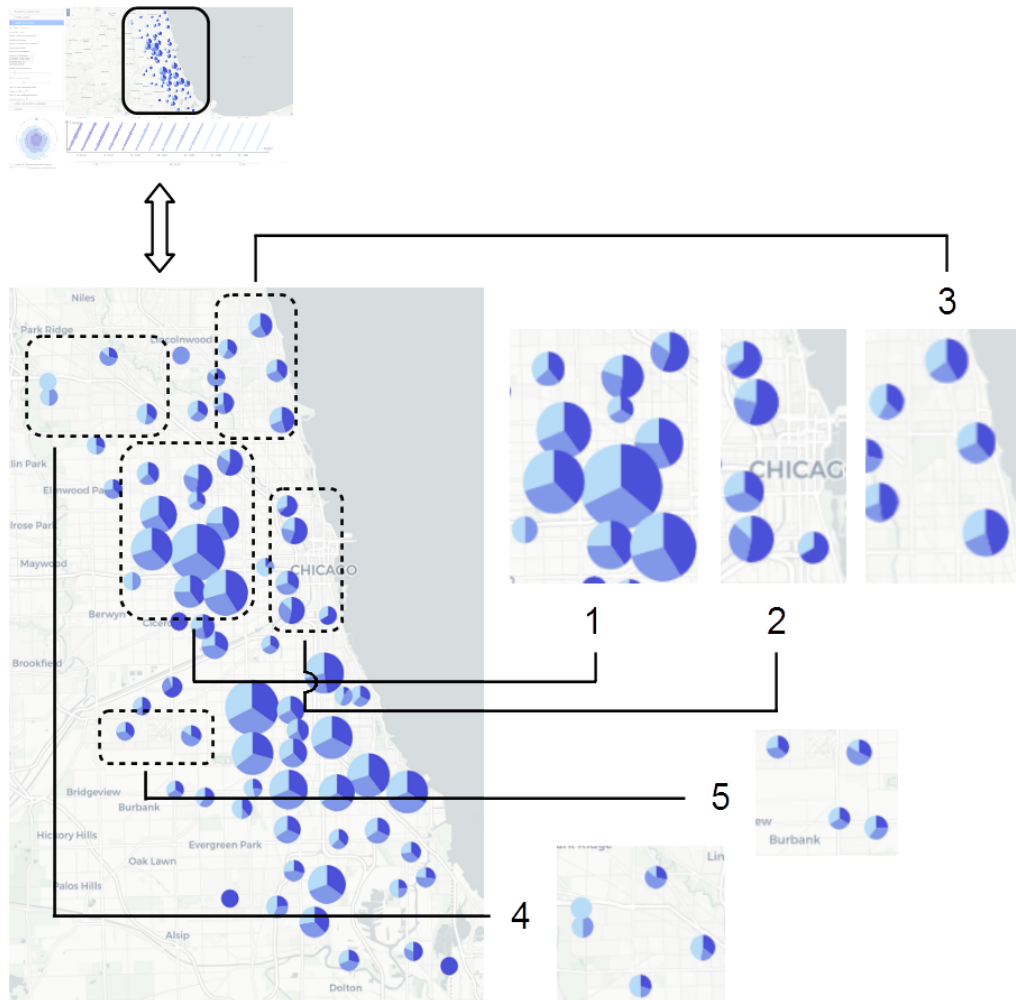


FIGURE 9.31 – Analyse de l'évolution des homicides du 1er janvier 2001 au 31 décembre 2015

On observe certaines zones où la diminution du nombre d'homicides entre 2001 et 2015 est plus forte que dans les autres secteurs de la ville. C'est le cas notamment de la zone située au niveau du centre-ville (Repère 2 de la Figure 9.31), et de la zone du *North Side* (Repère 3 de la Figure 9.31), où la proportion de la couleur sombre est très importante dans les entités graphiques correspondantes, tandis que la proportion de la couleur claire est très faible.

On observe également certaines zones, où le nombre d'apparitions d'événements reste faible, mais où les homicides sont en augmentation (Repère 4 de la Figure 9.31). D'autres zones présentent un nombre plus important d'événements entre 2005 et 2010 (Repère 5 de la Figure 9.31).

Une visualisation de la dimension temporelle des événements apparaissant dans le centre-ville de Chicago (Figure 9.32), représentant 370 des 7414 événements de la zone étudiée, montre également une diminution du nombre d'homicides entre 2001 et 2005. Le diagramme temporel mixte montre une diminution du nombre d'entités graphiques sur le diagramme à mesure que l'on progresse le long de l'axe des abscisses (Repère * de la Figure 9.32).

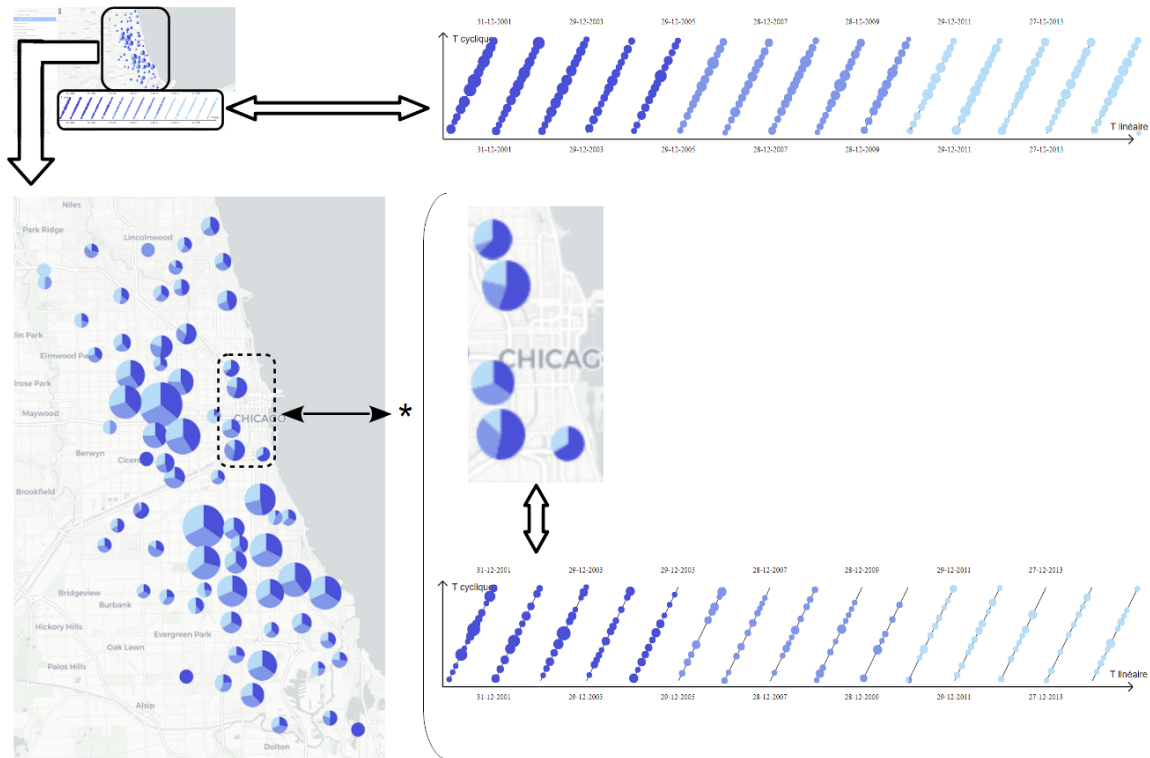


FIGURE 9.32 – Distribution temporelle linéaire des événements du centre-ville de Chicago

Pour compléter cette analyse, nous représentons les événements en fonction de leur lieu d'apparition en utilisant le découpage administratif des *Sides*, présenté dans la Figure 9.20, pour classer les événements. Nous comparons ensuite la distribution temporelle linéaire des événements apparaissant dans chaque *Side* (Figure 9.33).

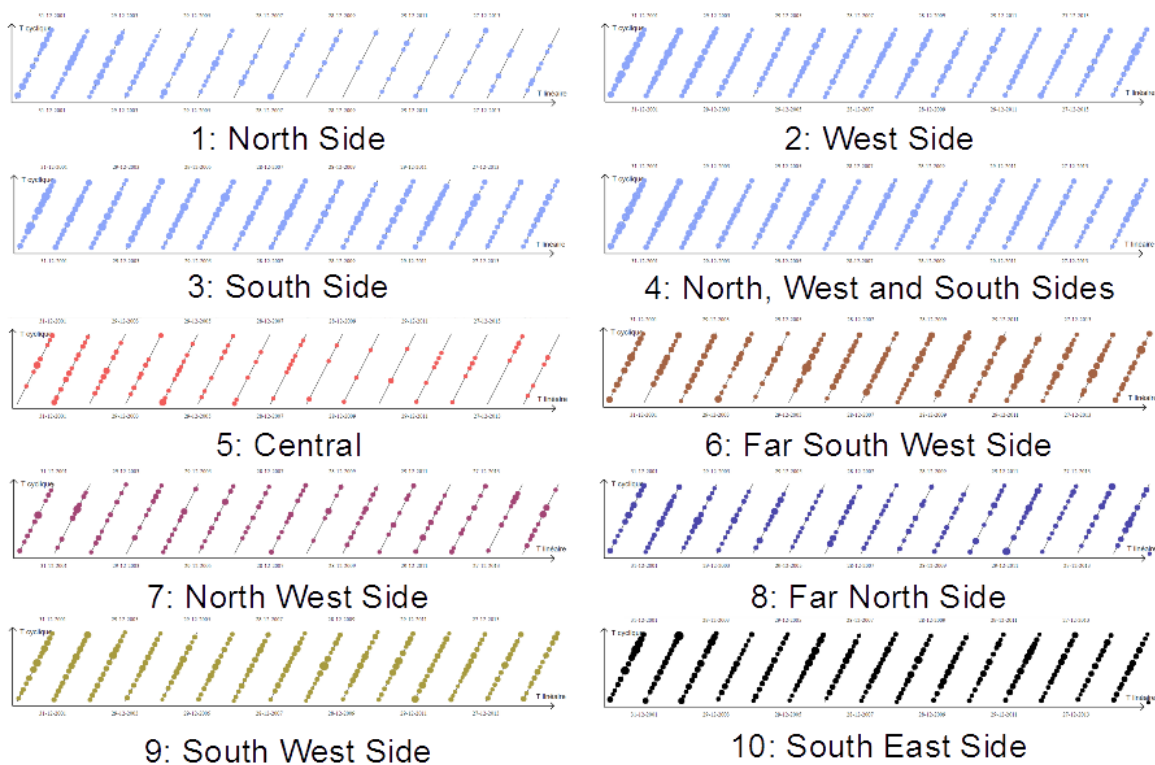


FIGURE 9.33 – Comparaison des distributions temporelles linéaires des homicides par zones

On retrouve dans ces diagrammes une nette diminution du nombre d'homicides dans le *North Side* (Repère 1 de la Figure 9.33) et dans le centre-ville (Repère 5 de la Figure 9.33), représenté par les fortes de couleur sombre dans les agrégats présents dans ces zones dans la Figure 9.30.

Les diagrammes représentant les événements des autres *Sides* ne permettent pas d'identifier des tendances significatives dans la distribution temporelle des homicides.

9.6.2.3 Crimes violents

La Figure 9.34 montre la représentation cartographique obtenue avec notre interface de géovisualisation une fois saisis les paramètres décrits précédemment.

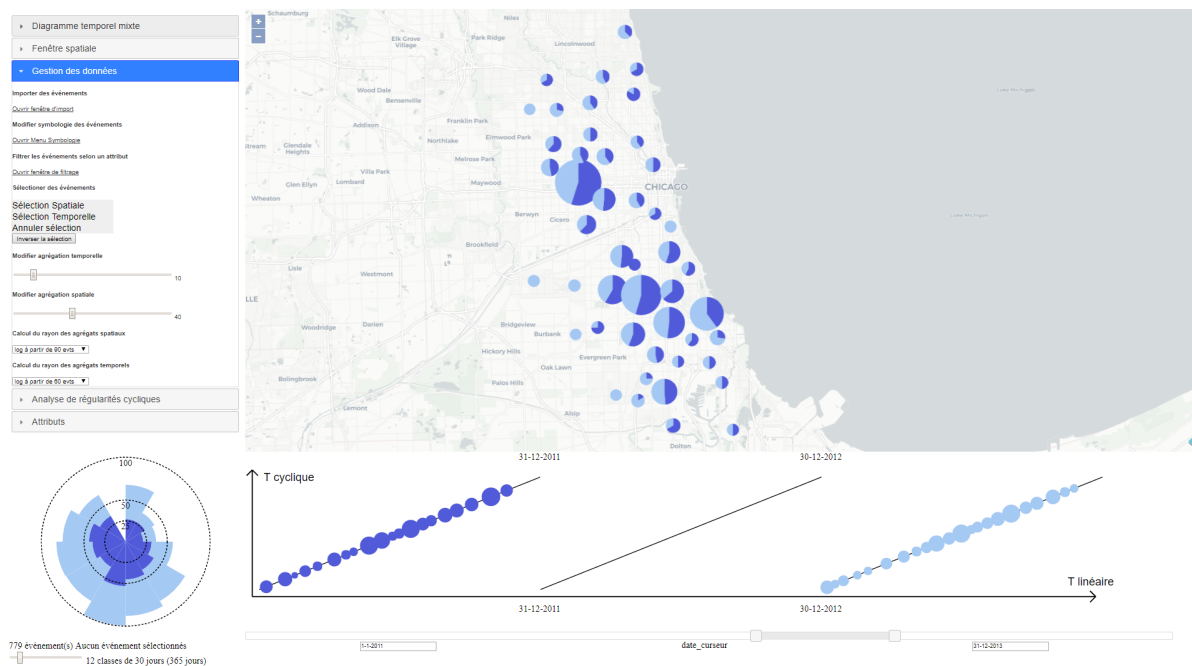


FIGURE 9.34 – Évolution 2011-2013 des crimes violents à Chicago

L'utilisation d'une représentation agrégée des événements permet d'identifier les zones de forte et de faible concentration d'événements grâce à la taille des entités graphiques représentées. La carte permet d'identifier la distribution spatiale du nombre d'événements sur les deux années, mais également sur la seule année 2011 ou la seule année 2013 (les événements d'une année peuvent être masquée dans la carte afin d'améliorer l'analyse).

On retrouve sur la Figure 9.34 de fortes concentrations d'événements dans les quartiers du *West Side* (Repère 1 de la Figure 9.35) et du *South Side* (Repère 2 de la Figure 9.35), conformément à ce qui est représenté dans la carte produite lors de l'étude de l'Université de Yale.

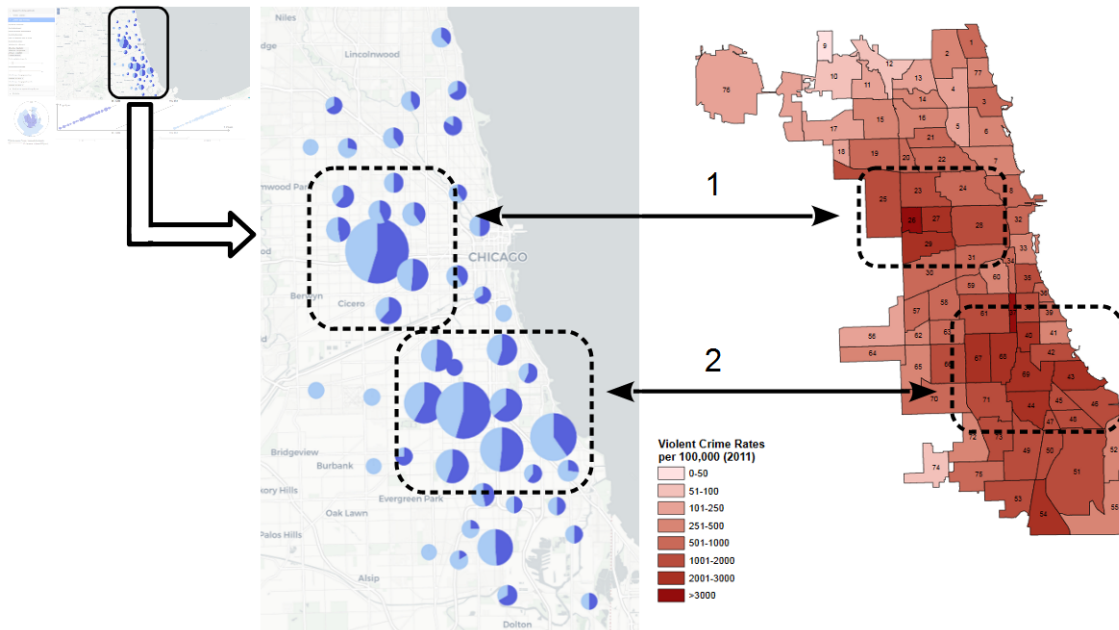


FIGURE 9.35 – Représentation de la distribution spatiale des crimes violents [Papachristos, 2013]

La représentation des événements selon deux valeurs, représentant les événements de 2011 pour la valeur sombre, et les événements de 2013 pour la valeur claire, permet d’identifier, pour chaque zone, si les événements sont en augmentation, en diminution ou en stagnation.

Les *Pie Chart* représentant les événements apparaissant dans le *West Side* (Repère 1 de la Figure 9.36) et le *South Side* (Repère 2 de la Figure 9.36) présentent une plus grande proportion de couleur sombre, ce qui correspond à une diminution du nombre de *crimes violents* entre 2011 et 2013, conformément à ce qui est représenté dans la carte produite lors de l’étude de l’Université de Yale.

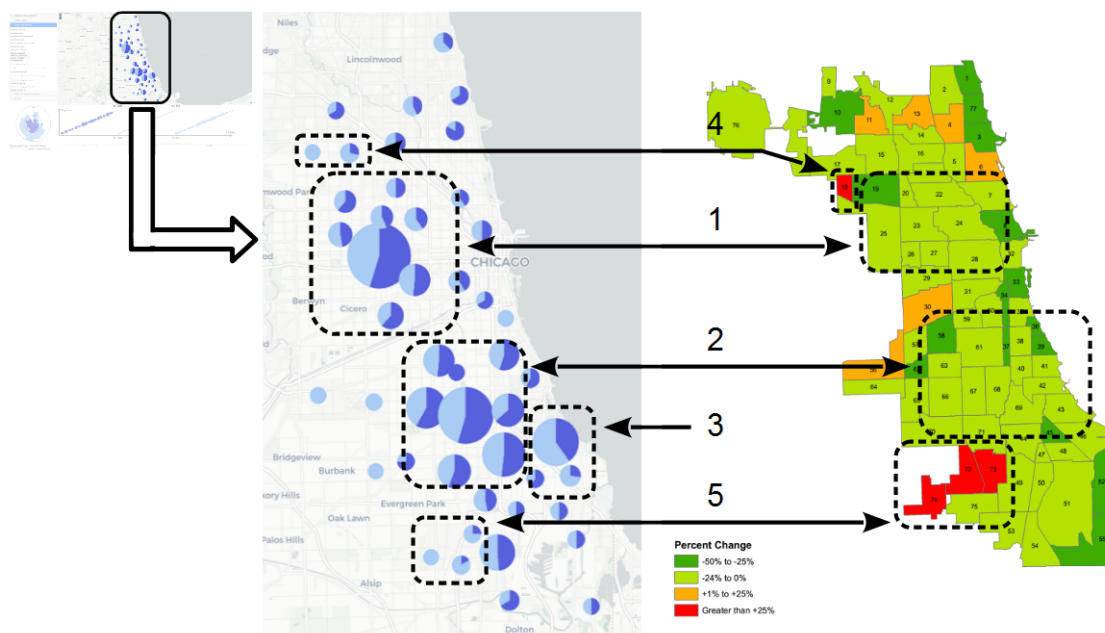


FIGURE 9.36 – Représentation de l’évolution des crimes violents selon leur position spatiale [Papachristos, 2013]

Une zone située à l'est du *South Side* (Repère 3 de la Figure 9.36) possède une proportion de couleur claire plus importante, ce qui correspond à une augmentation du nombre de *crimes violents*. Cette tendance ne se retrouve pas dans la carte produite par l'étude de l'Université de *Yale*.

Les deux zones présentant une forte augmentation du nombre de *crimes violents* dans la représentation de l'Université de *Yale* peuvent, quant à elle, être retrouvée dans notre représentation cartographique (Repères 4 et 5 de la Figure 9.36).

La carte produite par *Yale* présentant l'évolution du nombre de *crimes violents* par *Community area* (Repère 2 de la Figure 9.22), permet une visualisation des tendances dans l'évolution du nombre d'événements selon l'espace. Une telle carte permet d'identifier de fortes variations locales du nombre d'événements, qui peuvent être plus difficilement décelables au moyen d'un dispositif *small maps*, qui présenterait deux cartes de la distribution spatiale des événements, pour 2011 et pour 2013.

Il est ainsi possible d'identifier au moyen de cette carte la forte augmentation du nombre de crimes violents dans les *community areas* 18 (Repère 1 de la Figure 9.37), 72, 73 et 74 (Repère 2 de la Figure 9.37).

Cependant, cette forte augmentation du nombre d'événements concerne des zones qui, d'après la carte de la distribution spatiale des *crimes violents* en 2011 (Repère 2 de la Figure 9.22), ne présentent en 2011 qu'un faible nombre d'événements (Repères 3 et 4 de la Figure 9.37).

La carte de l'évolution du nombre de *crimes violents* doit donc être lue de manière conjointe avec la carte de la distribution spatiale des *crimes violents*, de manière à identifier l'évolution du nombre d'événements dans les *community areas* 18, 72, 73 et 74 comme significatives à une échelle locale, mais peu significative à l'échelle de l'ensemble de la zone d'étude.

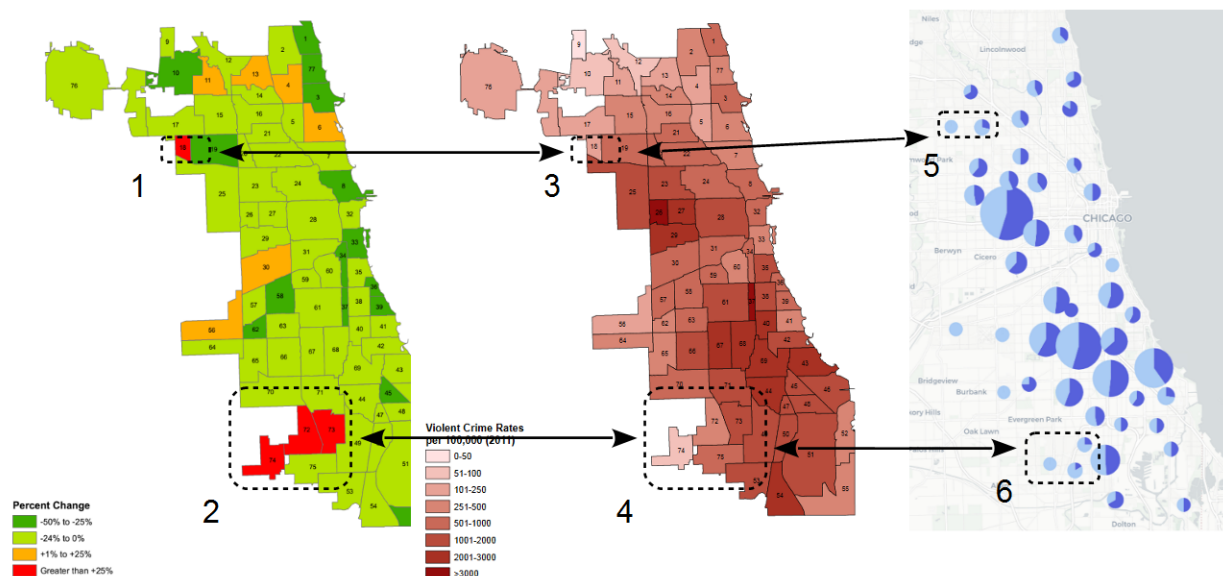


FIGURE 9.37 – Comparaison de notre représentation cartographique avec les représentations produites par l'étude de l'Université de *Yale* [Papachristos, 2013]

À l'inverse (Repères 5 et 6 de la Figure 9.37), la représentation offerte par l'outil permet de visualiser simultanément :

- une forte augmentation du nombre d'événements dans ces zones entre 2011 et 2013, représentée par une forte proportion de couleur claire dans les agrégats ;

- le faible nombre d'événements concernés par cette évolution, représenté par la faible taille des agrégats correspondants.

9.6.3 Synthèse

9.6.3.1 Événements de cambriolages

Une première analyse, effectuée selon une entrée temporelle, permet les observations suivantes :

- La distribution temporelle des cambriolages sur l'ensemble de la zone étudiée montre un pic d'intensité en début de journée entre 6h et 9h. Les cambriolages ont ensuite majoritairement lieu durant la journée entre 9h et 21h, tandis que les cambriolages survenant la nuit, entre 21h et 6h sont minoritaires.
- Certaines zones de la ville présentent une proportion de cambriolages nocturnes plus importante, notamment dans le *North Side*. Dans cette zone, la distribution temporelle des cambriolages présente deux pics d'intensité en début de journée entre 6h et 9h, et en début de soirée entre 18h et 21h. Les cambriolages ont ensuite majoritairement lieu la nuit entre 21h et 6h, tandis que les cambriolages survenant entre 9h et 18h sont minoritaires.
- Certaines zones de la ville, notamment à la frontière entre le *Far Southwest Side* et le *Southeast Side*, présentent une proportion de cambriolages nocturnes moins importante que la moyenne, et une proportion plus importante d'événements survenant le matin entre 6h et 12h.

Une seconde analyse, effectuée selon une entrée spatiale, permet d'identifier que :

- la distribution temporelle des cambriolages survenant dans le *North Side* correspond à celle trouvée dans la première analyse, et présente une proportion plus importante d'événements nocturnes que la moyenne, et une proportion moins importante d'événements survenant la journée ;
- à l'inverse du *North Side*, la distribution temporelle des cambriolages dans les autres *Sides* de Chicago, en dehors du centre-ville, correspond à la distribution temporelle des événements pour l'ensemble de la ville ;
- à l'inverse de l'ensemble [*North Side* ; *West Side* ; *South Side*], où les cambriolages ont majoritairement lieu en début de journée entre 6h et 9h, puis en journée entre 9h et 18h, les cambriolages survenant dans le centre-ville apparaissent majoritairement soit en pleine nuit entre 0h et 3h, soit au milieu de la journée entre 12h et 15h.

On observe ainsi deux zones présentant une distribution temporelle des cambriolages distincte de celle de l'ensemble de la ville : le *North Side* et le centre-ville.

Une hypothèse sur l'origine de la différence entre la distribution temporelle des événements du centre-ville, et celle des *Sides* l'environnant, peut être que :

- dans le centre-ville rassemblant des quartiers d'affaires, les lieux pouvant faire l'objet d'un cambriolage sont occupés majoritairement durant la journée ;
- à l'inverse dans l'ensemble [*North Side* ; *West Side* ; *South Side*], rassemblant des quartiers résidentiels, les lieux pouvant faire l'objet d'un cambriolage sont occupés durant la nuit.

Cependant, la distribution temporelle des événements du *North Side*, quartier très résidentiel [Wikipedia, 2018b], présentant une forte proportion d'événements nocturnes par rapport aux événements survenant durant la journée, s'oppose à cette hypothèse.

Il faut également préciser que le nombre d'événements survenant dans la zone du centre-ville est très peu important en comparaison au nombre d'événements survenant dans l'ensemble [*North Side* ; *West Side* ; *South Side*]. La comparaison de la distribution temporelle des cambriolages dans ces deux territoires peut ne pas être pertinente.

9.6.3.2 Événements d'homicides

Une première analyse, effectuée selon une entrée temporelle, permet d'identifier :

- une légère diminution du nombre d'homicides entre 2001 et 2015 sur l'ensemble de la ville de Chicago ;
- des zones de Chicago, comme le *West Side*, où cette diminution du nombre d'homicides au cours du temps reste faible ;
- d'autres zones, comme le *North Side* et le centre-ville, où cette diminution du nombre d'homicides au cours du temps est plus forte.

Une seconde analyse, effectuée selon une entrée spatiale, nous permet de confirmer la forte diminution du nombre d'homicides au cours du temps dans le *North Side* et le centre-ville. L'analyse ne nous permet cependant pas de distinguer d'autres tendances significatives dans les autres *Sides*.

9.6.3.3 Crimes violents

Bien que la perte de données lors de la création de notre jeu de données nous empêche de comparer directement nos résultats avec ceux présentés par l'étude de l'Université de *Yale*, notamment pour l'évolution du nombre d'événements, l'analyse réalisée avec *GrAPHiST* nous montre un avantage de notre interface par rapport aux représentations cartographiques présentées dans l'étude de l'Université de *Yale*.

La Figure 9.34 et la Figure 9.37 montrent qu'une représentation agrégée des événements spatio-temporels selon leur moment d'apparition relatif au temps linéaire, permet de présenter simultanément une distribution spatiale du nombre d'événements et l'évolution du nombre d'événements au cours du temps selon leur lieu d'apparition.

Le rendu visuel offert par l'outil permet à la fois d'identifier les zones où apparaissent le plus de crimes violents, d'analyser l'évolution du nombre d'événements, entre 2011 et 2013, selon différentes zones de l'espace, et d'identifier où les variations sont les plus fortes.

Notre approche permet ainsi de combiner les informations des deux cartes présentées dans l'étude de l'Université de *Yale* au sein d'une seule représentation cartographique.

Il faut cependant préciser que la comparaison de notre représentation graphique, avec la carte de l'étude présentant la distribution spatiale des *crimes violents*, présente un biais. Notre représentation graphique montre des données brutes, tandis que la carte de l'étude montre un nombre d'événements relatif, rapporté au nombre d'événements par *Community area*.

9.7 Synthèse générale

Les analyses présentées dans ce chapitre ont permis, pour chaque jeu de données étudié, de distinguer des patterns dans la distribution spatio-temporelle des événements. Ces patterns concernaient :

- des concentrations spatio-temporelles, relatives au temps cyclique, d'événements d'accidents routiers, d'inondations ou de cambriolages ;

- des évolutions, localisées dans l'espace, du nombre d'apparitions d'événements d'homicides et de *crimes violents* au cours du temps.

L'analyse a ainsi permis d'identifier des zones de l'espace où des événements présentent une distribution temporelle particulière, ainsi que des tendances localisées dans l'évolution de l'intensité d'un phénomène durant un intervalle temporel linéaire.

Les analyses ont été effectuées sur différentes échelles spatiales temporelles :

- Les événements d'accidents routiers et d'inondation ont été observés au travers d'une échelle temporelle cyclique d'une durée d'un an, mais selon différentes granularités temporelles : la distribution des accidents routiers au cours de l'année a été observée à travers une division de l'année en quatre saisons, tandis que la distribution des événements d'inondation au cours de l'année a été observée à travers une division de l'année en douze mois.
- Les événements de cambriolage ont été observés au travers d'une échelle temporelle cyclique d'une durée d'une journée.
- Les événements d'homicides et de *crimes violents* ont été observés au travers d'une échelle temporelle linéaire.
- Les événements d'inondation ont été observés à travers deux échelles spatiales, une échelle nationale (la France métropolitaine), et une échelle régionale (le sud-ouest de la France).

Pour chaque cas d'étude, hormis le cas des *crimes violents*, l'analyse a combiné une approche temporelle et une approche spatiale :

- dans un premier temps, l'approche temporelle a permis d'identifier quelles étaient les zones spatiales où les événements présentaient une distribution temporelle significative.
- dans un second temps, l'approche spatiale a permis de croiser les informations relatives à la distribution spatio-temporelle des événements, avec les informations relatives au découpage territorial utilisé (centre urbain et zones de montagnes pour les accidents routiers ; bassins versants pour les inondations ; quartiers de Chicago pour les événements de criminologie). Ce croisement d'information permet alors d'établir une hypothèse sur les causes de la distribution spatio-temporelle des événements.

L'absence de connaissances poussées sur les différentes problématiques abordées, ainsi que les limites posées par les jeux de données utilisés, nous empêchent de dégager de ces analyses des conclusions sur les cas d'étude abordées.

En revanche, ces analyses montrent l'opportunité que peut présenter notre approche pour l'analyse de données spatio-temporelles, effectuées à l'avenir par un personnel plus expert.

La comparaison de notre représentation cartographique avec des représentations plus classiques, pour le cas des *crimes violents*, a notamment montré l'intérêt de notre approche pour combiner, au sein d'une même représentation cartographique, des informations sur la distribution spatiale des événements et le suivi de l'évolution temporelle d'un phénomène spatial.

Enfin, ces analyses présentent le caractère généraliste de notre approche, permettant d'intégrer et d'analyser des données spatio-temporelles liées à différentes problématiques métiers.

Conclusion

Validation de la possibilité d'identifier des patterns par l'analyse exploratoire

Les analyses effectuées dans les deux premiers chapitres de cette partie (Chapitre 7 et Chapitre 8) ont montré qu'il est possible d'identifier les différents patterns, auxquels nous nous intéressons, au cours d'une analyse exploratoire de séries d'événements.

Analyse des patterns relatifs au temps linéaire

Identification de clusters relatifs au temps linéaire

L'analyse de l'ensemble des engagements du conflit (Section 7.4.1) a permis d'identifier les clusters correspondant aux différentes zones d'affrontement au cours de la guerre. L'analyse focalisée sur le *théâtre d'opérations de l'est* (Section 7.4.2.1), entre 1864 et 1865, a permis d'identifier d'autres clusters, correspondant aux différentes phases de l'offensive nordiste à travers le territoire confédéré. Les structures identifiées sur la carte correspondent à différentes phases du conflit, identifiées par les historiens, et correspondant à une concentration des combats dans des étendues restreintes du temps et de l'espace.

Identification de propagation spatiale des événements

L'analyse de l'ensemble des engagements du conflit (Section 7.4.1) a permis de mettre en évidence des propagations d'événements au cours du temps. Ces propagations, illustrant un déplacement de la ligne de front au cours de la guerre, correspondent à l'évolution du territoire contrôlé par l'*Union*, représentée dans la Figure 7.4. Les analyses focalisées sur le déroulement de l'*Overland Campaign* (Section 7.4.2.2) et de la campagne d'*Appomatox* (Section 7.4.2.3) ont également permis d'identifier des propagations spatiales d'événements au cours du temps, permettant de recréer le récit de l'avancement des troupes nordistes au cours de ces deux campagnes.

Identification d'une propagation globale des événements grâce aux *Pie Chart*

Les utilisations comparées des représentations agrégée et non agrégée des événements, au niveau du *théâtre d'opérations de l'est* (Section 7.4.1.5), ont permis de montrer l'avantage de la représentation des agrégats d'événements sous la forme de *Pie Chart* pour identifier une évolution globale du phénomène dans l'espace au cours du temps. Malgré les déplacements successifs de la ligne de front sur un axe nord-sud, le déplacement global des zones de conflit vers le sud au cours de la guerre pouvait être identifié au moyen de cette représentation.

Analyse des patterns relatifs au temps cyclique

L'analyse des jeux de données que nous avons créés (Chapitre 8) a montré qu'il était possible d'identifier des récurrences cycliques, et par conséquent des clusters relatifs à une échelle

cyclique, au moyen de *GrAPHiST*.

Ainsi, pour chaque jeu de données, hormis le jeu A_2 , l'objectif de l'analyse proposée par *GrAPHiST* a été atteint :

- l'existence de récurrences cycliques dans les séries d'événements, et par conséquent l'existence de clusters relatifs au temps cyclique, a été identifiée ;
- la durée de la période des cycles observés a été identifiée de façon précise ;
- ces cycles ont pu être localisés dans le temps et l'espace, mais la définition des contours de leur emprise spatiale ou temporelle manque de précision.

Utilisation de *GrAPHiST* pour l'analyse exploratoire de différents jeux de données réelles

Après avoir validé la possibilité d'identifier des patterns au moyen d'une analyse exploratoire effectuée avec *GrAPHiST*, nous avons appliqué notre proposition à différents jeux de données réelles. Ces cas d'étude avaient deux objectifs :

- présenter le caractère générique de l'approche, en analysant des jeux de données relatifs à des problématiques de nature différentes.
- présenter les perspectives de l'utilisation de *GrAPHiST* dans chacune de ces problématiques.

Analyse d'accidents routiers

Lors d'une première analyse (Section 9.4), nous avons étudié les accidents routiers apparaissant autour de la ville de *Grenoble*, entre 2005 et 2016. Cette analyse a eu pour but d'analyser si les accidents routiers, apparaissant durant les différentes saisons de l'année, étaient concentrés dans certaines zones de l'espace. L'analyse a permis d'identifier deux zones spatiales, *Grenoble* et les communes des massifs montagneux, où les saisons où surviennent le plus grand nombre d'accidents routiers ne sont pas les mêmes que pour l'ensemble de la région étudiée.

L'analyse a permis d'identifier des zones de l'espace où les événements présentent une distribution temporelle particulière sur une échelle cyclique d'une année.

Analyse d'épisodes d'inondations

Lors d'une seconde analyse (Section 9.5), nous avons étudié des épisodes d'inondation apparaissant sur le territoire de la France métropolitaine, entre 1900 et 2016. Cette analyse a eu pour but d'analyser si les épisodes d'inondation surviennent aux mêmes moments de l'année, dans les différentes régions de France.

L'analyse, effectuée sur l'ensemble de la France métropolitaine, a permis d'observer que les événements survenant dans le sud-ouest présentent une distribution temporelle distincte de celle des événements survenant dans l'ensemble de la zone d'étude. L'analyse, focalisée sur la région du sud-ouest, a permis d'identifier deux zones spatiales, où la majorité des inondations ne survient pas au même moment de l'année, que pour l'ensemble du sud-ouest de la France métropolitaine :

- les bassins versants des fleuves prenant leur source dans les Pyrénées : la Garonne et l'Adour ;
- les bassins versants des fleuves côtiers ou des fleuves prenant leur source dans le Massif central : la Dordogne et la Charente.

L'analyse a permis d'identifier des zones de l'espace où les événements présentent une distribution temporelle particulière sur une échelle d'une année. Ces zones ont été identifiées à une échelle nationale, puis à une échelle régionale. Cette distribution spatio-temporelle particulière a ensuite été mise en relation avec l'appartenance des événements correspondants à certains bassins versants.

Les analyses concernant les événements d'accidents routiers et d'inondations ont été effectuées au travers de la même échelle cyclique, mais selon des granularités temporelles différentes : les accidents routiers ont été analysés à travers une division de l'année en quatre saisons, et les inondations ont été analysés à travers une division de l'année en douze mois. L'analyse des événements d'inondation a ainsi été effectuée selon une échelle temporelle plus fine, permettant d'identifier des pics d'apparition d'événements moins étendus dans le temps.

Analyse de données criminologiques

Lors d'une troisième analyse (Section 9.6), nous étudions des données de criminologie, relatives aux crimes et délits survenus sur le territoire de la ville de *Chicago*, entre 2001 et 2017. Cette analyse s'est effectuée sur trois jeux de données différents :

- Dans un premier temps (Section 9.6.2.1), nous avons étudié des événements de cambriolages, survenus entre le 1^{er} janvier 2017 et le 1^{er} juillet 2017 ;
- Dans un second temps (Section 9.6.2.2), nous avons étudié des événements d'homicides, survenus entre 2001 et 2015 ;
- Enfin (Section 9.6.2.3), nous avons effectué une analyse de la distribution spatio-temporelle des *crimes violents*, regroupant les vols avec violence (*robbery*) et les homicides, entre 2011 et 2013 ;

Analyse d'événements de cambriolage

L'étude des événements de cambriolage a eu pour but d'analyser si ces événements surviennent aux mêmes moments de la journée, dans les différentes zones du territoire de Chicago. L'analyse a permis d'identifier deux zones présentant une distribution temporelle des cambriolages, au cours de la journée, distincte de celle de l'ensemble de la ville : le *North Side* et le centre ville.

L'analyse a permis d'identifier des zones de l'espace où les événements présentent une distribution temporelle particulière sur une échelle d'une journée.

Analyse d'événements d'homicides

L'étude des événements d'homicides a eu pour but d'analyser les tendances localisées dans l'évolution du nombre d'homicides entre 2001 et 2015 à Chicago. Nous avons observé :

- une légère diminution du nombre d'homicides entre 2001 et 2015 sur l'ensemble de la ville de Chicago ;
- certaines zones de Chicago, comme le *North Side* et le centre-ville, où la diminution du nombre d'homicides entre 2001 et 2015 est plus importante que sur l'ensemble de la ville.

L'analyse a permis d'identifier des tendances localisées dans l'évolution de l'intensité du phénomène sur le temps long.

Analyse d'événements de *crimes violents*

L'étude des *crimes violents* a eu pour but de comparer les représentations graphiques proposées par *GrAPHiST*, avec deux cartes produites lors d'une étude de l'*Institution for Social and*

Policy Studies de l'Université de *Yale* sur l'évolution du nombre de *crimes violents* entre 2011 et 2013 à Chicago [Papachristos, 2013](Figure 9.22).

L'analyse a montré l'intérêt de notre approche, qui combine, au sein d'une seule représentation cartographique, les informations contenues dans les deux cartes présentées dans l'étude de l'Université de *Yale*.

Critiques et travaux futurs

Les limites présentées par certaines représentations graphiques lors des expérimentations

Lors de nos expérimentations, il s'est avéré que le diagramme en dents de scie était plus efficace que le diagramme sinusoïdal pour rechercher des récurrences cycliques dans les données. Le diagramme sinusoïdal, au contraire, induit des erreurs de lecture du moment d'apparition des événements relatif au temps cyclique.

Pour ces raisons, nous avons majoritairement utilisé le diagramme en dents de scie, notamment pour la recherche de récurrences cyclique. Le diagramme temporel circulaire était privilégié pour l'analyse des clusters relatifs à une échelle temporelle cyclique. Le diagramme sinusoïdal, permettant d'observer la continuité temporelle des différentes instances de l'échelle cyclique, a été utilisé pour effectuer des analyses de patterns relatifs au temps linéaire. Cependant, la plupart de ces dernières ont également été effectuées au moyen du diagramme temporel en dents de scie.

Lors de nos expérimentations, il s'est également avéré que la représentation des événements sous une forme agrégée, dans la carte ou dans le diagramme, était plus efficace pour identifier des patterns que la représentation de chaque événement par une entité ponctuelle, notamment au sein d'importants jeux de données. Les analyses ont donc été principalement effectuées en utilisant une représentation agrégée des événements.

Travaux futurs : évaluer l'efficacité de *GrAPHiST* par des tests utilisateurs

Les expérimentations présentées dans les chapitres 7 et 8 ont montré que l'environnement *GrAPHiST* est fonctionnel : il est possible d'identifier les différents patterns que nous avons introduits dans le chapitre 2 au moyen de notre environnement.

Les cas d'étude du chapitre 9, et également celui du chapitre 7, ont montré le caractère générique de notre approche, l'analyse exploratoire permise par *GrAPHiST* ayant été appliquée à différents jeux de données, reliées à des problématiques de différentes natures.

Ces cas d'étude ont également montré différentes opportunités offertes par l'analyse exploratoire des dynamiques spatio-temporelles au moyen de l'outil :

- L'étude des batailles de la *Guerre de Sécession* (Chapitre 7) a montré qu'il était possible de recréer le récit d'un épisode historique à partir de l'analyse des patterns présents dans un jeu de données relatif à cet épisode de l'Histoire.
- L'étude des données d'accidents routiers (Section 9.4) présente un exemple d'analyse de la localisation de ces événements selon la saison de l'année au cours de laquelle ils apparaissent. Cette analyse peut aider à l'élaboration de politiques de sécurité routière visant à réduire le nombre d'accidents.
- L'étude des épisodes d'inondation (Section 9.5) présente un exemple d'analyse des rythmes annuels d'un phénomène naturel sur le temps long, ce qui peut aider à mettre en place des politiques de prévision des risques liés à ce phénomène.

- L'étude des cambriolages (Section 9.6.2.1) présente un exemple d'analyse de la localisation de ces événements selon le moment de la journée au cours duquel ils apparaissent. Cette analyse peut aider à la mise en place de politiques de sûreté visant à restreindre ce phénomène à court terme.
- L'étude des homicides (Section 9.6.2.2) présente un exemple d'analyse de l'évolution de la criminalité sur le temps long, à travers différentes zones de l'espace. Cette analyse peut aider à la mise en place de politiques publiques dans le but de réduire ce phénomène sur le long terme.

L'étude des *crimes violents* (Section 9.6.2.3) a, quant à elle, présenté l'intérêt de nos représentations graphiques par rapport à d'autres représentations plus classiques de la donnée spatio-temporelle, pour localiser des pics d'intensité et des tendances localisées dans l'évolution de l'intensité d'un phénomène.

La question qui se pose à présent est l'efficacité de notre approche dans l'analyse des phénomènes spatio-temporels : l'analyse exploratoire proposée par *GrAPHiST* présente-t-elle un intérêt par rapport aux analyses utilisées actuellement pour étudier les phénomènes spatio-temporels ? L'analyse proposée apporte-t-elle une meilleure précision ? Est-elle plus aisée à utiliser ? L'information y est-elle plus facilement interprétable ?

L'étude de l'efficacité de notre approche, en termes de facilité d'utilisation de *GrAPHiST* et d'interprétation des résultats, peut faire l'objet de tests utilisateurs. Ces tests impliqueraient plusieurs acteurs, qui auront pour tâche d'analyser une série de jeux de données prédéfinis, en essayant d'identifier des patterns connus de l'expérimentateur. Ces tests reviennent à répéter le même type d'expérimentations que celles décrites dans le chapitre 8, en augmentant le nombre d'utilisateurs. Ces tests pourraient alors donner une idée de la possibilité d'utilisation de notre approche par un plus large public.

L'étude de l'efficacité de notre approche exploratoire, par rapport à d'autres approches d'analyse utilisées pour étudier les phénomènes spatio-temporels, est plus complexe. Beaucoup de méthodes d'analyse des phénomènes spatio-temporels sont en effet développées de manière *ad hoc* [Andrienko and Andrienko, 2006]. La comparaison de notre proposition avec d'autres méthodes analytiques s'effectuerait donc dans le cadre de problématiques métiers définies.

Un des apports de fait de notre approche, pour chaque problématique, est dû au caractère exploratoire de l'approche, offrant à l'analyste la possibilité d'étudier des jeux de données inconnus de lui, mais dont l'analyse est susceptible d'apporter quelque chose à son domaine de recherche. Dans ce cas, l'analyste doit savoir si l'étude des différentes dynamiques que nous nous proposons d'analyser au travers de notre outil, présente un intérêt pour son domaine de recherche.

Une manière d'estimer l'efficacité de notre approche dans l'analyse des phénomènes spatio-temporels par rapport à d'autres méthodes, peut ainsi faire l'objet de tests qualitatifs. Au cours de ceux-ci nous pourrions présenter *GrAPHiST* aux experts de différents domaines de notre recherche, afin qu'ils estiment si l'analyse des dynamiques, proposées par les différents scénarios décrits dans le chapitre 6, peut être d'utilité dans leur domaine de recherche.

Conclusion générale

Rappel de nos objectifs

Dans un contexte d'accroissement constant des données spatio-temporelles disponibles, notre problématique est le manque d'approches génériques pour l'élaboration d'environnements de visualisation, permettant une analyse exploratoire des phénomènes spatio-temporels indépendante du thème traité. Pour répondre à cette problématique, notre objectif est de proposer des méthodes de modélisation des données spatio-temporelles, de représentation graphique, et d'interaction permettant la mise en place d'une analyse exploratoire des dynamiques spatio-temporelles, applicable à différents phénomènes.

Pour accomplir cet objectif, nous proposons *GrAPHiST* ((*environnement de*) *Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels*), un environnement de géovisualisation devant permettre une analyse de phénomènes de différentes natures selon différentes échelles spatiales ou temporelles, en mesure d'identifier les quatre dynamiques que nous présentons dans le chapitre (2).

Méthodologie mise en place

Notre premier travail (Repère 1 de la Figure 9.38) fut de définir, à travers l'établissement d'un état de l'art des méthodes de modélisation du changement dans l'espace, une approche de modélisation générique des données spatio-temporelles, afin de permettre l'étude de phénomènes de différentes natures au sein de *GrAPHiST*.

Notre second travail (Repère 2 de la Figure 9.38) fut l'établissement d'un état de l'art des approches existantes de représentation visuelle de la donnée spatio-temporelle, dans le but d'identifier les méthodes de représentation graphique et interactives, susceptibles de nous aider à concevoir notre environnement de visualisation, ainsi que les limites des environnements de visualisation actuels pour la recherche des dynamiques spatio-temporelles.

La troisième partie de notre travail (Repère 3) de la Figure 9.38) fut consacrée à l'élaboration et à l'implémentation de *GrAPHiST*. Les procédés de visualisation proposés dans *GrAPHiST* devaient dépasser les limites des environnements de géovisualisation actuels, identifiées dans l'état de l'art.

La quatrième partie de notre travail (Repère 4 de la Figure 9.38) consista à formaliser différents scénarios d'analyse, décrivant les actions devant être effectuées par l'utilisateur au moyen de *GrAPHiST*, pour identifier chaque dynamique présentée dans le chapitre 2.

La dernière partie de notre travail (Repère 5 de la Figure 9.38) consista à appliquer notre approche d'analyse exploratoire à différents cas d'étude, afin de valider notre proposition.

Le développement de *GrAPHiST* s'est effectué de manière empirique, selon une approche *bottom up*. La structure de l'environnement, les représentations graphiques utilisées, les possibilités d'interaction, les motifs graphiques recherchés et les stratégies d'utilisation ont évolué tout au long du développement, jusqu'à l'achèvement de la phase d'expérimentation (Repère 6 de la Figure 9.38). Ces modifications ont pour origine :

- l'identification de nouveaux besoins, ou de nouvelles opportunités de visualisation, lors du développement et de la phase d'expérimentation ;
- la découverte de méthodes existantes, pouvant servir de sources d'inspiration pour la création de nouveaux outils d'analyse, ouvrant de nouvelles perspectives pour l'environnement ;
- les difficultés rencontrées au cours de l'implémentation.

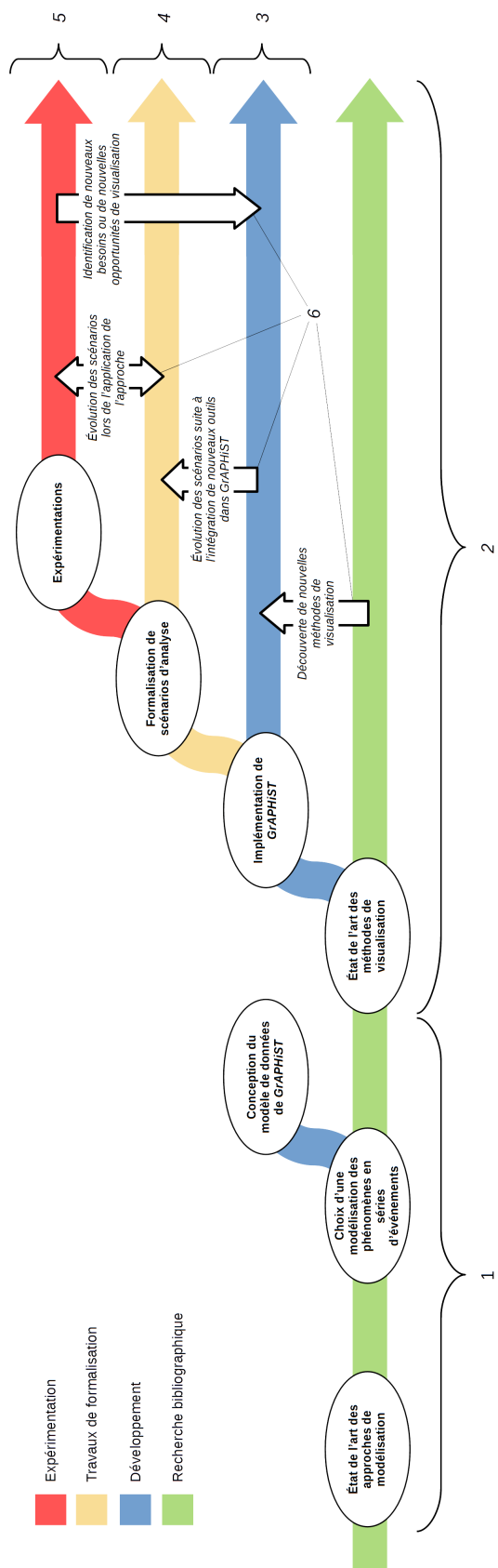


FIGURE 9.38 – Méthodologie de notre travail

Contributions de la thèse

La contribution de ce travail se situe à différents niveaux :

- au niveau de la modélisation des données spatio-temporelles ;
- au niveau de la visualisation des phénomènes spatio-temporels ;
- au niveau de la formalisation d'une méthode d'analyse exploratoire.

Modélisation des données spatio-temporelles

Malgré les différentes approches existant dans la littérature, les progrès dans la modélisation des phénomènes spatio-temporels restent insuffisants et manquent notamment d'approches génériques [Galton and Worboys, 2005, Haddad, 2009]. La modélisation des phénomènes reste majoritairement guidée par le domaine d'application étudié. Ces phénomènes sont le plus souvent décrits de manière *ad hoc* pour répondre à des besoins spécifiques au cas d'étude analysé, et non en se basant sur une approche théorique [Peuquet, 2002, Yuan, 2007, Haddad, 2009]. Les phénomènes les plus souvent modélisés sont les phénomènes naturels. D'autres phénomènes, comme les phénomènes sociaux, sont plus souvent modélisés de manière implicite, et non au travers d'entités spatio-temporelles explicitement définies [Haddad, 2009].

La première contribution de notre travail est une approche de modélisation générique des phénomènes spatio-temporels en séries d'événements, applicable à différents cas d'étude. Le modèle de données de *GrAPHiST* a pour objectif de permettre la modélisation d'événements, pour lesquels les données disponibles permettent la description de différentes phases de leur existence. Dans le cadre de cette thèse, nous nous sommes cependant limités à la représentation d'événements ponctuels dans le temps et l'espace.

Visualisation des phénomènes spatio-temporels

L'établissement d'un état de l'art des méthodes de représentations graphiques des données spatio-temporelles nous a permis de tirer les conclusions suivantes :

- L'utilisation d'environnements de géovisualisation à multi-fenêtrage apparaît comme la méthode la plus adaptée pour l'analyse exploratoire de données spatio-temporelles. Ces environnements peuvent intégrer d'autres méthodes de représentation (règles de symbologie, représentation en trois dimensions, *small-multiples*, animation) pour permettre à un utilisateur d'effectuer le lien entre les composantes temporelle et spatiale de la donnée.
- Les environnements de géovisualisation à multi-fenêtrage actuels présentent deux principales limites :
 - Les représentations du temps utilisées ne permettent pas une recherche des récurrences cycliques des phénomènes. Si ces environnements intègrent le temps cyclique, celui-ci est représenté au travers d'échelles usuelles, calendaires (heures de la journée, jour de la semaine, etc.). L'objectif de ces environnements n'est pas d'identifier les comportements cycliques d'un phénomène, mais d'analyser, à travers une représentation calendaire du temps, un phénomène pouvant être considéré à l'avance comme étant cyclique.
 - Les approches utilisées pour effectuer le lien entre les dimensions temporelle et spatiale passent majoritairement par une classification des données selon leur composante spatiale. Ce procédé permet l'analyse des événements à travers un découpage spécifique du territoire, mais nécessite d'énoncer a priori une hypothèse sur la distribution spatio-temporelle des événements.

Si certains environnements utilisent, à l'inverse, une approche passant par la classification et la représentation des données selon leur composante temporelle, peu d'environnements proposent d'allier les deux approches au sein d'une même interface de visualisation.

Notre proposition tente de pallier à ces deux limites.

L'interface de visualisation de *GrAPHiST* suit le principe du multi-fenêtrage, et comporte :

- une fenêtre spatiale composée d'une carte d'inventaire, à la manière des environnements *PerSE* [Swedberg and Peuquet, 2017] et *Growth Ring Map* [Bak et al., 2009, Andrienko et al., 2011] ;
- une fenêtre temporelle composée de deux diagrammes :
 - un diagramme temporel mixte, représentant le temps au travers d'une échelle linéaire et d'une échelle cyclique ;
 - un diagramme circulaire, inspirée du *Rose Chart* de Florence Nightingale [Nightingale, 1858], représentant le temps selon une échelle temporelle cyclique, divisée selon des intervalles temporels réguliers.

Notre seconde contribution est la formalisation d'un diagramme temporel mixte basé sur la forme d'un signal en dents de scie, inspiré des diagrammes en *tiles map*. Ce diagramme permet une lecture de la composante temporelle des données événementielles :

- selon le temps linéaire, en observant leur position le long de l'axe des abscisses ;
- selon une échelle cyclique, en observant leur position le long de l'axe des ordonnées.

En permettant, à l'instar de *Spiral Graph* [Weber et al., 2001], de modifier de manière dynamique et graduelle l'échelle cyclique représentée, ce diagramme autorise également la recherche et l'identification d'une réapparition cyclique des événements.

La troisième contribution de notre travail est une approche d'analyse des récurrences cycliques d'un phénomène, intégrant la dimension spatiale, à travers l'interface de *GrAPHiST*.

L'analyse de cycles dans les données s'effectue majoritairement sur un plan strictement temporel, notamment en utilisant des méthodes d'analyses quantitatives comme les transformées de Fourier. Peu d'analyses prennent en compte la dimension spatiale de ces cycles.

En intégrant l'analyse de récurrences cycliques, permise par le diagramme temporel mixte, au sein d'un environnement de géovisualisation, et en permettant d'effectuer le lien entre les composantes spatiale et temporelle des événements, *GrAPHiST* permet de prendre en compte la composante spatiale des événements lors de l'analyse des récurrences cycliques. Cette prise en compte de la dimension spatiale permet alors l'identification de comportements cycliques du phénomène, localisés dans l'espace.

L'état de l'art a montré que les procédés utilisés dans les environnements de géovisualisation, pour effectuer le lien entre les composantes temporelle et spatiale des données, peuvent être :

- orientée de l'espace vers le temps (*Time Wave* [Li, 2010], *PerSE* [Swedberg and Peuquet, 2017]). La composante spatiale des phénomènes est alors projetée dans une représentation temporelle.

-
- orientée du temps vers l'espace (*Growth Ring Map* [Bak et al., 2009, Andrienko et al., 2011]). La composante temporelle des phénomènes est alors projetée dans une représentation spatiale.

Notre quatrième contribution est de permettre de combiner ces deux approches, en offrant la possibilité d'analyser les phénomènes selon un point de vue orienté de l'espace vers le temps, ou du temps vers l'espace, dans un même environnement de visualisation. Les événements peuvent alors être représentés, au moyen de règles de symbologie, en fonction :

- de leur lieu d'apparition ;
- de leur moment d'apparition relatif à une échelle temporelle linéaire ;
- de leur moment d'apparition relatif à une échelle temporelle cyclique.

La combinaison de ces deux approches offre de nouvelles opportunités pour l'analyse exploratoire des phénomènes spatio-temporels :

- La projection de la composante temporelle des phénomènes dans une représentation spatiale permet une première phase d'analyse, durant laquelle l'utilisateur explore les données sans hypothèse énoncée *a priori* sur leur organisation spatio-temporelle. Lors de cette première phase, l'utilisateur peut identifier certaines zones de l'espace, pour lesquelles il considère que l'analyse peut être approfondie.
- La projection de la composante spatiale des phénomènes dans une représentation temporelle permet une seconde phase d'analyse, durant laquelle l'utilisateur peut explorer la composante temporelle du phénomène, à travers un découpage territorial formé par les zones spatiales identifiées précédemment.

Cette approche fut appliquée aux cas d'étude présentés dans le chapitre 9. Les événements sont d'abord analysés selon une entrée temporelle, qui permet une première observation de la distribution spatio-temporelle des événements. L'analyse est ensuite approfondie au travers d'un découpage territorial spécifique pour chaque cas d'étude.

La cinquième contribution de notre travail est une proposition de représentation des agrégats d'événements, spatialement ou temporellement proches, par une entité graphique unique prenant la forme d'un *Pie Chart*.

Cette représentation permet de réduire le nombre d'entités graphiques visibles à l'écran, rendant plus aisée l'identification des concentrations d'événements et permettant de résoudre le problème de superpositions d'entités graphiques proches, tout en représentant :

- le nombre d'événements agrégés ;
- la proportion des événements agrégés apparaissant dans chaque zone spatiale, ou durant chaque intervalle temporel, utilisé pour classer les événements selon une composante dimensionnelle.

Cette représentation offre la possibilité d'observer quels sont les intervalles temporels, ou les zones spatiales, les plus représentés au sein d'un agrégat d'événements, permettant ainsi l'identification visuelle de clusters dans la distribution spatio-temporelle des événements.

Les chapitres 7, 8 et 9 montrent que ce procédé permet une identification de ces clusters, y compris dans de larges jeux de données.

Le chapitre 7 (Section 7.4.1.5) montre également l'avantage des *Pie Charts* pour identifier un déplacement global, au cours du temps, de la zone d'apparition des événements dans le cas de larges jeux de données. En limitant le nombre d'entités graphiques à l'écran, tout en gardant visible la proportion des événements agrégés apparaissant durant les mêmes intervalles temporels, la représentation des agrégats d'événements spatialement proches sous la forme de *Pie Charts* permet d'identifier une dynamique globale du phénomène, là où un grand nombre d'entités graphiques, pouvant se superposer, gênent la lecture de la carte.

Formalisation d'une analyse exploratoire des dynamiques spatio-temporelles, indépendante du thème étudié

Une dernière contribution de notre travail est la formalisation d'une analyse exploratoire des dynamiques spatio-temporelles, applicable à différents phénomènes et problématique de recherche, et déclinée en plusieurs scénarios selon l'objectif poursuivi.

Ces scénarios décrivent les actions devant être effectuées par un utilisateur au moyen de *GrAPHiST* :

- dans le but d'identifier une dynamique parmi celles sur lesquelles nous nous focalisons, en fonction du phénomène étudié et de la problématique dans laquelle il prend place ;
- selon que l'analyse soit effectuée au travers d'une approche orientée de l'espace vers le temps, où l'analyste suspecte une organisation spatio-temporelle particulière du phénomène, ou au travers d'une approche orientée du temps vers l'espace, où l'analyste n'énonce pas a priori d'hypothèse sur la structure spatio-temporelle du phénomène.

L'analyse formalisée dans ces scénarios est indépendante du thème dans lequel prennent place les phénomènes étudiés. Ces scénarios peuvent ainsi être appliqués à l'étude de différents phénomènes, dans le cadre de différentes problématiques. Les analyses présentées dans les chapitres 7, 8 et 9 s'effectuent sur des jeux de données couvrant des problématiques différentes (reconstitution du récit d'un épisode de l'Histoire, analyse des rythmes des phénomènes d'accidents ou d'inondation, analyse de la distribution des événements de cambriolages durant la journée, analyse de l'évolution du nombre d'homicides sur une période de quinze ans), et montrent le caractère générique de l'analyse exploratoire formalisée.

Limites de notre approche

Limites dues au choix de se limiter aux événements ponctuels

La représentation d'événements ponctuels dans le temps et l'espace a influencé nos choix de représentation graphique. Si **l'utilisation de règles de symbologie, combinée à une représentation des événements proches par des agrégats prenant la forme de *Pie Charts***, permet l'analyse d'entités spatio-temporelles ponctuelles, **cette approche s'avère problématique pour l'analyse d'événements étendus dans le temps et l'espace.**

La représentation de la composante spatiale d'événements étendus dans l'espace empêche l'utilisation d'entités graphiques ponctuelles pour représenter des agrégats d'événements spatialement proches. Des entités graphiques surfaciques peuvent être superposées dans la carte dans le cas d'événements proches.

Certains environnements de géovisualisation utilisent la transparence pour répondre au problème de superposition d'entités surfaciques dans la carte. L'utilisation de la transparence rend

cependant difficile l'usage des variables de la couleur ou de la valeur pour représenter un événement en fonction de son moment d'apparition. Les couleurs ou les valeurs des entités superposées s'additionnant, il est impossible d'identifier les couleurs ou valeurs affectées à chaque entité.

Le même problème se pose pour la représentation d'événements étendus dans le temps dans le diagramme mixte. La composante temporelle de ces événements peut être représentée par des entités graphiques linéaires, le long de la courbe du diagramme (Figure 5.7). Cette représentation crée cependant des problèmes de superposition des entités représentant des événements temporellement proches. L'utilisation de la transparence pose alors les mêmes contraintes pour l'utilisation de la couleur ou de la valeur pour représenter le lieu d'apparition de ces événements.

La représentation des événements duratifs pose également la question de leur représentation dans le diagramme temporel circulaire : pour quel intervalle temporel, divisant l'échelle cyclique représentée, l'événement doit-il être comptabilisé ? Doit-il être représenté au travers d'une seule section du diagramme, ou l'événement doit-il être représenté au travers de l'ensemble des intervalles couverts par sa durée d'existence ?

Le choix s'est effectué dans le cadre de l'élaboration de notre environnement selon une approche *bottom up*. Nous cherchions à faire émerger une formalisation de règles de symbologie pour la représentation des événements spatio-temporels, à partir du développement empirique de *GrAPHiST*. Ce choix fut, à l'origine, motivé par :

- le format ponctuel de la majorité des données disponibles ;
- le besoin de limiter nos axes de travail lors du développement de *GrAPHiST*.

Ce choix a influencé les décisions prises par la suite lors de l'implémentation des différentes méthodes de représentation graphique. L'agrégation des événements proches, les règles de symbologie utilisées, et l'utilisation de la forme des *Pie Charts* pour représenter des agrégats d'événements furent déterminées par ce choix. **L'utilisation de ces méthodes exclut la représentation d'événements étendus dans le temps ou dans l'espace.**

Ce changement de point de vue sur les données à représenter s'est cependant effectué au cours du développement de *GrAPHiST*, avant l'utilisation des règles de symbologie soit formalisée, et non en amont du développement, lors d'une élaboration des spécifications de l'environnement. En résulte un décalage entre notre modèle de données, censé permettre la modélisation d'événements étendus dans le temps ou dans l'espace, et notre interface de visualisation, limitée à la représentation d'événements ponctuels.

Une autre conséquence du choix de se limiter aux événements ponctuels fut que, **malgré une modélisation permettant de décrire des événements spatio-temporels possédant différentes phases** au cours de leur existence, **les possibilités de représentations graphiques de ce type d'événements n'ont pas été étudiées au cours du développement de *GrAPHiST*.**

Cette limite de l'outil, en termes de représentations graphiques, ne représente cependant pas un problème dans le cadre de l'analyse des phénomènes spatio-temporels, modélisés sous la forme de séries événementielles. En effet, l'identification des dynamiques des phénomènes ne passe pas par l'observation de l'évolution individuelle de chaque événement dans le temps, mais par l'identification de structures dans la distribution spatio-temporelle de l'ensemble des événements.

L'absence de représentations graphiques, permettant de représenter l'évolution des différentes phases de chaque événement, est ainsi négligeable. L'enjeu principal des représentations gra-

phiques n'est pas la représentation des phases de chaque événement, mais la représentation des lieux et moments d'apparition de l'ensemble des événements composant le phénomène.

La représentation du temps cyclique dans le diagramme temporel sinusoïdal

Comme nous le précisons dans la section 5.3.3, la nature sinusoïdale d'une des formes prises par le diagramme temporel mixte peut entraîner des erreurs d'interprétation du moment d'apparition des événements, relatif à une échelle temporelle cyclique.

Ce dernier n'est pas représenté par la position de l'entité graphique correspondante le long de l'axe des ordonnées, comme c'est le cas dans le diagramme en dents de scie, mais par la position de l'entité graphique à l'intérieur d'une circonvolution de la courbe du diagramme. La lecture d'un moment temporel relatif au temps cyclique est ainsi rendue moins aisée que dans le diagramme en dents de scie.

Cette difficulté de lecture peut également être à l'origine d'erreurs d'interprétation dans la recherche de cycles dans les données événementielle : deux points de la courbe du diagramme, alignés parallèlement à l'axe des abscisses, peuvent ne pas représenter le même moment cyclique (Figure 5.18).

Enfin, le modèle sinusoïdal de la courbe peut fausser la lecture des espacements temporels entre les événements : une même longueur graphique représente un intervalle temporel plus important aux extrémités de la courbe qu'en son milieu.

Pour ces raisons, le **diagramme temporel sinusoïdal**, permettant néanmoins d'observer la continuité temporelle des différentes instances de l'échelle cyclique représentée, **est à réserver à l'analyse des événements au travers du temps linéaire.**

La validation de l'efficacité de l'approche

Les analyses de cas d'étude présentées dans les chapitres 7, 8 et 9 ont montré que l'environnement *GrAPHiST* permet l'identification des patterns décrits dans le chapitre 2, et ont montré que l'analyse exploratoire proposée peut être appliquée à différents phénomènes, dans le cadre de différentes problématiques de recherche.

L'objectif de ces expérimentations était de montrer que l'environnement *GrAPHiST* est fonctionnel. Nous souhaitons démontrer qu'il est possible d'identifier des dynamiques, lors de l'analyse de différents phénomènes spatio-temporels au travers de *GrAPHiST*. **En revanche, la question de l'efficacité de notre approche pour l'analyse des dynamiques des phénomènes spatio-temporels reste posée**, *GrAPHiST* n'ayant pas été confronté, ni à un panel plus large d'utilisateurs, ni à une application comparée avec d'autres méthodes d'analyse des phénomènes spatio-temporels.

Perspectives de recherche

Validation de l'efficacité de l'approche

La première perspective de recherche concernant notre proposition est la validation de l'efficacité de l'analyse exploratoire permise par *GrAPHiST*, pour l'étude des dynamiques des phénomènes spatio-temporels.

Pour répondre à cette question, il est nécessaire que l'environnement *GrAPHiST* fasse l'objet de tests utilisateurs. Ces tests peuvent porter :

- sur la facilité d'utilisation de l'environnement et de l'interprétation des résultats par un utilisateur ;

Ces tests impliquent de soumettre un panel d'utilisateur au même type d'expérimentation que celle effectuée dans le chapitre 8. Les utilisateurs devront analyser une série de jeux de données prédéfinis, afin d'y identifier des patterns connus de l'expérimentateur.

- sur l'intérêt de notre méthode d'analyse exploratoire, pour l'étude des dynamiques spatio-temporelles des phénomènes, par rapport aux méthodes existantes.

La comparaison de notre approche avec les méthodes existantes doit s'effectuer dans le cadre de problématiques de recherche définies. Ces tests peuvent être qualitatifs, et consister en une interview d'experts de différents domaines de notre recherche, sur l'intérêt qu'ils portent à l'identification des différentes dynamiques sur lesquelles nous nous sommes focalisés.

Élaboration de nouvelles méthodes de représentation graphique

Une perspective de recherche, pouvant être suivie dans la continuité de *GrAPHiST*, peut être **l'élaboration, et l'intégration dans l'environnement de visualisation de représentations graphiques adaptées aux événements non ponctuels**, permettant d'étendre l'analyse aux événements étendus dans le temps ou l'espace.

Comme nous le précisons auparavant, l'intégration d'événements surfaciques ou duratifs pose le problème de l'incompatibilité de l'utilisation de la transparence pour représenter des entités superposées, avec l'utilisation de la couleur ou de la valeur pour représenter une variable rattachée à ces entités.

Permettre la représentation d'événements, étendus dans l'espace et le temps, demande alors :

- soit la mise au point de nouvelles règles de symbologie, compatibles avec l'utilisation de la transparence, pour représenter les événements selon leur lieu ou selon leur moment d'apparition ;
- soit l'utilisation d'autres méthodes de représentation graphique des données spatio-temporelles, permettant de contourner la difficulté de combiner l'utilisation de règles de symbologie avec l'utilisation de la transparence.

Une proposition élaborée dans le cadre de la thèse, mais qui n'a pas été implémentée dans *GrAPHiST*, s'inscrit dans l'élaboration de nouvelles règles de symbologie. Le principe était d'utiliser la variable visuelle de l'orientation pour représenter les événements étendus dans l'espace, en fonction de leur moment d'apparition relatif à une échelle cyclique. Pour permettre de combiner cette règle de symbologie avec la transparence, la variable visuelle de l'orientation n'est pas appliquée sur la texture des entités graphiques, mais à un symbole graphique rattaché à chaque entité graphique, prenant la forme d'une flèche (Figure 9.39). Le procédé est le suivant :

- Un symbole de la forme d'une flèche est affecté à chaque entité dans la carte.
- Le centre de la flèche est positionné au barycentre de l'entité représentant l'événement.
- L'orientation de la flèche, par rapport à l'axe vertical, représente son moment d'apparition au travers de l'échelle cyclique utilisée. La flèche est, pour chaque événement, l'allégorie d'une horloge indiquant son moment d'apparition.
- La modification de l'échelle cyclique représentée modifie l'orientation des flèches dans la carte.

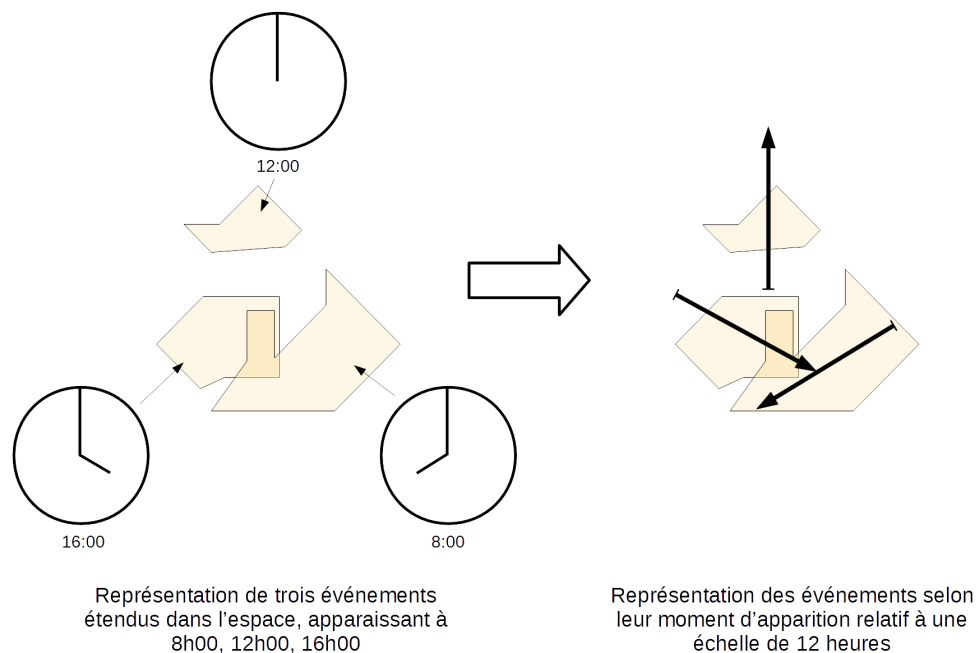


FIGURE 9.39 – Représentation du moment d'apparition des événements, au travers d'une échelle temporelle cyclique, au moyen de l'orientation d'un symbole flèche

Cette solution pourrait offrir la possibilité de représenter les événements selon leur moment d'apparition relatif à une échelle temporelle cyclique, sans que cette représentation soit impactée par une superposition d'événements surfaciques dans la carte.

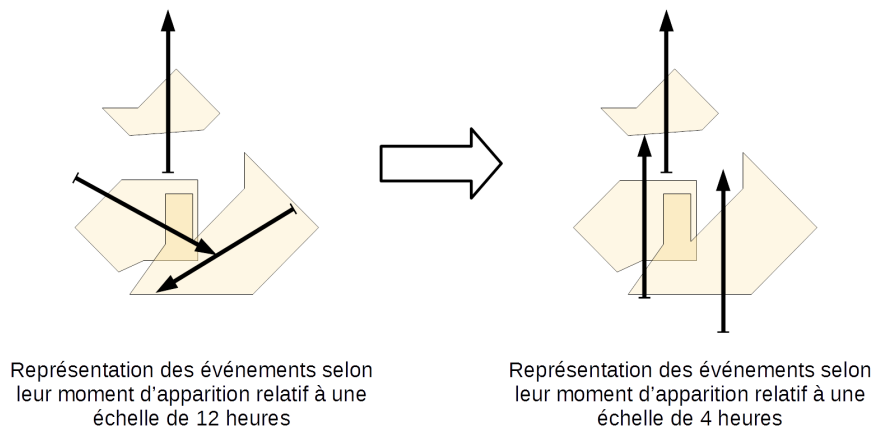


FIGURE 9.40 – Identification d'un cluster spatio-temporel relatif à une échelle cyclique, par identification d'un regroupement de symboles flèches pointant dans la même direction

Un cluster relatif à une échelle cyclique serait alors observable par une concentration dans l'espace de flèches pointant dans la même direction. La recherche de récurrence cyclique passerait par une modification de la durée de l'échelle cyclique, jusqu'à observer un alignement des flèches dans une même direction (Figure 9.40).

Une autre perspective d'amélioration des représentations utilisées dans *GrAPHiST*, peut être de représenter les agrégats d'événements dans la carte, non plus sous la forme de *Pie Charts*, mais sous la forme de *Rose Charts*.

La distribution temporelle des événements de chaque agrégat serait alors représentée, de la même manière que dans le diagramme temporel circulaire, directement sur la carte. Le moment d'apparition, relatif à l'échelle cyclique, des événements agrégés serait alors représenté non plus par la couleur, mais par l'orientation. Cette approche pourrait permettre :

- une meilleure identification du moment d'apparition des événements agrégés, selon l'échelle temporelle cyclique, l'orientation des secteurs des *Rose Charts* étant l'allégorie des différents secteurs d'une horloge ;
- une meilleure comparaison de la distribution temporelle des événements de chaque agrégat, représentée par chaque *Rose Chart* de la carte, avec la distribution temporelle de l'ensemble des événements, représentée par le *Rose Chart* du diagramme temporel circulaire.

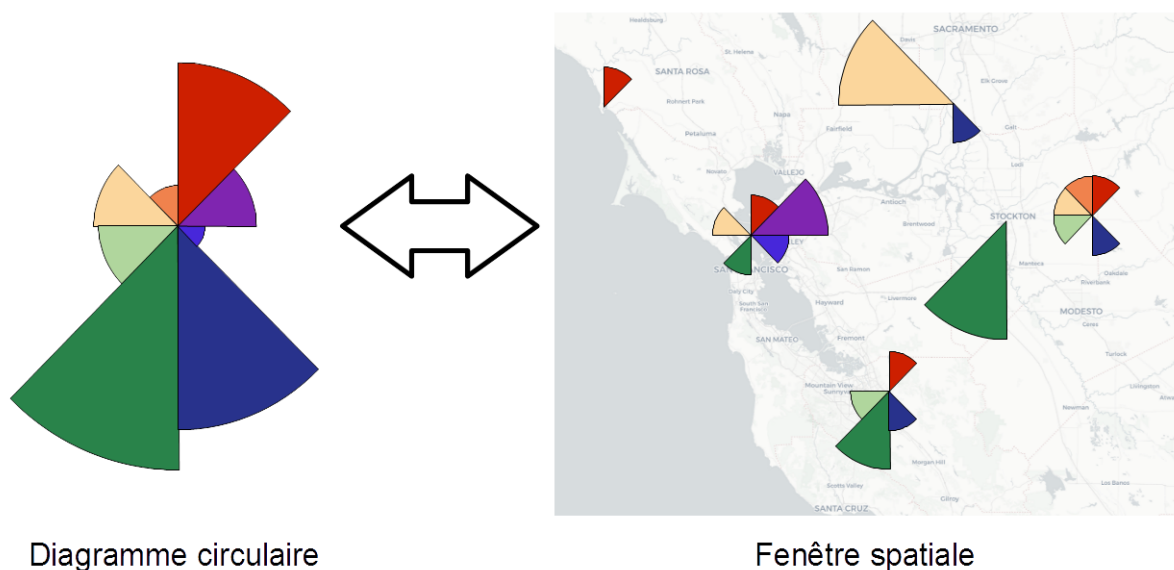


FIGURE 9.41 – Représentation des agrégats d'événements par des *Rose Charts*

Cette méthode pourrait cependant présenter plusieurs inconvénients :

- Les *Pie Charts* conservent la même surface graphique sur la carte, quelle que soit la distribution temporelle des événements agrégés. À l'inverse, les *Rose Charts* peuvent occuper une surface plus ou moins importante, selon que le nombre de secteurs présentant un rayon important est plus ou moins restreint. Les *Rose Charts* peuvent ainsi être moins visibles que les *Pie Charts* dans la carte.
- La surface graphique des *Pie Charts* restant constante, celle-ci est toujours centrée sur les coordonnées calculées pour le centre de l'agrégat d'événements. À l'inverse, si les événements représentés par un *Rose Chart*, apparaissent dans un même intervalle temporel, seuls quelques secteurs du *Rose Chart* sont visibles sur la carte. La surface graphique créée n'est alors plus centrée sur les coordonnées calculées pour le centre de l'agrégat d'événements, ce qui peut fausser l'identification de la localisation spatiale d'un cluster.

Perspectives d'amélioration de l'environnement *GrAPHiST*

D'autres perspectives peuvent concerner l'implémentation de l'environnement, dans le but d'améliorer son fonctionnement. Parmi celles-ci, certaines pourraient permettre de **rendre l'environnement utilisable par le plus grand nombre**.

GrAPHiST a été implémentée sous la forme d'une application Web, reliée à une base de données *PostgreSQL*. Le format Web permet l'utilisation de notre approche au moyen d'un simple

navigateur Internet, rendant plus aisée son utilisation par un large public, ce qui s'inscrit dans notre volonté d'offrir une méthode d'analyse généraliste des phénomènes spatio-temporels.

L'analyse d'un jeu de données nécessite cependant que ce dernier soit mis en forme, conformément au modèle de données décrit dans la section 5.1. Dans les chapitres 7, 8 et 9, nous mettons en forme les données analysées au moyen de scripts *SQL*, créés de manière *ad hoc* pour chaque jeu de données.

Afin d'améliorer le caractère *exportable* de l'environnement, **un des futurs travaux à effectuer consisterait à implémenter un outil permettant de mettre en forme les données à analyser de manière plus automatique.**

Une solution envisagée peut être l'utilisation d'une interface permettant à l'utilisateur d'importer des jeux de données bruts, et de spécifier quelles sont les composantes temporelles, spatiales et thématiques des données importées. En fonction des critères définis par l'utilisateur, cette interface peut faire appel à un script générique permettant de mettre en forme les données.

Une version plus aboutie de ce système pourrait également permettre l'import de jeux de données sous la forme de séries temporelles, rattachées à des entités spatiales. Un script générique pourrait alors recomposer ces données sous un format événementiel. Un jeu de données relatif à des mesures de pression atmosphérique effectuées par différents capteurs, disséminés dans l'espace, pourrait ainsi être recomposé sous la forme d'une série événementielle. Un événement serait défini par le franchissement d'une valeur seuil par la variable mesurée pour un capteur.

Une autre option envisagée pour améliorer le caractère exportable de notre approche, est **son implémentation sous la forme de *plug-in* d'un environnement de géovisualisation déjà existant, si possible largement utilisé, comme *QGIS*.** Ceci permettrait, entre autres, à l'utilisateur de combiner les outils proposés par notre approche, avec les analyses quantitatives implémentées dans ces environnements.

Ces nouveaux développements, en rendant *GrAPHiST* plus facilement utilisable par un plus grand nombre de personnes, peuvent également rendre plus aisée la mise en place de test utilisateurs. Dans le cas de domaine de recherche où l'analyse des données est d'ores et déjà effectuée au moyen d'environnements de géovisualisation, l'implémentation de notre approche sous la forme d'un *plug-in* pourrait également permettre la comparaison immédiate de notre méthode d'analyse exploratoire, avec les approches d'analyse existantes.

Perspectives d'application

Les expérimentations effectuées dans les chapitres 7 et 9 ont montré quelques perspectives d'application de l'analyse permise par l'environnement *GrAPHiST* :

- une analyse exploratoire similaire à celle effectuée dans le chapitre 7 peut aider à reconstruire le récit de certains épisodes de l'Histoire à partir de données brutes ;
- l'analyse des accidents routiers peut intervenir dans la mise en place de politiques de sécurité routière, adaptées à différentes régions de l'espace et à différents moments de l'année ;
- l'analyse des épisodes d'inondation peut faciliter la mise en place de politique de gestion du risque, en permettant l'analyse des mécanismes de ce phénomène selon différentes régions de l'espace ;
- l'analyse de la distribution spatio-temporelle des cambriolages selon une échelle temporelle d'une journée permet de mettre en place des politiques de sûreté, permettant d'agir rapidement sur ce phénomène ;

-
- l'analyse de l'évolution du nombre d'homicides au cours du temps, selon différentes régions de l'espace, peut intervenir dans la mise en place de politiques publiques et de sûreté, dans le but de réduire ce phénomène sur le long-terme.

D'autres perspectives d'application peuvent également être imaginées :

- une analyse similaire à celle présentée dans le chapitre 7 peut être effectuée sur des données d'épidémiologie, afin d'aider à l'identification des facteurs de propagation de certaines épidémies, en recréant le récit de la diffusion de celles-ci ;
- une analyse des récurrences cycliques spatialisées, similaires à celles effectuées dans le chapitre 8, peut être appliquée à des données météorologiques, afin d'aider à la prévision de certains phénomènes. Une analyse similaire peut être effectuée sur des données relatives aux phénomènes sismiques.

Ces multiples perspectives d'application sont un argument supplémentaire pour parfaire l'implémentation de *GrAPHiST*, dans le but de rendre celui-ci accessible et utilisable par le plus grand nombre d'analystes possible.

GrAPHiST peut également être un socle pour l'élaboration et l'implémentation future de nouvelles méthodes de représentation des données spatio-temporelles, qui pourraient étendre le champ des analyses possibles des phénomènes spatio-temporels.

S'intéresser aux problématiques de mobilité, par exemple, implique de faire évoluer les modalités de représentation offertes par *GrAPHiST*. Si l'on considère comme phénomène un ensemble de trajets d'individus effectués durant une emprise spatio-temporelle définie, chaque trajet peut être considéré comme un événement possédant :

- une phase décrivant un lieu et un moment de départ ;
- une phase décrivant un lieu et un moment d'arrivée ;
- plusieurs phases intermédiaires décrivant :
 - soit les différents lieux de passage de l'individu, et le moment de son passage dans chaque lieu ;
 - soit les différentes portions de trajet parcourues par l'individu, et l'intervalle temporel durant lequel il parcourt chaque portion.

Une solution pour étudier ce phénomène pourrait être d'intégrer à l'interface de *GrAPHiST* des représentations graphiques permettant de traiter le cas d'événements possédant :

- plusieurs phases au cours de leur existence ;
- une géométrie linéaire.

Les perspectives ouvertes par ce travail sont donc nombreuses, tant dans son expérimentation, dans les futures méthodes de représentations pouvant être intégrées au socle constitué par *GrAPHiST*, que dans les multiples domaines d'application de l'environnement.

Cinquième partie

Annexes

Chapitre 10

Manuel d'utilisation de *GrAPHiST*

Sommaire

10.1	Vue générale	394
10.2	Fenêtre temporelle	394
10.3	Interface de contrôle de la fenêtre temporelle	396
10.4	Interface de contrôle de la fenêtre spatiale	398
10.5	Import d'événements	398
10.6	Transformée de Fourier	400
10.7	Modification de la représentation graphique des événements	402
10.7.1	Diagramme temporel circulaire	402
10.7.2	Représentation des événements dans le diagramme mixte	402
10.7.3	Représentation des événements dans la carte	404
10.8	Sélection des événements	406
10.9	Utilisation de la symbologie	409
10.9.1	Représentation des événements selon leur lieu d'apparition	409
10.9.2	Représentation des événements selon leur moment d'apparition relatif à une échelle linéaire	411
10.9.3	Représentation des événements selon leur moment d'apparition relatif à une échelle cyclique	411

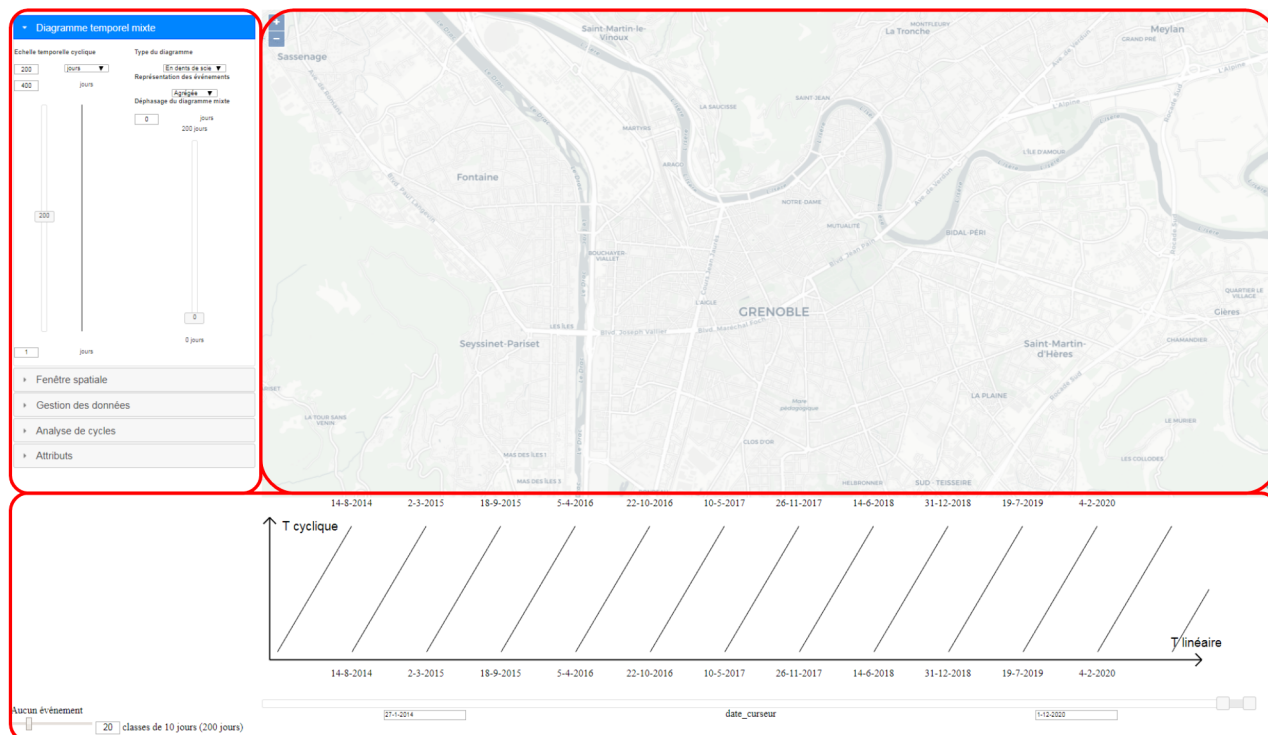
10.1 Vue générale

L'interface de *GrAPHiST* se compose (Figure 10.1) :

- d'une fenêtre spatiale ;
- d'une fenêtre temporelle ;
- d'une interface de contrôle.

Interface de contrôle

Fenêtre spatiale



Fenêtre temporelle

FIGURE 10.1 – Vue générale de l'interface de *GrAPHiST*

Le design de l'interface n'est pas dynamique et n'est pas adaptable à tout format d'écran. L'interface a été dessinée pour un écran 24 pouces, en utilisant le navigateur *Chrome*. L'environnement est également plus performant lorsque le navigateur *Chrome* est utilisé.

L'environnement *GrAPHiST* est consultable via cet URL : <http://lig-tdcge.imag.fr/TimeHelix-1.0/>.

10.2 Fenêtre temporelle

La fenêtre temporelle se compose :

- du diagramme temporel mixte (Repère 1 de la Figure 10.2) ;
- du diagramme temporel circulaire (Repère 2 de la Figure 10.2), visible seulement quand des événements ont été importés dans l'interface de visualisation.

Le diagramme indiqué par le Repère 3 de la Figure 10.2 est une timeline représentant l'étendue temporelle visualisable dans *GrAPHiST*. Celle-ci s'étant du 1er janvier 1800 au 1er janvier

2020.

L'étendue temporelle représentée dans le diagramme mixte (Repère 1 de la Figure 10.2), correspond à la zone grisée de la timeline indiquée par le Repère 3 (Repère 4 de la Figure 10.2).

Les bornes de cette étendue temporelle :

- sont représentées par les deux curseurs positionnés aux extrêmes de la zone grisée sur la timeline (Repère 4 de la Figure 10.2) ;
- correspondent aux dates indiquées dans les deux champs de texte situés en dessous de la timeline (Repères 5 et 6 de la Figure 10.2).

Cette étendue temporelle peut être modifiée :

- en faisant glisser les curseurs le long de la timeline ;
- en modifiant les dates indiquées dans les champs de texte.

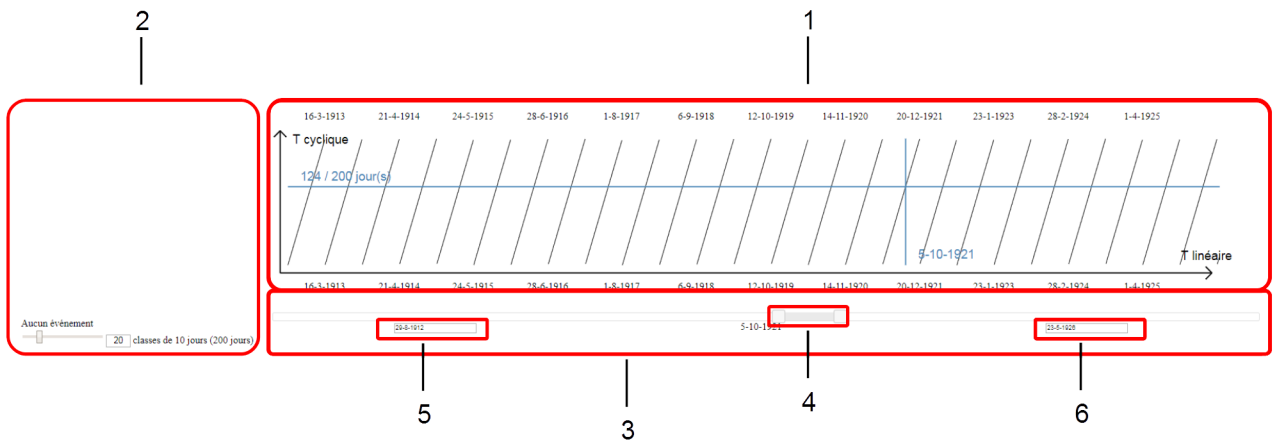


FIGURE 10.2 – Fenêtre temporelle de *GrAPHiST*

Lors d'un passage de la souris sur le diagramme temporel mixte, le moment temporel correspondant à la position du curseur est affiché sur le diagramme (Repère 7 de la Figure 10.3). Le diagramme indique la localisation temporelle de ce moment :

- selon une échelle temporelle linéaire (Repères 8 et 9 de la Figure 10.3) ;
- selon l'échelle temporelle cyclique utilisée (Repère 10 de la Figure 10.3).

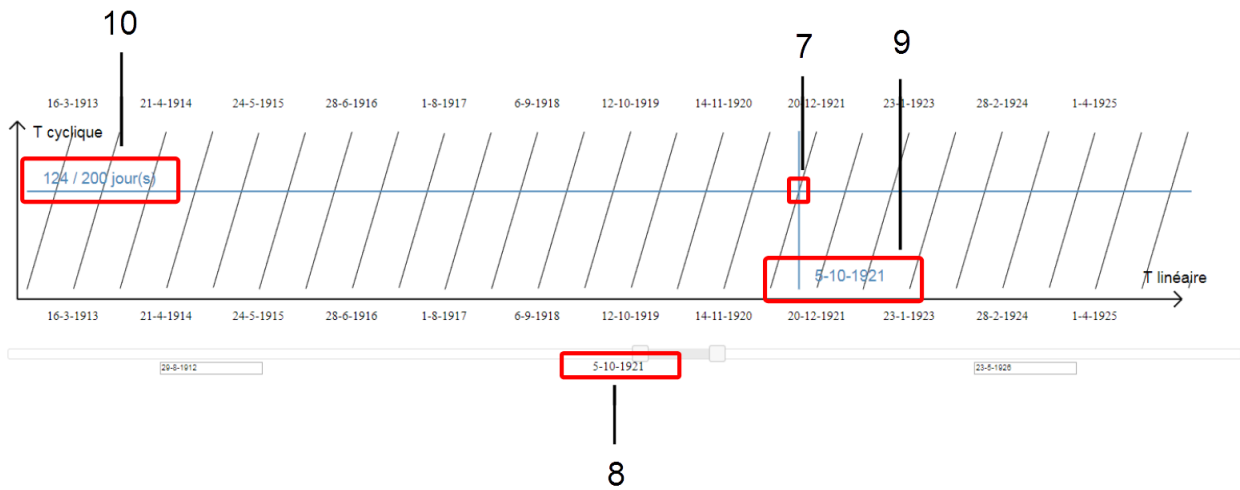


FIGURE 10.3 – Diagramme mixte de *GrAPHiST*

10.3 Interface de contrôle de la fenêtre temporelle

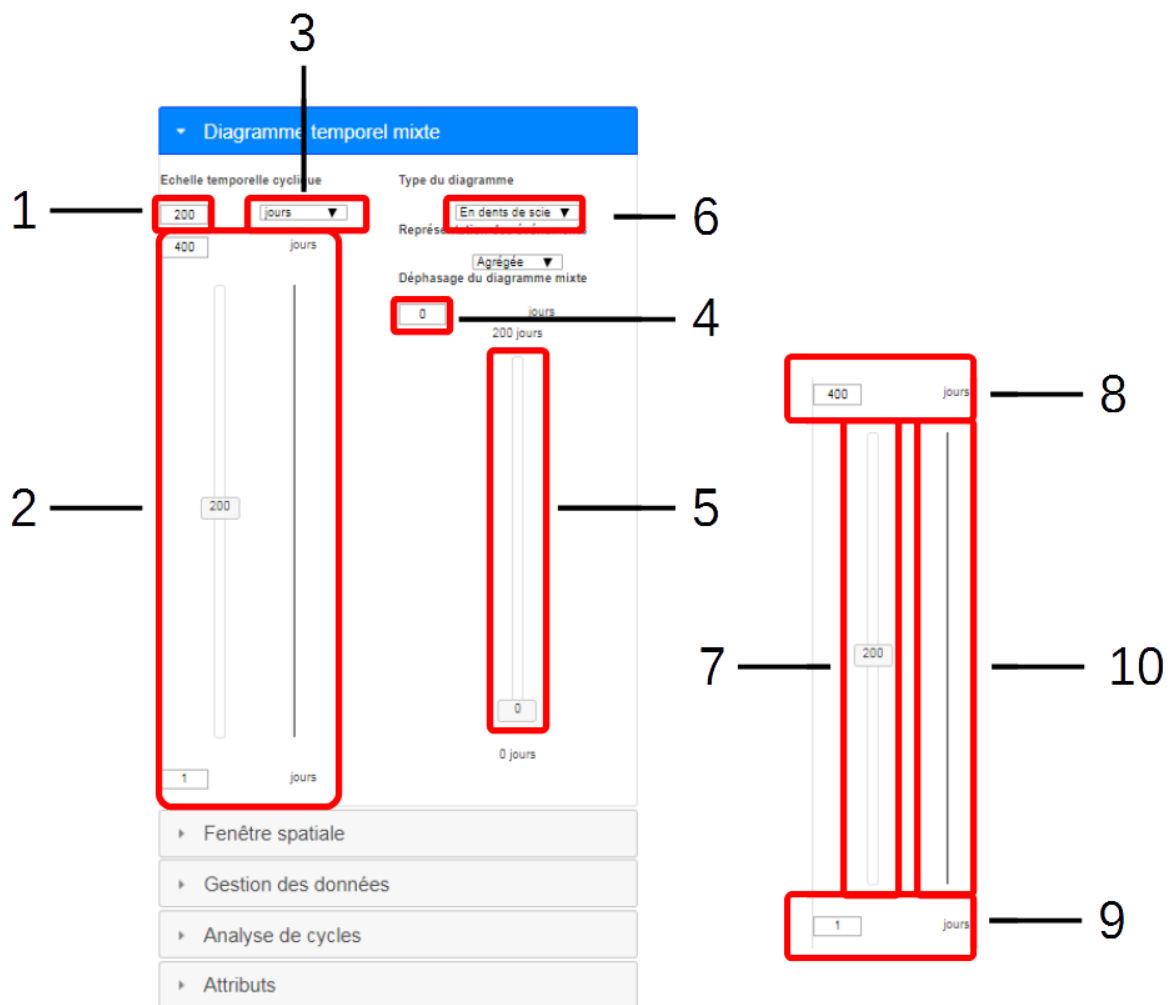


FIGURE 10.4 – Interface de contrôle de la fenêtre temporelle

La durée de l'échelle temporelle cyclique peut être modifiée au travers de l'interface de contrôle :

- en entrant une nouvelle durée au travers d'un champ de texte (Repère 1 de la Figure 10.4) ;
- au moyen d'un curseur (Repères 2 et 7 de la Figure 10.4). Au travers de ce dernier, la durée de l'échelle cyclique peut être modifiée de manière graduelle entre deux valeurs extrêmes, modifiables par l'utilisateur (Repères 8 et 9 de la Figure 10.4).

Après un import d'événements dans l'interface, les résultats renvoyés par la transformée de Fourier sont positionnés sur l'axe présenté par le Repère 10.

L'interface permet également de modifier le déphasage du diagramme temporel mixte, en entrant une valeur dans un champ de texte (Repère 4 de la Figure 10.4), ou au moyen d'un curseur (Repère 5 de la Figure 10.4).

L'utilisateur peut modifier la forme prise par la courbe du diagramme temporel mixte (Repère 6 de la Figure 10.4).

La modification de la durée de l'échelle cyclique peut s'effectuer de manière animée. Le bouton indiqué par le Repère 1 de la Figure 10.5 permet de lancer, ou d'arrêter une animation. Pour chaque pas de temps de l'animation, un pas de modification est ajouté à la durée de l'échelle cyclique, jusqu'à ce que celle-ci atteigne la durée indiquée par le Repère 8 de la Figure 10.4.

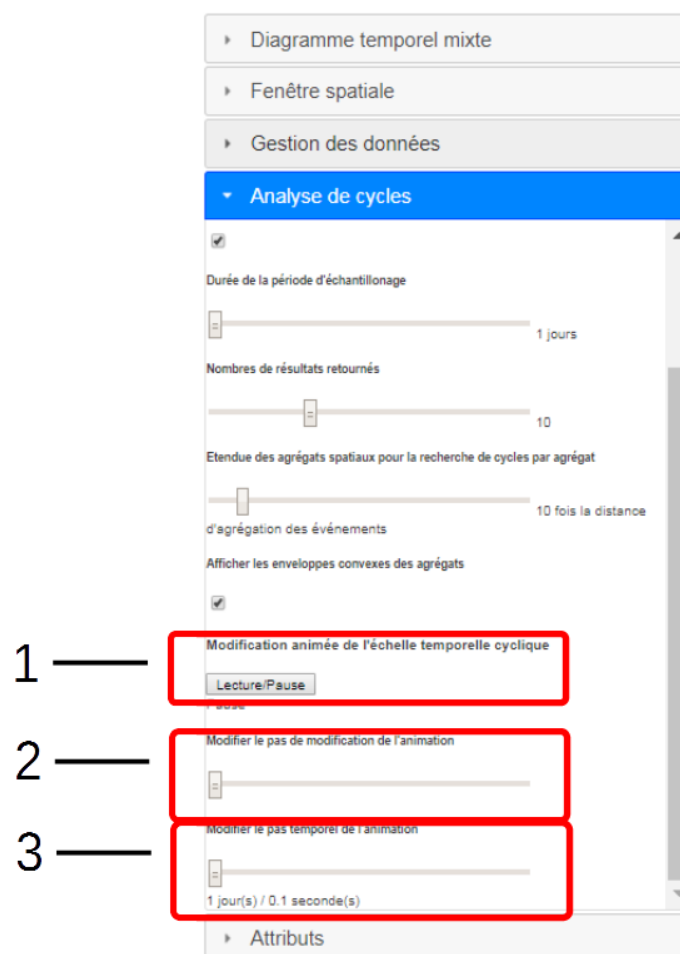


FIGURE 10.5 – Modification de la durée de l'échelle temporelle cyclique par l'animation

L'utilisateur peut également modifier le pas de modification (Repère 2 de la Figure 10.5) et le pas de temps utilisés par l'animation (Repère 3 de la Figure 10.5).

10.4 Interface de contrôle de la fenêtre spatiale

Le fond de carte de la fenêtre spatiale est modifiable au moyen de l'interface de contrôle (Figure 10.6).



FIGURE 10.6 – Interface de contrôle de la fenêtre spatiale

10.5 Import d'événements

La fenêtre d'import des événements peut être ouverte à travers l'interface de contrôle (Figure 10.7). L'import des événements peut être long pour certains jeux de données. Si le navigateur indique que la page ne répond plus, le script est encore en train de procéder.

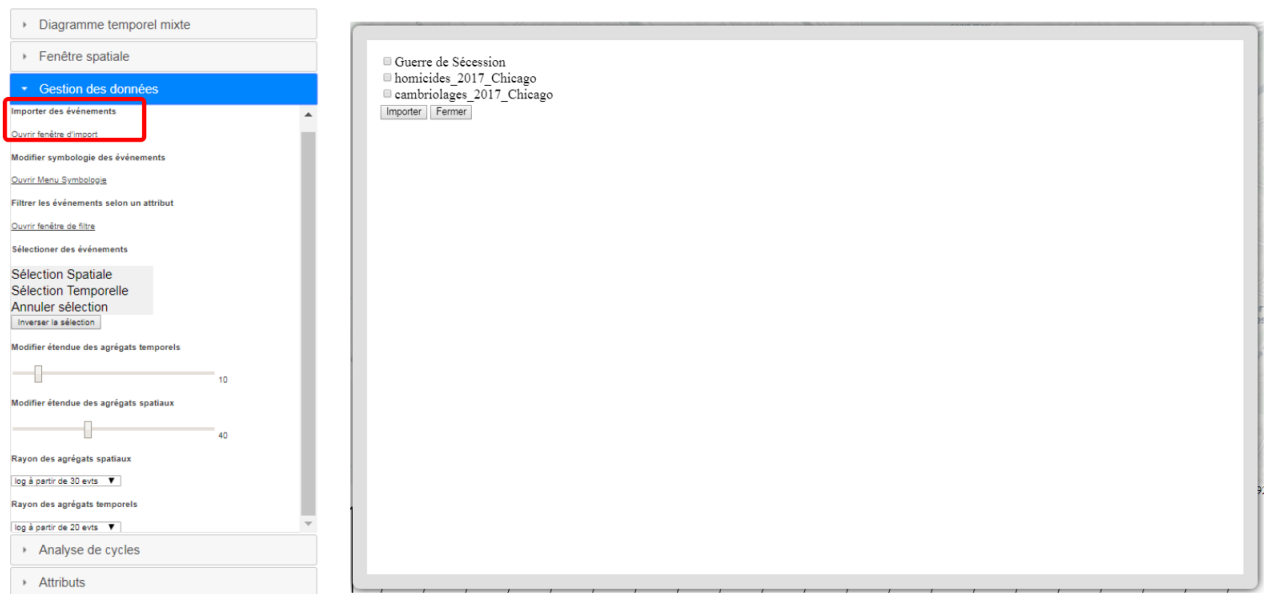


FIGURE 10.7 – Fenêtre d’import des événements

Le choix d’un phénomène modifie les emprises temporelle et spatiale représentées dans l’environnement.

Une fois un phénomène sélectionné au moyen de cette fenêtre, les événements à importer peuvent être filtrés selon leur composante thématique. Ce filtre s’effectue au travers d’une autre fenêtre, pouvant être ouverte depuis l’interface de contrôle (Figure 10.8).



FIGURE 10.8 – Fenêtre de filtre thématique des événements

10.6 Transformée de Fourier

Lors d'un import d'événements, une analyse par transformée de Fourier est effectuée automatiquement sur les données importées.

Le caractère automatique de l'analyse peut être désactivé dans l'interface de contrôle (Repère 1 de la Figure 10.9).



FIGURE 10.9 – Interface de contrôle (transformée de Fourier)

L'interface de contrôle permet également :

- de relancer une analyse par transformée de Fourier, une fois les événements importés (Repère 2) ;
- d'ouvrir une fenêtre affichant le signal créé à partir de la série événementielle, ainsi que le spectre calculé (Repère 3) ;
- de modifier la durée de la période d'échantillonnage du signal créé (Repère 4) ;
- de modifier le nombre de résultats renvoyés (Repère 5) ;

- dans le cas d'une représentation des événements selon leur moment d'apparition :
- de modifier l'étendue des agrégats spatiaux utilisés pour constituer les sous-groupes d'événements sur lesquels sont effectuées des transformées de Fourier (Repère 6) ;
- d'afficher ou non les enveloppes convexes de ces agrégats sur la carte (Repère 7).

Les résultats renvoyés par la transformée de Fourier sont représentés dans l'interface de contrôle (Figure 10.10). Un clic sur les entités graphiques représentant les résultats trouvés applique la valeur correspondante à la durée de l'échelle cyclique.

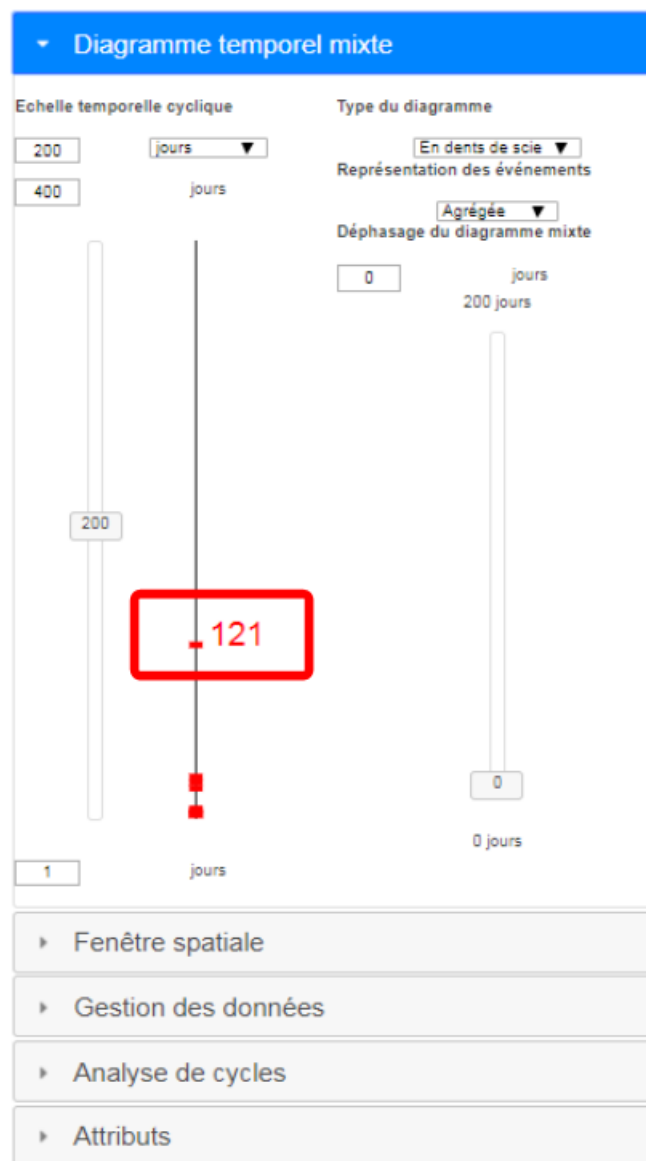


FIGURE 10.10 – Représentation des résultats de la transformée de Fourier dans l'interface de contrôle

10.7 Modification de la représentation graphique des événements

10.7.1 Diagramme temporel circulaire

Une fois les événements importés dans l'environnement, le diagramme circulaire est visible dans la fenêtre temporelle (Figure 10.11). Le nombre de sections du diagramme circulaire est modifiable :

- au moyen d'un curseur (Repère 1).
- au travers d'une entrée textuelle (Repère 2) ;

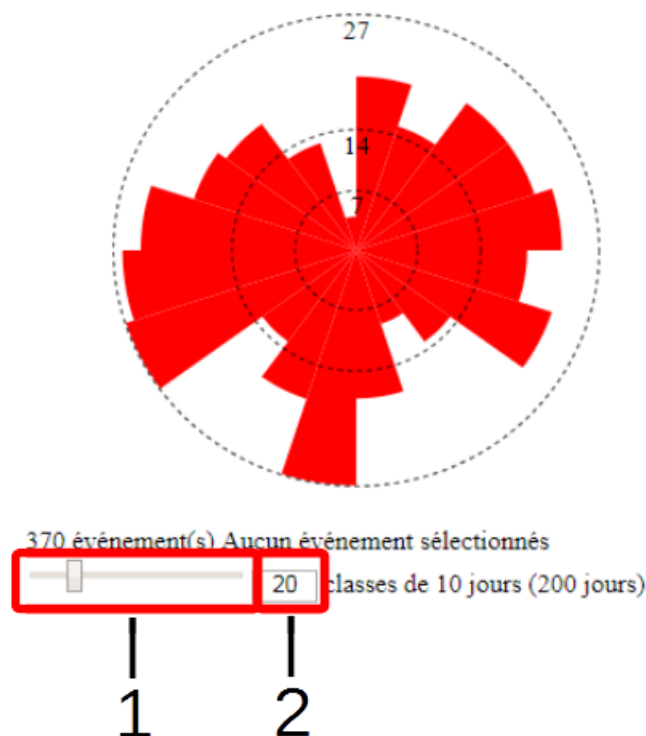


FIGURE 10.11 – Diagramme temporel circulaire de *GrAPHiST*

10.7.2 Représentation des événements dans le diagramme mixte

L'interface de contrôle permet de choisir si les entités graphiques du diagramme mixte représentent des événements de manière individuelle, ou représentent des agrégats d'événements temporellement proches (Repère 1 de la Figure 10.12). Pour que le changement soit appliqué, il est nécessaire de forcer la recréation de la courbe du diagramme temporel. Pour cela, il suffit de cliquer sur le champ de texte indiqué par le Repère 2 de la Figure 10.12, et d'appuyer sur *Entrée*.

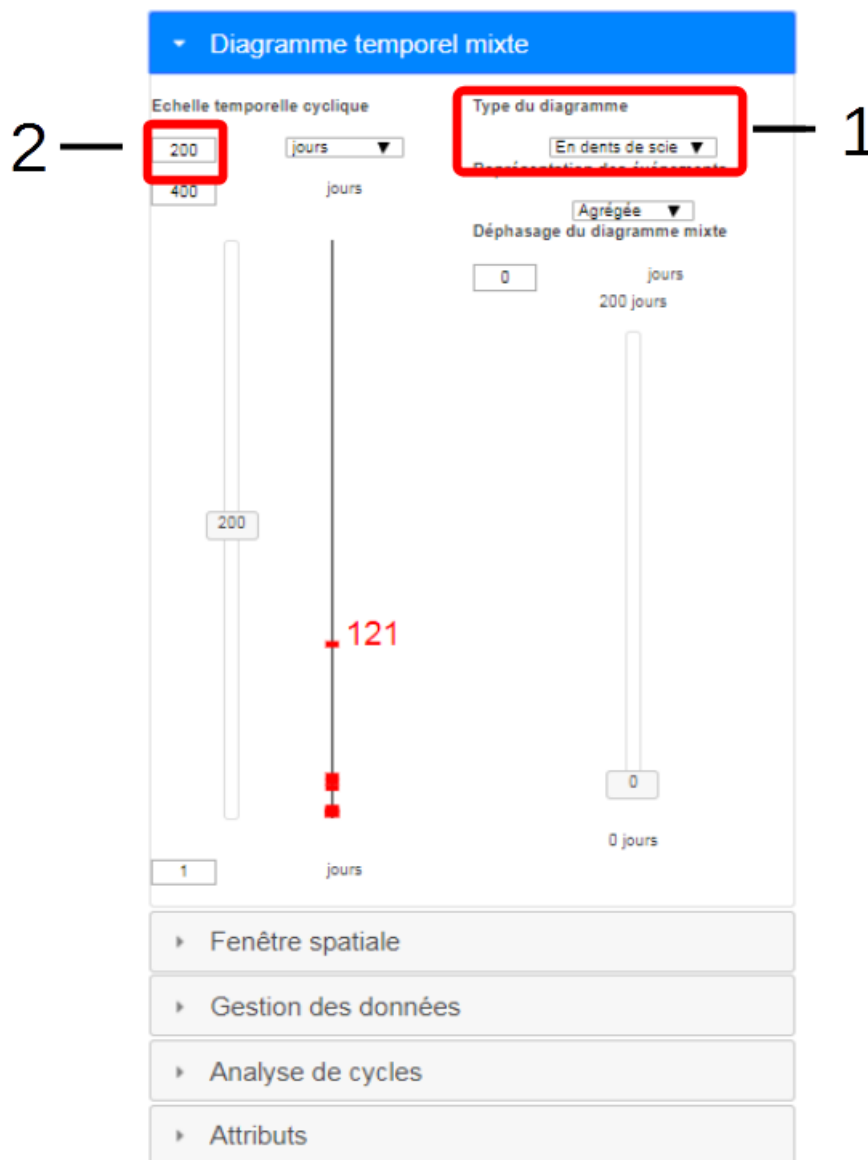


FIGURE 10.12 – Modification de la représentation temporelle des événements

L'interface de contrôle permet également de modifier les critères de représentation des agrégats. L'utilisateur peut modifier :

- l'étendue temporelle couverte par les agrégats du diagramme mixte (Repère 1 de la Figure 10.13) ;
- le nombre d'événements agrégés à partir duquel la taille de l'entité graphique représentant l'agrégat est calculée de manière logarithmique (Repère 2 de la Figure 10.13). Pour que le changement soit appliqué, il est nécessaire de forcer la recréation de la courbe du diagramme temporel. Pour cela, il suffit de cliquer sur le champ de texte indiqué par le Repère 2 de la Figure 10.12, et d'appuyer sur *Entrée*.

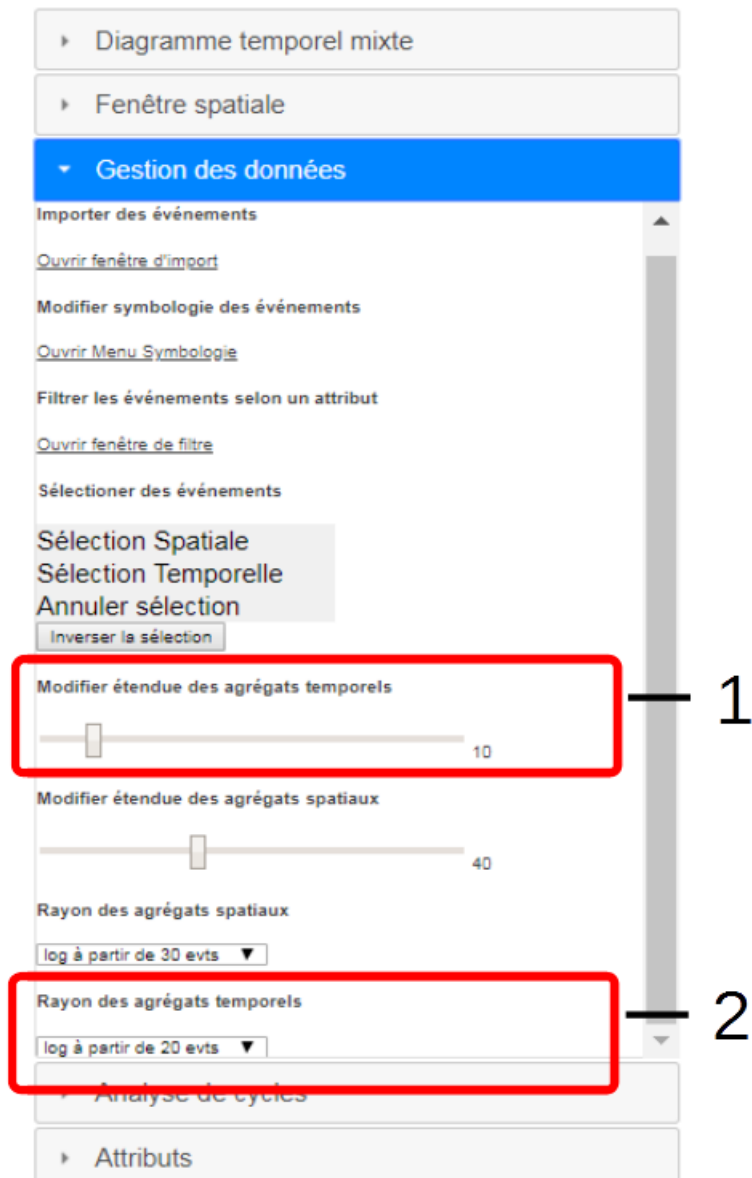


FIGURE 10.13 – Modification de la représentation des agrégats temporels

10.7.3 Représentation des événements dans la carte

L'interface de contrôle permet de choisir si les entités graphiques de la carte représentent des événements de manière individuelle, ou représentent des agrégats d'événements spatialement proches (Figure 10.14).

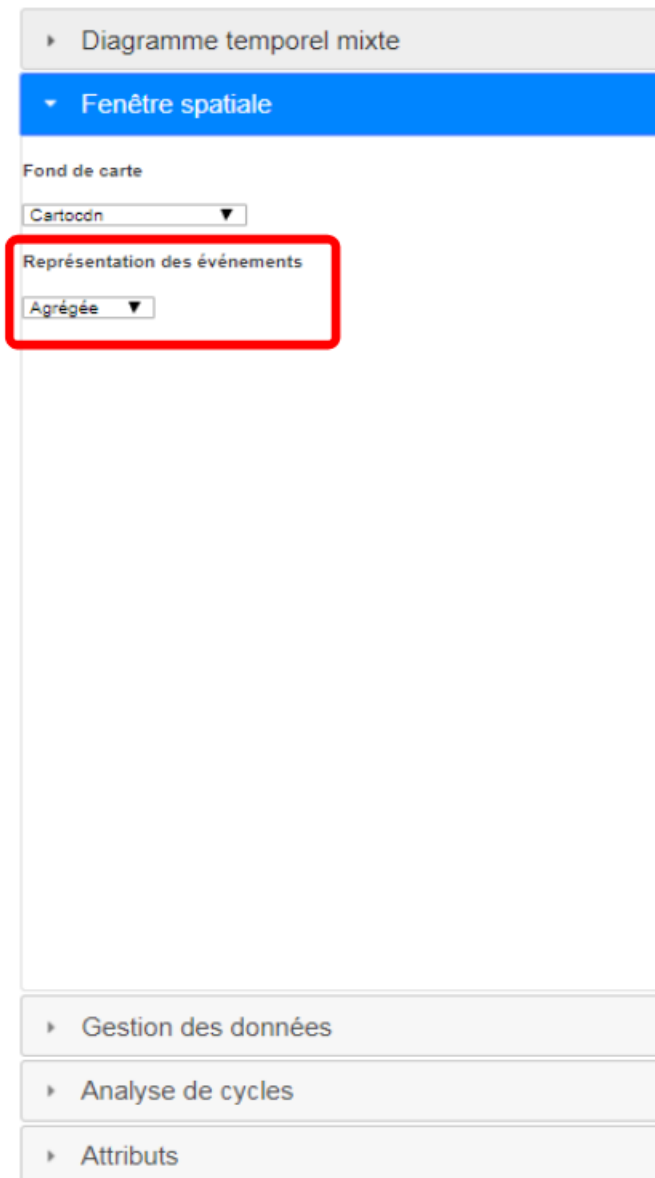


FIGURE 10.14 – Modification de la représentation spatiale des événements

L'interface de contrôle permet de modifier les critères de représentation des agrégats. L'utilisateur peut modifier :

- l'étendue spatiale couverte par les agrégats de la carte (Repère 1 de la Figure 10.15) ;
- le nombre d'événements agrégés à partir duquel la taille de l'entité graphique représentant l'agrégat est calculée de manière logarithmique (Repère 2 de la Figure 10.15). Pour que le changement soit appliqué, il est nécessaire de forcer la recréation des entités graphiques de la carte. Pour cela, il suffit d'effectuer un changement de l'échelle de la carte, au moyen des outils de zoom.

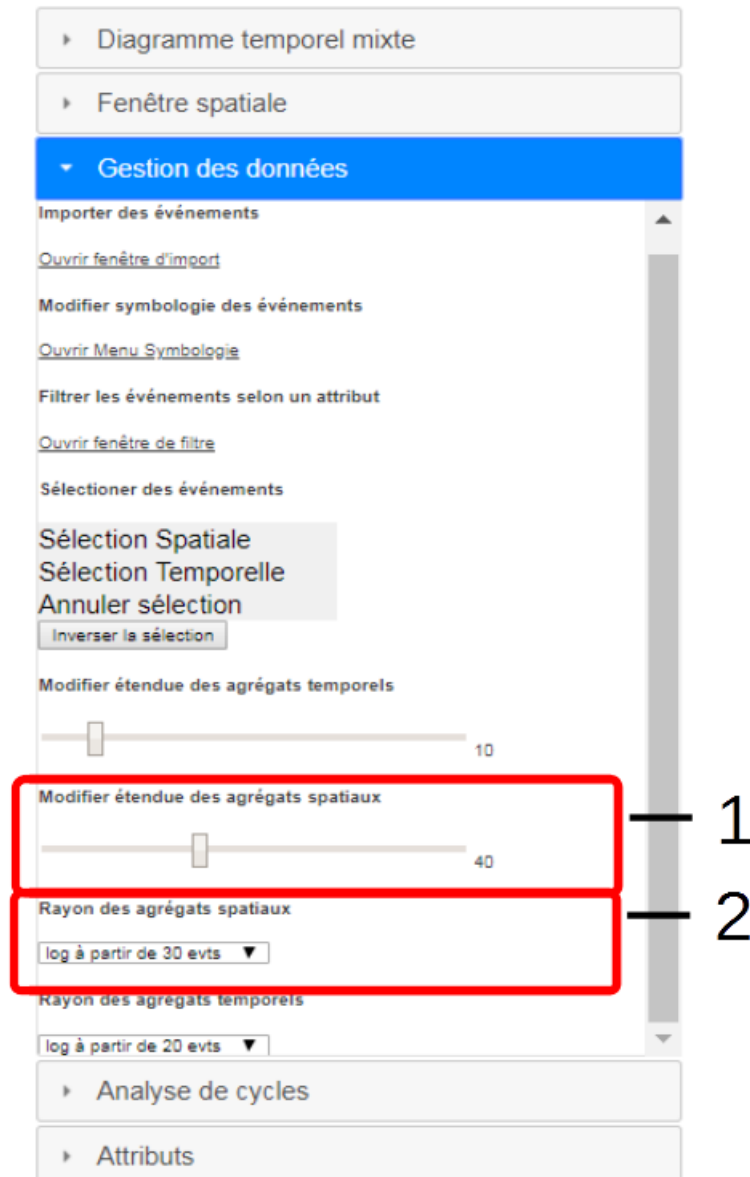


FIGURE 10.15 – Modification de la représentation des agrégats spatiaux

10.8 Sélection des événements

La sélection des événements s'effectue :

- en choisissant un outil de sélection dans l'interface de contrôle (Repère 1 de la Figure 10.16) ;
- en sélectionnant les événements au moyen d'un cliquer-glisser dans la fenêtre correspondante (la sélection présentée ici est une sélection spatiale, Repère 2 de la Figure 10.16).

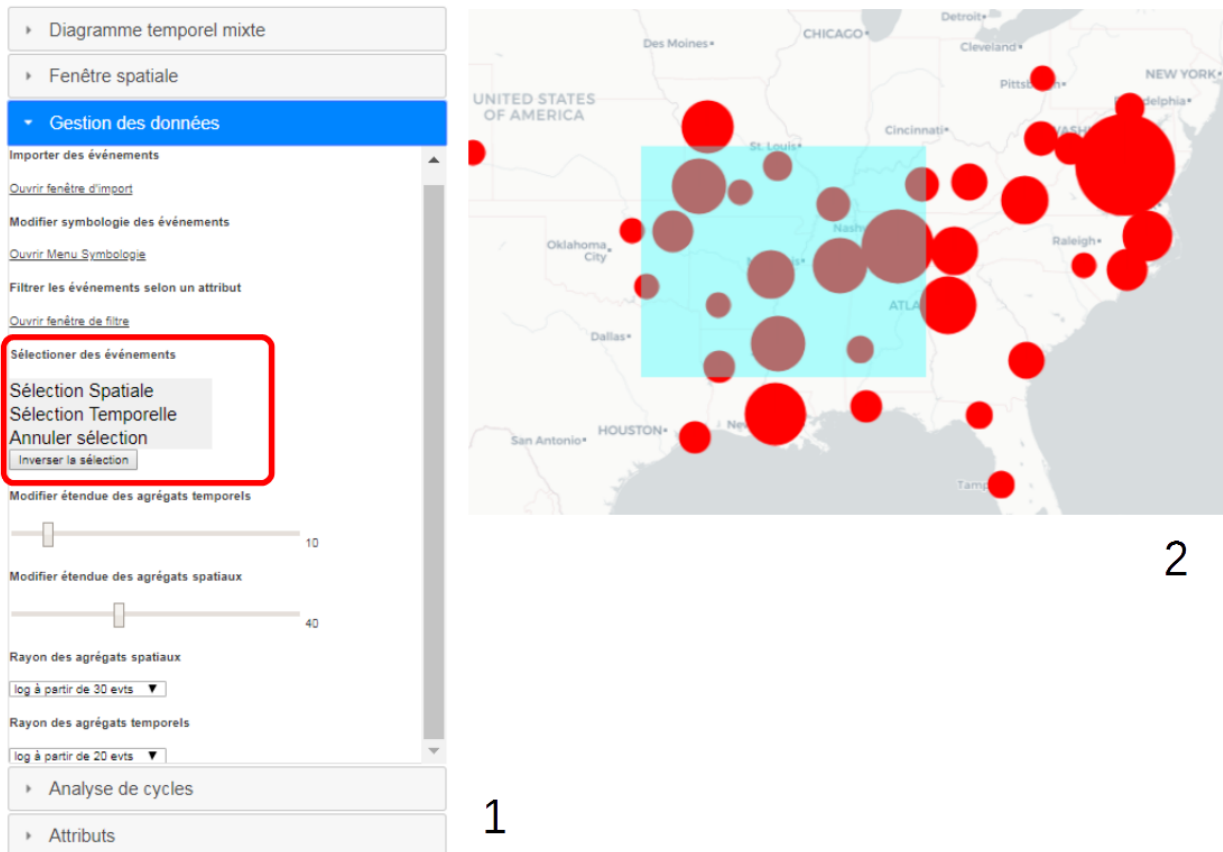


FIGURE 10.16 – Sélection des événements dans *GrAPHiST*

Une fois des événements sélectionnés, leur composante thématique est présentée de manière tabulaire dans l’interface de contrôle (Figure 10.17), et l’utilisateur peut masquer les événements non-sélectionnés dans les différentes fenêtres de l’interface, au moyen de boutons situés sous le diagramme temporel circulaire (Figure 10.18).

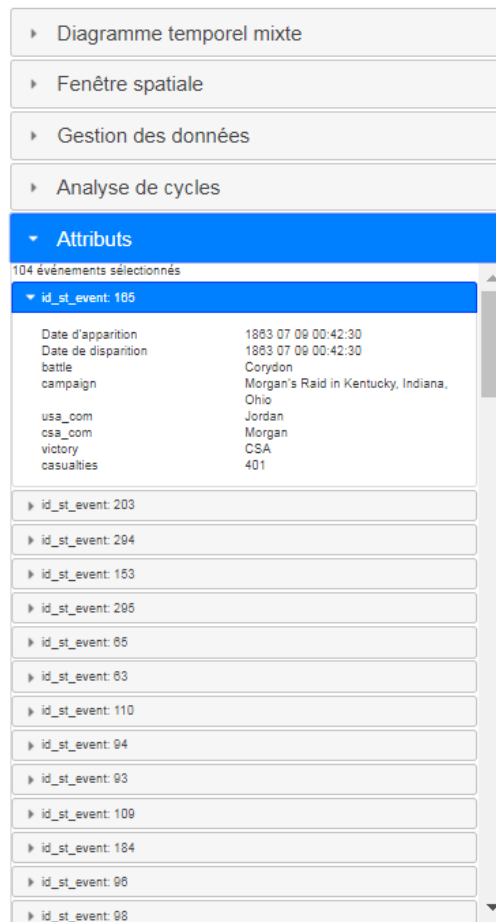


FIGURE 10.17 – Représentation de la composante thématique des événements dans *GrAPHiST*

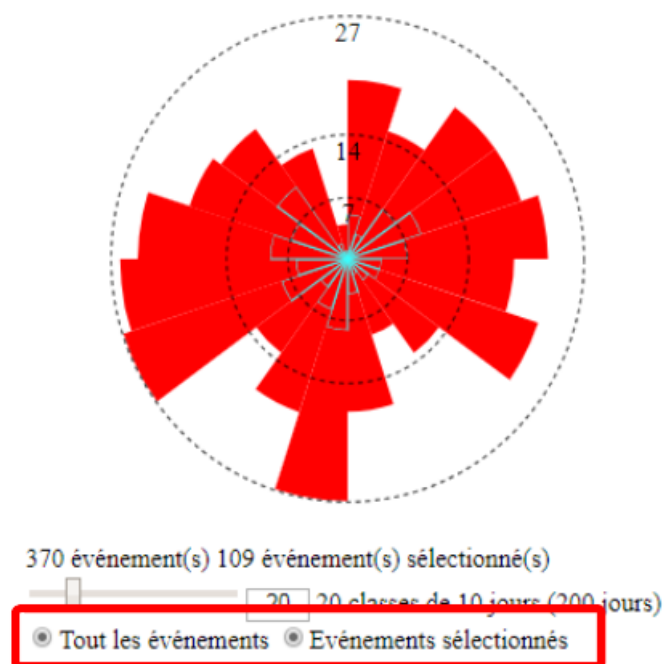


FIGURE 10.18 – Outils de *brushing* dans *GrAPHiST*

10.9 Utilisation de la symbologie

Via l'interface de contrôle (Repère 1 de la Figure 10.19), l'utilisateur ouvre une fenêtre permettant d'utiliser les règles de symbologie pour représenter les événements selon leur lieu ou leur moment d'apparition (Repère 2 de la Figure 10.19). Cette fenêtre permet d'opter pour une représentation des événements selon :

- leur lieu d'apparition (Repère 3 de la Figure 10.19) ;
- leur moment d'apparition relatif à une échelle linéaire (Repère 4 de la Figure 10.19) ;
- leur moment d'apparition relatif à une échelle cyclique (Repère 5 de la Figure 10.19).

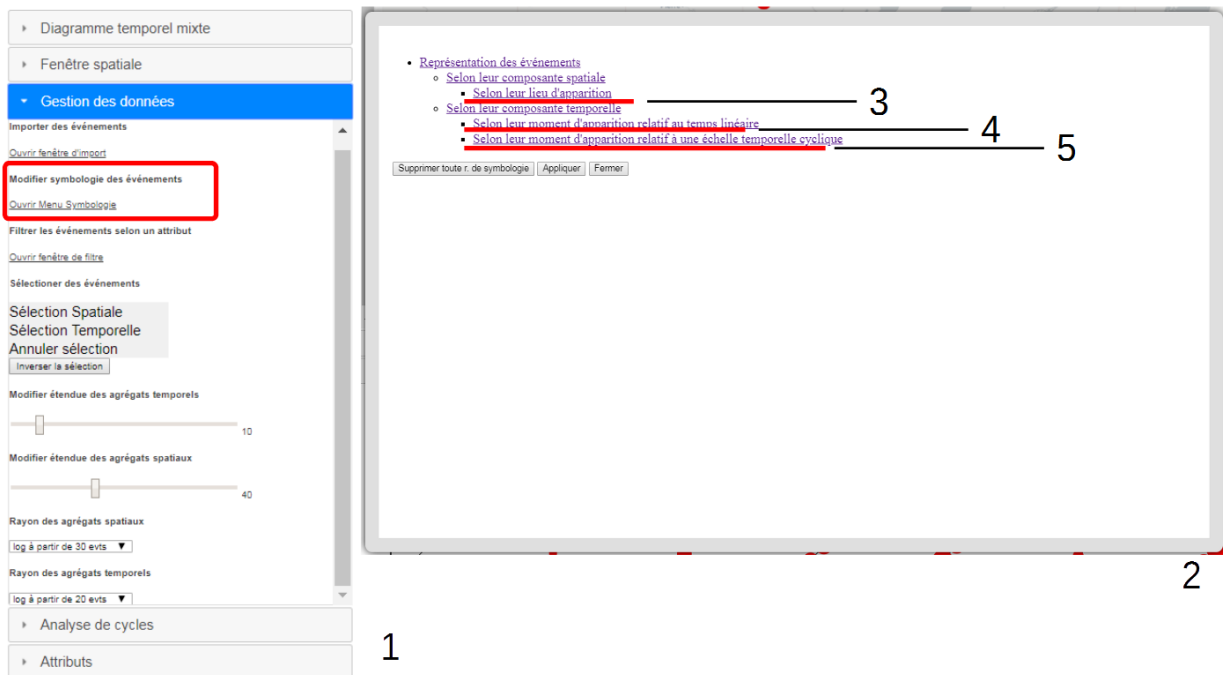


FIGURE 10.19 – Fenêtre de symbologie

10.9.1 Représentation des événements selon leur lieu d'apparition

Lorsqu'une représentation des événements selon leur lieu d'apparition est choisie, l'utilisateur doit créer des zones spatiales permettant de classer les événements. Ces zones peuvent être :

- créées à la volée par l'utilisateur ;
Pour cela, l'utilisateur clique sur "Ajouter Zone" (Repère 1 de la Figure 10.20). Cette action fait disparaître la fenêtre actuelle. L'utilisateur peut alors créer sur la carte une zone polygonale au moyen de la souris (Repère 2 de la Figure 10.20). Une fois la zone créée (double-clic au moyen de la souris), la fenêtre réapparaît.
- importées au moyen d'un *geojson* ;
L'utilisateur copie un *geojson* (au moyen d'un copier-coller) dans le champ de texte indiqué par le Repère 3. L'utilisateur charge ensuite le *geojson* au moyen du bouton indiqué par le Repère 4. Le *geojson* doit être au format présenté dans le fichier *default.geojson*, envoyé avec ce document.

Les zones créées à la volée peuvent également être récupérées sous un format *geojson*, en cliquant sur le bouton indiqué par le Repère 5. Le *geojson* créé est affiché dans le champ de texte indiqué par le Repère 3, et peut être récupéré par un copier/coller.

Les zones créées peuvent être supprimées (Repère 6 de la Figure 10.20). Le bouton indiqué par le Repère 7 permet de masquer le contour de ces zones dans la carte.

Les événements de chaque zone peuvent être masqués dans les différentes fenêtres de l'interface. Pour cela, il suffit de décocher les *checkbox* (Repère 8 de la Figure 10.20), correspondant aux zones dont les événements doivent être masqués.

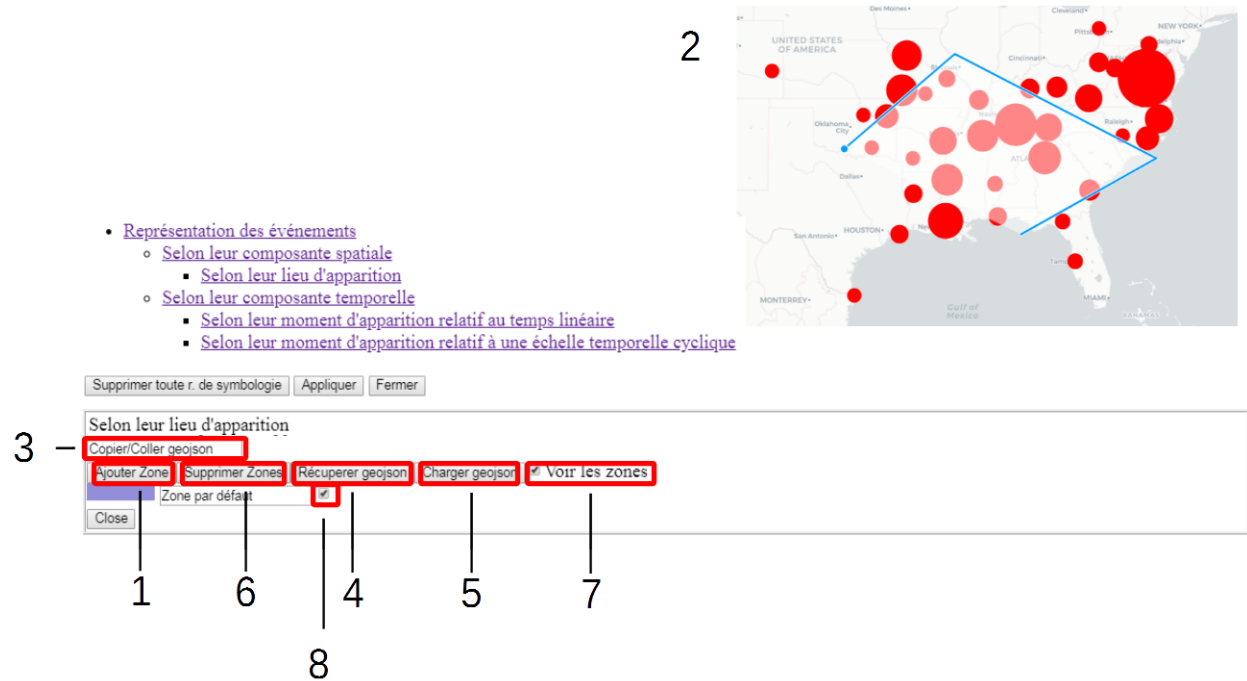


FIGURE 10.20 – Représentation des événements selon leur lieu d'apparition

Les couleurs affectées à chaque zone peuvent être modifiées en cliquant sur la vignette correspondante (Repère 1 de la Figure 10.21), puis en sélectionnant la couleur choisie dans une palette de couleur (Repère 2 de la Figure 10.21).

- Représentation des événements
 - Selon leur composante spatiale
 - Selon leur lieu d'apparition
 - Selon leur composante temporelle
 - Selon leur moment d'apparition relatif au temps linéaire
 - Selon leur moment d'apparition relatif à une échelle temporelle cyclique

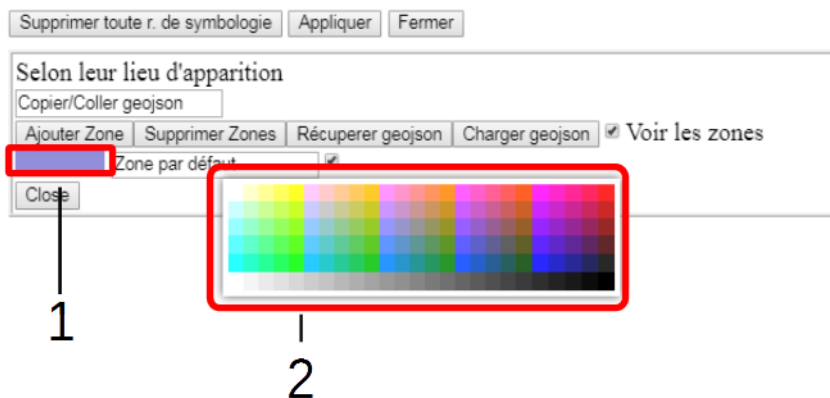


FIGURE 10.21 – Modification des couleurs affectées aux zones de classification spatiale

10.9.2 Représentation des événements selon leur moment d'apparition relatif à une échelle linéaire

Lorsqu'une représentation des événements selon leur moment d'apparition relatif à une échelle linéaire est choisie, l'utilisateur doit choisir le nombre d'intervalles temporels utilisés pour classer les événements (Figure 10.22).

- [Représentation des événements](#)
 - [Selon leur composante spatiale](#)
 - [Selon leur lieu d'apparition](#)
 - [Selon leur composante temporelle](#)
 - [Selon leur moment d'apparition relatif au temps linéaire](#)
 - [Selon leur moment d'apparition relatif à une échelle temporelle cyclique](#)

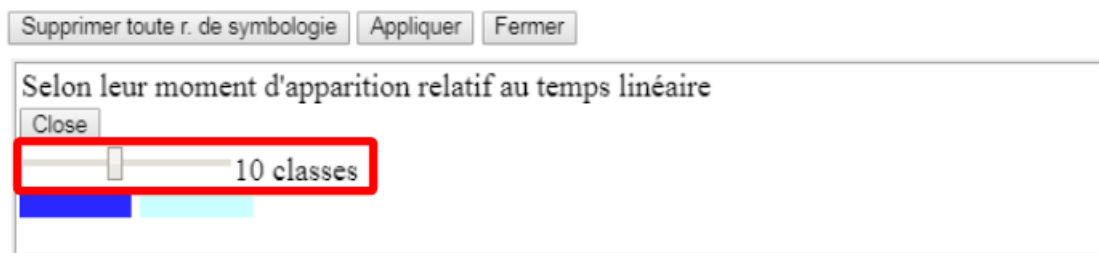


FIGURE 10.22 – Représentation des événements selon leur moment d'apparition relatif à une échelle temporelle linéaire

10.9.3 Représentation des événements selon leur moment d'apparition relatif à une échelle cyclique

Lorsqu'une représentation des événements selon leur moment d'apparition relatif à une échelle cyclique est choisie, l'utilisateur doit d'abord choisir le nombre de couleurs utilisées pour représenter les événements (Figure 10.23).

- [Représentation des événements](#)
 - [Selon leur composante spatiale](#)
 - [Selon leur lieu d'apparition](#)
 - [Selon leur composante temporelle](#)
 - [Selon leur moment d'apparition relatif au temps linéaire](#)
 - [Selon leur moment d'apparition relatif à une échelle temporelle cyclique](#)

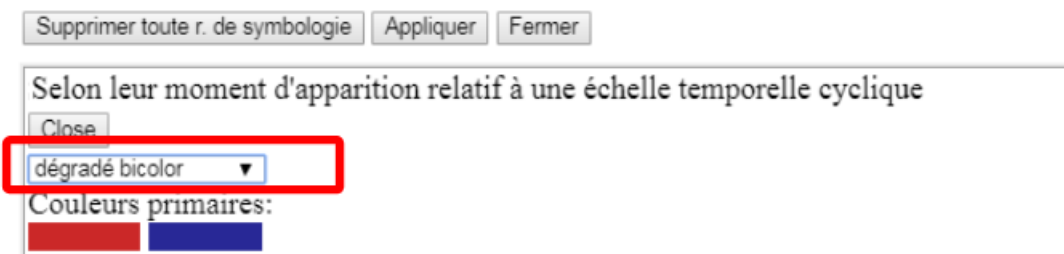


FIGURE 10.23 – Représentation des événements selon leur moment d'apparition relatif à une échelle temporelle cyclique

L'utilisateur choisi ensuite le nombre d'intervalles temporels utilisés pour classer les événements, en modifiant le nombre de sections utilisées dans le diagramme temporel circulaire (Figure 10.24).

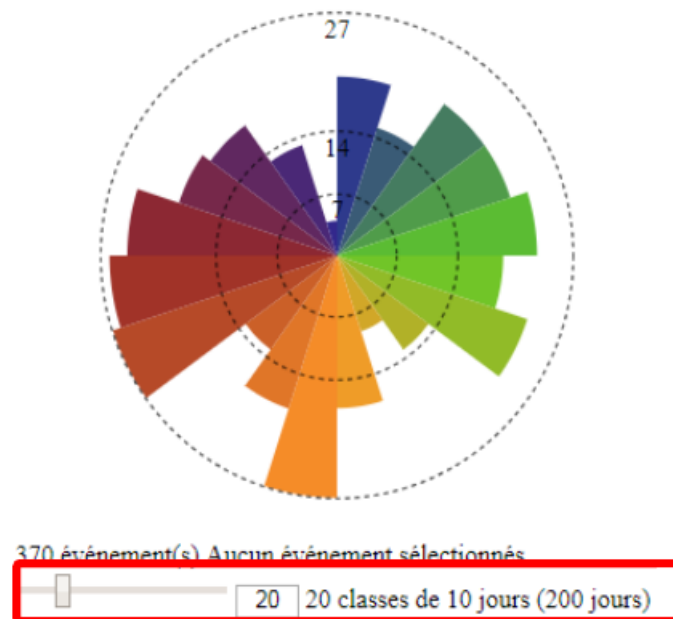


FIGURE 10.24 – Choix du nombre d'intervalles temporels utilisés pour classer les événements

Bibliographie

- [Aigner et al., 2007] Aigner, W., Miksch, S., Müller, W., Schumann, H., and Tominski, C. (2007). Visualizing time oriented data a systematic view. *Computers & Graphics*, 31(3) :401–409.
- [Aigner et al., 2011] Aigner, W., Miksch, S., Schumann, H., and Tominski, C. (2011). *Visualization of time oriented data*. Springer Science & Business Media.
- [Allen, 1983] Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11) :832–843.
- [Allen, 1984] Allen, J. F. (1984). Towards a general theory of action and time. *Artificial intelligence*, 23(2) :123–154.
- [Allen, 1995] Allen, R. (1995). Timelines as information system interfaces. In *Proceedings International Symposium on Digital Libraries*, pages 175–180.
- [Andrienko et al., 2011] Andrienko, G., Andrienko, N., Bak, P., Keim, D., Kisilevich, S., and Wrobel, S. (2011). A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages & Computing*, 22(3) :213–232.
- [Andrienko and Andrienko, 2004] Andrienko, N. and Andrienko, G. (2004). Interactive visual tools to explore spatio-temporal variation. In *Proceedings of the working conference on Advanced visual interfaces*, pages 417–420. ACM.
- [Andrienko and Andrienko, 2006] Andrienko, N. and Andrienko, G. (2006). *Exploratory analysis of spatial and temporal data : a systematic approach*. Springer Science & Business Media.
- [Andrienko et al., 2003a] Andrienko, N., Andrienko, G., and Gatalsky, P. (2003a). Exploratory spatio-temporal visualization : an analytical review. *Journal of Visual Languages & Computing*, 14(6) :503–541.
- [Andrienko et al., 2003b] Andrienko, N., Andrienko, G., and Gatalsky, P. (2003b). Visual data exploration using space-time cube. In *Proceedings of the 21st International Cartographic Conference*, pages 1981–1983.
- [Antoni et al., 2004] Antoni, J.-P., Klein, O., and Moisy, S. (2004). Cartographie interactive et multimédia : vers une aide à la réflexion géographique. *Cybergeo : European Journal of Geography*.
- [Armstrong, 1988] Armstrong, M. P. (1988). Temporality in spatial databases.
- [Arnaud and Davoine, 2009] Arnaud, A. and Davoine, P.-A. (2009). Cartographie des temporalités dans le domaine des risques. In *24th International Cartographic Conference, ICC09, Chile, November*.
- [Bak et al., 2009] Bak, P., Mansmann, F., Janetzko, H., and Keim, D. (2009). Spatiotemporal analysis of sensor logs using growth ring maps. *IEEE transactions on visualization and computer graphics*, 15(6).
- [Beard et al., 1990] Beard, D., Palaniappan, M., Humm, A., Banks, D., Nair, A., and Shan, Y.-P. (1990). A visual calendar for scheduling group meetings. In *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, pages 279–290. ACM.
- [Beard et al., 2008] Beard, K., Deese, H., and Pettigrew, N. R. (2008). A framework for visualization and exploration of events. *Information Visualization*, 7(2) :133–151.

- [Bennett and Fellbaum, 2006] Bennett, B. and Fellbaum, C. (2006). On what goes on : The ontology of processes and events. In *Formal Ontology in Information Systems : Proceedings of the Fourth International Conference (FOIS 2006)*, volume 150, page 4. IOS Press.
- [Berthold and Hand, 2007] Berthold, M. R. and Hand, D. J. (2007). *Intelligent data analysis : an introduction*. Springer.
- [Bertin, 1973] Bertin, J. (1973). *Sémiologie graphique*. Flammarion, Paris.
- [Bittner et al., 2004] Bittner, T., Donnelly, M., and Smith, B. (2004). Endurants and perdurants in directly depicting ontologies. *AI Communications*, 17(4) :247–258.
- [Bohm, 1980] Bohm, D. (1980). 1980 wholeness and the implicate order, london : Routledge and kegan paul.
- [Bonfim and Medeiros, 2008] Bonfim, C. and Medeiros, Z. (2008). Epidemiologia e geografia : dos primordios ao geoprocessamento. *Espaca Saude*, 10(1) :53–62.
- [Bony, 2001] Bony, J.-M. (2001). *Cours d'analyse : théorie des distributions et analyse de Fourier*. Editions Ecole Polytechnique.
- [Boyandin et al., 2012] Boyandin, I., Bertini, E., and Lalanne, D. (2012). A qualitative study on the exploration of temporal changes in flow maps with animation and small-multiples. In *Computer Graphics Forum*, volume 31, pages 1005–1014. Wiley Online Library.
- [Bruggmann et al., 2013] Bruggmann, A., Salvini, M. M., and Fabrikant, S. (2013). Cartograms of self-organizing maps to explore user-generated content. In *26th international cartographic conference*, pages 25–30.
- [Card et al., 1999] Card, S. K., Mackinlay, J. D., and Shneiderman, B. (1999). *Readings in information visualization : using vision to think*. Morgan Kaufmann.
- [Carlis and Konstan, 1998] Carlis, J. V. and Konstan, J. A. (1998). Interactive visualization of serial periodic data. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pages 29–38. ACM.
- [Castner, 1983] Castner, H. W. (1983). Research questions and cartographic design. *Graphic communication and design in contemporary cartography*. Chichester, UK : John Wiley & Sons, pages 87–113.
- [Catanas, 2002] Catanas, M. (2002). L'expérimentation dans la recherche en sciences sociales.
- [Cauvin et al., 2008] Cauvin, C., Escobar, F., and Serradj, A. (2008). Cartographie thématique 5. des voies nouvelles à explorer, paris, ed. hermes science, 320 pp. Technical report, ISBN 978-2-7462-2208-6.
- [Cécile Saint-Marc, 2012] Cécile Saint-Marc, P.-A. D. (2012). Évaluation de l'aléa volcanique du piton de la fournaise (île de la réunion) l'impact des coulées de lave. https://www.esrifrance.fr/iso_album/p15_mapbook2012.pdf.
- [Chesneau, 2006] Chesneau, E. (2006). Modèle d'amélioration automatique des contrastes de couleurs en cartographie-application aux cartes de risque. *Mémoire de thèse de doctorat en sciences de l'information géographique, IGN et Université de Marne-la-Vallée*.
- [Cheylan and Lardon, 1993] Cheylan, J.-P. and Lardon, S. (1993). Towards a conceptual data model for the analysis of spatio-temporal processes : the example of the search for optimal grazing strategies. *Spatial Information Theory A Theoretical Basis for GIS*, pages 158–176.
- [Cheylan et al., 1994] Cheylan, J.-P., Lardon, S., Mathian, H., and Sanders, L. (1994). Les problématiques liées au temps dans les sig. *Revue internationale de géomatique*, 4(3-4) :287–305.
- [Chi et al., 1998] Chi, E. H., Pitkow, J., Mackinlay, J., Pirolli, P., Gossweiler, R., and Card, S. K. (1998). Visualizing the evolution of web ecologies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 400–407. ACM Press/Addison-Wesley Publishing Co.

- [Choquet and Iksal, 2007] Choquet, C. and Iksal, S. (2007). Modélisation et construction de traces d'utilisation d'une activité d'apprentissage : une approche langage pour la réingénierie d'un eia. *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF)*, 14 :24–pages.
- [City of Chicago, 2017] City of Chicago, U. (2017). Chicago data portal. <https://data.cityofchicago.org/>.
- [Claramunt et al., 1998] Claramunt, C., Parent, C., and Thériault, M. (1998). Design patterns for spatio-temporal processes. In *Data Mining and Reverse Engineering*, pages 455–475. Springer.
- [Claramunt and Thériault, 1995] Claramunt, C. and Thériault, M. (1995). Managing time in gis : an event-oriented approach. *Recent Advances in Temporal Databases*, pages 23–42.
- [Claramunt et al., 1997] Claramunt, C., Thériault, M., and Parent, C. (1997). A qualitative representation of evolving spatial entities in two-dimensional topological spaces.
- [Cleveland, 1993] Cleveland, W. S. (1993). *Visualizing data*. Hobart Press.
- [Couclelis, 1992] Couclelis, H. (1992). People manipulate objects (but cultivate fields) : beyond the raster-vector debate in gis. *Theories and methods of spatio-temporal reasoning in geographic space*, pages 65–77.
- [Davoine, 2014] Davoine, P. (2014). Contributions géomatiques pour la gestion des risques naturels : modélisation, géovisualisation, acquisition.
- [Davoine et al., 2004] Davoine, P., Gensel, J., and Martin, H. (2004). A web-based multimedia framework for diffusing spatio-temporal information : Application to natural hazards. In *12th International Conference on Geoinformatics*, pages 149–156.
- [Davoine, 2016] Davoine, P.-A. (2016). Géovisualisation. https://ecolemagis2016.sciencesconf.org/conference/ecolemagis2016/pages/ecole_MAGIS_2016_Geovisualisation_.pdf.
- [Davoine et al., 2001] Davoine, P.-A., Brunet, R., Clavandier, G., Charrier, P., Favier, R., and Martin, H. (2001). Le projet sidira : une approche pluri-disciplinaire pour une meilleure appropriation des risques naturels via les nouvelles technologies de l'information. In *Actes du colloque-SIRNAT 2001*.
- [Davoine et al., 2012a] Davoine, P.-A., Moïsuc, B., and Gensel, J. (2012a). Genghis : un environnement pour la génération d'interfaces de visualisation spatio-temporelle.
- [Davoine et al., 2012b] Davoine, P.-A., Saint-Marc, C., Dimuro, A., Staudacher, T., Villanova-Oliver, M., and Boissier, P. (2012b). Constructions cartographiques pour la caractérisation de l'aléa volcanologique du piton de la fournaise. In *Actes de la Conférence internationale de Géomatique et Analyse Spatiale 2012*, page 129.
- [DiBiase et al., 1992] DiBiase, D., MacEachren, A. M., Krygier, J. B., and Reeves, C. (1992). Animation and the role of map design in scientific visualization. *Cartography and geographic information systems*, 19(4) :201–214.
- [Dragicevic and Huot, 2002] Dragicevic, P. and Huot, S. (2002). Spiraclock : a continuous and non-intrusive display for upcoming events. In *CHI'02 extended abstracts on Human factors in computing systems*, pages 604–605. ACM.
- [Dukaczewski, 2005] Dukaczewski, D. (2005). Entities–cartotropic method of selection of static and dynamic variables for temporal cartographic animations. In *Proceedings of the XXII rd ACI/ICA International Cartographic Conference*.
- [Edsall and Peuquet, 1997] Edsall, R. and Peuquet, D. (1997). A graphical user interface for the integration of time into gis. In *Proceedings of the 1997 American congress of surveying and mapping annual convention and exhibition, Seattle, WA*, pages 182–189.

- [Edsall et al., 1997] Edsall, R. M., Kraak, M.-J., MacEachren, A. M., and Peuquet, D. J. (1997). Assessing the effectiveness of temporal legends in environmental visualization. In *GIS/LIS*, volume 97, pages 28–30.
- [Elmasri and Lee, 1998] Elmasri, R. and Lee, J. Y. (1998). Implementation options for time-series data. In *Temporal Databases : Research and Practice*, pages 115–128. Springer.
- [Elmqvist et al., 2010] Elmqvist, N., Riche, Y., Henry-Riche, N., and Fekete, J.-D. (2010). Mélange : Space folding for visual exploration. *IEEE transactions on visualization and computer graphics*, 16(3) :468–483.
- [État français, 2017a] État français, Ministère de l’Écologie, d. D. d. e. d. l. (2017a). Base de données historiques sur les inondations. <http://bdhi.fr/appli/web/welcome>.
- [État français, 2017b] État français, D. i. à. s. r. (2017b). Observatoire national interministériel de la sécurité routière. <http://www.securite-routiere.gouv.fr/la-securite-routiere/1-observatoire-national-interministeriel-de-la-securite-routiere>.
- [État français, 2017c] État français, E. (2017c). data.gouv.fr. <http://www.data.gouv.fr/fr/>.
- [Fanea et al., 2005] Fanea, E., Carpendale, S., and Isenberg, T. (2005). An interactive 3d integration of parallel coordinates and star glyphs. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 149–156. IEEE.
- [Fonseca et al., 2002] Fonseca, F. T., Egenhofer, M. J., Agouris, P., and Câmara, G. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3) :231–257.
- [Forbus, 1984] Forbus, K. D. (1984). Qualitative process theory. *Artificial intelligence*, 24(1) :85–168.
- [Forlines and Wittenburg, 2010] Forlines, C. and Wittenburg, K. (2010). Wakame : sense making of multi-dimensional spatial-temporal data. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 33–40. ACM.
- [Frank, 2007] Frank, A. U. (2007). Ontology. In *Encyclopedia of Geographic Information Science*. (K.Kemp Ed.) Thousand Oaks, CA, Sage Publications.
- [Gadia and Vaishnav, 1985] Gadia, S. K. and Vaishnav, J. H. (1985). A query language for a homogeneous temporal database. In *Proceedings of the fourth ACM SIGACT-SIGMOD symposium on Principles of database systems*, pages 51–56. ACM.
- [Gadia and Yeung, 1988] Gadia, S. K. and Yeung, C.-S. (1988). A generalized model for a relational temporal database. In *ACM SIGMOD Record*, volume 17, pages 251–259. ACM.
- [Galton, 1990] Galton, A. (1990). A critical examination of allen’s theory of action and time. *Artificial intelligence*, 42(2-3) :159–188.
- [Galton, 2001] Galton, A. (2001). Space, time, and the representation of geographical reality. *Topoi*, 20(2) :173–187.
- [Galton, 2003] Galton, A. (2003). Integrating fields and objects in geographic information science. In *Workshop on Fundamental Issues in Spatial and Geographic Ontologies, Switzerland*.
- [Galton, 2004] Galton, A. (2004). Fields and objects in space, time, and space-time. *Spatial cognition and computation*, 4(1) :39–68.
- [Galton, 2006] Galton, A. (2006). On what goes on : The ontology of processes and events. *Frontiers in Artificial Intelligence and Applications*, 150 :4.
- [Galton, 2008] Galton, A. (2008). Experience and history : Processes and their relation to events. *Journal of Logic and Computation*, 18(3) :323–340.
- [Galton and Worboys, 2005] Galton, A. and Worboys, M. (2005). Processes and events in dynamic geo-networks. *GeoS*, 3799 :45–59.

- [Gasquet and Witomski, 2000] Gasquet, C. and Witomski, P. (2000). *Analyse de Fourier et applications : filtrage, calcul numérique, ondelettes*. Dunod.
- [Gatalsky et al., 2004] Gatalsky, P., Andrienko, N., and Andrienko, G. (2004). Interactive analysis of event data using space-time cube. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pages 145–152. IEEE.
- [Gautier et al., 2016] Gautier, J., Davoine, P.-A., and Cunty, C. (2016). Helical time representation to visualize return-periods of spatio-temporal events. In *19th AGILE International Conference on Geographic Information Science*, Helsinki, Finland. <http://hal.univ-grenoble-alpes.fr/hal-01609156>.
- [Goodchild, 1992] Goodchild, M. F. (1992). Geographical data modeling. *Computers & Geosciences*, 18(4) :401–408.
- [Green, 1998] Green, M. (1998). Toward a perceptual science of multidimensional data visualization : Bertin and beyond. *ERGO/GERO Human Factors Science*, 8.
- [Grenon and Smith, 2004] Grenon, P. and Smith, B. (2004). Snap and span : Towards dynamic spatial ontology. *Spatial cognition and computation*, 4(1) :69–104.
- [Guo et al., 2006] Guo, D., Chen, J., MacEachren, A. M., and Liao, K. (2006). A visualization system for space-time and multivariate patterns (vis-stamp). *IEEE transactions on visualization and computer graphics*, 12(6) :1461–1474.
- [Haddad, 2009] Haddad, H. (2009). *Une approche pour supporter l'analyse qualitative des suites d'actions dans un environnement géographique virtuel et dynamique : l'analyse "What-If" comme exemple*. PhD thesis, Université Laval.
- [Hägerstrand, 1970] Hägerstrand, T. (1970). What about people in regional science? *Papers in regional science*, 24(1) :7–24.
- [Hall, 1998] Hall, R. (1998). Psychology world. *Online Tutorial*. DOI= <http://web.mst.edu/~psyworld/extraneous.htm>, 1.
- [Harris, 2000] Harris, R. L. (2000). *Information graphics : A comprehensive illustrated reference*. Oxford University Press.
- [Harrison et al., 1994] Harrison, B. L., Owen, R., and Baecker, R. M. (1994). Timelines : an interactive system for the collection and visualization of temporal data. In *Graphics Interface*, pages 141–141. CANADIAN INFORMATION PROCESSING SOCIETY.
- [Hawley, 2004] Hawley, K. (2004). Temporal parts. *Stanford Encyclopedia of Philosophy*.
- [Heckert and Filliben, 2003] Heckert, N. A. and Filliben, J. J. (2003). Nist/sematech e-handbook of statistical methods ; chapter 1 : Exploratory data analysis. Technical report.
- [Henderson and Hollingworth, 1998] Henderson, J. M. and Hollingworth, A. (1998). Eye movements during scene viewing : An overview. In *Eye guidance in reading and scene perception*, pages 269–293. Elsevier.
- [Hewitt, 1983] Hewitt, K. (1983). Interpretations of calamity from the viewpoint of human ecology.
- [Hornsby and Egenhofer, 2002] Hornsby, K. and Egenhofer, M. J. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, 36(1) :177–194.
- [Hornsby and Cole, 2007] Hornsby, K. S. and Cole, S. (2007). Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS*, 11(4) :555–573.
- [Hunter and Williamson, 1990] Hunter, G. J. and Williamson, I. P. (1990). The development of a historical digital cadastral database. *International Journal of Geographical Information System*, 4(2) :169–179.

- [Irina Fabrikant et al., 2008] Irina Fabrikant, S., Rebich-Hespanha, S., Andrienko, N., Andrienko, G., and Montello, D. R. (2008). Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes. *The Cartographic Journal*, 45(3) :201–215.
- [Janet, 1885] Janet, P. (1885). La crise philosophique. 1865. 5. *J. Lemaitre*.
- [Jansen, 2008] Jansen, B. J. (2008). The methodology of search log analysis. *Handbook of research on Web log analysis*, pages 100–123.
- [Jespersen, 2012a] Jespersen, H. (2012a). Appomattox campaign. https://commons.wikimedia.org/wiki/File:Appomattox_Campaign_Overview.png.
- [Jespersen, 2012b] Jespersen, H. (2012b). Grant’s overland campaign. <https://commons.wikimedia.org/wiki/File:Overland-Richmond.png>.
- [Jespersen, 2012c] Jespersen, H. (2012c). Western theater of the american civil war. http://www.cwmaps.com/freemaps/ACW_Western_Overview.png.
- [Kaddouri et al., 2014] Kaddouri, L., Blaise, J.-Y., Davoine, P.-A., Mathian, H., and Saint-Marc, C. (2014). État des lieux des représentations dynamiques des temporalités des territoires. Research report, UMR 7300 ESPACE ; UMR 3495 MAP - GAMSAU ; Laboratoire d’Informatique de Grenoble. <https://hal.inria.fr/hal-01383450>.
- [Kakoudakis and Theodoulidis, 1996] Kakoudakis, I. and Theodoulidis, B. (1996). The tau time model. manchester : University of manchester (umist), 1996. Technical report, 24p. Technical Report TR-96-5, TimeLab.
- [Karam, 1994] Karam, G. M. (1994). Visualization using timelines. In *Proceedings of the 1994 ACM SIGSOFT international symposium on Software testing and analysis*, pages 125–137. ACM.
- [Katznelson, 2004] Katznelson, Y. (2004). *An introduction to harmonic analysis*. Cambridge University Press.
- [Kauppinen et al., 2008] Kauppinen, T., Väättäin, J., and Hyvönen, E. (2008). Creating and using geospatial ontology time series in a semantic cultural heritage portal. *The Semantic Web : Research and Applications*, pages 110–123.
- [Keim, 1996] Keim, D. A. (1996). Pixel-oriented database visualizations. *ACM Sigmod Record*, 25(4) :35–39.
- [Keim, 1997] Keim, D. A. (1997). Visual techniques for exploring databases. In *Knowledge Discovery in Databases (KDD’97)*.
- [Kincaid et al., 1985] Kincaid, C. M., Dupont, P. B., and Kaye, A. R. (1985). Electronic calendars in the office : an assessment of user needs and current technology. *ACM Transactions on Information Systems (TOIS)*, 3(1) :89–102.
- [Kinkeldey et al., 2014] Kinkeldey, C., MacEachren, A. M., and Schiewe, J. (2014). How to assess visual communication of uncertainty? a systematic review of geospatial uncertainty visualisation user studies. *The Cartographic Journal*, 51(4) :372–386.
- [Klir, 1985] Klir, G. (1985). Architecture of general systems problem solving.
- [Köbben and Yaman, 1996] Köbben, B. and Yaman, M. (1996). Evaluating dynamic visual variables. In *Proceedings of the seminar on teaching animated cartography, ACI/ICA, Madrid*, pages 45–51.
- [Kraak, 2003a] Kraak, M.-J. (2003a). Geovisualization illustrated. *ISPRS journal of photogrammetry and remote sensing*, 57(5-6) :390–399.
- [Kraak, 2003b] Kraak, M.-J. (2003b). The space-time cube revisited from a geovisualization perspective. In *Proc. 21st International Cartographic Conference*, pages 1988–1996.

- [Kraak et al., 1997] Kraak, M.-J., Edsall, R., and MacEachren, A. M. (1997). Cartographic animation and legends for temporal maps : Exploration and or interaction. In *Proceedings of the 18th International Cartographic Conference*, volume 1, pages 253–261.
- [Kraak and Ormeling, 2011] Kraak, M.-J. and Ormeling, F. (2011). *Cartography : visualization of spatial data*. Guilford Press.
- [Kuhn, 2001] Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15(7) :613–631.
- [Kumar et al., 1998] Kumar, V., Furuta, R., and Allen, R. B. (1998). Metadata visualization for digital libraries : interactive timeline editing and review. In *Proceedings of the third ACM conference on Digital libraries*, pages 126–133. ACM.
- [Langran, 1992] Langran, G. (1992). Time in geographic information systems, technical issues in geographic information systems.
- [Langran and Chrisman, 1988] Langran, G. and Chrisman, N. R. (1988). A framework for temporal geographic information. *Cartographica : The International Journal for Geographic Information and Geovisualization*, 25(3) :1–14.
- [Larousse, 2016] Larousse, E. (2016). Encyclopédie larousse en ligne. *En ligne*.
- [Lefebvre, 1974] Lefebvre, H. (1974). La production de l’espace. *L’Homme et la société*, 31(1) :15–32.
- [Li, 2010] Li, X. (2010). The time wave in time space : a visual exploration environment for spatio-temporal data.
- [Livet et al., 2010] Livet, P., Muller, J.-P., Phan, D., and Sanders, L. (2010). Ontology, a mediator for agent-based modeling in social science. *Journal of Artificial Societies and Social Simulation*, 13(1) :3.
- [Livet and Nef, 2009] Livet, P. and Nef, F. (2009). *Les êtres sociaux*. Hermann, Paris.
- [MacEachren, 1994] MacEachren, A. M. (1994). Time as a cartographic variable. *Visualization in geographical information systems*, 243.
- [MacEachren, 2004] MacEachren, A. M. (2004). *How maps work : representation, visualization, and design*. Guilford Press.
- [Mackinlay et al., 1991] Mackinlay, J. D., Robertson, G. G., and Card, S. K. (1991). The perspective wall : Detail and context smoothly integrated. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 173–176. ACM.
- [Mackinlay et al., 1994] Mackinlay, J. D., Robertson, G. G., and DeLine, R. (1994). Developing calendar visualizers for the information visualizer. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pages 109–118. ACM.
- [Marey, 1885] Marey, E. (1885). La methode graphique dans les sciences experimentales, with supplement, le development de las methode graphique par la photographie, g. Paris : G. Masson.
- [Marey, 1878] Marey, E.-J. (1878). *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine*. G. Masson.
- [Mathian and Sanders, 2014] Mathian, H. and Sanders, L. (2014). *Approches spatio-temporelles. Objets géographiques et processus de changement*. ISTE Edition.
- [Mathieu, 2015] Mathieu, F. (2015). *Le Phénomène. : Synthèse philosophique*. Frederic Mathieu.
- [McCarthy and Hayes, 1969] McCarthy, J. and Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Readings in artificial intelligence*, pages 431–450.
- [McDermott, 1982] McDermott, D. (1982). A temporal logic for reasoning about processes and plans. *Cognitive science*, 6(2) :101–155.

- [McIntosh and Yuan, 2005] McIntosh, J. and Yuan, M. (2005). Assessing similarity of geographic processes and events. *Transactions in GIS*, 9(2) :223–245.
- [Mennis et al., 2000] Mennis, J. L., Peuquet, D. J., and Qian, L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science*, 14(6) :501–520.
- [Minard, 1980] Minard, C. J. (1980). *Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813*. Graphics Press.
- [Mintz et al., 1997] Mintz, D., Fitz-Simons, T., and Wayland, M. (1997). Tracking air quality trends with sas/graph. In *Proceedings of the 22nd Annual SAS User Group International Conference (SUGI 97)*, pages 807–812.
- [Mitchell, 1999] Mitchell, A. (1999). *The ESRI guide to GIS analysis : geographic patterns & relationships*, volume 1. ESRI, Inc.
- [Moellering, 1988] Moellering, H. (1988). The proposed standard for digital cartographic data : report of the digital cartographic data standards task force. *The American Cartographer*, 15(1) :11–31.
- [Morrill, 1970] Morrill, R. L. (1970). *The spatial organization of society*. Duxbury Press.
- [Mourelatos, 1978] Mourelatos, A. P. (1978). Events, processes, and states. *Linguistics and philosophy*, 2(3) :415–434.
- [Müller and Schumann, 2003] Müller, W. and Schumann, H. (2003). Visualization for modeling and simulation : visualization methods for time-dependent data-an overview. In *Proceedings of the 35th conference on Winter simulation : driving innovation*, pages 737–745. Winter Simulation Conference.
- [Nicolas et al., 2014] Nicolas, S., Gyselinck, V., Vergilino-Perez, D., and Doré-Mazars, K. (2014). *Introduction à la psychologie cognitive*. In Press.
- [Nightingale, 1858] Nightingale, F. (1858). Diagram of the causes of mortality in the army in the east. *Public domain*. Retrieved February, 21 :2015.
- [NOAA, 2016] NOAA, NWS, P. T. W. C. (2016). Animated map : all earthquakes of the past 15 years. <https://www.youtube.com/watch?v=Ed0tG1fJHiY>.
- [Office of the Chief of Military History, 1989] Office of the Chief of Military History, U. S. A. (1989). The civil war 1861-1865. https://commons.wikimedia.org/wiki/File:Civil_war_1861-1865.jpg.
- [Ormeling, 1995] Ormeling, F. (1995). Teaching animated cartography. In *Proceedings of the Seminar on Teaching Animated Cartography ; Escuela Universitaria de Ingeniera Tecnica Topografica, Madrid, Spain*, pages 21–29. ICA.
- [Palen, 1999] Palen, L. (1999). Social, individual and technological issues for groupware calendar systems. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 17–24. ACM.
- [Papachristos, 2013] Papachristos, A. V. (2013). 48 years of crime in chicago : A descriptive analysis of serious crime trends from 1965 to 2013. *Chicago : ISPS Working Paper ISPS*, pages 13–023.
- [Pelekis et al., 2004] Pelekis, N., Theodoulidis, B., Kopanakis, I., and Theodoridis, Y. (2004). Literature review of spatio-temporal database models. *The Knowledge Engineering Review*, 19(3) :235–274.
- [Peterson, 1999] Peterson, M. P. (1999). Active legends for interactive cartographic animation. *International Journal of Geographical Information Science*, 13(4) :375–383.
- [Peuquet, 1988] Peuquet, D. J. (1988). Representations of geographic space : toward a conceptual synthesis. *Annals of the Association of American Geographers*, 78(3) :375–394.

- [Peuquet, 1994] Peuquet, D. J. (1994). It's about time : A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers*, 84(3) :441–461.
- [Peuquet, 2002] Peuquet, D. J. (2002). *Representations of space and time*. Guilford Press.
- [Plaisant et al., 1996] Plaisant, C., Milash, B., Rose, A., Widoff, S., and Shneiderman, B. (1996). Lifelines : visualizing personal histories. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 221–227. ACM.
- [Reitsma and Albrecht, 2005] Reitsma, F. and Albrecht, J. (2005). Implementing a new data model for simulating processes. *International Journal of Geographical Information Science*, 19(10) :1073–1090.
- [Reitsma and Dubayah, 2007] Reitsma, F. and Dubayah, R. (2007). Simulating watershed runoff with a new data model. *Hydrological processes*, 21(18) :2447–2457.
- [Reitsma, 2005] Reitsma, F. E. (2005). *A new geographic process data model*. PhD thesis.
- [Renolen, 1995] Renolen, A. (1995). Temporal maps and temporal geographical information systems (review of research).
- [Renolen, 1996] Renolen, A. (1996). History graphs conceptual modeling of spatiotemporal data gis frontiers in business and science, vol. 2. *Czech Republic, Brno : International Cartographic Association*.
- [Renolen, 2000] Renolen, A. (2000). Modelling the real world : conceptual modelling in spatio-temporal information system design. *Transactions in GIS*, 4(1) :23–42.
- [Robert Stewart, 2007] Robert Stewart, Jeremy Newcomb, C. B. (2007). The teaching american history grant. <http://teachingamericanhistory.org/past-programs/tah/>.
- [Saint-Marc, 2017] Saint-Marc, C. (2017). *Formalisation et géovisualisation d'événements historiques issus de risques naturels pour la compréhension des dynamiques spatiales*. PhD thesis, Université Grenoble-Alpes.
- [Saint-Marc et al., 2014] Saint-Marc, C., Davoine, P.-A., and Villanova-Oliver, M. (2014). Methods for mapping volcanic events overlaid across time. *Journal of Maps*, 10(2) :249–256.
- [Saint-Marc et al., 2017] Saint-Marc, C., Davoine, P.-A., Villanova-Oliver, M., Pams-Capoccioni, C., and Chenier, D. (2017). Géovisualisation de récits d'inondations et des effets dominos : De la méthode de géovisualisation à son évaluation dans un contexte expert. *Revue Internationale de Géomatique*, 27(4) :523–546.
- [Saint-Marc et al., 2018] Saint-Marc, C., Villanova-Oliver, M., Davoine, P.-A., Pams Capoccioni, C., and Chenier, D. (2018). Mapping the narratives of natural disasters and their domino effects. the case study of floods impacting railways. *International Journal of Cartography*, 4(1) :78–103.
- [Scharl and Tochtermann, 2009] Scharl, A. and Tochtermann, K. (2009). *The geospatial web : how geobrowsers, social software and the Web 2.0 are shaping the network society*. Springer Science & Business Media.
- [Schmidt et al., 2003] Schmidt, T. M., Petry, F. E., and Ladner, R. (2003). The object event calculus and temporal geographic information systems. *Lecture notes in computer science*, pages 799–814.
- [Shneiderman, 2003] Shneiderman, B. (2003). The eyes have it : A task by data type taxonomy for information visualizations. In *The Craft of Information Visualization*, pages 364–371. Elsevier.
- [Silva and Catarci, 2000] Silva, S. F. and Catarci, T. (2000). Visualization of linear time-oriented data : a survey. In *Web Information Systems Engineering, 2000. Proceedings of the First International Conference on*, volume 1, pages 310–319. IEEE.

- [Smith et al., 1994] Smith, B., Objects, V. A. F., Guarino, N., Vieu, L., and Pribbenow, S. (1994). Parts and wholes : Conceptual part-whole relations and formal mereology. In *11th European Conference on Artificial Intelligence, Amsterdam, 8 August 1994*, pages 15–23.
- [Smith and Varzi, 2000] Smith, B. and Varzi, A. C. (2000). Fiat and bona fide boundaries. *Philosophical and phenomenological research*, pages 401–420.
- [Snodgrass and Ahn, 1985] Snodgrass, R. and Ahn, I. (1985). A taxonomy of time databases. *ACM Sigmod Record*, 14(4) :236–246.
- [Solana, 2014] Solana, A. C. (2014). *Spatio-Temporal Narratives : Historical GIS and the Study of Global Trading Networks (1500-1800)*. Cambridge Scholars Publishing.
- [Spence, 2001] Spence, R. (2001). *Information visualization*, volume 1. Springer.
- [Swedberg and Peuquet, 2017] Swedberg, B. and Peuquet, D. (2017). Perse : visual analytics for calendar related spatiotemporal periodicity detection and analysis. *GeoInformatica*, 21(3) :577–597.
- [Tansel et al., 1993] Tansel, A. U., Clifford, J., Gadia, S., Jajodia, S., Segev, A., and Snodgrass, R. (1993). *Temporal databases : theory, design, and implementation*. Benjamin-Cummings Publishing Co., Inc.
- [Tessler, 1993] Tessler, F. (1993). Desktop calendars. *MacWorld*, 10(7) :104–109.
- [Thakur and Rhyne, 2009] Thakur, S. and Rhyne, T.-M. (2009). Data vases : 2d and 3d plots for visualizing multiple time series. In *International Symposium on Visual Computing*, pages 929–938. Springer.
- [The New York Times, 2012] The New York Times, U. (2012). Drought’s footprint. <https://archive.nytimes.com/www.nytimes.com/interactive/2012/07/20/us/drought-footprint.html>.
- [The Washington Post, 2011] The Washington Post, U. (2011). American civil war battles and casualties interactive map. <http://www.washingtonpost.com/wp-srv/lifestyle/special/civil-war-interactive/civil-war-battles-and-casualties-interactive-map/>.
- [Thériault et al., 1999] Thériault, M., Claramunt, C., and Villeneuve, P. (1999). A spatio-temporal taxonomy for the representation of spatial set behaviours. In *Spatio-temporal database management*, pages 1–18. Springer.
- [Tominski et al., 2005] Tominski, C., Schulze-Wollgast, P., and Schumann, H. (2005). 3d information visualization for time dependent data on maps. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 175–181. IEEE.
- [Tominski and Schumann, 2008] Tominski, C. and Schumann, H. (2008). Enhanced interactive spiral display. In *SIGRAD 2008. The Annual SIGRAD Conference Special Theme : Interaction ; November 27-28 ; 2008 Stockholm ; Sweden*, number 034, pages 53–56. Linköping University Electronic Press.
- [Tufte, 1983] Tufte, E. (1983). The visual display of quantitative information. number v. 914 in the visual display of quantitative information.
- [Tufte, 1991] Tufte, E. R. (1991). Envisioning information. *Optometry & Vision Science*, 68(4) :322–324.
- [Tukey, 1977] Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Reading, Mass.
- [Turdukulov et al., 2007] Turdukulov, U. D., Kraak, M.-J., and Blok, C. A. (2007). Designing a visual environment for exploration of time series of remote sensing data : In search for convective clouds. *Computers & Graphics*, 31(3) :370–379.
- [Van Wijk and Van Selow, 1999] Van Wijk, J. J. and Van Selow, E. R. (1999). Cluster and calendar based visualization of time series data. In *Information Visualization, 1999.(Info Vis’99) Proceedings. 1999 IEEE Symposium on*, pages 4–9. IEEE.

- [Wattenberg, 2002] Wattenberg, M. (2002). Arc diagrams : Visualizing structure in strings. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 110–116. IEEE.
- [Weber et al., 2001] Weber, M., Alexa, M., and Müller, W. (2001). Visualizing time-series on spirals. In *Infovis*, volume 1, pages 7–14.
- [Wikipedia, 2018a] Wikipedia (2018a). American civil war. https://en.wikipedia.org/wiki/American_Civil_War.
- [Wikipedia, 2018b] Wikipedia (2018b). Community areas in chicago. https://en.wikipedia.org/wiki/Community_areas_in_Chicago.
- [Wikipédia, 2018a] Wikipédia (2018a). Distribution de dirac. https://fr.wikipedia.org/wiki/Distribution_de_Dirac.
- [Wikipédia, 2018b] Wikipédia (2018b). Signal en dents de scie. https://fr.wikipedia.org/wiki/Signal_en_dents_de_scie.
- [Worboys, 1992] Worboys, M. (1992). A model for spatio-temporal information. In *Proceedings the 5th International Symposium on Spatial Data Handling, 1992*, volume 2, pages 602–611.
- [Worboys, 2005] Worboys, M. (2005). Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19(1) :1–28.
- [Worboys and Hornsby, 2004] Worboys, M. and Hornsby, K. (2004). From objects to events : Gem, the geospatial event model. In *GIScience*, volume 3234, pages 327–344. Springer.
- [Worboys, 2001] Worboys, M. F. (2001). Modelling changes and events in dynamic spatial systems with reference to socio-economic units.
- [Yuan, 1999] Yuan, M. (1999). Use of a three-domain representation to enhance gis support for complex spatiotemporal queries. *Transactions in GIS*, 3(2) :137–159.
- [Yuan, 2001] Yuan, M. (2001). Representing complex geographic phenomena in gis. *Cartography and Geographic Information Science*, 28(2) :83–96.
- [Yuan, 2007] Yuan, M. (2007). Gis approaches for geographic dynamics understanding and event prediction. In *Proceedings of SPIE*, volume 6578.

Table des figures

1.1	Relations topologiques entre objets spatiaux [Haddad, 2009]	31
1.2	Le modèle vecteur vs le modèle raster [Haddad, 2009]	32
1.3	Représentations linéaires et cycliques du temps	33
1.4	Les relations d'Allen [Allen, 1983]	34
1.5	Les temporalités caractérisant une entité temporelle [Hewitt, 1983][Allen, 1983, DiBiase et al., 1992, MacEachren, 1994]	35
1.6	Les temporalités caractérisent les relations temporelles entre entités [Hewitt, 1983][Allen, 1983, DiBiase et al., 1992, MacEachren, 1994]	35
1.7	Exemple d'une série temporelle	36
1.8	Exemple d'une série événementielle	36
1.9	Approches <i>SNAP</i> et <i>SPAN</i> de la modélisation de l'information spatio-temporelle [Mathian and Sanders, 2014]	37
1.10	Le Triad Model de Peuquet [Peuquet, 1994, Li, 2010]	41
1.11	Le Pyramid Model de Mennis [Mennis et al., 2000, Li, 2010]	42
1.12	<i>Space-Time Composite Model</i> [Langran and Chrisman, 1988, Renolen, 1995]	43
1.13	Modélisation du changement dans le <i>Space-Time Composite Model</i> [Langran and Chrisman, 1988, Mathian and Sanders, 2014]	44
1.14	Objets spatio-temporels [Worboys, 1992, Yuan, 1999]	44
1.15	Types de changements et catégorie des objets [Cheylan and Lardon, 1993]	46
1.16	Les processus spatio-temporels de Claramunt et Thériault [Claramunt and Thériault, 1995]	47
1.17	Du Triad Model au Pyramid Model [Peuquet, 1994, Mennis et al., 2000, Li, 2010]	51
1.18	Généricité de l'approche. Au centre, le Pyramid Model [Mennis et al., 2000].	55
2.1	Pics d'intensité du phénomène au travers d'une échelle temporelle linéaire	59
2.2	Pics d'intensité du phénomène relatifs à une échelle temporelle cyclique d'un an	60
2.3	Pics d'intensité du phénomène relatifs à une échelle temporelle cyclique d'une journée	60
2.4	Variations localisées du phénomène au cours du temps	61
2.5	Réapparition cyclique localisée d'un phénomène	61
2.6	Propagation spatiale d'un phénomène au cours du temps	62
2.7	Clusters spatio-temporels relatifs au temps linéaire	63
2.8	Identification de clusters relatifs au temps linéaire pour l'analyse de tendances	63
2.9	Clusters spatio-temporels relatifs au temps cyclique	64
2.10	Propagation spatiale d'un phénomène au cours du temps	65
2.11	Récurrences cycliques spatialisées d'événements	66
2.12	Carte de cas de choléra pendant l'épidémie de 1854 à Londres	69
3.1	L'horaire de train de Marey [Marey, 1885] [Tufte, 1991]	85
3.2	Sémiologie graphique et variables visuelles proposées par Bertin [Bertin, 1973]	86

3.3	Représentation de la trace d'un objet selon les règles préconisées par Bertin [Bertin, 1973]	87
3.4	Pertinence des variables visuelles dans des cartes dynamiques [Köbben and Yaman, 1996, Dukaczewski, 2005, Kaddouri et al., 2014]	88
3.5	Évaluation de l'aléa volcanique du Piton de la Fournaise. L'impact des coulées de lave [Cécile Saint-Marc, 2012]	88
3.6	Sémiologie graphique pour la représentation du temps dans les cartes selon Vasilev [Cauvin et al., 2008]	89
3.7	Le temps représenté par une variable spatiale. Projection de la carte de Minard dans le cube spatio-temporel [Minard, 1980, Marey, 1878, Kraak, 2003b]	91
3.8	Cube spatio-temporel pour la représentation des événements [Gatalsky et al., 2004]	92
3.9	Observation des taches solaires par Christoph Scheiner	93
3.10	États américains touchés par la sécheresse de 1960 à 1999 [The New York Times, 2012]	93
3.11	Copie d'écran d' <i>EventViewer</i> [Beard et al., 2008]	94
3.12	Copie d'écran de <i>Value Flow Map</i> [Andrienko and Andrienko, 2004]	95
3.13	Time Lapse des principaux tremblements de Terre sur les cent dernières années, et sur les quinze dernières années [NOAA, 2016]	96
3.14	Historique des coulées de lave du Piton de la Fournaise de 1972 à 2011 [Davoine, 2014]	99
3.15	Géovisualisation au prisme de 3 dimensions [MacEachren, 1994]	101
3.16	Outils de l'interactivité et services rendus [Davoine, 2014]	101
3.17	Formalisme des environnements de géovisualisation [Davoine et al., 2001, Davoine et al., 2004, Kraak and Ormeling, 2011]	102
3.18	Fonctionnement d'une interface de géovisualisation, l'exemple de GenGHIS [Davoine et al., 2012a, Kaddouri et al., 2014]	103
3.19	De la Triade de Peuquet à la géovisualisation [Kraak and Ormeling, 2011, Kaddouri et al., 2014]	103
4.1	Éruptions du Piton de la Fournaise [Saint-Marc et al., 2014]	109
4.2	Différentes formes de légendes temporelles	111
4.3	Modèle détaillé de classification des représentations temporelles [Aigner et al., 2007, Li, 2010]	113
4.4	Taxonomie des tâches de visualisation [Aigner et al., 2011]	113
4.5	Principe de la ligne de temps, ou <i>timeline</i>	114
4.6	Capture d'écran de l'environnement <i>SPHERE</i> [Davoine, 2014]	114
4.7	Principe de la roue du temps, ou <i>timewheel</i>	115
4.8	Capture d'écran de l'environnement <i>Silhouette Graph</i> [Harris, 2000]	115
4.9	Combinaison de la <i>timeline</i> et de la <i>timewheel</i> dans <i>TEMPEST</i> [Edsall and Peuquet, 1997]	116
4.10	Identification d'un cluster dans la distribution temporelle linéaire des événements dans une <i>timeline</i>	117
4.11	Identification d'un cluster dans la distribution temporelle linéaire des événements dans une <i>timeline</i> avec agrégation des événements temporellement proches	117
4.12	Identification d'un cluster dans la distribution temporelle linéaire des événements dans une <i>timeline</i> avec agrégation des événements selon un pas d'échantillonnage	118
4.13	Identification d'un cluster dans la distribution temporelle des événements selon une échelle cyclique dans une <i>timewheel</i>	118
4.14	Diagramme des causes de mortalité dans les forces expéditionnaires dans la Guerre de Crimée [Nightingale, 1858]	119
4.15	<i>Sector Graph</i> de Harris [Harris, 2000]	120
4.16	Identification visuelle de récurrences cycliques dans les <i>timelines</i>	121

4.17	Difficulté de l'identification d'une récurrence cyclique au travers d'une <i>timewheel</i>	122
4.18	<i>Arc Diagrams</i> : identification de récurrences cycliques dans les timelines, assistée par des analyses mathématiques [Wattenberg, 2002]	123
4.19	Identification de récurrences cycliques dans Cycle Plot [Cleveland, 1993]	125
4.20	Parallel Glyphs [Fanea et al., 2005]	126
4.21	Visualisation de données temporelles dans un diagramme en spirale [Weber et al., 2001]	127
4.22	Identification de récurrences cycliques dans un diagramme en spirale [Weber et al., 2001]	128
4.23	Visualisation de données temporelles sur un diagramme en <i>tiles map</i> [Mintz et al., 1997, Aigner et al., 2011]	129
4.24	Visualisation d'une récurrence cyclique dans l'environnement <i>Cluster Calendar</i> [Van Wijk and Van Selow, 1999]	130
4.25	Représentation du temps dans <i>Time Coil</i> [Edsall and Peuquet, 1997]	131
4.26	Représentation du temps dans <i>Time Wave</i> [Li, 2010]	131
4.27	Alignements sans récurrences cycliques dans un diagramme de type <i>Time Wave</i>	132
4.28	Structure de Time Wave [Li, 2010]	135
4.29	Symbologie des événements dans <i>Time Wave</i> [Li, 2010]	136
4.30	Visualisation du lien entre les composantes temporelle et spatiale des données, une représentation orientée de la carte vers le diagramme temporel [Li, 2010]	137
4.31	Structure de <i>PerSE</i> [Swedberg and Peuquet, 2017]	138
4.32	Diagramme temporel cyclique de <i>PerSE</i> [Swedberg and Peuquet, 2017]	139
4.33	Classification spatiale et représentation des événements selon les diagrammes de Voronoi [Swedberg and Peuquet, 2017]	141
4.34	Échelles temporelles cycliques paramétrables dans <i>Time Wave</i> (capture d'écran) [Li, 2010]	143
4.35	Représentation de l'atteinte de la température maximale quotidienne dans <i>Time Wave</i> [Li, 2010]	144
4.36	Calendriers géorgien et hégirien dans <i>PerSE</i> [Swedberg and Peuquet, 2017]	145
4.37	La recherche de récurrences cycliques dans les données temporelles au moyen de Spiral Graph [Weber et al., 2001]	146
4.38	Analyse des dynamiques spatio-temporelles par classification spatiale des événements dans un diagramme de type <i>Time Wave</i>	148
4.39	Captures d'écran de <i>Wakame</i> [Forlines and Wittenburg, 2010] et de <i>Data Vases</i> [Thakur and Rhyne, 2009]	149
4.40	Utilisation d'un cube spatio-temporel pour la représentation d'entités spatio-temporelles occurrentes [Gatalsky et al., 2004, Tominski et al., 2005, Turdukulov et al., 2007]	150
4.41	Capture d'écran de <i>VIS-STAMP</i> [Guo et al., 2006]	151
4.42	Utilisation de dispositif <i>small-multiples</i> pour la représentation d'entités occurrentes [Tufte, 1983]	152
4.43	Représentation d'événements au travers du temps linéaire et du temps cyclique dans <i>Growth Ring Map</i> [Bak et al., 2009, Andrienko et al., 2011]	153
4.44	Utilisation de <i>Pie Charts</i> pour représenter des événements dans un cube spatio-temporel [Andrienko and Andrienko, 2006]	156
4.45	Représentation d'articles <i>Wikipédia</i> à travers une série de <i>Pie Charts</i> [Bruggmann et al., 2013]	157
5.1	Phénomènes, événements, et sous-parties spatio-temporelles	173
5.2	Modélisation des événements dans <i>GrAPHiST</i>	175
5.3	Modélisation des événements dans <i>GrAPHiST</i> , lien avec le Pyramid Model de Mennis [Mennis et al., 2000]	175
5.4	Structure générale de <i>GrAPHiST</i>	176

5.5	Classification des événements par phénomène	178
5.6	Représentation tabulaire de la composante thématique des événements	179
5.7	Première implémentation du diagramme temporel mixte [Gautier et al., 2016]	181
5.8	Diagramme temporel mixte sinusoïdal	182
5.9	Changement d'échelle dans le diagramme sinusoïdal	182
5.10	Diagramme temporel mixte en dents de scie	183
5.11	Changement d'échelle dans le diagramme en dents de scie	184
5.12	Position des entités graphiques dans le diagramme sinusoïdale	185
5.13	Position des entités graphiques dans le diagramme en dents de scie	185
5.14	Procédé de création des agrégats temporels	186
5.15	Affichage de l'étendue temporelle des agrégats lors d'un survol	186
5.16	Représentation des agrégats temporels selon différents paramètres pour le calcul de la taille des entités	187
5.17	Visualisation de la continuité du temps linéaire dans les deux types de diagrammes temporels mixte	188
5.18	Visualisation des moments d'apparition dans le temps cyclique dans les deux types de diagrammes temporels mixte	189
5.19	Caractère non constant du rapport entre intervalle temporel représenté et longueur graphique dans le diagramme sinusoïdal	189
5.20	Division de l'échelle temporelle cyclique en différents intervalles dans les diagrammes temporels	190
5.21	Diagramme circulaire d'effectif	191
5.22	Procédé de création des agrégats spatiaux	194
5.23	Affichage de l'étendue spatiale des agrégats lors d'un survol	195
5.24	Représentation des agrégats spatiaux selon différents paramètres pour le calcul de la taille des entités	196
5.25	Sélection et <i>brushing</i> des événements selon un critère dimensionnel	197
5.26	Création des zones de classification spatiale	199
5.27	Représentation d'agrégats d'événements selon leur lieu d'apparition	201
5.28	Représentation de l'effectif des différentes classes d'événements, selon leur intervalle d'apparition relatif à une échelle cyclique	203
5.29	Division de l'étendue temporelle étudiée en intervalles temporels linéaires	204
5.30	Représentation d'agrégats d'événements selon leur moment d'apparition linéaire	205
5.31	Représentation de l'effectif des différentes classes temporelles d'événements, selon leur intervalle d'apparition relatif à une échelle cyclique	206
5.32	Représentation du moment temporel cyclique selon deux, quatre ou six couleurs	207
5.33	Affectation des couleurs aux sections du diagramme temporel circulaire, dans le cas de quatre couleurs de base	208
5.34	Représentation des événements selon huit intervalles temporels cycliques	209
5.35	Structure de la base de données	210
5.36	Fonctionnement de <i>GrAPHiST</i>	211
6.1	Résultats des interactions de l'utilisateur avec l'interface de <i>GrAPHiST</i> , dans le cadre d'une recherche d'un pattern spatio-temporel	217
6.2	Identification d'une concentration temporelle selon le temps linéaire	218
6.3	Modification de la représentation des agrégats dans le diagramme mixte pour l'identification de concentrations temporelles selon le temps linéaire	219
6.4	Identification de concentrations temporelles selon le temps cyclique dans le diagramme mixte	219
6.5	Identification de concentrations temporelles selon le temps cyclique dans le diagramme circulaire	219

6.6	Modification du nombre de sections dans le diagramme circulaire pour l'identification de concentrations temporelles	220
6.7	Identification d'une concentration spatiale d'événements, représentés par des entités ponctuelles	221
6.8	Modification de la représentation des agrégats dans la carte pour l'identification de concentrations spatiales	221
6.9	Identification d'un cycle strict sur le diagramme temporel mixte	222
6.10	Identification d'un cycle, apparaissant durant un intervalle temporel étendu, sur le diagramme temporel mixte	222
6.11	Identification d'un cycle sur le diagramme temporel mixte et de la concentration temporelle cyclique correspondante sur le diagramme circulaire	223
6.12	Différentes manifestations d'une concentration temporelle identifiée sur le diagramme circulaire, dans le diagramme mixte	223
6.13	Recherche d'une récurrence cyclique au moyen du diagramme temporel sinusoïdal	224
6.14	Influence de l'orientation des alignements graphiques dans la recherche d'une récurrence cyclique au moyen du diagramme temporel en dents de scie	225
6.15	Recherche d'une récurrence cyclique au moyen du diagramme temporel en dents de scie	225
6.16	Identification de la formation d'une concentration temporelle sur le diagramme circulaire, dans le cadre de la recherche d'une récurrence cyclique	226
6.17	Analyse de cycles assistée par les transformées de Fourier	231
6.18	Composantes des diagrammes de séquences décrivant les scénarios d'analyse . . .	233
6.19	Exemple d'un diagramme de séquence	234
6.20	Classification spatiale des événements en deux zones, créées à main levée	235
6.21	Identification de clusters relatifs au temps linéaire, selon une entrée spatiale . . .	237
6.22	Identification de clusters relatifs à une échelle cyclique, selon une entrée spatiale .	239
6.23	Classification des événements selon des plages de distance sur l'axe nord-sud . . .	240
6.24	Identification d'une propagation d'événements, selon une entrée spatiale	241
6.25	Analyse de cycles par transformée de Fourier, selon des zones de classification spatiale	242
6.26	Identification d'une récurrence cyclique spatialisée, selon une entrée spatiale . . .	245
6.27	Classification des événements selon leur moment d'apparition, relatif au temps linéaire ou relatif à une échelle cyclique	247
6.28	Identification de clusters relatifs au temps linéaire, selon une entrée temporelle . .	249
6.29	Identification de clusters relatifs à une échelle cyclique, selon une entrée temporelle	251
6.30	Identification de propagation d'événements, selon une entrée temporelle	252
6.31	Identification de la formation d'un cluster spatio-temporel dans la carte dans le cadre de la recherche d'une récurrence cyclique spatialisée	255
6.32	Identification de récurrences cycliques spatialisées, selon une entrée temporelle . .	258
7.1	Guerre de Sécession ou guerre civile américaine [Larousse, 2016]	278
7.2	Implémentation des données relatives à la guerre de Sécession dans notre modèle	279
7.3	Batailles de la guerre de Sécession selon le nombre de victimes [The Washington Post, 2011]	280
7.4	Territoires perdus par la <i>Confédération</i> lors de la guerre de Sécession [Office of the Chief of Military History, 1989]	281
7.5	Campagne de l'Overland, de mai à juin 1864 [Jespersen, 2012b]	282
7.6	Campagne d'Appomattox , du 2 au 9 avril 1865 [Jespersen, 2012a]	283
7.7	Visualisation des batailles de la guerre de Sécession	285
7.8	Distribution temporelle des combats de la guerre de Sécession selon une échelle cyclique d'une année	286

7.9	Analyse des différentes phases de la guerre de Sécession, avec une représentation non-agrégée	286
7.10	Analyse du déroulement des campagnes sur le théâtre de l'Ouest [Jespersen, 2012c]	287
7.11	Visualisation des batailles de la guerre de Sécession (représentation agrégée) . . .	288
7.12	Analyse des différentes phases de la guerre de Sécession, avec une représentation agrégée	288
7.13	Comparaison représentations non-agrégée et agrégée	289
7.14	Visualisation des engagements du théâtre de l'Est, du <i>1er mai 1864</i> au <i>30 avril 1865</i>	290
7.15	Analyse des engagements du théâtre de l'Est, du <i>1er mai 1864</i> au <i>30 avril 1865</i> .	291
7.16	Visualisation des engagements du <i>1er mai 1864</i> au <i>1er juillet 1864</i> , et comparaison avec un récit existant de l' <i>Overland Campaign</i> [Jespersen, 2012b]	292
7.17	Visualisation des engagements du <i>2 avril 1865</i> au <i>9 avril 1865</i> , et comparaison avec un récit existant de la campagne d' <i>Appomattox</i> [Jespersen, 2012c]	293
8.1	Emprise de notre zone d'étude dans la fenêtre spatiale	300
8.2	Zone d'apparition des événements fictifs	301
8.3	Intervalles d'apparition des événements fictifs	302
8.4	Implémentation des données fictives dans notre modèle	303
8.5	Identification des caractéristiques d'une série d'événements cycliques	304
8.6	Distribution spatio-temporelle du jeu A_1	307
8.7	Spectre de Fourier du jeu A_1	308
8.8	Identification d'un cluster spatio-temporel dans le jeu A_1	308
8.9	Visualisation de la composante temporelle du cluster spatio-temporel dans le jeu A_1	309
8.10	Modification du déphasage du diagramme temporel mixte	309
8.11	Identification de la période de la série événementielle cyclique dans le jeu A_1 . . .	309
8.12	Identification de l'emprise temporelle de la série événementielle cyclique dans le jeu A_1	310
8.13	Modification des paramètres de représentation des événements du jeu A_1	310
8.14	Visualisation d'une récurrence spatialisée dans le jeu A_1	311
8.15	Comparaison des résultats obtenus avec les données de référence pour le jeu A_1 .	311
8.16	Analyse du jeu A_2	312
8.17	Comparaison des résultats obtenus avec les données de référence pour le jeu A_2 .	312
8.18	Événements du jeu A_2 représentés selon une échelle cyclique de 581 jours	313
8.19	Événements du jeu A_2 représentés selon une échelle cyclique de 570 jours	313
8.20	Analyse du jeu A_3	314
8.21	Comparaison des résultats obtenus avec les données de référence pour le jeu A_3 .	314
8.22	Analyse du jeu A_4	315
8.23	Comparaison des résultats obtenus avec les données de référence pour le jeu A_4 .	315
8.24	Analyse du jeu B_5	316
8.25	Comparaison des résultats obtenus avec les données de référence pour le jeu B_5 .	316
8.26	Analyse du jeu B_6	317
8.27	Comparaison des résultats obtenus avec les données de référence pour le jeu B_6 .	317
8.28	Analyse du jeu B_7	318
8.29	Comparaison des résultats obtenus avec les données de référence pour le jeu B_7 .	318
8.30	Distribution spatio-temporelle du jeu B_8	319
8.31	Identification d'un cluster spatio-temporel dans le jeu B_8	320
8.32	Visualisation de la composante temporelle du cluster spatio-temporel dans le jeu B_8	320
8.33	Identification de la période de la série événementielle cyclique dans le jeu B_8 . . .	321
8.34	Identification de l'emprise temporelle de la série événementielle cyclique dans le jeu B_8	321
8.35	Modification des paramètres de représentation des événements du jeu B_8	322

8.36	Visualisation d'une récurrence spatialisée dans le jeu B_8	322
8.37	Comparaison des résultats obtenus avec les données de référence pour le jeu B_8	323
8.38	Précision des résultats obtenus par rapport au nombre d'événements	326
9.1	Implémentation des données relatives aux accidents routiers dans notre modèle	332
9.2	Implémentation des données relatives aux inondations dans notre modèle	334
9.3	Implémentation des données de criminologie dans notre modèle	335
9.4	Division de l'échelle temporelle de 365 jours en quatre sections	338
9.5	Zones créées pour la classification spatiale des événements d'accidents routiers	339
9.6	Visualisation des accidents routiers dans la région de Grenoble, du <i>1er janvier 2005</i> au <i>1er janvier 2016</i>	339
9.7	Visualisation des accidents routiers autour de la ville de Grenoble	340
9.8	Représentation des accidents routiers autour de la ville de Grenoble, selon leur lieu d'apparition	341
9.9	Distribution des accidents routiers par zone	341
9.10	Division de l'échelle temporelle d'un an en douze sections avec six couleurs primaires	343
9.11	Zones créées à partir des bassins versants pour la classification spatiale des événements d'inondation	344
9.12	Visualisation des cas d'inondations sur le territoire de la France métropolitaine, du <i>1er janvier 1900</i> au <i>1er janvier 2016</i>	344
9.13	Distribution temporelle des inondations durant l'année	345
9.14	Représentation des inondations selon une échelle cyclique d'un an	345
9.15	Analyse des inondations selon une échelle cyclique d'un an	346
9.16	Représentation des inondations du sud-ouest de la France selon une échelle cyclique d'un an	347
9.17	Analyse des inondations du sud-ouest de la France selon une échelle cyclique d'un an	348
9.18	Bassins versants du sud-ouest	348
9.19	Analyse de la distribution temporelle cyclique des événements dans les bassins versants du sud-ouest	349
9.20	Division administrative de la ville de Chicago [City of Chicago, 2017]	351
9.21	Division de l'échelle temporelle de 24h en huit sections avec six couleurs primaires	352
9.22	Cartes de la répartition des crimes violents à Chicago en 2011 (1) et de leur évolution entre 2011 et 2013 (2) [Papachristos, 2013]	353
9.23	Visualisation des événements de cambriolages du <i>1er janvier 2017</i> au <i>1er juillet 2017</i>	355
9.24	Distribution temporelle des événements de cambriolages	356
9.25	Distribution spatio-temporelle des événements de cambriolages	356
9.26	Distribution temporelle des événements de la zone correspondant au Repère 2	357
9.27	Distribution temporelle des événements de la zone correspondant au Repère 5	358
9.28	Représentation des événements de cambriolages par <i>Side</i>	358
9.29	Comparaison des distributions temporelles cycliques des cambriolages par zones	359
9.30	Visualisation des données relatives aux homicides du <i>1er janvier 2001</i> au <i>31 décembre 2015</i>	360
9.31	Analyse de l'évolution des homicides du <i>1er janvier 2001</i> au <i>31 décembre 2015</i>	361
9.32	Distribution temporelle linéaire des événements du centre-ville de Chicago	362
9.33	Comparaison des distributions temporelles linéaires des homicides par zones	362
9.34	Évolution 2011-2013 des crimes violents à Chicago	363
9.35	Représentation de la distribution spatiale des crimes violents [Papachristos, 2013]	364
9.36	Représentation de l'évolution des crimes violents selon leur position spatiale [Papachristos, 2013]	364

9.37	Comparaison de notre représentation cartographique avec les représentations produites par l'étude de l'Université de <i>Yale</i> [Papachristos, 2013]	365
9.38	Méthodologie de notre travail	378
9.39	Représentation du moment d'apparition des événements, au travers d'une échelle temporelle cyclique, au moyen de l'orientation d'un symbole flèche	386
9.40	Identification d'un cluster spatio-temporel relatif à une échelle cyclique, par identification d'un regroupement de symboles flèches pointant dans la même direction	386
9.41	Représentation des agrégats d'événements par des <i>Rose Charts</i>	387
10.1	Vue générale de l'interface de <i>GrAPHiST</i>	394
10.2	Fenêtre temporelle de <i>GrAPHiST</i>	395
10.3	Diagramme mixte de <i>GrAPHiST</i>	396
10.4	Interface de contrôle de la fenêtre temporelle	396
10.5	Modification de la durée de l'échelle temporelle cyclique par l'animation	397
10.6	Interface de contrôle de la fenêtre spatiale	398
10.7	Fenêtre d'import des événements	399
10.8	Fenêtre de filtre thématique des événements	399
10.9	Interface de contrôle (transformée de Fourier)	400
10.10	Représentation des résultats de la transformée de Fourier dans l'interface de contrôle	401
10.11	Diagramme temporel circulaire de <i>GrAPHiST</i>	402
10.12	Modification de la représentation temporelle des événements	403
10.13	Modification de la représentation des agrégats temporels	404
10.14	Modification de la représentation spatiale des événements	405
10.15	Modification de la représentation des agrégats spatiaux	406
10.16	Sélection des événements dans <i>GrAPHiST</i>	407
10.17	Représentation de la composante thématique des événements dans <i>GrAPHiST</i>	408
10.18	Outils de <i>brushing</i> dans <i>GrAPHiST</i>	408
10.19	Fenêtre de symbologie	409
10.20	Représentation des événements selon leur lieu d'apparition	410
10.21	Modification des couleurs affectées aux zones de classification spatiale	410
10.22	Représentation des événements selon leur moment d'apparition relatif à une échelle temporelle linéaire	411
10.23	Représentation des événements selon leur moment d'apparition relatif à une échelle temporelle cyclique	411
10.24	Choix du nombre d'intervalles temporels utilisés pour classer les événements	412

GrAPHiST

Une approche d'analyse exploratoire pour l'identification des dynamiques des phénomènes spatio-temporels.

Les données permettant de décrire des phénomènes spatio-temporels sont de plus en plus nombreuses. Ces nouvelles données peuvent alors être éloignées de celles habituellement observées pour l'étude de certains phénomènes. Leur analyse, selon une approche hypothético-déductive telle qu'elle est majoritairement effectuée en statistique et dans les *SIG*, peut ainsi passer sous silence certaines informations insoupçonnées, mais pertinentes, sur les dynamiques de ces phénomènes spatio-temporels.

Il peut alors être intéressant de simplement donner à voir les données, pour observer ce qu'elles ont à montrer, avant de les analyser. Ce principe est celui de l'analyse exploratoire : le procédé est de permettre à un utilisateur d'effectuer une exploration libre des données, au moyen de représentations visuelles, afin de mettre en lumière des structures ou des relations insoupçonnées. Aujourd'hui, l'analyse exploratoire est notamment possible au moyen d'environnements de visualisation, intégrant différentes représentations graphiques et cartographiques interactives.

Les environnements de visualisation sont majoritairement développés de manière *ad hoc*, dans le cadre d'une thématique particulière. Or l'émergence constante de nouvelles données incite à promouvoir des méthodes d'analyse applicables à des phénomènes de différentes natures. En fonction de la problématique dans laquelle s'insèrent ces derniers, les dynamiques sur lesquelles va se focaliser l'analyse diffèrent. Analyser un phénomène météorologique dans un but de prévision implique de s'intéresser aux récurrences cycliques du phénomène. Analyser l'évolution d'une population pour la mise en place de politiques publiques implique d'analyser ce phénomène sur le temps long et selon différentes zones de l'espace.

Notre objectif est de proposer une méthode d'analyse exploratoire des phénomènes spatio-temporels et de leurs dynamiques, indépendante du thème traité. Pour cela, nous proposons un environnement de géovisualisation, *GrAPHiST* (*Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels*), permettant l'analyse de différentes dynamiques, selon différentes échelles spatiales et temporelles (linéaires ou cycliques). Développer cet environnement implique de s'interroger sur la modélisation du changement dans l'espace, la nature des dynamiques spatio-temporelles à étudier, et les outils visuels et interactifs permettant de les identifier.

Ainsi, les contributions de notre recherche se situent à plusieurs niveaux :

- une modélisation générique des phénomènes spatio-temporels, sous la forme de séries événementielles ;
- de nouvelles méthodes de représentations graphiques et interactives, autorisant la recherche et l'identification des dynamiques spatio-temporelles, notamment :
 - l'introduction de diagrammes temporels interactifs permettant la recherche visuelle de récurrences cycliques dans les données spatio-temporelles ;
 - l'utilisation de règles de symbologie permettant la visualisation des relations entre les composantes temporelle et spatiale des phénomènes ;
 - de nouvelles méthodes de représentations des agrégats d'événements proches, permettant d'identifier des structures dans leur distribution spatio-temporelle ;
- la formalisation d'une approche d'analyse exploratoire des dynamiques spatio-temporelles, déclinée en plusieurs scénarios selon l'objectif poursuivi.

Nous validons notre approche en l'appliquant à l'analyse de différents jeux de données. L'objectif est de vérifier la possibilité d'identifier des dynamiques, relatives au temps linéaire ou cyclique, au moyen de *GrAPHiST*, et d'illustrer le caractère générique de l'approche, ainsi que les opportunités d'analyse offertes par l'environnement.

An exploratory analysis approach for the identification of dynamics of spatio-temporal phenomena.

Datasets allowing the description of spatio-temporal phenomena are becoming ever more numerous. These new data can be very different from those usually observed for studying spatio-temporal phenomena. An analysis through a hypothetic-deductive approach, like is mainly done in statistic and *GIS* domains, can lead to miss some unsuspected, but relevant, information about the dynamics of these spatio-temporal phenomena.

It can be interesting then, to just present the data, to observe what they have to show, before analysing them. This is the principle of the exploratory data analysis : the process is to allow a user to freely explore data, through visual representations, in order to highlight unsuspected structures or relationships. Today, exploratory data analysis is possible through visualization environments, which integrate different graphic or cartographic interactive representations.

Visualization environments are mainly developed in an *ad hoc* manner, in the context of a particular thematic field. However, the constant appearance of new data encourages promoting analysis methods, which could be applied to several types of phenomena. According to the domain related to these phenomena, the analysis will be focused on different dynamics. Analysing a meteorological phenomenon, in a forecasting purpose, implies a focus on the cyclic recurrences of the phenomenon. Analysing the increase of a population, for the purpose of deciding public policies, implies an analysis of the phenomenon on a long-term, through different spatial areas.

Our objective is to propose a method for the exploratory analysis of spatio-temporal phenomena and their dynamics, which would be independent of the topic. In order to achieve this, we propose a geovisualization environment, *GrAPHiST* (*Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels* ; *Geovisualization for spatio-temporal phenomena analysis*), allowing the analysis of several dynamics, through different spatial and temporal (linear or cyclic) scales. Developing this environment implies to focus on how spatial changes are modelled, on the nature of the spatio-temporal dynamics we have to study, and on the visual and interactive tools, which allow the identification of these dynamics.

So, the contributions of our research can be found at several levels :

- a generic modelling approach of spatio-temporal phenomena, in the form of event series ;
- new graphical and interactive representation methods, which allow the investigation and the identification of spatio-temporal dynamics, including :
 - the introduction of interactive temporal diagrams, which allow the visual research of cyclic recurrences in spatio-temporal data ;
 - the use of symbology rules, which allow the visualization of relationships between the spatial and temporal components of phenomena ;
 - new methods to represent aggregated closed events, which allow to identify structures in their spatio-temporal distribution ;
- the formalization of an exploratory approach for the spatio-temporal dynamics analysis, divided into several scenarios, according to the purpose of the analysis.

We validate our proposition by applying it to the analysis of several datasets. The objective is to verify the possibility to identify dynamics, related to linear or cyclic time, through the use of *GrAPHiST*, and to illustrate the generic aspect of the approach, as well as the analysis opportunities given by the environment.