



HAL
open science

Actes de la conférence SAGEO 2019

Sandro Bimonte, Sylvie Lardon

► **To cite this version:**

Sandro Bimonte, Sylvie Lardon. Actes de la conférence SAGEO 2019. Sandro Bimonte; Sylvie Lardon. SAGEO 2019 - Spatial analysis and geomatics, Nov 2019, Clermont-Ferrand, France. 2019. hal-02396759

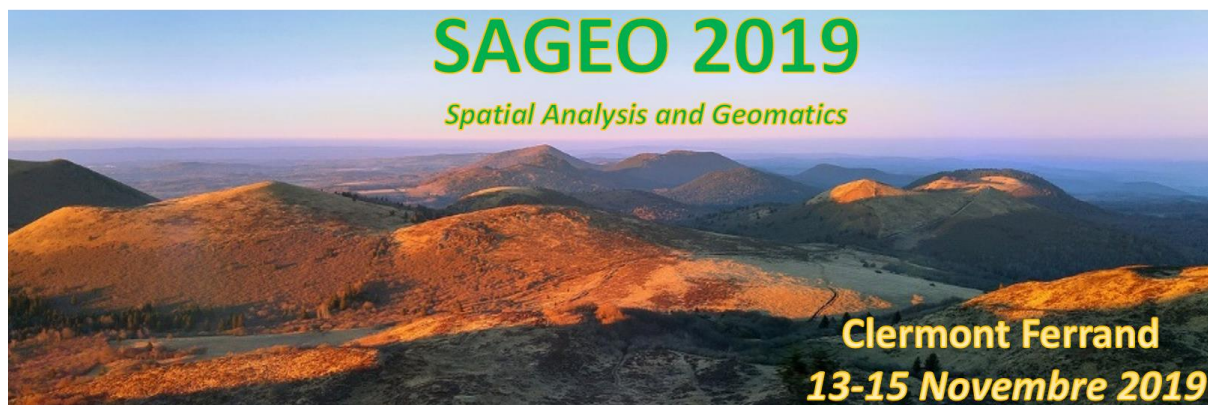
HAL Id: hal-02396759

<https://hal.science/hal-02396759>

Submitted on 6 Dec 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.



Actes de la conférence

Éditeurs

Sandro Bimonte, Université Clermont Auvergne, IRSTEA, Clermont-Ferrand

Sylvie Lardon, INRA & AgroParisTech, UMR Territoires, Clermont-Ferrand

Nous remercions le **GDR CNRS MAGIS**, **IRSTEA**, **INRA**, **Fédération des Recherches en Environnement**, **Université Clermont Auvergne**, **la Région Auvergne Rhône-Alpes**, **UMR Territoires**, **AFIGEO**, **Ville de Clermont Ferrand**, et **Clermont Auvergne Métropole** pour leur soutien.



Conférence internationale francophone

Spatial Analysis and GEomatics (SAGEO 2019)

Éditorial

Sous l'égide du GdR MAGIS du CNRS, la conférence internationale francophone SAGEO est un événement annuel majeur dans le paysage de la Géomatique, de l'Analyse Spatiale et des Sciences de l'Information Géographique. Ses objectifs sont de présenter des travaux de recherche récents et de réunir des chercheurs de disciplines diverses (informatique, géographie, archéologie, urbanisme, sciences de l'environnement, agronomie, humanités spatiales, etc.). Il s'agit également de promouvoir les échanges entre les différents acteurs du monde de la géomatique (chercheurs et enseignants-chercheurs, ingénieurs, gestionnaires, aménageurs, entrepreneurs...) qui œuvrent au développement et à la valorisation des recherches au sein et auprès des organismes publics, des entreprises et des collectivités territoriales.

L'édition 2019 de SAGEO se veut un forum ouvert à une majorité d'acteurs de la communauté francophone en géomatique. Dans ce contexte, nous avons souhaité proposer divers types de communications qui mettent en relief la richesse et la complémentarité des travaux et des applications.

- 3 conférences invitées traitant des différentes dimensions de la géomatique :
 - Pr. Beniamino Murgante. School of Engineering – University of Basilicata, Potenza – Italie. **The evolution of spatial analysis and modeling in decision processes.**
 - Dr. Thierry Ramadier. UMR SAGE, CNRS Strasbourg. **Articuler cognition spatiale et cognition environnementale : les représentations socio-spatiales.**
 - Dr. Patrick Taillandier. MIAT (Mathématiques et Informatique Appliquées de Toulouse), INRA. **Modélisation à base d'agents des systèmes complexes.**
- 18 présentations sur divers thèmes : mobilité, risques, agriculture numérique, analyse spatiale, représentations spatiales, stockage et flux
- 7 démonstrations logicielles
- 9 posters

Une session de rencontres et d'échanges entre laboratoires de recherche spécialisés en Géomatique et organismes (entreprises spécialisées mais aussi gestionnaires du territoire et utilisateurs) pour développer de possibles partenariats a été organisée en collaboration avec AFIGEO. Cette session fait l'objet d'un exposé « **Collaborations entre académique et non-académique : Moyens et outils** ». Audrey Raibon (CCI Puy de Dôme), Solène Vallin (Université Clermont Auvergne) et Julie Gratien (Auvergne-Rhône-Alpes Entreprises), en présence du *Lit Grandes Cultures Auvergne*, *ESRI France*, *Planet Observer*, *Centre Régional Auvergne-Rhône-Alpes de l'Information Géographique*. *Auvergne – Rhône – Alpes*, *InAirTECH* et *ISO GEO*.

Cette session industrielle anime les différentes communautés autour de la géomatique. Nous remercions les responsables des ateliers qui participent activement au foisonnement scientifique de ce domaine pluridisciplinaire

Responsables du Comité Scientifique

Sylvie Servigne, INSA, LIRIS, Lyon

Didier Josselin, CNRS, UMR ESPACE, Avignon

Sylvie Lardon, INRA & AgroParisTech, UMR Territoires, Clermont-Ferrand

Sandro Bimonte, Université Clermont Auvergne, IRSTEA, Clermont-Ferrand

Comité d'Organisation

Sandro Bimonte, Université Clermont Auvergne, IRSTEA, TSCF

Michelle Borel, INRA, UMR Territoires

Julien Chadeyron, Université Clermont Auvergne

Nicolas Dumoulin, IRSTEA, LISC

Sylvie Lardon, INRA & AgroParisTech, UMR Territoires

Emilie Mirault-Colin, IADT

François Pinet, IRSTEA TSCF

Comité des Relations Internationales

Roland Billen, Département de Géographie de l'Université de Liège, Belgique

Azedine Boulmakoul, Université Hassan II, Maroc

Rodolphe Devillers, Department of Geography Memorial University of Newfoundland, Canada

Ali Frihida, Department TIC, Tunis National School of Engineering, Tunisie

Chrispin Pettang, Ecole Polytechnique de Yaoundé, Cameroun

Khaled Rezeg, Laboratoire d'INformatique Intelligente (LINFI), Université de BISKRA, Algérie

Comité de Programme

Ali Hassan

André Miralles

Rezeg Khaled

R Billen

Ali Frihida

Christophe Claramunt

Marlene Villanova-Oliver

Florence LeBer

Anne Ruas

Jerome Gensel

Cyril Ray

Rémy Thibaud

Malika Madelin

Isabelle Mougnot

Eric Desjardin

Laurent Etienne

Francois Bavaud

Carmen Brando

Jean-François Girres

Mauro Gaio

Sebastien Rey-Coyrehourcq

David Sheeren

Marion LeTexier

Florence Sedes

Michel Mainguenaud

Jean-Pierre Chanet

Alain Bouju

Eric Daudé

Eric Guilbert

Guillaume Touya

Géraldine Delmondo

Sidonie Christophe

Thomas Louail

Anne Puissant

Gilles Gesquiere

Marie-Hélène Desede-Marceau

Christian Sallaberry

Helene Mathian

Mireille Batton-Hubert

Eric Kergosien

Cyril Derunz

Gaetan Palka

Matthieu Noucher

Maguelonne Teisseire

Rodolphe Devillers

Mathieu Roche

Cecile Duchene

Didier Leibovici

Thomas Devogele

Therese Libourel

Lucile Sautot

Sonia Chardonnel

Sommaire

Leborgne A., Meyer A., Giraud H., Le Ber F., Marc-Zwecker S. Un graphe spatio-temporel pour modéliser l'évolution de parcelles agricoles...	8
Prudhomme C., Cruz C., Boochs F. Modélisation sémantique et logique pour une simulation multi-agent dans le contexte de gestion de catastrophe.....	21
Karasiak N., Dejoux J.F., Monteil C., Sheeren D. Estimer l'apport de la phénologie dans la classification des essences forestières.....	32
Hamrouni Y., Paillassa E., Chéret V., Monteil C., Sheeren D. Cartographie à large échelle des peupleraies à partir de séries temporelles d'images Sentinel-2.....	45
Duroudier S., Chardonnel S., Mericskay B., Andre-Poyaud I., Bedel O., Depeau S., Devogele T., Etienne L., Lepetit A., Moreau C., Pelletier N., Ployon E., Tabaka K. Données hétérogènes de mobilités quotidiennes : protocole de diagnostic qualité et d'apurement à partir de la base MOB'IKIDS.....	58
Fillières-Riveau G., Barra V., Favreau J.M., Touya G. Enrichissement d'orthophotographie par des données OpenStreetMap pour l'apprentissage machine.....	71
Chagnaud C., Garat P., Davoine P.A., Brun-Trigaud G. Classification d'aires de dispersion à l'aide d'un facteur géographique : application à la dialectologie.....	84
Merchez L., Mathian H., Le Gall J. Mesurer les "vides alimentaires" dans un contexte urbain hétérogène, réflexion méthodologique et application à Lyon-Saint-Etienne.....	97
Viry M., Villanova-Oliver M. L'Ontologie d'Alerte Choucas : de la modélisation des connaissances à un outil support d'un raisonnement géovisuel. Application à la recherche de victime en haute-montagne.....	110
Fusco J., Marraccini E., Debolini M. Intensification, periurbanization and specialization as significant short-term land system dynamics in the Mediterranean basin.....	122
Menin A., Chardonnel S., Davoine P.A., Ortega M., Dublé E., Nedel L. eSTIME : une approche visuelle, interactive et modulable pour l'analyse multi-points de vue des mobilités quotidiennes.....	135

Escuriet M. Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière. Comprendre, cartographier et explorer le passé pour maîtriser son environnement.....	148
Stoll B., Hart M., Gaspar C. Contribution de la géomatique à l'observation des nids de tortues marines de Tetiaroa.....	161
Pinet F. La programmation logique pour les entrepôts de données spatiales.....	174
Gautier J., Davoine P.A., Cuntty C. Diagrammes temporels mixtes et analyse harmonique pour l'identification de cycles.....	182
Sautot L., Chraïbi E., Fize J., Peillet S., Journaux L., Cernesson F. Qui a peur du changement climatique ?	194
Lardon S., Marraccini E., Rizzo D., Benoit M. Combiner cartes à dires d'acteurs et lectures de paysage pour analyser les circuits agricoles et alimentaires de la Vallée de la Levrière. Un itinéraire méthodologique innovant.....	205
Bahoken F., Le Campion G., Jégou L., Maisonobe M., Côme E. Panorama d'un geoweb des flux et réseaux.....	218
Amarouch S., Durand J.B., Josselin D., Garcin L., Brachet N., Audard F., Carpentier-Postel S. Estimer des flux de navetteurs avec un modèle gravitaire Application géomatique en région Provence Alpes Côte d'Azur.....	232
Depeau S., Bedel O., Cherel P., André-Poyaud I., Catheline Y., Chardonnel S., Gombaud J., Jambon F., Lepetit A., Mericskay B., Quesseveur E., Leprince F. MK-MOBIBACK : un dispositif hybride et intégré pour enquêter finement les mobilités quotidiennes des familles.....	238
Bimonte S., Johany F., Noiret G., Lebert I., René-Martellet M., Dernas S. WIKISemio : un outil de testing collaboratif de cartes : application au risque de piqûre de tiques.....	241
Gautier J., Davoine P.A., Cuntty C. GrAPHiST: un environnement de visualisation pour l'analyse exploratoire de données spatio-temporelles.....	258
Côme E., Bapaume T., Jégou L., Bahoken F., Maisonobe M., Roelandt N., Le Campion G. Arabesques, application d'exploration et de géovisualisation de données de flux et de réseaux.....	265

Brahem M., Chachoua M., El Hafyani H., Kedad Z., Ktaish A., Mehanna S., Ray C., Taher Y., Thibaud R., Yeh L., Zeitouni K. Polluscope - Vers un observatoire participatif de l'exposition individuelle à la pollution de l'air et de ses effets sanitaires.....	275
Edoh-Alove E., Besnard A., Brenon C., Bimonte S., Fontaine B., Hassan A., Hifdi Y., Preud'homme R., Rakotoniaina I., Sakka A. Analyse en ligne des données de biodiversité en milieu agricole.....	281
Stoll B., Longine M. Géomorphologie à Tetiaroa : Prémices d'un observatoire de données scientifiques et environnementales.....	287
Stoll B., Bossin H. Observatoire de données de Tetiaroa : Contribution à la modélisation de l'environnement des moustiques.....	293
Ondzigue Mbenga R., Devogele T., Maghendji S., Peralta V., Ngoungou E.B. Un système d'information géographique décisionnel pour la surveillance épidémiologique de la tuberculose en Afrique Subsaharienne : Cas du Gabon.	299
Monkam-Djambou L.C., Kouamou G., Pettang C. Web mapping comme outil support dans la dynamique participative au développement territorial.....	307
Maillé E., Duvillier C. Un modèle statique de risque d'incendie de forêt induit et subi.....	313
Métral C., Falquet G. Une technique de spécification abstraite pour la publication de géodonnées liées.....	319
Koteich B., Saux E., Grancher A. Approche ontologique pour la modélisation de carte contextuelle.....	325
Gharbi A., De Runz C., Akdag H., Faïz S. Règles d'association pour l'analyse de l'évolution d'un territoire.....	331

Un graphe spatio-temporel pour modéliser l'évolution de parcelles agricoles

Aurélie Leborgne, Adrien Meyer, Henri Giraud, Florence Le Ber, Stella Marc-Zwecker

*ICube UMR 7537, Université de Strasbourg, CNRS, ENGEES, 67400 Illkirch, France
{aurelie.leborgne, stella, henri.giraud}@unistra.fr, florence.leber@engees.unistra.fr*

RÉSUMÉ. Dans cet article nous présentons une variante du graphe spatio-temporel défini par (Del Mondo et al., 2010), permettant de tenir compte d'entités spatio-temporelles nombreuses, imprécises, et non connectées. Le graphe proposé est appliqué aux données issues du registre parcellaire graphique, où les agriculteurs déclarent leurs cultures, en respect de la politique agricole européenne. Une méthode de simplification est proposée, afin de réduire la taille du graphe. Notre objectif final est d'utiliser ce modèle pour représenter et fouiller les grandes masses de données spatio-temporelles disponibles en sciences de l'environnement.

ABSTRACT. This article describes a variation of the spatio-temporal graph model defined by (Del Mondo et al., 2010). This variation allows to represent numerous spatio-temporal entities, that can be vague and disconnected. This model is used to represent data from the Land Parcel Identification System where farmers declare their crops with respect to European agricultural policy. A simplification method is described, that allows to reduce the graph size. Our final aim is to use this model for representing and analyzing big spatio-temporal data in environment sciences.

MOTS-CLÉS : Données spatio-temporelles, graphe, modélisation, agriculture

KEYWORDS: Spatio-temporal data, graph, modeling, agriculture

1. Introduction

Actuellement, des quantités massives de données spatio-temporelles hétérogènes peuvent être accumulées. Elles peuvent être issues de grandes bases de données géographiques, de séries d'images satellite, de capteurs, *etc.* Ces données doivent être fouillées afin d'extraire des informations pertinentes, intéressantes et utilisables dans diverses applications, comme la compréhension d'un phénomène particulier, la mise en évidence d'une même évolution à différents endroits, la prédiction d'un phénomène, *etc.* De manière plus précise, pour extraire de la connaissance de ces données, il est nécessaire de connaître les différents éléments qui caractérisent les phénomènes (le *quoi*), la manière dont ces éléments se répartissent dans l'espace (le *où*), ainsi que le moment où ces éléments apparaissent (le *quand*) (Peuquet, Wentz, 1994).

Il existe de nombreuses manières de modéliser les données spatio-temporelles. Le modèle *snapshot* (Armstrong, 1988; Chen *et al.*, 2013) permet d'obtenir un ensemble de couches indépendantes, où une couche décrit les états spatiaux des zones à un temps donné. Le modèle *Space-Time Composite* (Langran, Chrisman, 1988) peut être vu comme une extension du modèle *snapshot*. Il permet de prendre en compte le changement d'une entité en lui ajoutant des attributs qui décrivent ce changement dans le temps. Par ailleurs, le modèle *Spatio-Temporal Object* (Worboys, 1994) permet de représenter l'évolution d'une zone grâce à trois dimensions. Les deux premières dimensions permettent de représenter les empreintes spatiales, et la troisième, l'évolution des empreintes spatiales dans le temps. En revanche, ces modèles ne peuvent représenter que des changements soudains.

D'autres approches s'attachent à représenter les changements liés à l'identité. Elles exploitent, en particulier, des modèles de trajectoires d'entités dans l'espace et dans le temps (Hornsby, Egenhofer, 2002). Toutefois, ces modèles ne permettent pas d'analyser correctement l'évolution de l'espace et du temps du fait qu'ils ne gèrent pas simultanément des informations sur le *où*, le *quand* et le *quoi*.

Dans le cadre de notre projet, qui s'intéresse aux évolutions de l'usage agricole des sols, nous nous focalisons sur les représentations à l'aide de graphes, qui sont des outils puissants permettant de modéliser un phénomène dynamique de manière assez naturelle. Dans la majorité des graphes, les sommets modélisent les entités dans un domaine donné et les arêtes, les relations existant entre deux entités. Dans un premier temps, les recherches se sont focalisées sur les relations spatiales (Jiang, Claramunt, 2004). Certains auteurs se concentrent sur les relations spatio-temporelles entre deux entités. Plus précisément, une arête entre deux entités consécutives, au sens temporel du terme, permet de représenter le fait que l'espace occupé par la première entité est lié à l'espace occupé par la deuxième entité (Khiali *et al.*, 2018; Aydin, Angryk, 2016). (Sriti *et al.*, 2005) utilisent également un graphe orienté en combinant un graphe avec des relations spatiales et un graphe avec des relations spatio-temporelles. À aucun moment, ces dernières méthodes ne prennent en compte le *quoi*, c'est-à-dire l'identité des entités. Signalons aussi les travaux de (Degenne, Lo Seen, 2016) qui utilisent les graphes pour simuler des phénomènes spatio-temporels : les arêtes représentent ici

des interactions entre entités. Concernant les applications agricoles, notons encore les travaux de (Guyet, Moinard, 2014) où les graphes servent à la fois pour fouiller puis simuler des parcelles agricoles.

Finalement, une des seules méthodes possibles, permettant de modéliser nos données dans le but de les fouiller efficacement tout en suivant leur évolution, était l'utilisation de graphes spatio-temporels (Del Mondo *et al.*, 2010) dans lesquels trois types de relations existent (spatiale (*où?*), temporelle (*quand?*) et filiation (*quoi?*)).

L'article s'organise comme suit. Dans la section 2 nous présentons les données et la problématique visée, tandis que la section 3 s'attache à la définition d'un graphe spatio-temporel modélisant l'évolution des parcelles agricoles. Les sections 4 et 5 traitent du calcul et de la simplification du graphe. La section 6 conclue cet article et dresse quelques perspectives de travail.

2. Données utilisées

Les évolutions actuelles de l'agriculture, entre agriculture biologique et industrielle, les risques de dégradation de la biodiversité et des ressources en eau, conduisent de nombreux chercheurs à s'intéresser à la compréhension et à la prévision des évolutions de paysages agricoles (Poggi *et al.*, 2018). Par exemple, la disparition de prairies permanentes au profit de grandes cultures conduit d'une part à la dégradation des ressources en eau, d'autre part à l'augmentation des risques d'érosion des terres. Inversement, l'installation de haies permet de préserver les sols et peut favoriser les auxiliaires de cultures. Pour mener ces études, différentes méthodes sont développées, et différentes sources de données sont exploitées : l'imagerie satellitaire, les enquêtes de terrain, et grâce au mouvement d'ouverture des données mis en oeuvre par l'état, des données administratives, tel que le Registre Parcellaire Graphique (RPG).

Le RPG est une base de données géographique mise en place par la France afin de répondre à l'exigence de l'Union Européenne pour tous les Etats Membres de localiser et d'identifier les parcelles agricoles en contrepartie des aides versées au titre de la politique agricole commune (PAC). Le RPG a été mis en place en 2002 et compile les données issues des déclarations de surfaces agricoles effectuées chaque année par les agriculteurs eux-mêmes. Ainsi tous les ans, chaque exploitant doit dessiner et mettre à jour les îlots de culture qu'il exploite, et déclarer les cultures qui y sont pratiquées et leurs surfaces. Le RPG contient plus de 7 millions de parcelles, la figure 1 en montre un extrait pour la plaine du Rhin.

La version diffusée dans le cadre du service public de mise à disposition des données de référence est dite "anonymisée" puisque certaines informations déclarées par les exploitants sont retirées. Les attributs contenus dans la version 2.0 du RPG et associés à chacune des parcelles correspondent à :

- un identifiant unique de la parcelle,
- un code permettant de connaître la culture principale sur la parcelle,



Figure 1. Extrait du millésime 2017 du RPG sur la plaine d'Alsace

- un code permettant de connaître à quel groupe de cultures appartient la culture de la parcelle,
- deux codes permettant de connaître les cultures dérobées, c'est-à-dire les cultures intercalées entre deux cultures principales.

Bien que le RPG constitue une source précieuse d'informations fournissant des détails sur l'état et l'évolution de l'occupation du sol et des structures foncières, il doit cependant être utilisé avec précaution car la situation donnée par le RPG est approximative, en particulier à cause de l'absence de certaines surfaces agricoles (surfaces des exploitations non aidées par exemple), mais aussi à cause de la variabilité d'une année sur l'autre (évolution des aides, surfaces non déclarées pour une raison relative à la vie d'une exploitation), ou encore de modifications effectuées par l'exploitant alors même qu'il n'y a aucun changement sur le terrain. Notons par ailleurs qu'il existe un outil dédié pour exploiter ces données, appelé RPG Explorer (Levasseur *et al.*, 2016).

Dans le RPG, les parcelles sont des polygones, représentés par la liste de leurs sommets géoréférencés. Dans la suite chacun des polygones est considéré comme une entité connexe, sans trou et dotée d'une frontière et d'un intérieur. Pour une entité A , on note δA sa frontière, A° son intérieur. On note aussi A^e son extérieur, et pour certaines situations liées à l'imprécision des données, ΔA représentera la *frontière épaisse* de l'entité (voir section 3.2).

3. Un graphe spatio-temporel de parcelles agricoles

Nous nous appuyons sur le modèle de graphe spatio-temporel qui est défini par (Del Mondo *et al.*, 2010) comme l'union de trois sous-graphes (voir figure 2) :

- **Le graphe des relations spatiales**, qui représente les interactions spatiales entre entités à un moment donné. Les relations utilisées dans (Del Mondo *et al.*, 2010) sont les relations de la théorie RCC8 (Region Connection Calculus) (Randell *et al.*, 1992).
- **Le graphe des relations spatio-temporelles**, qui représente les interactions spatiales entre entités à des moments successifs. De même que pour le graphe spatial, sont utilisées ici les relations RCC8.
- **Le graphe des relations de filiation**, qui représente la transmission de l'identité à entre entités à des temps différents ; deux types de filiation sont considérées, la continuation et la dérivation, notées respectivement γ et δ sur la figure 2.

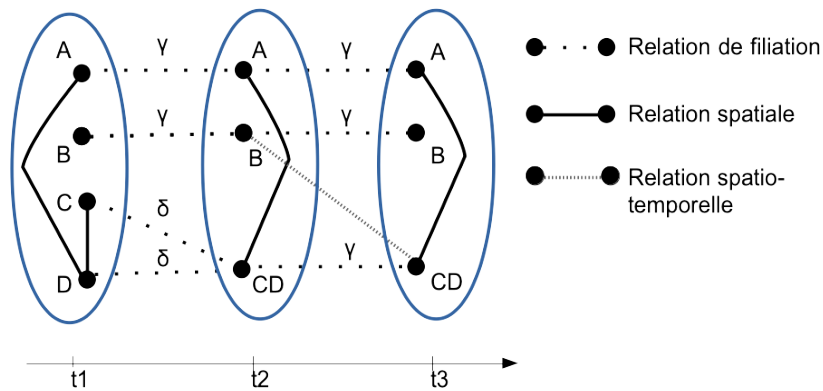


Figure 2. Le modèle de graphe spatio-temporel proposé par (Del Mondo *et al.*, 2010)

Les différents types de relations pour notre application sont précisés ci-dessous.

3.1. Choix des relations spatiales

Les huit relations de base de la théorie RCC8 (Randell *et al.*, 1992) permettent de décrire de façon exhaustive et disjointe (une seule relation est vérifiée) la position relative de deux entités. Ces relations sont les suivantes : $TPP(x, y)$ (x est partie propre tangentielle de y), $TPPI(x, y)$ (x contient y comme partie propre tangentielle), $NTPP(x, y)$ (x est une partie propre non-tangentielle de y), $NTPPI(x, y)$ (x contient y comme partie propre non tangentielle), $PO(x, y)$ (x recouvre partiellement y), $EC(x, y)$ (x est connecté extérieurement à y), $DC(x, y)$ (x est déconnecté de y), et $EQ(x, y)$ (x et y sont égales).

Notons que, pratiquement, peu de relations RCC8 sont vérifiées : de fait, il n'y a pas de recouvrement complet entre des parcelles distinctes, ce qui exclut les relations EQ , TPP , $NTPP$ et leurs inverses. De plus, la relation DC est représentée implicitement dans le graphe par l'absence d'arête. Pour enrichir le modèle, et tenir compte des particularités des données du RPG (les parcelles sont rarement connexes), nous avons ajouté des relations de proximité. Nous introduisons tout d'abord une distance

entre deux entités, mesurée comme le minimum des distances euclidiennes $d(x, y)$ entre les points x et y des deux frontières des entités.

DÉFINITION 1. — Soient A et B deux parcelles, la distance entre A et B est définie par $d(A, B) = \min_{x \in \delta A, y \in \delta B} d(x, y)$

Sur cette base, nous distinguons deux types de relation de proximité, une proximité absolue et une proximité relative à la taille des parcelles. Ces deux notions peuvent rendre compte de phénomènes différents, d'une part l'existence de zones séparant deux parcelles (chemins, haies) et d'autre part de l'influence d'une parcelle sur une autre, qui dépend de la taille relative des parcelles (par ex. diffusion de gènes entre cultures). La relation de proximité absolue permet de relier deux à deux les parcelles se situant à moins d'une distance b l'une de l'autre. Dans le cadre de notre étude, $b = 2$ mètres. Les parcelles A et B sont *absolument proches* ($p_a(A, B)$) si et seulement si :

$$0 < d(A, B) \leq b \quad (1)$$

La relation de proximité relative permet de relier une parcelle à une autre si leur distance respecte un seuil proportionnel à l'aire de la seconde. Formellement, la parcelle A est *relativement proche* de B , ($p_r(A, B)$) si et seulement si :

$$0 < d(A, B) \leq \frac{\sqrt{R_B}}{N} \quad (2)$$

où R_B est l'aire de la parcelle B et N est un nombre arbitraire. Cette relation n'est pas symétrique : une petite parcelle peut être proche d'une parcelle plus importante, au sens où elle subit l'influence de cette grande parcelle sur la majeure partie de sa surface, mais l'inverse n'est pas forcément vrai (Lavigne *et al.*, 2008).

3.2. Choix des relations spatio-temporelles

Une relation spatio-temporelle repose sur la même base de définition que les relations spatiales, mais entre des entités définies à des temps différents. Les données RPG étant annuelles, nous considérons uniquement des relations entre entités de deux années successives : sont calculées les relations entre les parcelles des temps t et $t + 1$ ayant des empreintes spatiales non disjointes. Contrairement au cas des relations spatiales, l'ensemble des relations RCC8 est exploité pour les relations spatio-temporelles : en effet, EQ correspond à l'invariance, TPP et $NTPP$ à l'agrandissement, $TPPI$ et $NTTPI$ au rétrécissement d'une parcelle entre deux temps.

Par ailleurs, pour tenir compte de l'imprécision temporelle des données (la même parcelle, à deux dates successives, pouvant ne pas recouvrir exactement la même surface, en raison de l'inexactitude des mesures de délimitation sur le terrain), nous proposons d'utiliser une extension de RCC8 pour des régions à frontières épaisses (Cohn, Gotts, 1996 ; Clementini, Di Felice, 1997), que nous appelons *relations spatio-temporelles floues*. Les modèles proposés comptent 44 à 46 relations. Chaque relation peut être rattachée à une ou plusieurs relations de RCC8. Compte tenu des caractéristiques des données, nous avons considéré l'ensemble des configurations proposées,

puis nous l'avons divisé en sous-ensembles, affecté chacun à une seule relation RCC8 (voir le groupe correspondant à la relation *EC* sur la figure 3).



Figure 3. Relations spatio-temporelles floues : les 4 configurations rattachées à la relation *EC*

Notons qu'il est tout à fait possible d'appliquer ce modèle de relations floues pour prendre en compte une imprécision spatiale, liées aux caractéristiques des données, mais cela n'est pas nécessaire pour les données RPG.

3.3. Relations de filiation

Nous discernons deux types de relations de filiation en accord avec (Del Mondo *et al.*, 2010) :

- **La filiation de type continuation** : deux parcelles *A* et *B* sont en relation de continuation si *A* et *B* ont la même identité, c'est-à-dire qu'il s'agit de la même entité mais à des temps différents. Par conséquent, une entité ne peut recevoir au plus qu'une seule relation de continuation.
- **La filiation de type dérivation** : deux parcelles *A* et *B* sont en relation de filiation de type dérivation si *B* dérive de *A* et potentiellement d'autres entités. Une partie de l'information contenue par *A* est présente dans *B*.

L'*identité* d'une parcelle est décrite dans le *RPG*, et elle est conservée d'une édition à l'autre du *RPG* (pour les éditions postérieures à 2015), lorsque la géométrie de la parcelle n'est pas modifiée. Un exemple de graphe de relations de filiation est présenté sur la Figure 4 pour un groupe de parcelles à deux temps différents.

4. Calcul des relations

Le calcul des relations spatiales entre deux entités *A* et *B* s'appuie sur une matrice d'intersection (Egenhofer, Al-Taha, 1992) de la forme :

$$MI = \left\{ \begin{array}{ccc} \dim(A^\circ \cap B^\circ) & \dim(A^\circ \cap \delta B) & \dim(A^\circ \cap B^e) \\ \dim(\delta A \cap B^\circ) & \dim(\delta A \cap \delta B) & \dim(\delta A \cap B^e) \\ \dim(A^e \cap B^\circ) & \dim(A^e \cap \delta B) & \dim(A^e \cap B^e) \end{array} \right\}$$

où *dim* est le nombre maximum de dimensions de l'ensemble résultant de l'intersection des ensembles intérieurs (X°), frontières (δX) ou extérieurs (X^e) des entités *A* et *B*. Notons donc que $0 \leq \dim \leq 2$. Par ailleurs, le symbole \emptyset est utilisé lorsque $\dim = 0$ et $-\emptyset$ lorsque $\dim > 0$. La figure 5 présente un exemple de masque correspondant à la relation topologique de recouvrement partiel. Ceci nous permet d'établir

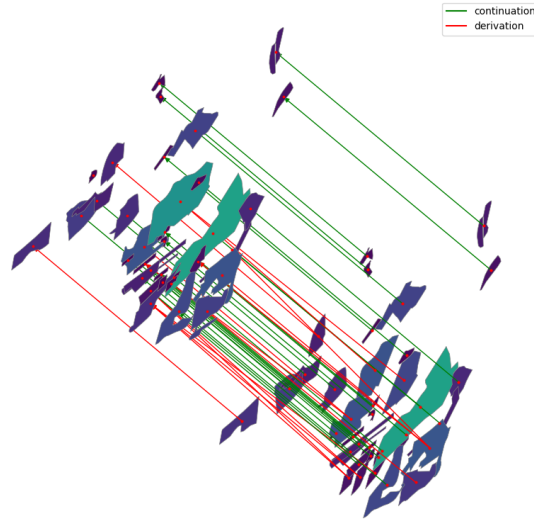


Figure 4. Visualisation d'un graphe de relations de filiation entre parcelles à deux temps différents ; les relations de continuation sont figurées en rouge, celles de dérivation en vert ; les couleurs des parcelles n'ont pas de signification

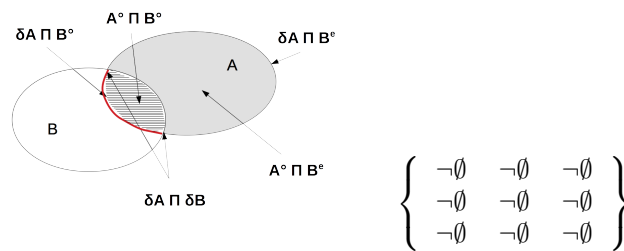


Figure 5. Exemple de la relation PO, intersections et matrice associée

les relations entre les parcelles en comparant, grâce à un ET logique, la matrice MI obtenue aux différents masques de référence.

Pour les relations de proximité présentées au paragraphe 3.1, nous utilisons la même méthode en manipulant un polygone dilaté d'une distance fixée pour la relation à calculer et les parcelles concernées.

Pour les relations spatio-temporelles floues, est utilisée la même approche, mais en considérant une frontière épaisse et en exploitant les matrices définies dans (Clementini, Di Felice, 1997). La figure 6 présente un exemple de masque correspondant à une relation topologique floue rattachée à EC.

Pour mettre en œuvre cette méthode, et pallier les problèmes de calcul liés à la grande dimension des données, nous avons introduit un index spatial sous la forme

Un graphe spatio-temporel de parcelles

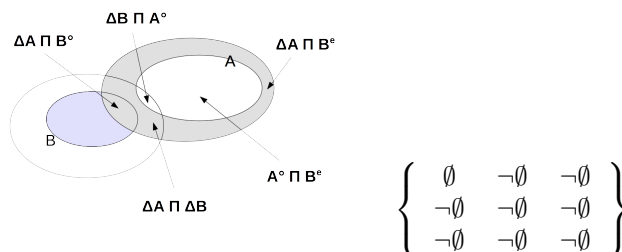


Figure 6. Exemple d'une relation EC floue (frontière épaisse), intersections et matrice associée

d'un R-arbre tel que décrit dans (Guttman, 1984). Chaque feuille de l'arbre représente une entité dans son rectangle englobant minimal; au niveau supérieur, les entités proches sont regroupées et représentées avec leur rectangle englobant minimal dans un nœud; le processus est réitéré à chaque niveau de l'arborescence, où les nœuds représentent un nombre croissant d'entités. Cette structure peut aussi être vue comme une approximation de plus en plus grossière de l'ensemble des données, ce qui nous permet de limiter le nombre de couples d'entités à tester. Une requête sur le R-arbre a une complexité moyenne en $O(\log n)$, avec n le nombre de parcelles (versus n^2 pour l'approche naïve), ce qui permet de calculer les relations entre les parcelles en un temps raisonnable. Par exemple, 76740 polygones sont traités en 2 minutes par la méthode exploitant les R-arbres alors que la méthode naïve prend 3766 minutes, soit plus de deux jours.

5. Simplification du graphe

Un graphe spatio-temporel de parcelles est constitué d'un ensemble de sommets (représentant les parcelles, avec leurs attributs) et de de trois ensembles d'arêtes, représentant respectivement les relations spatiales, spatio-temporelles et de filiation entre parcelles. Compte tenu de la taille des données traitées, le nombre de relations explose rapidement, et pour faciliter l'exploitation du graphe, il est nécessaire de le simplifier. Pour cela, nous avons choisi de nous focaliser sur la sémantique des données. Plus précisément, nous avons choisi de regrouper en un seul sommet les sommets qui sont spatialement en relation (dans notre cas, des sommets liées par les relations *PO*, *EC* ou de proximité) et qui partagent les mêmes valeurs d'attributs (ici le même type de culture pour l'année considérée).

La procédure consiste à parcourir le graphe et à tester si des sommets en relation portent les mêmes valeurs d'attributs. Si c'est le cas, les sommets peuvent être remplacés par un seul sommet, et il est alors nécessaire d'adapter les relations entrantes et sortantes des anciens sommets vers le nouveau. Cependant, il ne peut pas y avoir plus d'une relation spatiale entre deux sommets. La figure 7 décrit une situation où une parcelle *A* est en relation avec deux parcelles *U* et *V*, qui sont elles-mêmes en relation. *U* et *V* possèdent les mêmes attributs et sont donc regroupées. Le sommet résultant, noté

U' , est relié par deux relations spatiales avec A , $relAU$ et $relAV$ qu'il faut réduire en une relation résultante. Comme montré sur la figure 7, $relAU$ est une relation EC et $relAV$ une relation PO , donc AU' est une relation PO . Nous pouvons dire, en effet, que PO est *plus prioritaire* que EC lors d'un conflit entre deux relations. On note $relation_1 \oplus relation_2$ (respectivement \ominus) pour " $relation_1$ est *plus prioritaire (ou égal) que* $relation_2$ " (respectivement *moins prioritaire que*). Dans le cadre de notre application, les priorités entre les relations spatiales sont $PO \oplus EC \oplus p_a \oplus p_r$.

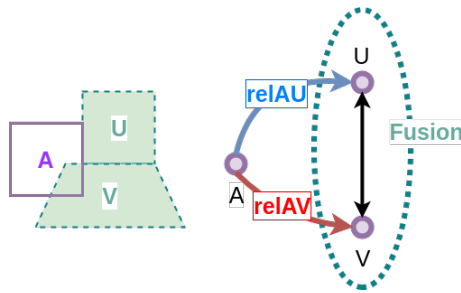


Figure 7. Regroupement de parcelles, le sommet A est en relation avec les sommets U et V qui fusionnent

Algorithme 1 : Simplification de graphe spatial

Data : graphe étudié, ensemble des critères, ensemble des relations à simplifier

Result : Le graphe simplifié sur nos critères

```

1 soit  $L(X, Y, rel1)$  un lien de type rel1 entre les sommets X et Y
2 soit  $X \nrightarrow Y$  une notation indiquant que X et Y n'ont pas de liens
3 forall liens  $L(U, V, relUV)$  devant être simplifié do
4   forall liens  $L(A, V, relAV)$  entrant sur V do
5     if  $A \nrightarrow U$  then
6       ajouter  $L(A, U, relAV)$  au graphe
7     else if  $relAV$  prioritaire sur  $relAU$  then
8       mettre à jour lien( $A, U, relAV$ )
9   end
10  forall liens  $L(V, B, relVB)$  sortant de V do
11    if  $U \nrightarrow B$  then
12      ajouter  $L(U, B, relVB)$  au graphe
13    else if  $relVB$  prioritaire sur  $relUB$  then
14      mettre à jour lien( $U, B, relVB$ )
15  end
16  Fusionner les données de U et V dans U
17  supprimer  $U \rightarrow V$ 
18  supprimer  $V \rightarrow U$ 
19  supprimer V
20 end

```

Un graphe spatio-temporel de parcelles

L'algorithme 1 permet de simplifier le graphe spatial. La figure 8 montre la différence entre le graphe spatial initial et le graphe simplifié pour un petit nombre de parcelles (sont représentées en bleu les relations de proximité relative, en vert les relations de proximité absolue et en orange les relations *EC*).



Figure 8. A gauche, les relations spatiales pour un extrait du RPG, à droite, après simplification : les parcelles en orange et en jaune sont des regroupements

L'étape suivante, non décrite ici, consiste à simplifier les relations spatio-temporelles : d'une part en rétablissant les relations spatio-temporelles adéquates pour les sommets fusionnés spatialement, d'autre part en fusionnant des sommets de temps différents portant la même sémantique (par exemple, une parcelle qui, à des temps successifs, porte une succession de cultures reconnue). Les relations de filiation devront également être rétablies (par exemple, en remplaçant les relations de continuation arrivant aux parcelles fusionnées spatialement par des relations de dérivation).

6. Conclusion et perspectives

Dans ce travail, nous proposons d'utiliser un graphe spatio-temporel permettant de modéliser l'évolution de données agricoles dans le temps et l'espace. Ce modèle de graphe a été défini dans (Del Mondo *et al.*, 2010) dans un contexte plus général, mais avec des applications réduites. Nous l'avons adapté afin de traiter des grandes données issues du registre parcellaire graphique (RPG).

Nous proposons tout d'abord d'étendre les relations spatiales en définissant deux types de proximité adaptées aux particularités des phénomènes sur les parcelles agricoles. Nous proposons aussi d'utiliser un modèle de relations floues pour tenir compte de l'incertitude sur la localisation des mêmes parcelles dans le temps. Grâce aux identités des parcelles fournies par le RPG, nous avons également construit les relations de filiation afin d'obtenir un graphe spatio-temporel complet, comme défini par (Del Mondo, 2011).

Nous exploitons une méthode efficace pour détecter les relations entre parcelles et réduire le temps de calcul. Finalement nous nous sommes intéressés à la simplification du graphe spatio-temporel. Nous présentons un algorithme de simplification du graphe spatial avec regroupement des sommets voisins portant les mêmes attributs et suppression des relations redondantes. Ces stratégies nous ont ainsi permis de réduire la complexité des calculs de détection et d'améliorer notre modélisation.

Malgré cela, le graphe construit est complexe, même sur des ensembles de données restreints, et il devient rapidement difficile à visualiser. Des travaux sont en cours pour améliorer cet aspect. Au-delà nous travaillerons au développement de méthodes de fouille permettant de résumer les graphes, ou d'en extraire les informations pertinentes dans un cadre applicatif donné.

Bibliographie

- Armstrong M. P. (1988). Temporality in spatial databases. In *GIS/LIS 88 Proceedings: Accessing the World, Falls Church, VA*, vol. II, p. 880–889. American Society for Photogrammetry and Remote Sensing.
- Aydin B., Angryk R. A. (2016). A graph-based approach to spatiotemporal event sequence mining. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, p. 1090–1097.
- Chen J., Wu H., Li S., Liao A., He C., Peng S. (2013). Temporal logic and operation relations based knowledge representation for land cover change web services. *ISPRS journal of photogrammetry and remote sensing*, vol. 83, p. 140–150.
- Clementini E., Di Felice P. (1997). Approximate topological relations. *International Journal of Approximate Reasoning*, vol. 16, n° 2, p. 173–204.
- Cohn A. G., Gotts N. M. (1996). The "Egg-Yolk" representation of regions with indeterminate boundaries. In P. A. Burrough, A. U. Frank (Eds.), *Geographic objects with indeterminate boundaries*, p. 171-187. London, Taylor & Francis.
- Degenne P., Lo Seen D. (2016). Ocelet: Simulating processes of landscape changes using interaction graphs. *SoftwareX*, vol. 5, p. 89–95.
- Del Mondo G. (2011). *Un modèle de graphe spatio-temporel pour représenter l'évolution d'entités géographiques*. Thèse de doctorat non publiée, Université de Bretagne occidentale-Brest.
- Del Mondo G., Stell J. G., Claramunt C., Thibaud R. (2010). A graph model for spatio-temporal evolution. *Journal of Universal Computer Science*, vol. 16, p. 1452–1477.
- Egenhofer M. J., Al-Taha K. K. (1992). Reasoning about gradual changes of topological relationships. In *Theories and methods of spatio-temporal reasoning in geographic space*, p. 196–219.
- Guttman A. (1984). R-trees: A dynamic index structure for spatial searching. In *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data*, p. 47–57. New York, NY, USA, ACM.
- Guyet T., Moinard Y. (2014). Programmation par ensembles réponses pour simuler l'assolement d'un paysage. In *Actes de la conférence RFIA 2014, Rouen, France*.

Un graphe spatio-temporel de parcelles

- Hornsby K., Egenhofer M. J. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, vol. 36, n° 1-2, p. 177–194.
- Jiang B., Claramunt C. (2004). Topological analysis of urban street networks. *Environment and Planning B: Planning and design*, vol. 31, n° 1, p. 151–162.
- Khiali L., Ienco D., Teisseire M. (2018). Object-oriented satellite image time series analysis using a graph-based representation. *Ecological informatics*, vol. 43, p. 52–64.
- Langran G., Chrisman N. R. (1988). A framework for temporal geographic information. *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 25, n° 3, p. 1–14.
- Lavigne C., Klein E. K., Mari J.-F., Le Ber F., Adamczyk K., Monod H. *et al.* (2008). How do genetically modified (GM) crops contribute to background levels of GM pollen in an agricultural landscape? *Journal of Applied Ecology*, vol. 45, n° 4, p. 1104–1113.
- Levasseur F., Martin P., Bouty C., Barbottin A., Bretagnolle V., Théron O. *et al.* (2016). RPG Explorer: A new tool to ease the analysis of agricultural landscape dynamics with the Land Parcel Identification System. *Computers and Electronics in Agriculture*, vol. 127, p. 541–552.
- Peuquet D. J., Wentz E. (1994). An approach for time-based analysis of spatiotemporal data. *Unknown Journal*, p. 489–504.
- Poggi S., Papaix J., Lavigne C., Angevin F., Le Ber F., Parisey N. *et al.* (2018). Issues and challenges in landscape models for agriculture: from the representation of agroecosystems to the design of management strategies. *Landscape Ecology*, vol. 33, n° 10, p. 1679–1690.
- Randell D. A., Cui Z., Cohn A. G. (1992). A Spatial Logic based on Regions and Connection. In *Proceedings 3rd international conference on knowledge representation and reasoning*.
- Sriti M., Thibaud R., Claramunt C. (2005). A fuzzy identity-based temporal gis for the analysis of geomorphometry changes. In *Journal on data semantics iii*, p. 81–99.
- Worboys M. F. (1994). A unified model for spatial and temporal information. *The Computer Journal*, vol. 37, n° 1, p. 26–34.

Modélisation sémantique et logique pour une simulation multi-agent dans le contexte de gestion de catastrophe

Claire Prudhomme¹, Christophe Cruz², Frank Boochs¹

1. Institut i3mainz de l'Université des Sciences appliquées,
Lucy-Hillebrand-Str. 2, 55128 Mainz, Germany
{claire.prudhomme,frank.boochs}@hs-mainz.de

2. Laboratoire d'Informatique de Bourgogne (LIB) - EA 7534, Université de Bourgogne Franche-Comté, Bâtiment i3M rue Sully, 21000 Dijon, France
christophe.cruz@ubfc.fr

RÉSUMÉ. La gestion de catastrophe nécessite une préparation à la fois individuelle et collaborative entre les divers intervenants. Les exercices collaboratifs visent à entraîner les intervenants à appliquer les plans préparés ainsi qu'à identifier les problèmes et points d'améliorations potentiels. Ces exercices étant coûteux, la simulation informatique est un outil intéressant permettant d'évaluer la préparation à travers une plus grande diversité de contexte. Cependant, les travaux de recherche centrés sur la simulation et la gestion de catastrophe sont spécialisés sur un problème particulier plutôt que sur l'évaluation globale des plans préparés. Cette limite s'explique par le défi que constitue la réalisation d'un modèle de simulation capable de représenter et de s'adapter à une large diversité de plans provenant de divers disciplines. Les travaux présentés dans ce papier adressent ce défi en adaptant le modèle de simulation en fonction des informations et plans de gestion de catastrophes intégrés dans une base de connaissance. Le modèle de simulation créé est ensuite programmé automatiquement afin d'exécuter des expériences de simulation visant à améliorer les plans d'actions. Le résultat des expériences est analysé afin de générer de nouvelles connaissances et savoir permettant d'enrichir les plans de gestion de catastrophe dans un cycle vertueux. Ce papier présente une preuve de concept sur le plan national français Novi, pour lequel les expériences de simulation ont permis de savoir quel est l'impact de la répartition des médecins sur l'application du plan ainsi que d'identifier leur répartition.

ABSTRACT. Disaster management requires both individual and collaborative preparedness among the various stakeholders. Collaborative exercises aim to train stakeholders to apply the plans prepared and to identify potential problems and areas for improvement. As these exercises are costly, computer simulation is an interesting tool to

SAGEO'2019 - Clermont Ferrand, 13-15 novembre 2019

evaluate preparation through a wider variety of contexts. However, research on simulation and disaster management focuses on a particular problem rather than on the overall assessment of the plans prepared. This limitation is explained by the challenge of creating a simulation model that can represent and adapt to a wide variety of plans from various disciplines. The work presented in this paper addresses this challenge by adapting the simulation model based on disaster management information and plans integrated into a knowledge base. The simulation model created is then automatically programmed to perform simulation experiments to improve action plans. The results of the experiments are analyzed in order to generate new knowledge and know-how to enrich disaster management plans in a virtuous cycle. This paper presents a proof of concept on the French national Novi plan, for which simulation experiments have made it possible to know the impact of the distribution of doctors on the application of the plan as well as to identify their distribution.

MOTS-CLÉS : Gestion de catastrophe, Simulation multi-agent, Technologies du Web sémantique, Modélisation

KEYWORDS: Disaster management, Multi-agent simulation, Semantic Web technologies, Modeling

1. Introduction

La gestion de catastrophe est un domaine pluri-disciplinaire qui nécessite un travail collaboratif afin de répondre aux besoins produits par les conséquences d'une catastrophe. Ce travail collaboratif est préparé et organisé à travers la création de plans. Ces plans ont pour objectif de formaliser le fonctionnement des réponses à apporter lors de la gestion de catastrophe en fonction des risques propres à chaque localité (ex. municipalité). Les exercices à des fins d'évaluation et d'amélioration ont un coût financier important limitant la complexité et la récurrence de leur mise en place. C'est pourquoi, la simulation informatique est un support de décision souvent utilisé dans ce contexte afin d'évaluer les plans à travers une variété de scénarii de complexité variable.

2. Travaux connexes

Parmi les différents modèles de simulation existants, le modèle multi-agent est le modèle le plus utilisé dans ce contexte à travers les comportements et interactions de ses intervenants (Mishra *et al.*, 2018). Cependant, les intervenants dans la gestion de catastrophe et leur comportement varient d'une localité à une autre. Ils varient également au sein d'une même localité en fonction de la situation de catastrophe et des besoins. La réalisation d'un modèle de simulation capable de s'adapter aux diverses organisations et planifications implique

une variabilité des entités, de leurs comportements et de leurs interactions. Cela revient à une adaptation et une variation de modèle. Afin de permettre la variation de modélisation, certaines approches proposent un méta-modèle comme base de modélisation (par exemple, (Kruchten *et al.*, 2007) pour les catastrophes affectant les infrastructures critiques, (An, Christley, 2011) pour le domaine biomédical, (Boufedji *et al.*, 2017) pour les systèmes multi-agent en général). En plus d'être utilisé comme base de modélisation, le méta-modèle est combiné aux approches de programmation automatique afin de produire la simulation d'un modèle. Bien que ces approches facilitent le développement de simulation, elles ne traitent pas le problème d'adaptation de la modélisation en fonction des données métiers de gestion de catastrophe, qui requiert l'explicitation des connaissances en modèle.

3. Méthodologie

Afin d'apporter une réponse, l'approche de (Prudhomme, Cruz, Boochs, 2019) propose un système basé sur les connaissances expertes pour permettre l'évaluation de la préparation. Ce système utilise les technologies du web sémantique afin d'intégrer et de représenter les connaissances, les données métiers du domaine de la gestion de catastrophe, ainsi que les connaissances de simulation multi-agent. Ces connaissances permettent d'adapter le modèle aux divers scénarii de catastrophe ainsi qu'aux connaissances métiers liées à la préparation de la gestion de catastrophe pour une localité donnée.

Cette proposition repose sur un système d'information géographique et sémantique, une base de connaissance et une librairie contenant les composants de base implémentés d'un système multi-agent. La génération des simulations pour évaluer la préparation à la gestion de catastrophe repose sur trois processus. Le premier intègre les données et connaissances métiers de gestion de catastrophe dans une ontologie de ce domaine, appelée "semDM". Ce premier processus est basé sur l'approche de (Prudhomme, Homburg *et al.*, 2019) qui correspond à l'extension de l'approche (Prudhomme *et al.*, 2017) et est inclus dans le système d'information. Celui-ci, utilisant *GeoSPARQL*, permet d'interpréter sémantiquement et de manipuler divers types de données dont les données géospatiales. Le second processus crée le modèle de simulation correspondant aux données intégrées, selon un méta-modèle de simulation multi-agent. Ce méta-modèle est représenté par une ontologie, appelé "semMAS", qui définit le modèle de simulation multi-agent en fonction de l'ontologie "semDM". Ces deux ontologies ("semDM" et "semMAS") sont contenues dans la base de connaissance et permettent de représenter les deux domaines de connaissances intervenant dans ce processus. Enfin le dernier processus est une approche de programmation automatique basée sur le méta-modèle de simulation et utilisant la librairie de simulation multi-agent implémentée selon le méta-modèle.

Le modèle de simulation programmé permet l'exécution d'expériences d'application des plans dans lesquels les informations non définies sont considérées

comme des variables dont nous souhaitons étudier l'impact sur la résolution des objectifs du plan. L'étude de l'impact de ces variables permet d'évaluer l'importance de cette variable sur l'ensemble du plan et de produire des améliorations des plans. Cette étude est réalisée par un module qui analyse les résultats d'expérimentation, effectue de nouvelles expérimentations si besoin pour raffiner les observations et tirer des conclusions. Ces conclusions se traduisent par un enrichissement de la base de connaissance afin d'améliorer les plans préparés. La figure 1 présente une vue d'ensemble de cette méthodologie.

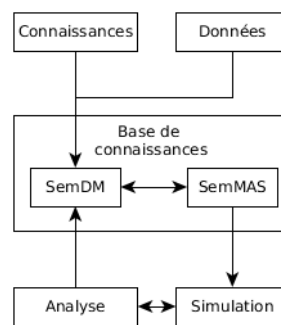


Figure 1. Vue d'ensemble de la méthodologie

4. Preuve de concept

La preuve de concept de cette méthodologie a été appliquée sur un plan français de gestion de catastrophes en cas de victimes nombreuses. Ce cas d'étude est décrit dans la première sous-section. La seconde sous-section présente sa modélisation dans le domaine de gestion de catastrophes, puis le modèle multi-agent de simulation correspondant. Le modèle conceptuel multi-agent résulte de l'interprétation automatique du modèle de gestion de catastrophes. Le modèle conceptuel multi-agent a l'avantage d'être indépendant de toute plateforme, ce qui permet de le partager et d'utiliser la plateforme souhaitée pour exécuter la simulation. Cependant, l'exécution de la simulation requiert un modèle de programmation correspondant au modèle conceptuel et à la plateforme choisie. Enfin, la dernière sous-section présente les résultats de la simulation et l'enrichissement de la base de connaissance résultant de l'application de la méthodologie.

4.1. Description du cas d'étude

4.1.1. Le plan évalué

Le cas d'étude concerne le plan français national, appelé *Plan Novi*, visant à gérer un grand nombre de victimes (plus de 100 victimes). Ce plan est basé

sur un système de triage des victimes visant à optimiser leur gestion. Il consiste à définir un poste médical avancé à la fois proche de la catastrophe et proche de route permettant un accès rapide aux secours. Ce poste médical avancé est un point intermédiaire entre la zone de catastrophe (parfois également appelée zone de danger) et l'hôpital, cela peut être un bâtiment existant ou un système modulaire (tentes). Il vise à trier les victimes afin de les répartir de manière adaptée dans les hôpitaux et éviter une surcharge d'un hôpital qui pourrait engendrer une moins bonne prise en charge des victimes. Les sauveteurs ou autres intervenants dans la gestion de catastrophe (ex. la police ou le groupe d'intervention de la gendarmerie nationale en cas de terrorisme) habilités à intervenir sur la zone de catastrophe ont pour objectif de secourir les victimes, ce qui consiste à rechercher, identifier et évacuer les victimes de la zone de danger vers le poste médical avancé. Cette évacuation est appelée *Petite Noria* ou *Noria de ramassage*. Les victimes acheminées au poste médical avancé y reçoivent des premiers soins de stabilisation et sont de nouveau triées afin d'être évacuées vers un hôpital. L'évacuation des victimes vers un hôpital est appelée *Grande Noria* ou *Noria d'évacuation* et est réalisée majoritairement par des ambulances de la structure mobile d'urgence et de réanimation (SMUR).

4.1.2. Données et informations du cas d'étude

Le cas d'étude consiste à évaluer le plan Novi dans la petite ville de Montbard (Côte d'Or, France). Ce cas d'étude a donc requis l'intégration des données géographiques de la ville telles que celles représentant les routes et les bâtiments. Les données intégrées dans le cadre de la gestion de catastrophe sont les effectifs des pompiers (24), des médecins (10) et des ambulances (4).

4.2. Génération des expériences de simulation

La génération de simulations est basée sur les connaissances et informations de gestion de catastrophe contenues dans une base de connaissances. Le plan Novi ainsi que les données précédemment présentés ont été intégrés à la base de connaissance afin de générer le modèle conceptuel de simulation, puis le modèle programmé et enfin les expériences de simulation.

4.2.1. Du modèle de gestion de catastrophe au modèle programmé de simulation

Les modèles de gestion de catastrophe et de simulation multi-agent contenus dans la base de connaissance ont été présentés dans (Prudhomme, Cruz, Boochs, 2019). Grâce à la définition des concepts de simulation multi-agent par rapport à ceux de gestion de catastrophe. L'analyse logique de la base de connaissance permet de peupler le modèle conceptuel multi-agent. La plateforme GAMA (Taillandier *et al.*, 2019) a été choisie pour exécuter les expériences de simulation en raison de son avantage à créer un environnement basé sur des données géographiques réelles. Cette plateforme utilise son propre

langage *GAML* basé sur le langage de programmation Java, les concepts du langage *GAML* ont été ajoutés à la base de connaissance afin de permettre une programmation automatique du modèle conceptuel multi-agent pour cette plateforme. Afin d'illustrer le résultat de la méthodologie appliquée à ce cas

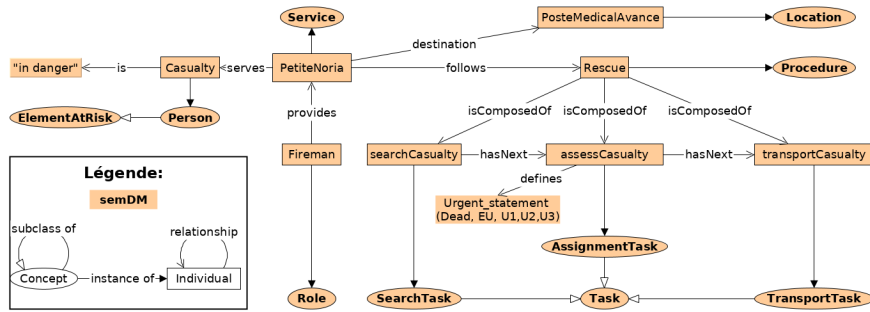


Figure 2. Illustration du contenu de la base de connaissance représentant la gestion de catastrophe

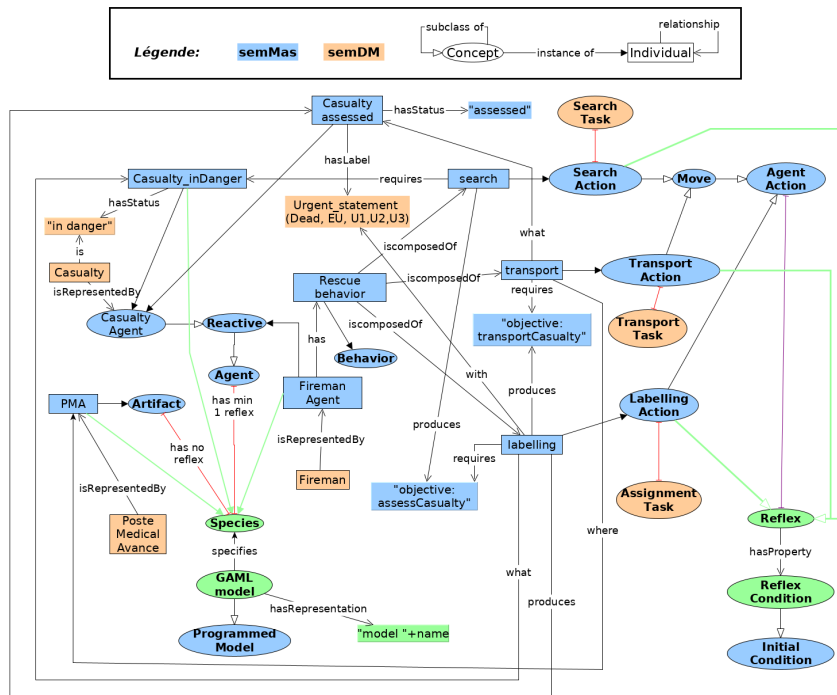


Figure 3. Illustration du contenu de la base de connaissance représentant la modélisation multi-agent ainsi que ses relations de contraintes (en rouge) et de classes équivalentes (en violet)

d'étude, la figure 2 illustre une portion simplifiée de la base de connaissance après intégration des connaissances de gestion de catastrophe et la figure 3 illustre la modélisation multi-agent correspondante après l'application du raisonnement logique. Dans cet exemple, le rôle de pompier donne naissance à un agent réactif de type pompier. Cet agent possède les capacités (ensemble d'actions possibles) correspondant aux tâches que peut accomplir un pompier en cas de catastrophe. Chacune de ses capacités repose sur des actions d'agent ou *reflex* en GAML déjà implémentés. Quelques concepts spécifiques à GAML sont représentés dans la figure 3 en vert. Le peuplement des concepts de simulation multi-agent dans la base de connaissances permet ainsi de créer et instancier des agents dynamiquement.

4.2.2. Génération des expériences de simulation

Afin de simuler une catastrophe dont le nombre de victimes est supérieur à 100, la salle Paul Éluard de Montbard a été sélectionné comme lieu de catastrophe pour sa capacité d'accueil de 750 personnes et son usage fréquent pour l'accueil d'évènements. L'école la plus proche de la salle Paul Éluard a été définie comme poste médical avancé en raison de sa proximité avec la zone de catastrophe, son accès facilitant les entrées et sorties de victimes, ainsi que ses équipements d'urgences déjà sur place. Le plan Novi présente une répartition des médecins entre l'hôpital et le poste médical avancé sans préciser de proportion de répartition, ce qui engendre une prise de décision au moment de la gestion d'une catastrophe pouvant impacter son efficacité. Par conséquent, l'effectif des médecins envoyés au poste médical avancé représente la variable observée et étudiée dans ce cas d'étude. Les autres variables d'entrée de cette simulation sont le nombre de victimes et leur répartition parmi les cinq échelons d'urgence de victimes associés au triage du plan Novi (Urgence dépassée/décédé (D), les urgences absolues: extrême urgence (EU) et blessé grave (U1), les urgences relatives: blessé sérieux (U2) et blessé léger(U3)).

La capacité de la salle Paul Éluard étant de 750 personnes, le nombre de victimes a été fixé en proportion de sa capacité, soit une configuration correspondant à $\frac{1}{3}$ de sa capacité (250 victimes), une configuration correspondant à $\frac{2}{3}$ de sa capacité (500 victimes), et une configuration correspondant à sa capacité totale (750 victimes). Pour chacune de ces configurations la répartition de victimes parmi les différents échelons d'urgence a été tirée aléatoirement. La répartition des médecins entre le poste médical avancé et l'hôpital a été étudiée en pourcentage avec une définition de configuration tous les 10%. Cent expérimentations ont été exécutées pour chaque combinaison de configuration. La figure 4 présente la visualisation d'une de ces expérimentations.

4.3. Résultats

La méthodologie appliquée à ce cas d'étude fournit dans un premier temps des résultats d'expérimentations de simulation présentés dans la première sous-

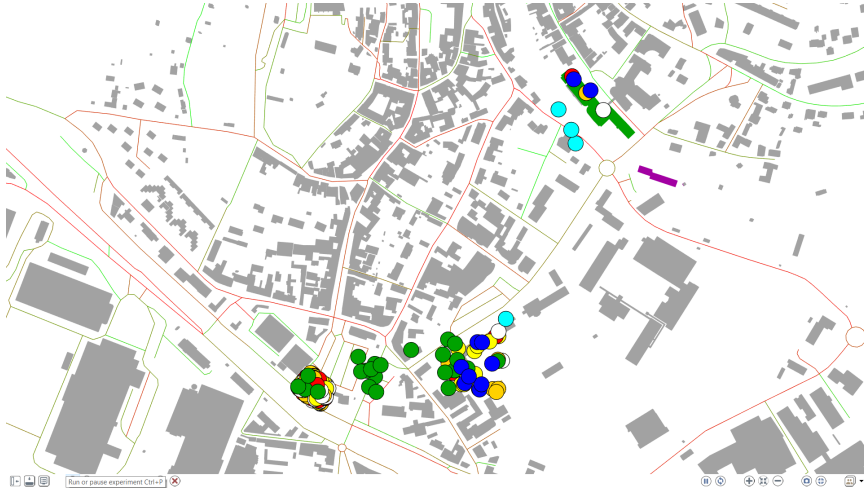


Figure 4. Vue de la simulation

section. Puis, ces résultats sont analysés afin de produire de nouvelles connaissances permettant d'enrichir la base de connaissance. Les nouvelles connaissances résultant de ce cas d'étude sont présentées dans la seconde sous-section.

4.3.1. Résultats des expérimentations

Les expériences de simulation ont permis d'observer l'impact de la proportion des médecins assignés au poste médical avancé. Le graphique de la figure 5 montre le résultat en terme de décès en fonction de la proportion des médecins assignés au poste médical avancé. Chacune de ces valeurs est exprimée en pourcentage. Le graphique représente trois courbes, une pour chaque configuration du nombre de victimes (250, 500, et 750 victimes). L'observation de ces trois courbes montre que plus le nombre de victimes est grand, plus le nombre de décès l'est également. Cette évolution naturelle traduit le manque d'effectifs des médecins par rapport à la quantité de victimes.

Dans chacune des configurations de victimes, la proportion optimale de médecins assignés au poste médical avancé est de 80%.

Une seconde expérimentation de la méthodologie a été appliquée sur un cas d'étude similaire, c'est-à-dire sur l'application du plan Novi mais dans un autre contexte (une autre ville, et donc d'autres données en entrées). La figure 6 montre les résultats obtenus pour ce second cas d'étude. Pour ce second cas d'étude, dans la configuration de 50 victimes, la proportion optimale de médecins assignés au poste médical avancé est de 60%. En revanche, dans les configurations de 300 et 550 victimes, la proportion optimale de médecins est de 50%.

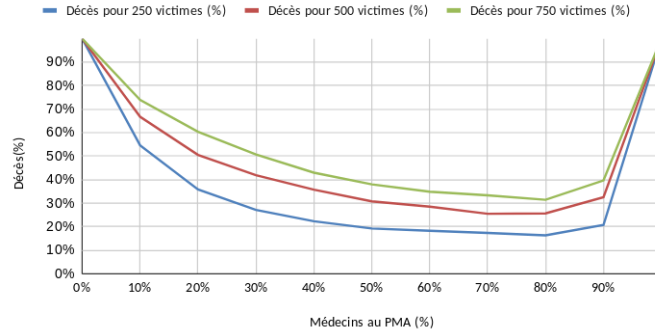


Figure 5. Graphique représentant le pourcentage de décès des victimes en fonction du pourcentage de médecins au poste médical avancé du premier cas d'étude

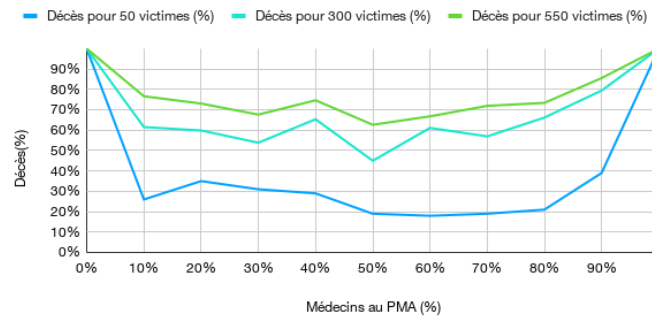


Figure 6. Graphique représentant le pourcentage de décès des victimes en fonction du pourcentage de médecins au poste médical avancé du second cas d'étude

4.3.2. Savoir et connaissances enrichissant la base de connaissance

L'analyse des résultats précédemment présentés permet d'enrichir la base de connaissance. Pour le premier cas d'étude sur la ville de Montbard, le résultat de la proportion optimale de médecins assignés au poste médical avancé ayant montré une constance de valeurs pour chaque configuration de victimes, permet d'enrichir la base de connaissance par l'ajout d'un fait traduisant le besoin d'envoyer 80% des médecins au poste médical avancé.

Pour le second cas d'étude, les résultats montrent une constance de valeurs dans deux configurations successives sur trois (configurations de 300 et 550 victimes). Ces résultats permettent d'enrichir la base de connaissance par la règle logique présentée par l'équation 1.

$$\begin{aligned}
 & \text{hasQuantity}(\text{Casualty}, nbc) \sqcap \text{greaterEqThan}(nbc, 300) \sqcap \\
 & \text{lowerEqThan}(nbc, 550) \rightarrow \text{requiredProportion}(\text{ampManagement}, 0.5) \quad (1)
 \end{aligned}$$

Les résultats, montrant une variation de valeur optimale dans la configuration de 50 victimes comparée à la configuration de 300 victimes, permettent tout d'abord d'enrichir la base de connaissance par la règle logique présentée dans l'équation 2.

$$\begin{aligned} & hasQuantity(Casualty, nbc) \sqcap equals(nbc, 50) \\ \rightarrow & requiredProportion(ampManagement, 0.6) \end{aligned} \quad (2)$$

Puis, le système utilise une extension de fonctionnalité de règle, appelée un *built-in*, permettant le calcul d'une interpolation linéaire afin d'enrichir la base de connaissance. Dans ce contexte le *built-in* prend les 5 valeurs suivantes en paramètre: x , le nombre de victimes de la catastrophe; v_{v0} , le nombre minimal de victimes (ici 50); v_{v1} , le nombre maximal de victimes (ici 300); v_{p0} , le pourcentage de médecins associé au nombre minimal de victimes (ici 0,6); et v_{p1} , le pourcentage de médecins associé au nombre maximal de victimes (ici 0,5). À partir de ces paramètres le *built-in*, appelé *LinearInterpo*, calcule une estimation de la proportion de médecins adaptée au nombre de victimes, lorsque celui-ci est compris entre 50 et 300 (exclus). L'équation 3 présente le calcul de cette estimation à travers la fonction $f(t)$, où $t = \frac{x-v_{v0}}{v_{v1}-v_{v0}}$.

$$f(t) = (1 - t)v_{p0} + t \times v_{p1} \quad (3)$$

Basée sur ce *built-in*, la base de connaissance est enrichie de la règle logique présentée dans l'équation 4.

$$\begin{aligned} & hasQuantity(Casualty, nbc) \sqcap greaterThan(nbc, 50) \sqcap lowerThan(nbc, 300) \\ \rightarrow & requiredProportion(ampManagement, LinearInterpo(nbc, 50, 300, 0.6, 0.5)) \end{aligned} \quad (4)$$

5. Conclusion

Ce papier a permis d'illustrer dans les grandes lignes la génération automatique d'expérimentations de simulation basée sur les connaissances et informations de gestion de catastrophes visant à améliorer les plans d'actions et la préparation à la gestion de catastrophe. Bien que ce cas d'étude soit limité à l'étude d'une seule variable, il permet d'illustrer le fonctionnement du système. Les expériences de simulation permettent d'évaluer les plans préparés et d'étudier les bonnes décisions à prendre relatives au plan. Cette étude permet ainsi d'enrichir la base de connaissance sur la gestion de catastrophe propre à un contexte d'étude. Cet enrichissement de la base de connaissance pourrait également permettre à l'avenir de supporter la mise en place des plans lors d'une catastrophe, et d'aider à la décision lors des réponses à produire après une catastrophe.

Remerciement

Ces travaux de recherche sont soutenus par le ministère fédéral allemand de l'éducation et de la recherche ("Bundesministerium für Bildung und Forschung")[numéro de subvention 03FH032IX4].

Nous remercions également la ville de Montbard et plus particulièrement madame Porte, maire de Montbard, pour sa collaboration vers de plus amples cas d'étude.

Bibliographie

- An G., Christley S. (2011). Agent-based modeling and biomedical ontologies: a roadmap. *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 3, n° 4, p. 343–356.
- Boufedji D., Guessoum Z., Brandão A., Ziadi T., Mokhtari A. (2017). Towards a mas product line engineering approach. In *International workshop on engineering multi-agent systems*, p. 161–179.
- Kruchten P., Woo C., Monu K., Sotoodeh M. (2007). A human-centered conceptual model of disasters affecting critical infrastructures. In *Proceedings of the 4th international conference on information systems for crisis response management (iscram)*, p. 327–344.
- Mishra D., Kumar S., Hassini E. (2018). Current trends in disaster management simulation modelling research. *Annals of Operations Research*, p. 1–25.
- Prudhomme C., Cruz C., Boochs F. (2019). Semantic and logic modeling of disaster simulation for multi-agent systems. *International Journal of Modeling and Optimization*.
- Prudhomme C., Homburg T., Ponciano J.-J., Boochs F., Cruz C., Roxin A.-M. (2019). Interpretation and automatic integration of geospatial data into the semantic web. *Computing*, p. 1–27. Consulté sur <https://doi.org/10.1007/s00607-019-00701-y>
- Prudhomme C., Homburg T., Ponciano J.-J., Boochs F., Roxin A., Cruz C. (2017, avril). Automatic integration of spatial data into the semantic web. In *Webist 2017*. Porto, Portugal. Consulté sur <https://hal-univ-bourgogne.archives-ouvertes.fr/hal-01493390>
- Taillandier P., Gaudou B., Grignard A., Huynh Q.-N., Marilleau N., Caillou P. *et al.* (2019). Building, composing and experimenting complex spatial models with the gama platform. *GeoInformatica*, vol. 23, n° 2, p. 299–322.

Estimer l'apport de la phénologie dans la classification des essences forestières

Nicolas Karasiak¹, Jean-François Dejoux²,
Claude Monteil¹, David Sheeren¹

1. DYNAFOR, Université de Toulouse, INRA, Castanet-Tolosan, France
nicolas.karasiak@inra.fr

2. CESBIO, Université de Toulouse, CNES/CNRS/INRA/IRD/UPS,
Toulouse, France

n/a

RÉSUMÉ. L'idée de créer une cartographie des essences forestières de manière automatique est presque aussi ancienne que la télédétection. Si les capteurs n'ont cessé d'évoluer de par leur résolution spatiale, spectrale et temporelle, force est de constater qu'il n'y a toujours pas de carte réalisée de manière automatique aujourd'hui. Cet article se focalise sur l'apport de la phénologie (i.e. le cycle annuel de la végétation) des différentes essences de feuillus dans la cartographie supervisée des forêts. Pour ce faire, l'ensemble des images du satellite Sentinel-2 a été utilisé pour les années 2017 et 2018, et des observations in situ ont eu lieu tous les 10 jours au printemps et à l'automne. Les premiers résultats montrent une adéquation entre les données terrains et la sélection des dates pour cartographier les feuillus. Cependant, lors d'une étude pour chaque espèce, l'interprétation des dates en lien avec la phénologie reste ardue.

ABSTRACT. Mapping tree species is an old question, almost as long as remote sensing. Spatial, spectral and temporal resolution are constantly improving, but there is still no automatic map of tree species at a national scale. This paper studies the contribution of phenology (i.e. vegetation annual cycle) of different broadleaved species in supervised per pixel classification. To achieve this analysis, every image from Sentinel-2 satellite was used for 2017 and 2018. Moreover, a survey to collect in situ phenology was made during spring and autumn. The first results highlight that the observations from the phenology survey are consistent with the selection of dates to spatially predict all broadleaved. However, when studying each species, the understanding of selected dates in relation to phenology remains more complicated.

MOTS-CLÉS : Sentinel-2, Série temporelle, satellite, phénologie, cartographie, forêt

KEYWORDS: Sentinel-2, Time Series, satellite, phenology, mapping, forest

1 Introduction

La forêt joue un rôle majeur dans la préservation de la biodiversité mondiale (Thompson *et al.*, 2011). L'organisme de l'ONU dédié à l'alimentation et à l'agriculture (la FAO) définit la forêt comme un espace d'au moins 0,5 hectare (5 000 m²) avec des arbres d'une hauteur minimum de 5 mètres et un couvert forestier de plus de 10%. Rentre dans cette définition une multitude d'espaces allant des plantations mono-spécifiques à cycle court, aux forêts primaires de plusieurs centaines d'années et avec une diversité importante. Or, plus une forêt est diversifiée en essences, plus elle pourra, par exemple, bloquer la menace d'insectes et de pathogènes invasifs (Guo *et al.*, 2019). Cartographier les espèces qui composent la forêt est par conséquent indispensable.

Depuis plusieurs décennies, l'IGN cartographie les grands types de forêts françaises en photo-interprétant des images à très haute résolution spatiale (Boureau, 2008). Cependant, il s'avère que ces images s'adaptent assez mal à la classification supervisée et ne donnent pas des résultats satisfaisants (Immitzer *et al.*, 2012).

Outre la résolution spatiale, il est possible d'améliorer la résolution spectrale, cela signifie qu'il y aura plus de bandes dans un même domaine spectral (bleu, vert, rouge, proche infra-rouge). Les images de ces travaux proviennent de missions aéroportées et ne couvrent qu'un petit territoire en raison du coût d'acquisition et de l'absence de satellite de ce type (Ghiyamat, Shafri, 2010). Les résultats encourageants sont assez variés et dépendent de plusieurs facteurs : nombre d'essences, algorithme, résolution spatiale et spectrale.

Concernant la résolution temporelle, la plupart des travaux utilisent des satellites à faible résolution spatiale comme MODIS ou Landsat (Pasquarella *et al.*, 2018; Aragonés *et al.*, 2019). Si les résultats sont aussi encourageants, les limitations induites par la taille d'un pixel ne permettent pas de cartographier les forêts tempérées qui sont à la fois variées en essences, disséminées sur le territoire et parfois de taille restreinte.

Au-delà de cartographier les espèces et de constater les erreurs, il est primordial de comprendre pourquoi certaines essences sont mieux prédites que d'autres. L'une des explications est que selon la saison, des différences peuvent apparaître entre les essences. Des nombreuses études ont montré que l'ajout de dates au printemps ou à l'automne augmente la qualité des modèles (Fassnacht *et al.*, 2016). Il est cependant difficile de quantifier l'apport de la phénologie dans la cartographie des essences tant les espèces et les dates d'acquisition sont différentes entre les études.

L'arrivée en 2016 du satellite Sentinel-2 avec ses 5 jours de résolution temporelle et ses nombreuses bandes dans l'infrarouge devrait apporter des éléments de réponses sur la possibilité de cartographier les essences forestières. Aujourd'hui, seul un faible nombre de travaux utilise au moins une dizaine d'acquisitions pour classer les essences à partir d'un satellite à haute résolution

spatiale (Sheeren *et al.*, 2016; Grabska *et al.*, 2019), alors que des comportements spectraux différents entre espèces pourraient s’observer à certaines dates clés (Fassnacht *et al.*, 2016).

Malgré les avancées technologiques (capteurs, algorithmes), il est aujourd’hui difficile d’estimer le réel potentiel de la télédétection dans la cartographie des essences. La plupart des travaux antérieurs n’ont pas tenu compte l’indépendance spatiale des données d’entraînement. Ainsi, plus l’autocorrélation spatiale est importante plus la qualité prédictive estimée sera biaisée par rapport à la qualité réelle (Mannel *et al.*, 2011). Quelques chercheurs se sont néanmoins intéressés à ce biais en télédétection des essences forestières dès les années 90 (Congalton, 1991). Plus récemment, c’est dans une revue de littérature que Fassnacht *et al.* (2016) ont souligné le peu de travaux ayant pris en compte ce biais.

À notre connaissance, l’apport de la phénologie dans la classification des feuillus et la compréhension de l’interaction entre les phénomènes observés in situ et l’observation satellite n’ont jamais fait l’objet à partir d’un capteur de haute résolution spatiale, spectrale et temporelle. Afin d’estimer au mieux l’apport de la phénologie, et comme le nombre important de dates et donc de variables spectrales peut réduire la qualité des modèles, un protocole de sélection de caractéristiques a été réalisé de manière systématique pour estimer les meilleures dates.

Cet article a pour objectifs (i) de réaliser une cartographie des essences forestières, (ii) d’identifier les bandes spectrales et les dates les plus importantes, pour finalement (iii) voir s’il y a une cohérence entre les résultats précédemment obtenus et les observations réalisées in situ à l’automne et au printemps.

2 Matériel

2.1 Zone d’étude

La zone d’étude se situe au sud ouest de Toulouse. Elle couvre un territoire de 24 km x 24 km (Figure 1). Cette limitation correspond à l’emprise du satellite Formosat-2 sur laquelle le CNES a un suivi régulier avec des acquisitions privilégiées à cet endroit grâce à l’Observatoire Spatial Régional fondé en 2006 par le CESBIO.

La Garonne traverse la zone d’étude à l’est du nord au sud et influence la composition du sol et la topographie. Le climat est typique du subatlantique avec un automne ensoleillé, un été chaud et sec et un hiver doux et pluvieux (température annuelle moyenne $> 13^{\circ}\text{C}$; précipitation = 656 mm). La forêt recouvre un peu plus de 10% de la scène soit 53 km².

2.2 Images

Les images utilisées proviennent des satellites Sentinel-2 et sont composées de 4 bandes à 10-m de résolution spatiale (bleu, vert, rouge et proche infra-rouge), et de 6 bandes à 20-m allant du red-edge (portion du spectre juste après

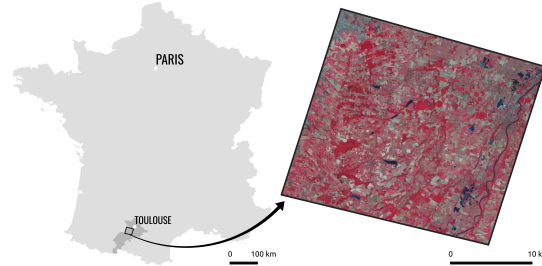


FIGURE 1 – Zone d'étude, image en fausse couleur infrarouge.

le rouge mais avant le proche infra-rouge) au proche et moyen infra-rouge. La résolution radiométrique est de 12 bits. Les images ont été téléchargées via la plateforme THEIA. Ces données sont disponibles directement au niveau 2A (i.e. produit ortho-rectifié et en réflectance) et incluent un masque de nuages et d'ombres. Ces masques proviennent de l'algorithme MAJA qui inclut la correction atmosphérique en combinant des critères multitemporel et multispectral pour estimer l'épaisseur optique des aérosols (Hagolle *et al.*, 2015).

Le deuxième satellite Sentinel-2B qui vient doubler la résolution temporelle à 5 jours a débuté sa mission courant été 2017, ce qui explique le plus faible nombre d'acquisition au printemps 2017 (Figure 2).

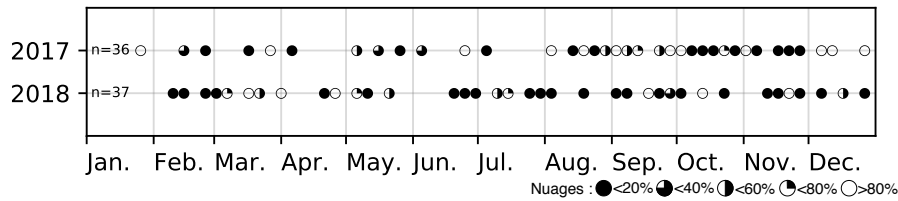


FIGURE 2 – Liste des dates auxquelles les images Sentinel-2 ont pu être acquises pour 2017 et 2018. La symbologie correspond au pourcentage de nuages détectés dans les références terrain.

Le masque des forêts est issu de la BDForêt v.2 (IGN[®]). Cette donnée a été produite à partir de la BDOrtho (50cm) datant de 2013. Des améliorations ont été effectuées par photo-interprétation à partir d'images Sentinel-2 et concernent uniquement les coupes forestières.

2.3 Références terrain

Quatre campagnes de relevés ont été effectuées entre novembre 2013 et janvier 2017 afin d'identifier et de référencer les essences forestières de la zone d'étude. Les 72 principaux massifs forestiers ont été visités. Seules les espèces dominantes de feuillus ou de résineux ont été sélectionnées. Pour s'assurer de l'homogénéité de la donnée, les références ont été acquises à partir d'un récepteur GPS Garmin GPSMap 62st (3-5 m de précision) au centre d'une zone couvrant approximativement 576 m² (i.e. neuf pixels contigus du satellite For-

TABLE 1 – Liste des essences forestières et leur effectif en pixel ($n = 1626$). Le nombre de peuplements a été calculé à partir de la donnée (IGN BDForêt® v.1)

Espèce	Effectif	Peuplement
Feuillus		
Bouleau (<i>Betula pendula</i>)	67	3
Chêne (<i>Quercus robur/pubescens/petraea</i>)	247	12
Chêne rouge(<i>Quercus rubra</i>)	205	9
Peuplier (<i>Populus spp.</i>)	139	7
Frêne (<i>Fraxinus excelsior</i>)	130	5
Robinier (<i>Robinia pseudoacacia</i>)	88	9
Saule (<i>Salix spp.</i>)	52	4
Eucalyptus* (<i>Eucalyptus spp.</i>)	143	5
Résineux		
Pin laricio (<i>Pinus nigra subsp. laricio</i>)	201	9
Pin maritime (<i>Pinus pinaster</i>)	157	8
Pin noir (<i>Pinus nigra</i>)	52	2
Sapin blanc (<i>Abies alba</i>)	53	4
Sapin douglas (<i>Pseudotsuga menziesii</i>)	50	8
Cyprès (<i>Cupressus</i>)	42	1

*Seul feuillus à feuilles persistantes présent dans la zone d'étude.

mosat, soit par pixel $8\text{ m} \times 8\text{ m}$, voir Sheeren *et al.* (2016)). Seuls les pixels au centre de ces zones ont été utilisés pour la classification.

Au total il a été récolté 1626 références terrain pour 14 espèces, dont 8 feuillus et 6 conifères (Table 1). Pour certaines essences, l'identification a été limitée au genre en raison de l'existence de nombreux cultivars (comme pour le peuplier) ou de la difficulté à déterminer l'espèce exacte (chêne, saule ou eucalyptus). La distribution par classe est déséquilibrée mais reflète l'abondance des essences dans ces forêts.

2.4 Suivi phénologique

Parmi l'ensemble des références présentées précédemment, il a été choisi deux peuplements distincts pour les 7 essences à feuillage caduc. Cette sélection s'est effectuée à l'automne 2017 en ne gardant que les peuplements sains (aucune maladie significative n'a été observée).

Afin d'avoir une mesure constante dans le temps et dans l'espace, une placette de 100m^2 destinée à observer la phénologie a été localisée dans chaque peuplement sélectionné. Cette placette a été matérialisée sur le terrain en implantant un piquet en son centre, correspondant au centre d'un pixel Sentinel-2. La localisation s'est faite à l'aide d'un GPS différentiel de type GPS Geo7X de Trimble avec une précision de l'ordre du mètre.

Ainsi, depuis l'automne 2017, un suivi régulier est effectué tous les 10 jours et ce, simultanément au passage d'un des deux satellites Sentinel-2. Les relevés

s'effectuent au printemps et à l'automne. Les différents stades phénologiques des essences (débourrement, développement des feuilles, floraison ou encore sénescence) ont été transcrits en s'appuyant sur le protocole BBCH (Badeau *et al.*, 2017). Le code BBCH est un nombre composé de deux chiffres, le premier représentant le stade phénologique, le second l'accomplissement du stade. Par exemple, le code 15 signifie que l'arbre est au stade 1 (développement des feuilles) et que la majorité des feuilles (5 pour 50%) sont déployées. Afin d'être le plus homogène possible entre les observations terrain et les observations satellite, la phénologie notée correspond à l'observation moyenne des individus de la placette.

Pour chacun des 14 plots, le protocole s'est déroulé en 4 étapes :

1. Photographie de la parcelle
2. BBCH moyen de la ligneuse dominante
3. Commentaire sur l'état du sous-bois
4. Photographie hémisphérique (DHP) de la canopée.

3 Méthode

Un schéma simplifié du processus est présenté en figure 3.

3.1 Pré-traitement

Pour chacune des deux années, une Série Temporelle d'Images Satellites (SITS) a été générée en utilisant une interpolation linéaire pour chaque date à laquelle un nuage a été détecté. Le but est de supprimer les données bruitées (i.e. pixel nuageux ou ombré) et de retrouver une valeur spectrale de réflectance supposée cohérente. Plus précisément, les pixels invalides ont été remplacés en interpolant linéairement les valeurs des deux pixels valides les plus proches temporellement (avant et après).

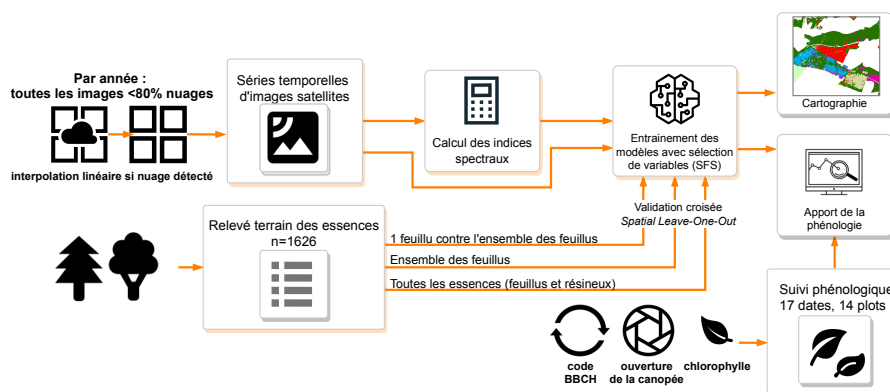


FIGURE 3 – Protocole de classification, répété pour chaque année (2017 et 2018) en veillant à bien utiliser les mêmes échantillons de validation et d'entraînement.

En plus des données spectrales, 10 indices spectraux reflétant entre autre l'activité de la végétation (teneur en chlorophylle, anthocyanine et caroténoïdes) ont été calculés pour chaque année.

3.2 Apprentissage des modèles

3.2.1 Algorithme

Pour entrainer le modèle en vue de reconnaître les essences à partir de chaque SITS, l'algorithme SVM (Support Vecteur Machine) a été choisi car il est réputé pour être efficace avec un faible nombre de données d'entraînement et un nombre important de caractéristiques. Le noyau utilisé est de type gaussien (Radial Basis Function, RBF), paramétré en validation croisée comme suit : $C = \{1^{-2}, 1^{-1}, \dots, 1^3\}$ et $\gamma = \{1^{-3}, 1^{-8}, \dots, 1^3\}$. Les variables spectrales ont été standardisées en amont de l'apprentissage.

3.2.2 Sélection des meilleures caractéristiques

Pour estimer l'apport de chaque bande spectrale (et donc de chaque date) à la qualité du modèle, nous avons opté pour une sélection de caractéristiques de type Sequential Feature Selection (SFS). L'ajout des bandes se fait de manière itérative, une à une, et la meilleure caractéristique est ajoutée à un ensemble initialement vide. Ce protocole a été utilisé de manière systématique pour évaluer les performances du modèle à prédire les différentes essences selon une difficulté d'apprentissage croissante :

1. **Chaque feuillus est classé contre l'ensemble des autres feuillus** (1 essence contre 1 classe comprenant les 6 autres essences de feuillus)
2. **Les 7 différents feuillus sont classés simultanément**
3. **Les 14 essences sont classées simultanément** à partir des valeurs spectrales uniquement.

3.2.3 Validation et apprentissage des modèles

Les différents modèles ont été évalués et appris en utilisant une validation croisée. Étant donné le biais optimiste possible lié à l'autocorrélation spatiale dans les jeux de données Mannel *et al.* (2011), nous avons opté pour une validation croisée de type spatialisée (Le Rest *et al.*, 2014).

Une validation croisée spatialisée consiste à sélectionner des échantillons de validation qui soient distants d'au moins d mètres des échantillons d'entraînement. Pour calculer cette distance d , nous avons utilisé l'indice de Moran sur chaque bande spectrale et moyenné le résultat, cela nous donne un indice allant de -1 à 1, ou 1 correspond au maximum d'autocorrélation spatiale, et 0 le cas idéal sans autocorrélation.

Nous avons fait varier de manière linéaire la distance de 1 pixel (10-m) à 100 pixels (1000-m) pour estimer la distance à laquelle la variabilité entre les valeurs spectrales du voisinage défini (de 1 à 100 pixels) est égale à la

variabilité totale du jeu de données. Dale, Fortin (2014) considère qu'en dessous de *Moran's $i=0.2$* , l'autocorrélation spatiale est quasi nulle.

Nous obtenons alors une distance de 360-m en 2017 et 250-m en 2018. Afin de comparer les résultats de ce travail avec les précédents travaux obtenus à l'aide du satellite Formosat-2, nous avons choisi de garder une distance de $d=340\text{-m}$.

Dans le cas de l'apprentissage de l'ensemble des essences, cette validation croisée est composée de 42 plis (soit 42 jeux de validation et d'entraînement) car la classe ayant le plus faible effectif, celle du cyprès, est composée de 42 échantillons. Chaque pli contient 14 références (une par essence) servant à valider le modèle validation. Le résultat des 42 prédictions est moyenné pour avoir une estimation globale de l'erreur.

4 Résultats

4.1 Classification : Un feuillus contre tous

Afin de bien comprendre l'apport des dates dans la classification des feuillus, chaque essence a été classée en la comparant à l'ensemble des feuillus (classification à deux classes et non en gardant les 7 classes initiales).

Pour les deux années, aucune essence n'est classée avec un κ inférieur à 0.9 (Figure 4 pour 2018). Le frêne est l'essence la moins bien reconnue lors des deux années (respectivement $\kappa = 0.90$ et 0.91). Les autres essences ont des prédictions proches ou égales à $\kappa = 1$, ce qui est très élevée pour le problème posé.

Parmi les 37 dates disponibles en 2018 (370 bandes spectrales), et les 36 dates disponibles en 2017, il faut entre 2 et 5 bandes spectrales à des dates différentes pour prédire au mieux chaque essence. Si les dates sélectionnées varient d'une année sur l'autre, le choix des bandes spectrales est assez stable entre années pour le peuplier (bande 11), le bouleau et le chêne (bande 12), et pour le saule (bande 4, rouge).

Concernant le bouleau, le frêne ou le robinier, on observe qu'en 2018 les dates choisies sont plus précoces de plusieurs mois. À l'inverse pour le peuplier, le saule ou le chêne, des dates plus tardives ont été préférées.

En 2017 la fin de saison est privilégiée lors de la sélection (18, 23 et 28 octobre) par plusieurs essences (chêne, frêne, robinier, peuplier), alors qu'en 2018, il s'agit plutôt du début de saison, avec comme préférence le 21 avril (chêne rouge, robinier, peuplier). Seul le chêne bénéficie de manière significative des dates de l'automne 2018.

4.2 Classification de l'ensemble des feuillus

La classification de l'ensemble des feuillus s'est effectuée avec l'aide d'indices spectraux et des séries temporelles avec l'ensemble des bandes ou seulement les bandes initialement à 10m. L'indice kappa se situe en moyenne entre 0.7 et 0.9,

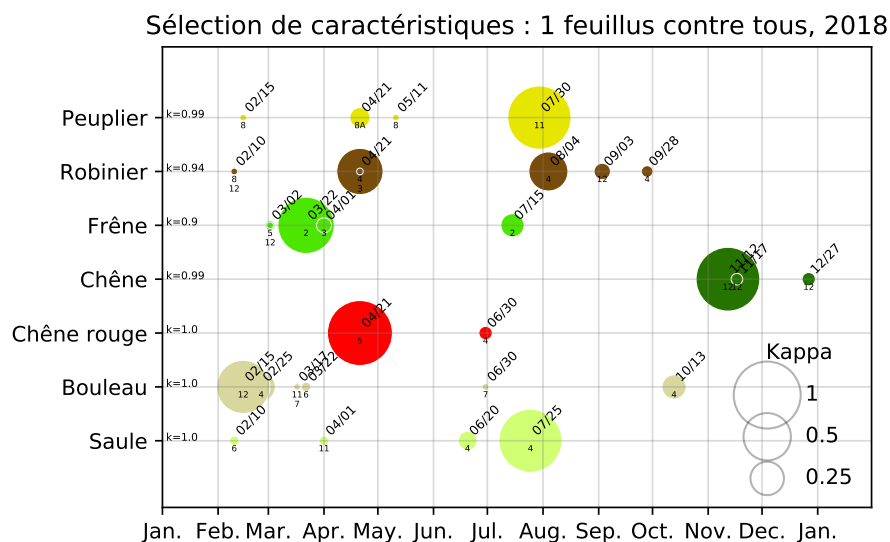


FIGURE 4 – Le kappa de la meilleure combinaison des dates est inscrit sur la gauche pour chaque essence (pour le robinier en 2018 : $k = 0.94$). La taille de chaque cercle est proportionnelle à l'apport en kappa de la bande spectrale. Le numéro sous la ligne correspond à la bande sélectionnée.

avec un maximum de performance et de stabilité interannuelle pour l'indice de chlorophylle LChloC ($\kappa^{2017} = 0.85 \pm 0.13$, $\kappa^{2018} = 0.87 \pm 0.12$). L'indice ACORVI (un NDVI avec dont le rouge a été augmenté de 0.05) a de moins bonnes performances en plus d'être instable d'une année sur l'autre (NDVI : $\kappa^{2017} = 0.84 \pm 0.14$, $\kappa^{2018} = 0.77 \pm 0.18$).

Pour l'année 2017, une tendance s'exprime avec trois périodes sélectionnées : début de saison avec fin mars et début avril, milieu de saison avec le mois de juin et tout début juillet et en fin de saison avec octobre et novembre.

Pour l'année 2018, les dates sélectionnées sont début mai, fin juin, début juillet, et fin novembre.

4.3 Classification de l'ensemble des références

La sélection de caractéristiques (réduction de la dimension des données, i.e. du nombre de variables) permet d'améliorer significativement la prédiction (table 2) passant pour l'année 2017 de $\kappa = 0.57 \pm 0.12$ à $\kappa = 0.88 \pm 0.11$ et pour 2018 de $\kappa = 0.67 \pm 0.13$ à $\kappa = 0.90 \pm 0.08$. L'ajout de 6 bandes du red-edge (cf.2.2) et de l'infra-rouge ré-échantillonnées à 10-m pour les besoins de cette étude montre aussi un apport certain en terme de qualité lors de la sélection de caractéristiques ($\Delta\kappa = 0.05$).

TABLE 2 – Qualité de la prédiction de l'ensemble des essences (feuillus et résineux) en utilisant la totalité de la série temporelle, ou avec sélection de caractéristiques (SFS) sur les 4 bandes spectrales à 10-m ou sur l'ensemble des bandes. Entre parenthèses est spécifié le nombre de bandes nécessaires pour obtenir la qualité optimale.

Année	Série temporelle (10m)	Série temporelle (10m+20m)	SFS (10m)	SFS (10m+20m)
2017	$k = 0.51 \pm 0.11$	$k = 0.57 \pm 0.12$	$k = 0.84 \pm 0.09$ (23)	$k = 0.88 \pm 0.11$ (34)
2018	$k = 0.66 \pm 0.12$	$k = 0.67 \pm 0.13$	$k = 0.84 \pm 0.12$ (45)	$k = 0.90 \pm 0.08$ (47)

5 Discussion

5.1 Explication des dates sélectionnées pour l'ensemble des feuillus

L'apport des données multitemporelles avec la sélection de caractéristiques est indéniable pour cartographier les essences forestières. Nos résultats confirment les travaux de Bolyn *et al.* (2018); Grabska *et al.* (2019) et montrent l'importance des dates au printemps avec Sentinel-2. Selon les indices utilisés, nous voyons que pour 2017 les dates du printemps sélectionnées concernent fin mai et début juin, alors que pour 2018 il s'agit plutôt de fin mars et de mi-avril. L'une des explications de la non-sélection des dates entre avril et fin mai en 2017 est qu'une seule image très ennuagée était disponible à ces dates. De manière générale, Hill *et al.* (2010) avaient montré qu'à partir d'un capteur de 2-m de résolution spatiale, la meilleure prédiction était réalisée avec une date mi-mars, une à l'été et une fin octobre, cependant seulement 5 dates étaient disponibles au total.

Les mauvaises performances de la série temporelle utilisée dans sa globalité peuvent s'expliquer par du bruit dû à des nuages non détectés et à la dimension de la donnée. Cependant, la très haute qualité des résultats de la classification des feuillus ($\kappa > 0.9$) fait penser à une mauvaise représentativité du jeu de données terrain. En effet, lors de la prédiction des 7 meilleurs modèles des 7 essences de feuillus sur un masque feuillus/résineux sur notre zone d'étude, seulement 53% des pixels ont été classés pour une seule essence, 15% par au moins deux essences, et 32% des pixels n'ont jamais été affectés par une quelconque essence. Cela peut s'expliquer par des références terrain qui ne représentent pas suffisamment la diversité spectrale des forêts.

5.2 Explication des dates et bandes sélectionnées par feuillus

En plus de réduire le nombre de dimensions, la sélection de caractéristiques apporte des informations essentielles à la compréhension de l'algorithme.

5.2.1 Sélection en début de saison

On peut observer que la bande red-edge (B5 – 705nm) est sélectionnée en premier lieu pour classer le chêne rouge. D'après les observations in situ, la date du 21 avril 2018 correspond au début de saison du chêne rouge, c'est-à-dire qu'au minimum 10% des feuilles étaient sorties. Cependant, on observe que

la bande 5 évolue positivement depuis début mars à l'inverse du rouge. Cela pourrait s'expliquer par le sous-étage : les feuilles tombées l'année précédente voient leur réflectance augmenter dans le red-edge en se décomposant.

Pour le robinier, au 11 mai 2018 les deux parcelles suivies étaient en fleur et cela concorde avec le choix de deux bandes à cette date : la B11 (infra-rouge) et B3 (vert). La détection de la floraison de cette espèce s'était déjà faite de manière précise par drone (Carl *et al.*, 2017).

Pour le bouleau, la bande 12 (SWIR, liée à la teneur en eau) a été sélectionnée au 15 février 2018, date à laquelle où aucun sous-étage ni début débourrement n'ont été observés. L'algorithme détecte donc un comportement lié à la structure ou au sol, mais pas à l'espèce. Pour l'année 2017, la bande 12 a elle aussi été sélectionnée à deux reprises, fin septembre et fin octobre. Si la saison sélectionnée diffère entre les deux années, il semblerait que la teneur en eau soit la méthode la plus simple pour identifier les bouleaux.

5.2.2 Sélection en fin de saison

Concernant le chêne, essence proche spectralement du bouleau (Sheeren *et al.*, 2016), la bande 12 est elle aussi privilégiée mais plutôt à la fin de l'automne (fin octobre et début novembre), soit vers la fin la senescence.

Enfin, si l'on compare de manière globale les bandes sélectionnées pour chaque indice spectral avec les observations terrains, on note que les bandes sélectionnées par l'algorithme sont à la fin avril/mi-mai, ce qui correspond au début de saison observée in situ. À la fin du mois d'avril les feuilles commencent à se déployer, mi-mai l'ensemble des essences avaient fini le déploiement des feuilles. L'autre période sélectionnée est mi-novembre, fin de saison observée in situ : les feuilles étaient toutes tombées ou étaient toutes colorées.

5.3 Qualité de prédiction

Si les résultats sont élevés lors de la cartographie de l'ensemble des essences, il s'avère que la répartition des feuillus varie parfois de manière importante d'une année sur l'autre. En 2017, 19.2% des feuillus prédits étaient des saules contre 12.6% en 2018, alors que cette essence n'a aucune confusion observée lors de l'apprentissage. Les autres essences varient aussi de plusieurs points de pourcentage à l'exception du peuplier. Cela montre que nos échantillons ne sont pas suffisamment représentatifs de la diversité des forêts étudiées.

6 Conclusion

Grâce aux deux satellites Sentinel-2, il a été possible d'obtenir un nombre conséquent d'images en 2017 et 2018. La sélection des caractéristiques a apporté non seulement une meilleure qualité en réduisant le nombre de dates nécessaires, mais a aussi permis d'analyser les dates et les bandes spectrales privilégiées par l'algorithme. Si de manière générale les dates choisies coïncident avec le début et la fin de saison comme il a été observé in situ, l'interprétation intraspécifique est plus complexe. Pour certaines espèces comme le robinier il a été facile de

mettre en relation les choix de l'algorithme avec la floraison, alors que pour d'autres essences il est plus difficile de comprendre certains choix comme celle de la fin de saison pour le chêne.

Malgré ces limites de compréhension, la qualité globale de prédiction est très élevée alors même qu'une validation croisée rigoureuse a évité de sélectionner les échantillons spatialement autocorrélés. Cependant, on observe que la répartition des essences n'est pas tout à fait la même d'une année sur l'autre. Cela montre que, malgré des résultats encourageants, il persiste un biais optimiste sans doute lié à la représentativité de nos références terrain.

S'il est possible de comprendre certaines relations entre les choix de l'algorithme et la phénologie, une analyse plus approfondie sera menée dans un article plus long et à partir des données phénologiques récoltées à chaque observation in situ (photos, ouverture de la canopée, sous-étage). Enfin, pour pallier aux mauvaises performances des séries temporelles sans sélection de caractéristiques, il serait intéressant d'étudier le potentiel des synthèses mensuelles produites par le pôle THEIA, qui, si les résultats sont concluants, devraient faciliter la cartographie à plus large échelle.

Matériel supplémentaire:

La cartographie des essences forestières de 2018 ainsi que de nombreux graphiques sont consultables à cette adresse :
https://nkarasiak.github.io/supplementary_sageo2019/.

Bibliographie

- Aragones D., Rodriguez-Galiano V., Caparros-Santiago J., Navarro-Cerrillo R. (2019). Could land surface phenology be used to discriminate mediterranean pine species? *International Journal of Applied Earth Observation and Geoinformation*, vol. 78, p. 281–294.
- Badeau V., Ducatillon C., Lebourgeois F. (2017). *Les plantes au rythme des saisons: guide d'observation phénologique*. (OCLC: 1027717693)
- Bolyn C., Michez A., Gaucher P., Lejeune P., Bonnet S. (2018). Forest mapping and species composition using supervised per pixel classification of sentinel-2 imagery. *Biotechnol. Agron. Soc. Environ.*
- Boureau J.-G. (2008). *Manuel d'interprétation des photographies aériennes infrarouges : Application aux milieux forestiers et naturels*. Nogent-sur-Vernisson, La Documentation Française.
- Carl C., Landgraf D., Maaten-Theunissen M. van der, Biber P., Pretzsch H. (2017, octobre). Robinia pseudoacacia L. Flower Analyzed by Using An Unmanned Aerial Vehicle (UAV). *Remote Sensing*, vol. 9, n° 11, p. 1091.

- Congalton R. G. (1991, juillet). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, vol. 37, n° 1, p. 35–46.
- Dale M. R., Fortin M.-J. (2014). *Spatial Analysis: A Guide For Ecologists* (2^e éd.). Cambridge, Cambridge University Press.
- Fassnacht F. E., Latifi H., Stereńczak K., Modzelewska A., Lefsky M., Waser L. T. *et al.* (2016, 12). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, vol. 186, p. 64–87.
- Ghiyamat A., Shafri H. (2010). A review on hyperspectral remote sensing for homogeneous and heterogeneous forest biodiversity assessment. *International Journal of Remote Sensing*, vol. 31, n° 7, p. 1837–1856.
- Grabska E., Hostert P., Pflugmacher D., Ostapowicz K. (2019, mai). Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing*, vol. 11, n° 10, p. 1197.
- Guo Q., Fei S., Potter K. M., Liebhold A. M., Wen J. (2019, avril). Tree diversity regulates forest pest invasion. *Proceedings of the National Academy of Sciences*, vol. 116, n° 15, p. 7382–7386.
- Hagolle O., Huc M., Pascual D., Dedieu G. (2015, 09 03). A multi-temporal and multi-spectral method to estimate aerosol optical thickness over land, for the atmospheric correction of formosat-2, landsat, vens and sentinel-2 images. *Remote Sensing*, vol. 7, n° 3, p. 2668–2691.
- Hill R. A., Wilson A., George M., Hinsley S. (2010, février). Mapping tree species in temperate deciduous woodland using time-series multi-spectral data. *Applied Vegetation Science*, vol. 13, n° 1, p. 86–99.
- Immitzer M., Atzberger C., Koukal T. (2012, septembre). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, vol. 4, n° 9, p. 2661–2693.
- Le Rest K., Pinaud D., Monestiez P., Chadoeuf J., Bretagnolle V. (2014, juillet). Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global Ecology and Biogeography*, vol. 23, n° 7, p. 811–820.
- Mannel S., Price M., Hua D. (2011, octobre). Impact of reference datasets and autocorrelation on classification accuracy. *International Journal of Remote Sensing*, vol. 32, n° 19, p. 5321–5330.
- Pasquarella V. J., Holden C., Woodcock C. (2018). Improved mapping of forest type using spectral-temporal landsat features. *Remote Sensing of Environment*, vol. 210, p. 193–207.
- Sheeren D., Fauvel M., Josipović V., Lopes M., Planque C., Willm J. *et al.* (2016, septembre). Tree Species Classification in Temperate Forests Using Formosat-2 Satellite Image Time Series. *Remote Sensing*, vol. 8, n° 9, p. 734.
- Thompson I. D., Okabe K., Tylianakis J. M., Kumar P., Brockerhoff E. G., Schellhorn N. A. *et al.* (2011, décembre). Forest Biodiversity and the Delivery of Ecosystem Goods and Services: Translating Science into Policy. *BioScience*, vol. 61, n° 12, p. 972–981.

Cartographie à large échelle des peupleraies à partir de séries temporelles d'images Sentinel-2

Yusra Hamrouni^{1,2}, Éric Paillassa³, Véronique Chéret¹, Claude Monteil¹, David Sheeren¹

1. DYNAFOR, Université de Toulouse, INRA, Castanet-Tolosan, France
yusra.hamrouni@inra.fr

2. Conseil National du Peuplier, 75116 Paris, France

3. Centre National de la propriété Forestière, Institut pour le Développement Forestier, 33075 Bordeaux, France

RÉSUMÉ. La fréquence de mise à jour des cartes forestières nationales est insuffisante pour suivre les espèces et plus particulièrement les peupleraies. Pour cette essence à courte rotation (15 à 20 ans), ce rythme s'avère peu adapté et ne permet pas d'assurer un suivi régulier. Cet article présente une approche de cartographie à large échelle des peupleraies à partir d'images satellites Sentinel-2. La question de disponibilité des données de référence a été abordée et l'apprentissage actif peut être une option envisageable face au manque d'échantillons d'entraînement.

ABSTRACT. Update frequency of national forest maps is insufficient for effective monitoring of tree species and, more especially, for poplar plantations. This update rate proves to be unsuitable because of the short rotation cycle (15 to 20 years) of poplars. This paper presents a new approach for large-scale poplar plantations mapping with Sentinel-2 time series. The availability of reference data is addressed and active learning technique could be an option where training samples are limited.

MOTS-CLÉS : Classification, séries temporelles, sentinel-2, large échelle, peupleraies, apprentissage actif

KEYWORDS: Classification, temporal time series, sentinel-2, large scale, poplar plantations, active learning

1. Introduction

La connaissance précise de la localisation des peuplements forestiers et leur composition en essences est une donnée incontournable pour les gestionnaires réalisant des opérations sylvicoles. C'est aussi un préalable indispensable pour étudier la réponse des forêts au changement climatique et prédire la répartition potentielle des essences dans le futur.

Bien qu'il existe des cartes forestières et des bases de données nationales, les connaissances sur ces espèces restent encore très lacunaires et plus particulièrement pour la filière peuplier. Le peuplier est la deuxième essence de feuillus plantée en France après le Chêne Sessile. C'est l'un des arbres à cycle de production le plus rapide avec une durée moyenne de 15 à 20 ans. Pour une telle essence à courte rotation, les sources d'informations cartographiques existantes s'avèrent insuffisantes voire inadaptées pour assurer un suivi régulier soit à cause d'une marge d'erreur importante, d'une durée d'acquisition trop longue ou encore d'une mise à jour difficile. Afin de mieux sécuriser et développer cette filière, la production de données cartographiques fiables et régulières s'avère cruciale, dans un contexte de tension sur la ressource bois, tension due à un renouvellement des peupleraies en recul permanent depuis plus de 20 ans.

La télédétection est utilisée depuis longtemps pour suivre les milieux forestiers Boyd, Danson (2005). Néanmoins, l'identification des espèces est un problème complexe qui reste encore ouvert Fassnacht *et al.* (2016). La démarche de cartographie opérationnelle est toujours la photo-interprétation qui requiert d'importants moyens humains et financiers. Elle n'est pas automatisée et dépend des campagnes de levés aéroportés ce qui rend difficile un suivi régulier des peuplements. Le contexte actuel de disponibilité en images satellites à moindre coût apparaît pertinent et favorable pour réinterroger les capacités de la télédétection à répondre à ce problème. L'arrivée des satellites Sentinel dans le cadre du programme Copernicus de l'Agence Spatiale Européenne offre l'opportunité d'explorer l'approche hypertemporelle pour suivre les couverts forestiers et en particulier les peupleraies. Ces données constituent une révolution dans le domaine de la télédétection avec un accès gratuit, une haute résolution spatiale (10m, 20m), une richesse spectrale (13 bandes) et polarimétrique (2 polarisations) et une très haute répétitivité temporelle (5 à 6 jours).

C'est dans cette optique que s'inscrit la présente étude qui consiste à *i*) exploiter le potentiel des séries temporelles Sentinel-2 pour la détection des peupleraies, *ii*) évaluer la pertinence d'utiliser des références issues des bases de données nationales par rapport à des échantillons actualisés et validés visuellement et *iii*) définir une stratégie de classification à large échelle avec une approche d'apprentissage par transfert permettant d'aboutir, à terme, à une production nationale dans le cadre du pôle THEIA ¹.

1. Pôle Thématique Surfaces Continentales : <https://www.theia-land.fr/>

2. Matériels

2.1. Sites d'étude

Pour ce travail, trois grands bassins populicoles ont été identifiés dans la vallée de la Garonne, la région Grand-Est et la région Centre ce qui correspond à cinq tuiles (carrés de 100 km de côté) Sentinel-2 (Figure 1). Ces sites retenus couvrent la variabilité nationale en termes de modes de gestion des peupleraies et de conditions climatiques.

Les tuiles traitées pour cet article sont la T31TCJ qui couvre les départements du Tarn-Et-Garonne (82), du Gers (32) et de la Haute Garonne (31), et la tuile T31UEQ englobant principalement les départements de l'Aisne (02) et de la Marne (51)(en gras sur la figure 1).

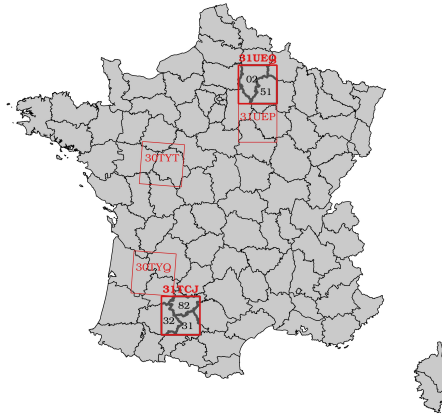


FIGURE 1 – Sites d'étude.

2.2. Images satellites optique : Sentinel-2

Toutes les images Sentinel-2 ont été exploitées pour l'année 2017 sur les deux tuiles d'étude : T31TCJ et T31UEQ. Ce sont des produits en réflectance de surface distribués par la plateforme THEIA, orthorectifiés et corrigés des effets atmosphériques (Niveau 2A).

2.3. Données de référence

Afin de réaliser une classification supervisée à partir des données Sentinel-2, un jeu de références d'entraînement et de validation a été créé selon deux approches :

2.3.1. Première approche : 100% BD Forêt@IGN

Les échantillons de référence ont été extraits à partir de la BD Forêt@IGN qui est établie à partir d'orthophotographies acquises entre 2004 et 2013 pour les cinq départements étudiés. Ces références représentent toutes les classes de feuillus (dont peupleraies) disponibles par département.

2.3.2. Deuxième approche : photo-interprétation pour le peuplier et BD Forêt@IGN pour les autres classes

Le décalage entre l'année d'acquisition des images Sentinel-2 (2017) et les années de référence de la BD Forêt@IGN (2004-2013) peut être à l'origine d'erreurs dans les échantillons (coupes, abandons...). Ayant un court cycle de production (15-20 ans), les peupleraies sont les plus susceptibles de contenir du bruit.

La deuxième approche consiste donc à valider visuellement les échantillons de la classe peuplier à l'aide d'orthophotographies et d'images Sentinel-2 par année et par département. Pour les autres classes de feuillus, les références ont été échantillonnées à partir de la BD Forêt@IGN. Comparée à la première, cette deuxième approche exige un travail supplémentaire d'interprétation, mais assure un jeu de données fiable pour la classe peuplier.

3. Méthodes

3.1. Classification supervisée département par département

Pour chaque département des deux tuiles d'étude (T31TCJ et T31UEQ) une classification supervisée a été effectuée selon les deux approches d'échantillonnage détaillées en 2.3. L'objectif étant d'évaluer la qualité de prédiction des peupliers selon le type des références (visuellement validées ou non).

Toutes les images Sentinel-2A/B disponibles (quelque soit le pourcentage de nuages) ont été téléchargées sur les deux tuiles (39 images pour la T31TCJ et 30 pour la T31UEQ). Pour une même année, les images acquises ne sont pas forcément synchrones d'une tuile à l'autre, soit en raison de la présence de nuages ou de contraintes d'acquisition propres au capteur. C'est pourquoi, un *gap-filling* temporel a été appliqué. Il s'agit de remplacer chaque pixel invalide (détecté dans le masque de nuage) par une valeur interpolée (interpolation linéaire) avec les dates valides de la série temporelle Inglada (2016). Les images ont été ensuite rééchantillonnées avec un pas de 10 jours commun à toutes les tuiles. À partir de la série temporelle rééchantillonnée, les valeurs aberrantes des données de référence ont été retirées.

Le jeu des pixels de référence restant a été réparti en 50% pour l'entraînement et 50% pour la validation. La séparation s'est faite en adoptant un échantillonnage stratifié par classe et par peuplement. En effet, les pixels uti-

Cartographie des peupleraies à partir d'images Sentinel-2

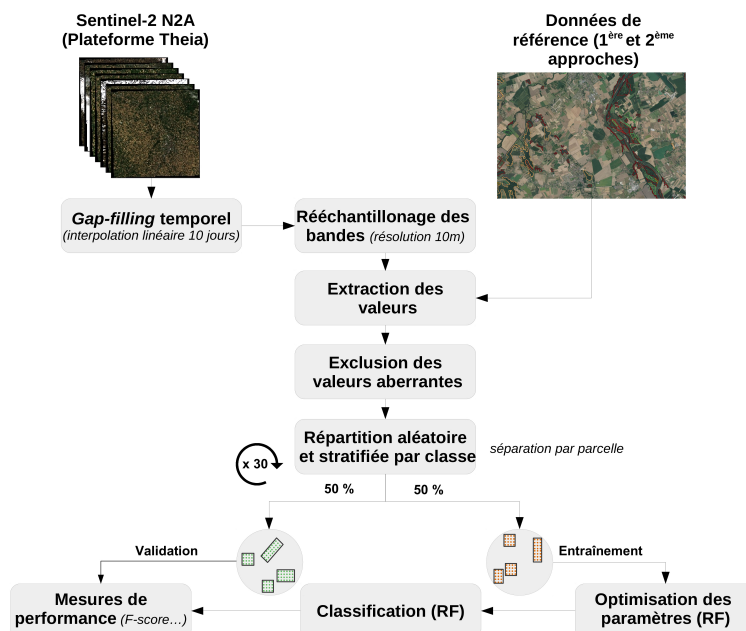


FIGURE 2 – Schéma méthodologique.

lisés pour l'apprentissage appartiennent à des peuplements qui ne sont pas utilisés pour la validation pour limiter l'autocorrelation spatiale. Ce tirage est répété 30 fois afin de quantifier la variabilité liée à la sélection aléatoire. Enfin, sur les 30 jeux entraînement/validation, une classification supervisée par forêts aléatoires (*Random Forest*) a été réalisée avec une optimisation préalable des hyperparamètres de l'algorithme (nombre d'arbres, profondeur maximale...).

3.2. Apprentissage actif pour une classification à large échelle

Dans l'approche précédente, un modèle est créé pour chaque département de chaque tuile avec des données de référence qui lui sont spécifiques. Afin d'établir des cartographies à une échelle plus large que celle du département, deux solutions peuvent s'envisager. La première est la génération de plusieurs modèles locaux et la fusion des résultats cartographiques. La seconde est la création d'un modèle local et son adaptation progressive pour rendre les prédictions performantes à l'échelle nationale.

La première option considère que des échantillons d'entraînement sont disponibles partout et suffisamment représentatifs de toutes les classes. En exploitant les bases de données forestières disponibles à l'échelle nationale (approche 2.3.1), il se peut toutefois que la qualité des références soit insuffisante en raison

d'un rythme de mise à jour inadapté au peuplier (essence à rotation courte). Dans ce cas, la photo-interprétation pourrait constituer une alternative pour construire un jeu de référence actualisé et précis. Néanmoins, cette approche s'avère peu réaliste pour une production à large échelle dans un contexte opérationnel en raison du coût et du temps qu'elle représente. Sans évoquer le biais de l'opérateur photo-interprète, une autre contrainte de la photo-interprétation est la redondance. En effet, plusieurs pixels voisins peuvent contenir la même information spectrale qui peut avoir tendance à biaiser le modèle appris et réduire sa capacité à généraliser (*surapprentissage*).

La deuxième option envisagée consiste à choisir un modèle robuste localement et à l'appliquer ensuite sur une zone plus large ou complètement distincte. Selon la stratégie adoptée, c'est une option qui peut limiter le besoin en références pour apprendre mais qui requiert toutefois une adaptation du modèle local. En effet, si la nouvelle zone à classer présente un espace de variables avec une distribution assez différente du modèle initial (en raison de conditions d'acquisition différentes, du relief...), les prédictions du modèle initial seront peu performantes sur la nouvelle zone Crawford *et al.* (2013). Il est donc indispensable de modifier le modèle ce qui nécessite, du point de vue de l'apprentissage automatique, de faire appel à des techniques d'adaptation de domaine. En télédétection, les approches d'adaptation de domaine se différencient en quatre groupes Tuia *et al.* (2016) :

1. *Adaptation de domaine par sélection d'attributs invariants* : elle consiste à identifier les bandes spectrales ou les attributs de l'image les plus robustes au changement et de ne conserver que ces variables pour réduire le décalage entre les deux domaines (source et cible).

2. *Adaptation de la distribution* : dans ce cas un espace commun aux deux domaines est créé à partir des deux distributions pour s'adapter au classifieur qui reste inchangé.

3. *Adaptation du classifieur par des approches semi-supervisées* : le modèle de classification entraîné sur les données étiquetées (classes connues) de la source est modifié pour s'adapter à la distribution des données de la cible sans avoir accès à leurs étiquettes.

4. *Adaptation du classifieur par apprentissage actif* : le classifieur entraîné avec les données étiquetées de la source est enrichi par des échantillons de la cible. C'est un cas particulier de l'approche précédente dans lequel le classifieur est autorisé à s'entraîner de manière itérative avec des références étiquetées de la cible, en nombre limité. Ces références sont sélectionnées selon leur potentiel d'amélioration du modèle initial pour bien classer le domaine cible. Dans cette quatrième catégorie, l'apprentissage actif permet de guider le processus d'adaptation de domaine en sélectionnant une taille optimale de références pour garantir de bonnes performances de classification Tuia *et al.* (2009).

L'hypothèse derrière cet apprentissage actif est qu'un algorithme est capable d'atteindre de meilleurs résultats de classification en misant non pas sur la

quantité d'échantillons (sélectionnés de façon aléatoire) mais plutôt sur leur qualité (les plus pertinents) Settles (2009). C'est une méthode utilisée depuis longtemps en traitement automatique du langage naturel Yang *et al.* (2009); Figueroa *et al.* (2012) et qui commence à montrer son potentiel en télédétection Tuia *et al.* (2009); Rougier (2016); Ben Slimene (2017).

En apprentissage actif, l'algorithme calcule une mesure d'information des échantillons à partir de leurs attributs et les ordonne selon cette valeur. Il sélectionne ensuite de façon itérative les échantillons les plus informatifs et demande à l'expert du domaine de leur attribuer une étiquette (classe) pour réapprendre un nouveau modèle.

L'un des critères d'information les plus utilisés est l'incertitude Settles (2009). Il s'agit de choisir les échantillons pour lesquels le modèle est le plus incertain, après prédiction. L'incertitude peut être mesurée par plusieurs méthodes parmi lesquelles l'entropie. Pour chaque échantillon, cette mesure prend en compte les probabilités d'appartenance à chacune des classes possibles du modèle. Lorsque le degré de certitude est élevé, l'entropie est faible (i.e. la probabilité d'appartenance est forte pour une classe spécifique). Une forte incertitude apparaît lorsque les valeurs de probabilité sont distribuées dans un grand nombre de classes et qu'aucune de celles-ci ne se démarque des autres (i.e. entropie élevée). Les échantillons sélectionnés par l'algorithme sont donc ceux qui ont la valeur d'entropie la plus élevée. Après étiquetage de ceux-ci, le modèle initial est ré-entraîné pour s'adapter et ce, de façon itérative jusqu'à atteindre une condition d'arrêt (ex. nombre maximal d'itérations, score maximal...). Dans ce travail, l'apprentissage actif est appliqué dans deux cas :

- en partant d'un modèle entraîné sur le département de la Marne (51) de la tuile T31UEQ pour prédire le département du Tarn-Et-Garonne (82) de la tuile T31TCJ, soit du Nord vers le Sud;
- dans le sens opposé, en partant d'un modèle entraîné sur le département 82 et enrichi par apprentissage actif pour prédire le département 51.

L'entropie est utilisée comme mesure d'incertitude et deux critères d'arrêt ont été fixés : un F-score maximal de la classe peuplier de 99% et un nombre d'échantillons maximal à rajouter de 1000 avec 10 échantillons par itération (100 itérations).

4. Résultats et discussions

Les tables 1 et 2 récapitulent les résultats obtenus respectivement avec la première et la deuxième approche de choix des données de référence, sans apprentissage actif (soit un modèle par département). Pour rappel, l'enjeu est de savoir si la BD Forêt® de l'IGN, disponible à l'échelle nationale, est suffisamment fiable pour générer des échantillons d'entraînement ou s'il faut adopter une autre stratégie qui est ici la photo-interprétation.

TABLE 1 – Classification 2017 avec l'approche 100% BD Forêt@IGN

Tuile	Département	Nb classes	Nb échantillons d'entraînement par classe	Précision globale moyenne _(*30)	F-score moyen Peuplier _(*30)
T31TCJ (S.O)	82	6	1000	78.0	93.8 ± 1.2
	32	4	1000	79.6	88.6 ± 2.2
	31	4	250	80.9	84.7 ± 8.3
T31UEQ (N.E)	51	4	1000	59.9	68.7 ± 7.8
	02	6	1000	71.3	76.6 ± 3.6

Les résultats montrent que l'utilisation de données de référence mises à jour par photo-interprétation (2^{ème} approche) permet de mieux identifier la classe peuplier avec une amélioration du F-score (moyenne harmonique du taux de rappel et précision) de 5 à 20% en moyenne et selon l'année d'édition de la BD Forêt@IGN.

TABLE 2 – Classification 2017 avec l'approche Photo-interprétation + BD Forêt@IGN

Tuile	Département	Nb classes	Nb échantillons d'entraînement par classe	Précision globale moyenne _(*30)	F-score moyen Peuplier _(*30)
T31TCJ (S.O)	82	6	1000	79.1	98.9 ± 0.6
	32	4	1000	82.5	96.0 ± 0.2
	31	4	250	86.4	99.6 ± 0.4
T31UEQ (N.E)	51	4	1000	69.2	90.9 ± 2.4
	02	6	1000	79.1	89.9 ± 3.9

L'évolution de la précision globale moyenne est moins importante. Elle tient compte cette fois de toutes les classes qui elles, n'ont pas été actualisées contrairement à la classe peuplier.

Le résultat cartographique du département 82 obtenu avec la première approche est visuellement moins précis que celui avec photo-interprétation avec une sur-détection des peupleraies. Pour ce premier cas, le modèle a été entraîné mais aussi validé avec des échantillons de peuplier non mis à jour, qui comportent notamment des coupes entre l'année de référence de la BD Forêt@IGN (2013) et celle des images Sentinel-2 (2017). La performance de la classe peuplier est élevée (F-score = 93.8% pour le 82) mais elle n'est pas représentative de la réalité. Les erreurs résultent d'une confusion avec du sol nu qui n'apparaissent plus dans le deuxième cas (références mises à jour par photo-interprétation).

Lorsque l'apprentissage actif est appliqué, les performances varient suivant le sens adopté (i.e. selon le département sur lequel repose le modèle initial). Dans un premier temps, le modèle initial est construit sur la base des références actualisées (approche 2.3.2) du département 51 puis, enrichi progressivement avec des échantillons du 82 par apprentissage actif tout en restant performant sur le département initial.

Cartographie des peupleraies à partir d'images Sentinel-2

Le modèle initial (51) est fondé sur quatre classes : peuplier, chêne, forêt de feuillus fermés et forêt de feuillus ouverts. Le département cible (82) en possède deux de plus : robinier et châtaignier, qui n'existent pas dans le modèle initial.

Lorsque le modèle 51 est appliqué directement sur le département 82, sans adaptation (i.e. pas d'apprentissage actif), la classe peuplier est bien identifiée mais est sur-détectée, notamment à cause de confusions avec le chêne et les forêts de feuillus fermés (Figure 3 en colonne). Autrement dit, certains échan-

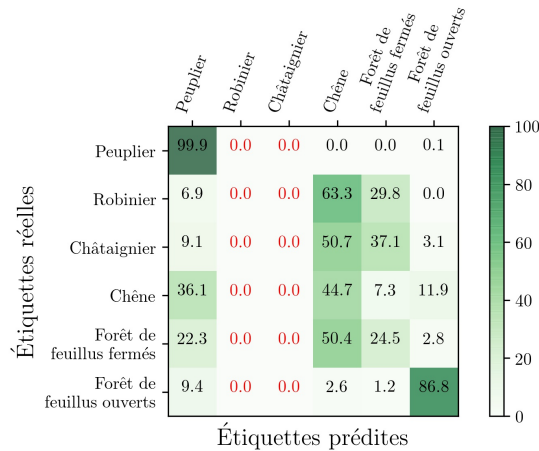


FIGURE 3 – Matrice de confusion (en %) obtenue en prédisant le 82 à partir du modèle appris sur le 51 (soit à $t=0$: avant ajout d'échantillons du 82).

tillons de chêne et de forêt de feuillus fermés du 82 ont une signature spectrale plus proche des peupliers du 51 que de leurs classes respectives.

En ce qui concerne les classes de robinier et de châtaignier qui sont absentes du modèle initial (51), elles sont principalement confondues avec du chêne et des forêts de feuillus fermés et elles ne sont évidemment jamais prédites par le modèle (Figure 3).

Lorsque l'apprentissage actif est mis en œuvre en rajoutant progressivement des échantillons du 82, les performances de classification augmentent mais varient selon les classes observées.

La figure 4 illustre l'évolution du F-score de la classe peuplier en fonction du nombre total d'échantillons rajoutés du département 82. On constate que l'apprentissage actif permet d'augmenter plus rapidement la qualité de prédiction des peupliers qu'une sélection aléatoire des échantillons (courbe verte vs rouge, figure 4). On observe également qu'au-delà de 200 échantillons rajoutés, on atteint un score maximal pour cette classe, très élevé (F-score de 96% contre 87% par sélection aléatoire).

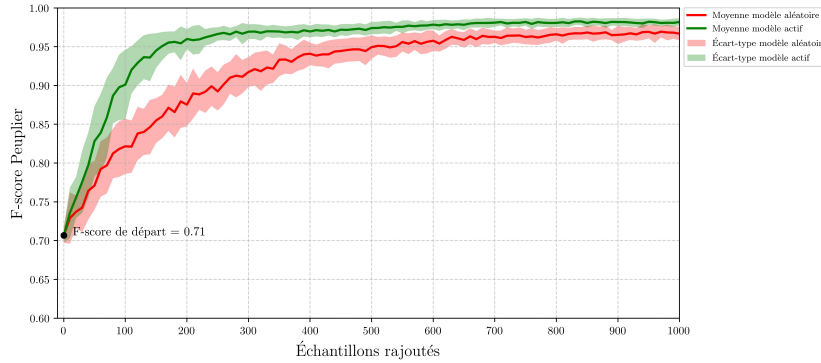


FIGURE 4 – Évolution du F-score Peuplier moyen (*30) en fonction du nombre d'échantillons : du département (51) vers le département (82).

L'évolution de la précision globale de la classification en fonction du nombre total d'échantillons et de la classe est illustrée en figure 5. En analysant le nombre d'échantillons rajoutés par classe et par itération, on peut voir que les échantillons de châtaignier et de robinier (absents initialement) sont les plus fréquemment sélectionnés par l'algorithme qui les considère comme incertains, et qui demande donc un étiquetage pour mieux les classer. Cette adaptation progressive rend le modèle initial transférable au département 82 ce qui augmente la précision globale.

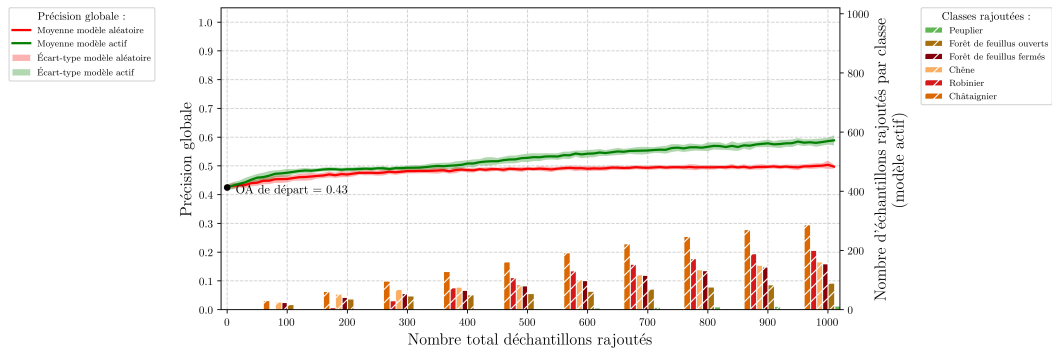


FIGURE 5 – Évolution de la précision globale moyenne(*30) en fonction du nombre d'échantillons : Modèle actif VS Modèle aléatoire du (51) au (82).

Dans l'exemple du châtaignier de la figure 6, le F-score a évolué de 0 à 40% à la 60^{ème} itération avec l'ajout 200 échantillons (sélectionnés parmi 600 au total) et a atteint 52% avec 310 échantillons parmi les 1000 au total (dernière itération). Par comparaison, le F-score châtaignier du modèle aléatoire

Cartographie des peupleraies à partir d'images Sentinel-2

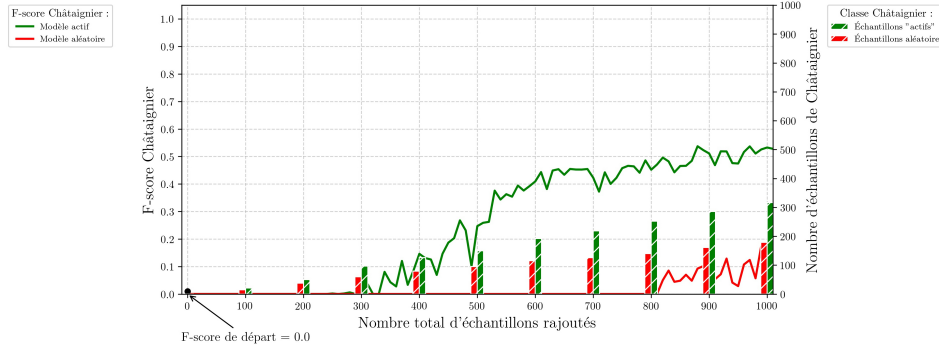


FIGURE 6 – Évolution du F-score châtaignier en fonction du nombre d'échantillons de la classe : Modèle actif VS Modèle aléatoire du (51) au (82).

n'a commencé à évoluer qu'à partir de la 80^{ème} itération avec 130 échantillons supplémentaires (parmi 800) pour atteindre 18% à la dernière itération.

Malgré l'existence des classes de chêne et de forêt de feuillus fermés dans le modèle initial, certains échantillons appartenant à ces classes ont été sélectionnés par apprentissage actif. Leur taux de confusion initial étant élevé (Figure 3), le modèle a cherché à améliorer leur reconnaissance. La classe de feuillus fermés est très hétérogène. Elle renferme un mélange de plusieurs essences dont la composition peut varier d'une zone à l'autre.

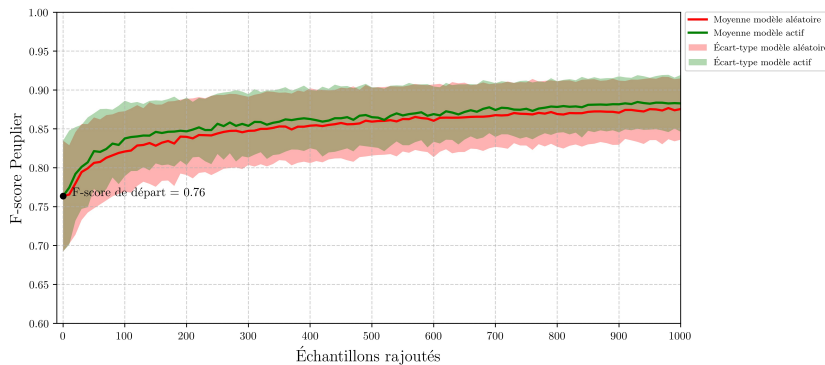


FIGURE 7 – Évolution du F-score Peuplier moyen (*30) en fonction du nombre d'échantillons : du département (82) vers le département (51).

En ce qui concerne la classe peuplier, très peu d'échantillons supplémentaires ont été sélectionnés par apprentissage actif (5/1000 pixels au total). L'amélioration du F-score n'est donc pas liée à l'ajout d'échantillons de peupliers mais plutôt à la diminution des confusions avec le chêne et les feuillus fermés, beaucoup plus présents dans le jeu d'échantillons enrichi.

Lorsque l'apprentissage actif est appliqué dans le sens opposé (du 82 vers le 51), les résultats sont différents. Les valeurs de F-score peuplier sont très similaires à celles obtenues avec des références sélectionnées aléatoirement (Figure 7).

Comme le montre la figure 8, les échantillons les plus informatifs sélectionnés appartiennent aux quatre classes à prédire, y compris les peupliers. La classe de feuillus fermés est la plus représentée et aussi la plus difficile à classer. Son hétérogénéité (mélange d'essences) entraîne des confusions avec des classes

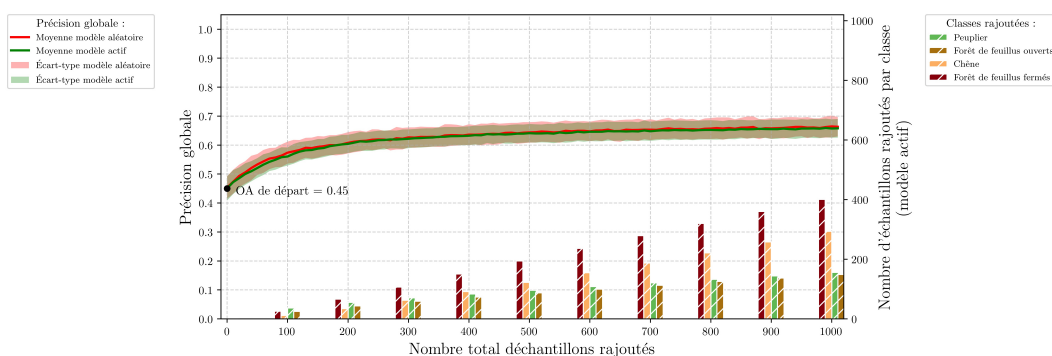


FIGURE 8 – Évolution de la précision globale moyenne(*30) en fonction du nombre d'échantillons : Modèle actif VS Modèle aléatoire du (82) au (51).

pures qui en réalité, pourraient correspondre à des prédictions cohérentes avec le terrain (mais nous n'avons pas pu le vérifier).

L'apprentissage actif permet toujours d'améliorer la performance du modèle mais cette fois, de manière équivalente à un échantillonnage aléatoire ce qui lui fait perdre son intérêt.

5. Conclusion

Dans cette étude, nous avons présenté une première approche de cartographie des peupleraies à large échelle à partir de données optiques Sentinel-2. Les premiers résultats montrent des performances très élevées pour séparer les peupleraies des autres classes de feuillus toutes confondues. Nous avons également mis en évidence l'impact du choix des références d'entraînement et de validation des modèles. Afin d'obtenir des prédictions fiables, la BD Forêt de l'IGN n'est pas suffisante et la photo-interprétation s'est avérée nécessaire.

Dans une optique de production à large échelle, une stratégie d'apprentissage par transfert, en s'appuyant sur des techniques d'apprentissage actif, a montré son intérêt pour minimiser le nombre d'échantillons en garantissant des performances optimales. Cependant, elle peut être pénalisée par les références

bruitées (classes hétérogènes, nuages non détectés...) qui risquent d'être sélectionnées par l'algorithme. Le sens dans lequel le transfert est réalisé a donc un impact sur les performances prédictives.

Dans la suite, l'apprentissage actif sera testé en ne conservant que les classes pures. L'approche sera également appliquée et validée sur d'autres tuiles Sentinel-2 pour aboutir, à terme, à un modèle adapté à l'échelle nationale.

Bibliographie

- Ben Slimene I. (2017). *Apprentissage actif pour la classification des occupations du sol sur larges étendues à partir d'images multispectrales à haute résolution spatiale : Application en milieu cultivé, Lebna (Cap-Bon Tunisie)*. Thesis, Montpellier.
- Boyd D. S., Danson F. M. (2005, mars). Satellite remote sensing of forest resources: Three decades of research development. *Progress in Physical Geography*, vol. 29, n° 1, p. 1-26.
- Crawford M. M., Tuia D., Yang H. L. (2013, mars). Active Learning: Any Value for Classification of Remotely Sensed Data? *Proceedings of the IEEE*, vol. 101, n° 3, p. 593-608.
- Fassnacht F. E., Latifi H., Stereńczak K., Modzelewska A., Lefsky M., Waser L. T. *et al.* (2016, décembre). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, vol. 186, p. 64-87.
- Figuroa R. L., Zeng-Treitler Q., Ngo L. H., Goryachev S., Wiechmann E. P. (2012). Active learning for clinical text classification: Is it better than random sampling? *Journal of the American Medical Informatics Association : JAMIA*, vol. 19, n° 5, p. 809-816.
- Inglada J. (2016, février). *Otb Gapfilling, A Temporal Gapfilling For Image Time Series Library*. Zenodo.
- Rougier S. (2016). *Apport des images satellites à très haute résolution spatiale couplées à des données géographiques multi-sources pour l'analyse des espaces urbains*. Strasbourg.
- Settles B. (2009). *Active learning literature survey*. Computer Sciences Technical Report n° 1648. University of Wisconsin-Madison.
- Tuia D., Persello C., Bruzzone L. (2016, juin). Domain Adaptation for the Classification of Remote Sensing Data: An Overview of Recent Advances. *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, n° 2, p. 41-57.
- Tuia D., Ratle F., Pacifici F., Kanevski M. F., Emery W. J. (2009, juillet). Active Learning Methods for Remote Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, n° 7, p. 2218-2232.
- Yang B., Sun J.-T., Wang T., Chen Z. (2009). Effective multi-label active learning for text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '09*, p. 917. Paris, France, ACM Press.

Données hétérogènes de mobilités quotidiennes : protocole de diagnostic qualité et d'apurement à partir de la base MOBI'KIDS

**Duroudier S.¹, Chardonnel S.¹, Mericskay B.², Andre-Poyaud I.¹, Bedel O.³,
Depeau S.², Devogele T.⁴, Etienne L.⁴, Lepetit A.², Moreau C.⁴, Pelletier N.²,
Ployon E.¹, Tabaka K.¹**

1. Univ. Grenoble Alpes, CNRS, Science Po Grenoble, PACTE 38000 Grenoble, France, sylvestre.duroudier@univ-grenoble-alpes.fr
2. UMR ESO, CNRS/Université Rennes 2, Place du recteur Henri Le Moal, 35043 Rennes Cedex, France
3. Alkante, 4 rue Alain Colas, 35530 Noyal sur Vilaine
4. Laboratoire d'informatique LIFAT, 3 place Jean Jaurès, 41000 Blois

RESUME. Cet article a pour objectif de présenter la méthodologie de diagnostic qualité et d'apurement des données expérimentée à partir d'une enquête de mobilité individuelle (programme MOBI'KIDS). Une première section revient sur la démarche suivie et pointe l'enjeu de l'évaluation de la qualité de données hétérogènes issues d'une méthode mixte et longitudinale de collecte (suivis GPS, enquêtes, observations). Une deuxième section établit un diagnostic qualité selon l'origine (GPS, algorithme, enquête) et la nature des erreurs (complétude, précision, cohérence). Ces typologies permettent, dans une troisième section, la définition d'une chaîne de traitements reproductible visant à améliorer la qualité interne et externe des données.

ABSTRACT. This paper aims at proposing a data quality diagnosis and clearance methodology experimented on an individual mobility survey (MOBI'KIDS program). The first section presents the theoretical approach to highlight the issue of a data quality diagnosis applied on heterogeneous data collected from mixed methods (GPS tracks, surveys, observations). Secondly, two typologies of major errors are discussed according to their origin (GPS, algorithm, survey) and their nature (completeness, accuracy, consistency). A processing chain is thirdly defined to improve both internal and external data quality in order to the perspective of a replicable methodology.

Mots-clés : mobilité quotidienne, traces GPS enrichies, méthode mixte, qualité des données

KEYWORDS: Daily mobility, semantic GPS data, mixed method, data quality

1. Introduction

Avec le développement de dispositifs mobiles capables de collecter des données de haute résolution spatiale et temporelle sur de longues périodes, l'analyse des données individuelles de mobilité quotidienne a largement évolué (Stopher, 2009 ; Shoval et al., 2014, Drevon et al., 2014). L'un des principaux enjeux méthodologiques actuels consiste à hybrider les méthodes traditionnelles des « small data » apportant des informations sémantiques riches, avec des méthodes relevant des « big data » permettant des collectes plus massives et plus longues (Chen et al., 2016).

C'est dans cette perspective que le programme MOBI'KIDS¹ (MK) analyse la mobilité des enfants et de leurs parents afin de comprendre dans quelles conditions l'autonomie de déplacement se met en place. Une enquête réalisée auprès d'une cohorte d'enfants et de parents permet de reconstituer les trajectoires quotidiennes décrivant spatialement, temporellement et sémantiquement la succession de leurs activités et leurs déplacements. L'approche mixte de cette enquête combine différents dispositifs de collecte (suivi GPS, enquêtes mobilité, entretiens semi-directifs), qui permettent la constitution d'un corpus de données originales et riches, mais dont l'hybridation pose un ensemble de défis et d'enjeux tant sur le plan technique que méthodologique (Lenormand *et al.*, 2014 ; Lord *et al.*, 2018).

La plus-value de ces méthodes mixtes est que le corpus est enrichi progressivement par différentes techniques de collecte. Cette diversité des méthodes mobilisées implique un travail spécifique de mise en cohérence des données soulevant avec d'autant plus d'acuité la question de leur qualité. Aussi, l'évaluation de la qualité interne et externe des données collectées apparaît indispensable et nécessite de formaliser une approche reproductible et transposable au-delà du programme MK. Cette étape, préalable à l'analyse même des données, est souvent peu ou non évoquée dans la littérature, alors qu'elle est un gage scientifique d'autant plus crucial que les corpus sont complexes et multi-sources. S'insérant dans le programme MK en cours, cet article a pour objectif de discuter cette étape de la recherche en présentant la méthodologie de diagnostic et les préconisations d'apurement des données expérimentée sur le corpus MK.

Une première section revient sur la démarche suivie et pointe l'enjeu de l'évaluation de la qualité de données hétérogènes issues d'une méthode mixte et longitudinale de collecte. Une deuxième section établit un diagnostic qualité selon l'origine (GPS, algorithme, enquête) et la nature des erreurs (complétude, précision, cohérence). Ces typologies permettent, dans une troisième section, la définition d'une chaîne de traitements reproductible visant à améliorer la qualité interne et externe des données.

¹ Les travaux présentés dans cet article sont financés par le projet MOBI'KIDS « Le rôle des cultures éducatives urbaines (CEU) dans l'évolution des mobilités quotidiennes et des contextes de vie des enfants. Collecte et analyse de traces géolocalisées et enrichies sémantiquement » (ANR-16-CE22-0009).

2. De la démarche de MOBI’KIDS à l’enjeu de l’apurement

2.1. MOBI’KIDS : position de recherche

L’analyse de la mobilité et plus spécifiquement de l’autonomie des enfants est aujourd’hui envisagée selon différentes approches. D’un côté, des recherches qualitatives interrogent les enfants sur leurs expériences subjectives et sensibles de mobilité ; de l’autre, des recherches quantitatives (questionnaire, accéléromètre) mesurent l’intensité de l’activité physique des enfants. Enfin, des approches géomatiques analysent les facteurs géographiques et matériels des environnements dans lesquels évoluent les enfants (Depeau et Quesseveur, 2014 ; Kytä *et al.*, 2018). Encore peu de travaux envisagent les dimensions de l’apprentissage de la mobilité chez les enfants par des approches hybrides (Christensen *et al.*, 2011 ; Depeau *et al.*, 2017). L’ambition de MK est de caractériser les pratiques quotidiennes des enfants dans leurs contextes socio-spatiaux. L’objectif est d’analyser l’évolution des pratiques en suivant les mêmes enfants à deux périodes, à la fin de l’école primaire et au début du collège, selon une approche à la fois descriptive des pratiques (avec une fine résolution spatiale et temporelle) et compréhensive des expériences et des représentations de l’enfant.

2.2 Les données attendues

Répondre à cet objectif scientifique requiert un corpus de positions spatio-temporelles relatives aux déplacements et aux activités des enquêtés modélisées sous la forme d’une suite ordonnée de trajets et de lieux enrichis sémantiquement. D’une part, les lieux correspondent aux arrêts prolongés, formalisés par des points, et caractérisés par des heures de début et de fin ainsi que par les activités qui y sont réalisées (domicile, école, travail, course, loisirs...). D’autre part, les trajets correspondent aux déplacements entre deux lieux, modélisés par des lignes, et caractérisés par leurs bornes temporelles, le mode de transport utilisé et les modalités d’accompagnement.

Dans cette perspective, une enquête a été menée auprès d’une cohorte de 89 familles (soit 182 personnes à raison d’un enfant et un parent par famille²). Le recrutement s’est fait par l’intermédiaire des écoles primaires (classes de CM1-CM2), dans le but de suivre tous les enfants d’une même classe ayant potentiellement des interactions sociales y compris en dehors de l’école. Situés dans la métropole de Rennes, les terrains d’étude se composent de 3 écoles du centre de Rennes et 2 écoles de la commune périurbaine d’Orgères. Si la cohorte a déjà été enquêtée pendant les deux périodes (primaire et collège), les analyses présentées dans cet article portent uniquement sur les traces d’une partie de la cohorte (150 individus) lors de la première période.

² Exceptionnellement, dans certaines familles, deux enfants ont été suivis.

2.3 Un protocole hybride et intégré de collecte de données hétérogènes

Dans la pratique, le protocole de collecte consiste pour chaque personne à porter un enregistreur GPS pendant cinq jours consécutifs. En complément, plusieurs questionnaires en présentiel sont menés auprès de l'ensemble des individus de la cohorte. Le premier vise à renseigner le profil général des familles enquêtées. Le second, passé immédiatement après le suivi GPS, consiste en un examen exhaustif du séquençage des traces collectées des personnes enquêtées *via* une tablette (Depeau *et al.*, 2019) afin de les renseigner sémantiquement (mode de déplacement et d'accompagnement, activités...). Enfin, les individus enquêtés sont interrogés sur leurs perceptions des mobilités enfantines et divers éléments attenants aux attitudes éducatives.

Le phasage de ce protocole (Fig.1) oblige à séquencer *a priori* les traces brutes issues de la collecte GPS afin de mener rapidement un entretien en face à face avec les enquêtés. Le but est d'enrichir les données sur la base d'une trace prétraitée, c'est-à-dire structurée sous la forme d'une séquence ordonnée de positions spatiales correspondant à des situations d'immobilité (des arrêts dans des lieux) et à des situations de mobilité (des trajets). Cela impose de choisir à l'amont de l'enrichissement des paramètres de séquençage déterminant les valeurs qui définissent l'emprise spatio-temporelle d'un arrêt et d'un trajet. Répondant donc à des hypothèses théoriques et à des arbitrages en matière d'agrégation, l'algorithme de séquence s'appuie sur les règles suivantes : un arrêt est un ensemble de positions regroupées dans un rayon de moins de 50m de distance sur un temps d'au moins 300s ou à moins de 300s d'une mise en veille du *datalogger*³ ; un trajet correspond à un ensemble de positions consécutives entre 2 arrêts.

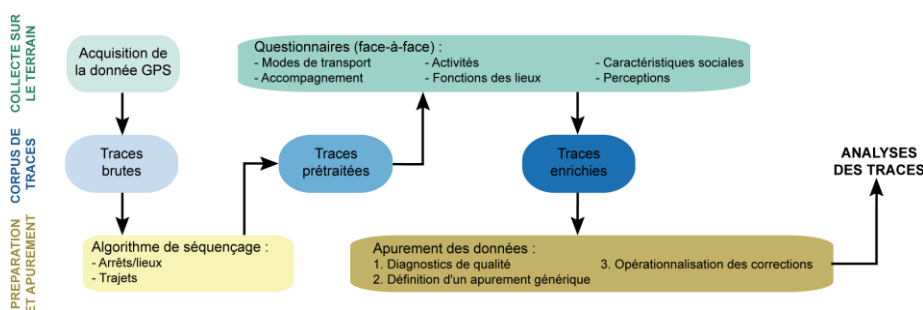


FIGURE 1. Saisir les traces de la mobilité quotidienne dans MOBI'KIDS : entre collecte sur le terrain, préparation et apurement des données.

³ Le *datalogger* utilisé dans le projet enregistre la position géographique des individus toutes les secondes. Impliquant de très gros volumes de données (potentiellement jusqu'à 600 000 informations par semaine par capteur), cette fréquence d'enregistrement est affinée par un accéléromètre qui limite les données collectées lors des longues périodes d'inactivité.

Appliqué au corpus traité dans cet article, ce séquençage permet d’identifier à partir des 7,6 millions de positions brutes enregistrées pour les 150 individus enquêtés, un ensemble de traces prétraitées constituées de 3794 trajets et 2381 arrêts.

Par ailleurs, conscients des limites induites par un séquençage paramétré *a priori* et de manière homogène pour l’ensemble des contextes d’enregistrement, les enquêteurs pouvaient annoter les écarts (notamment les imprécisions spatiales, temporelles) relevés par les personnes enquêtées entre les arrêts et trajets détectés et ceux effectivement réalisés.

2.4. Enjeux de la qualité des données de l’enquête

A ce stade du programme, bien avant l’analyse du corpus, le principal enjeu relève du problème bien connu de l’évaluation de la qualité des données (Devillers et Jeansoulin, 2005 ; Guptill et Morisson, 2013). Celui-ci se pose de manière singulière dans MK en raison du caractère inédit des données recueillies et des objectifs analytiques du projet de recherche. Il est donc nécessaire de définir un protocole général de traitement des données qui permette d’en diagnostiquer la qualité tout en préparant la base en vue d’analyses ultérieures.

Suivant plusieurs définitions (Aalders, 2002 ; Devillers, 2004), on distingue ici la qualité interne et la qualité externe des données. Concernant la qualité interne, il s’agit d’identifier les différences entre les données attendues et réelles, c’est-à-dire de quantifier et de qualifier les erreurs sémantiques, géographiques et temporelles selon leur origine (modes de collecte et compilation ; Servigne *et al.*, 2005) et leur nature (complétude et exhaustivité, cohérence, précision et indexation ; Guptill et Morrison, 1995 ; David et Fasquel, 1997). Sur ce plan, le programme MK réalise un grand écart entre d’une part les méthodes de nettoyage et de séquençage de données GPS (Biljecki *et al.* 2013 ; Lin et Hsu, 2014 ; Dalumpines *et al.*, 2017), et d’autre part les procédures d’apurement de données d’enquêtes classiques (statistiques publiques, enquêtes ménages-déplacements ; Certu, 2008 ; INED, 2019). De manière concomitante, la qualité externe caractérise l’adéquation entre les données et les besoins de ses utilisateurs. Il s’agit alors d’opérer des actions (agrégation, suppression, recodage, étiquetage, etc.) qui diffèrent selon les objectifs analytiques (autonomie des enfants, co-présence...). Finalement, ces évaluations des qualités interne et externe ouvrent des pistes de définition d’une méthode générique d’apurement de la base de données qui tienne également compte des spécificités des analyses envisagées.

3. Typologie et diagnostic de la qualité interne des données de l’enquête

3.1. Typologie des erreurs selon l’étape du protocole

3.1.1. Erreurs liées à la collecte de données GPS

Une première cause d’erreurs relève des dispositifs d’enregistrement des positions géographiques collectés au cours de l’enquête. On distingue d’abord les contraintes techniques des *dataloggers* (Neatt *et al.*, 2016) : le temps de chauffe

(TTFF), le déchargement de la batterie, la mise en veille ou la défaillance de l'accéléromètre. Ainsi, le TTFF peut conduire à la perte d'informations entre le début d'une mobilité (où le *datalogger* est en veille due à une immobilité prolongée) et l'enregistrement des positions une fois le signal GPS actif. Un autre ensemble d'erreurs relève des limites techniques du dispositif selon les environnements pratiqués par les enquêtés (Fig.2) : perte de signal à l'intérieur des bâtiments ou dans le métro, nombre insuffisant de satellites captés, réverbération des signaux lors « d'effets de canyon urbain ». Ce type de cas génère des positions GPS aberrantes qui se répercutent sur les trajets : traces décalées ou hors du réseau routier, vibrations et sinuosités, formes rectilinéaires, etc.

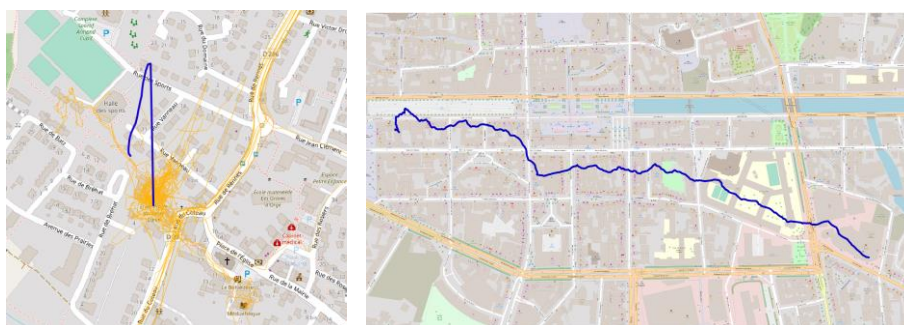


FIGURE 2. *Effet de la perte momentanée du signal satellite (à gauche), effet de canyon urbain et décalage de la trace (à droite).*

Une analyse simple des positions GPS collectées, reposant sur le nombre de satellites captés ou l'indicateur DOP (*dilution of precision* ; Kim et Park, 2017), permet une première estimation de la qualité des données. Ainsi, 0,86% des positions réparties sur 25% des trajets ont capté strictement moins de 4 satellites.

Une seconde analyse s'inspirant des règles de filtrage de Neatt et al. (2016) a porté sur les écarts de distance, de temps, et les vitesses moyennes associées, pour chaque couple de positions successives. Ces mesures permettent d'identifier 18901 couples de positions qui ont un écart temporel supérieur à la période d'échantillonnage (1s). De plus, 394 couples se distinguent par une vitesse supérieure à 180km/h, tandis que 3424 cas correspondent à une distance spatiale supérieure à 50m. Selon ces critères, près de 19 279 couples de positions distinctes présentent une aberration, soit 0,25% du total, et peuvent faire l'objet d'un filtrage.

3.1.2. Forces et limites de l'algorithme de séquençage des traces brutes

Deuxièmement, et indépendamment des erreurs liées aux dispositifs d'enregistrement, l'algorithme de séquençage est susceptible de générer d'autres erreurs. En effet, cet algorithme se base sur un ensemble de règles spatiales et temporelles (cf. 2.3) parfois inadaptées devant l'hétérogénéité des conditions d'enquêtes : par exemple entre un milieu urbain dense et un milieu ouvert, ou

lorsque l’amplitude spatiale des déplacements est réduite (domicile proche de l’école ou au-dessus d’un centre commercial). La multitude de situations contribue ainsi à générer des faux positifs ou négatifs (Fig.3) : par exemple des arrêts détectés sur des autoroutes en raison de bouchons ou encore des trajets au sein de lieux (domiciles, écoles) voire de « grands lieux » (centre commerciaux, parcs, cimetières).

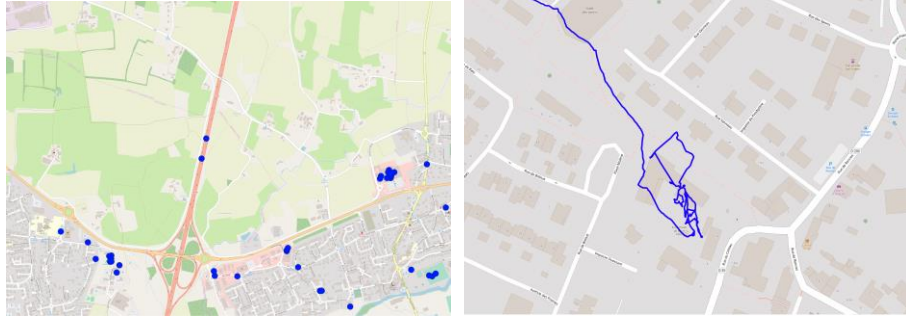


FIGURE 3. Faux lieux sur l’autoroute (à gauche) et faux trajet à l’école (à droite).

3.1.3. Les biais de l’enquête

Troisièmement, la phase d’enquête (Fig.1), comporte des biais susceptibles d’introduire de nouvelles erreurs (Beaud et Weber, 1998). Dans la pratique, les conditions de passation des enquêtes sont inégales, parfois contraintes ou peu adéquates (lieux publics bruyants, disponibilité et attention variables des enquêtés...). A cela s’ajoute l’hétérogénéité d’expérience des enquêteurs mobilisés, qui implique différentes sensibilités aux informations importantes à retenir et à saisir dans l’examen des traces. En outre, l’entretien peut être long (plus d’une heure) et altérer la précision des réponses de certaines personnes enquêtées. En résumé, ces erreurs sont de différentes natures : caractérisations incohérentes des activités ou de certains modes des trajets, erreurs de saisie dans les questions ouvertes, utilisation fréquente des notes correctives, etc.

3.2. Diagnostic de l’intégrité des données selon la nature des erreurs

3.2.1. Complétude spatio-temporelle et sémantique des données

Un premier type de problème correspond à l’incomplétude des données collectées. Dans le cadre du programme MK, le problème se pose à la fois dans la nécessaire continuité spatio-temporelle des informations et dans l’ajout de données sémantiques par le processus d’enquête.

D’une part, concernant les données issues de la collecte GPS, il est essentiel de s’assurer de leur continuité spatio-temporelle dans la perspective d’analyser les agendas individuels. Pour cela, plusieurs vérifications sont mises en œuvre sur l’ensemble de la cohorte (Fig.4) : absence de données enregistrées, alternance stricte

arrêt-trajet, adresses identiques d'arrivée et de départ de trajets successifs, continuité des identifiants dans les tables.

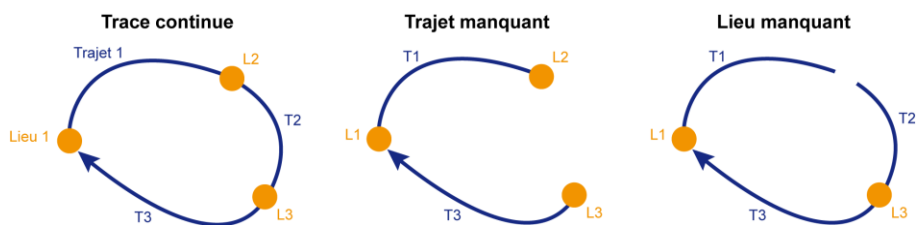


FIGURE 4. Exemples d'erreurs de non continuité des traces.

Table	Critères	Conditions	Nombre	En %
Trajets	Mode de transport et accompagnement	non renseigné	467	12,31
	Lieu de départ	non renseigné	291	7,67
	Lieu d'arrivée	non renseigné	304	8,01
Lieux	Type de lieu	non renseigné	288	12,08
	Fréquence d'usage	non renseigné	390	16,36

TABLE 1. Exemples d'incomplétudes sémantiques

D'autre part, l'incomplétude des données est liée à l'absence de renseignements sémantiques des trajets et des lieux ayant fait l'objet d'un enrichissement par l'enquête (Tab.1). On dénombre par exemple 12,3% des trajets non renseignés selon les modes de transport et d'accompagnement, quand 16,4% des lieux ne sont pas caractérisés par une fréquence d'usage.

3.2.2. Précision géographique et temporelle

Un second type de problème relève de la précision des traces. Sur le plan temporel, les données collectées sont de bonne facture puisque chaque position est caractérisée à la seconde près. Cependant, chaque arrêt et trajet est défini seulement par ses bornes de début et de fin, ce qui implique une perte des temporalités détaillées au sein de cet intervalle.

Par contre, l'imprécision géographique est plus importante au sein des données séquencées puisque la localisation des arrêts ou le tracé des trajets dépendent des positions GPS et de l'algorithme de séquençage. Ceux-ci peuvent donc être imprécis et situés à distance des lieux ou des infrastructures (Fig.5) : la distance entre des traces ayant utilisé la même route peut dépasser plusieurs centaines de mètres. De même, la distance des points au barycentre d'une école varie entre 84m et 450m.



FIGURE 5. Imprécision des trajets à l'entrée d'Orgères (à gauche) et des lieux associés à l'une des écoles (à droite).

3.2.3. Cohérence logique et sémantique

Un troisième type d'erreurs correspond aux incohérences logiques et sémantiques dans – et entre – les tables enrichies. Mises en exergue notamment par des requêtes sur les variables qui caractérisent les trajets et les arrêts, ces incohérences sont diverses (Tab.2). Une partie correspond aux longueurs, durées et vitesses aberrantes : par exemple près de 6% des trajets présentent des durées inférieures à 30s ou supérieures à 3h. De même, d'autres erreurs concernent les activités, les modes de transport et les modes d'accompagnement. Ces incohérences peuvent également être révélées par la combinaison de plusieurs variables : on observe par exemple des « enfants conducteurs d'une voiture », des « enfants à l'école le dimanche », ou encore des « piétons se déplaçant à plus de 20km/h ».

TABLE 2. Exemples d'incohérences logiques et sémantiques

Table	Critères	Conditions	Nombre	En %
Trajets	Durée	< 30sc et > 3h	225	5,93
	Vitesse	< 1km/h et > 100km/h	151	3,98
	Statut familial et mode de transport	Enfant et conducteur	36	0,95
	Mode de transport et vitesse	Marche et > 20km/h	63	1,66
Lieux	Durée	< 300sc	247	10,37
	Date de fin	Supérieur à la date de rendu des GPS	74	3,11
	Type de lieu et durée	Non-domicile et > 12h	162	6,80

4. Mise en application de l'apurement et des corrections des données

Cette classification associée à une quantification des erreurs constitue un préalable à la définition d'une procédure d'apurement visant à améliorer la qualité générale des données. Dans la mesure où l'enquête MK se déroule en deux temps, il est nécessaire que le protocole d'apurement soit à la fois reproductible et qu'il tienne également compte de la diversité des besoins d'analyse. Cette partie présente en ce

sens les principales lignes directrices et quelques résultats préliminaires du travail en cours de définition et de mise en œuvre de ce protocole d'apurement.

La clé de voute du protocole consiste en un apurement générique, appliqué à l'ensemble de la base de données relationnelle et qui vise à améliorer la qualité interne. Cet apurement s'organise en plusieurs étapes :

- Étape 1 : Identification semi-automatique des erreurs de complétude, de cohérence et de précision grâce à des requêtes prédéfinies collectivement (cf. 3.2 ; Tab.1 et 2). Ces requêtes (menées en SQL, dans R et dans des SIG) sont appliquées à toutes les tables de la BDD, y compris celles ne comportant pas de géométrie spatiale.
- Étape 2 : Définition des actions correctives (Tab.3). Après un étiquetage des erreurs dans les tables, d'autres actions sont envisagées tel que la création dans la BDD d'un type d'objet « inconnu » en plus des trajets et des lieux, qui regroupe les temps invalides selon la complétude ou la cohérence. En outre, plusieurs actions géométriques visent à améliorer la précision spatiale des traces, notamment leur appariement au réseau routier à l'aide d'un SIG. Enfin, d'autres actions portent sur la sémantique : soit la modification si des notes correctives permettent de rectifier les informations, soit le recodage des modalités (notamment pour les réponses aux questions ouvertes), soit la désignation de l'individu comme invalide s'il n'a aucune sémantique.
- Étape 3 : Essais d'implémentation et chaînage des actions correctives. Il s'agit d'abord de tester la mise en œuvre technique et les effets des actions correctives envisagées : en les implémentant dans la BDD, puis en réitérant l'étape 1 de diagnostic. Par exemple, la suppression des positions GPS erronées selon le nombre de satellites permettra, sans re-segmenter, d'améliorer la précision et la cohérence des lieux et des trajets. Ces essais indépendants permettent ensuite de chaîner les corrections en tenant compte des relations entre les différentes tables de la BDD. Par exemple, les « temps inconnus », pour lesquels les données sont incomplètes ou insuffisamment fiables, sont d'abord capturés à partir de l'absence de traces brutes, puis complétées par les mini-trajets (moins de 10m), et enfin terminés par les trajets et les lieux ayant une sémantique lacunaire.
- Étape 4 : Opérationnalisation de la chaîne de corrections dans la BDD. Menée dans un environnement Postgresql, cela doit aboutir à la création d'une version apurée qui servira de base aux différentes analyses. Cependant, la mise en œuvre de ce protocole pourra être renouvelée après chaque nouvelle phase du suivi longitudinal (enfants scolarisés au collège).

A l'issue de ces 4 étapes, une phase de pré-analyse de ce corpus apuré consiste à valider la possibilité de créer des « indicateurs » pour les analyses thématiques envisagées. Ainsi, afin de mesurer l'évolution de l'autonomie de déplacements des enfants entre les différents groupes d'enfants et les phases d'enquêtes, il est nécessaire de créer des indicateurs synthétiques sur les modes de transport et d'accompagnement, sur les itinéraires empruntés ou les séquences d'activité.

TABLE 3. Actions correctives de l’apurement générique

	Spatial (sources : <i>datalogger</i> et séquençage)	Temporel (sources : <i>datalogger</i> et séquençage)	Sémantique (source : enquête)
Complétude	Exemple : Pas de géométrie associée à la trace du lieu ou du trajet. Action : étiquetage des individus, définition d'un temps invalide dans les séquences trajets-lieux.	Exemple : Discontinuité dans la succession des activités ou dans l'alternance des lieu-trajet Actions : balisage des "moments" concernés, définition d'un temps invalide dans les séquences trajets-lieux.	Exemple : Absence de renseignement dans les tables des trajets et des lieux. Actions : étiquetage des ind., recodage si possible à l'aide des notes des lieux ou trajets, suppression si l'enquête n'a aucune information.
Coherence	Exemple : Trajet avec une distance trop courte ; lieu inexistant ou non indexé à une adresse. Actions : étiquetage des ind., définition d'un temps invalide, correction des lieux par les coordonnées et/ou adresses.	Exemple : Durée trop courte ; Traces hors des périodes de collecte. Actions : étiquetage des ind., définition d'un temps invalide selon la durée, suppression des traces « hors délai ».	Exemple : Enfants conducteurs, enfants à l'école le dimanche, piétons à plus de 20km/h Actions : étiquetage des ind., recodage si possible à l'aide des notes correctives des lieux ou trajets.
Précision	Exemple : Localisations diffuses des trajets et des lieux. Actions : agrégation et appariement au réseau routier ou aux parcelles/adresses		Exemple : Diversité des noms déclarés des lieux Actions : étiquetage des ind., création de dictionnaires, recodage dans les tables

5. Conclusion : Vers un apurement adapté et reproductible

Le programme MOBI’KIDS repose sur un protocole méthodologique mixte de collecte de données de mobilités quotidiennes. Celles-ci sont à la fois hétérogènes et multidimensionnelles, spatio-temporelles et sémantiques : des traces numériques par suivi GPS, des renseignements sur les mobilités et lieux fréquentés à partir de questionnaires. Outre les erreurs qu’impliquent ces étapes de collecte, la mise en relation de ces données requiert, au préalable des analyses, la mise en œuvre d’un protocole de diagnostic qualité et d’apurement.

Dans cette perspective, cet article avait pour objectif de présenter la démarche exploratoire suivie dans MOBI’KIDS de diagnostic de la qualité des données collectées. La discussion permet d’établir un état des lieux des erreurs sémantiques,

géométriques et temporelles selon leurs origines (dispositif technique, biais d'enquête, algorithme de séquençage) ou leurs natures (imprécision, incomplétude, incohérence). Ce travail exploratoire de diagnostic ouvre la voie à la définition et la mise en œuvre d'un protocole d'apurement de la base de données relationnelle. Ce protocole est avant tout confronté à un enjeu d'objectivité et de reproductibilité. Il s'agit de construire une chaîne méthodologique qui pourra être déployée aux différents temps de l'enquête longitudinale MOBI'KIDS, puisque les enfants sont également suivis au collège. Mais plus largement, l'enjeu est de proposer une méthode reproductible à d'autres cas d'analyses de mobilités quotidiennes croisant de multiples et diverses informations.

Remerciements : Ces travaux sont financés par MOBIKIDS (ANR-16-CE22-0009)

Bibliographie

- Aalders H. J. G. L. (2002), The Registration of Quality in a GIS, *Spatial Data Quality* (W. Shi, P. Fisher, and M. F. Goodchild, Eds), Taylor & Francis, p. 186-199.
- Beaud S., Weber, F. (1998). *Guide de l'enquête de terrain : produire et analyser des données ethnographiques*, La Découverte, Paris.
- Biljecki F., Ledoux H., Oosterom P. van (2013). Transportation mode-based séquençage and classification of movement trajectories. *International Journal of Geographical Information Science* 27, 385–407.
- Chen C., M J., Susilo Y., Liu, Y., Wang M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285–299.
- Christensen P., Mikkelsen M.R., Nielsen T.A.S., Harder H. (2011). Children, Mobility, and Space: Using GPS and Mobile Phone Technologies in Ethnographic Research. *Journal of Mixed Methods Research* 5, 227–246.
- CERTU (2008). *L'enquête ménages déplacements « standard Certu »*, Guide méthodologique.
- Dalumpines R., Scott D.M. (2017). Making mode detection transferable: extracting activity and travel episodes from GPS data using the multinomial logit model and Python. *Transportation Planning and Technology* 40, 523–539.
- David B., Fasquel P. (1997). Qualité d'une base de données géographique : concepts et terminologie. *Bulletin d'information de l'IGN*, n°67.
- Depeau S., Bedel O., Chérel P., André-Poyaud I., Chardonnel S., Gombaudo J., Jambon, F., Lepetit A., Mericskay B., Quesseveur E. (2019) MK-MOBIBACK : un dispositif hybride et intégré pour enquêter finement les mobilités quotidiennes des familles, Communication Acceptée, Conférence SAGEO 2019.
- Depeau S., Chardonnel S., André-Poyaud I., Lepetit A., Jambon F., Quesseveur E., Gombaudo J., Allard T., Choquet C.-A. (2017). Routines and informal situations in children's daily lives. *Travel Behaviour and Society* 9, 70–80.

- Depeau, S. & Quesseveur, E. (2014). A la recherche d'espaces invisibles de la mobilité : usages, apports et limites des techniques GPS dans l'étude des déplacements urbains à l'échelle pédestre. *Netcom*, vol 28, n°1-2, p. 35-54.
- Devillers R. (2004). *Conception d'un système multidimensionnel d'information sur la qualité des données géospatiales*. Thèse de doctorat, Université de Marne la Vallée.
- Devillers R., Jeansoulin R. (2005). *Fundamentals of Spatial Data Quality*, ISTE, London, Newport Beach.
- Drevon, G., Jambon, F., Chardonnel, S., Christophe, S., André-Poyaud, I., Davoine, PA, Lutoff, Céline (2014). Évaluation comparée de l'apport de l'assistance GPS aux enquêtes de mobilité. *Netcom*, (28–1/2), 13–34.
- Guptill S.C., Morrison J.L. (1995). *Elements of Spatial Data Quality*, Elsevier, Oxford.
- Kim N. H., Park, C. H. (2017). Simulation Analysis of GPS/GLONASS Absolute Positioning Performance in an Urban Canyon Environment. *International Journal of Computer Theory and Engineering*, 9(1).
- Kyttä M., Oliver M., Ikeda E., Ahmadi E., Omiya I., Laatikainen T. (2018). Children as urbanites: mapping the affordances and behavior settings of urban environments for Finnish and Japanese children. *Children's Geographies* 16, 319–332.
- Lenormand M., Picornell M., Cantú-Ros O.G., Tugores A., Louail T., Herranz R., Barthelemy M., Frías-Martínez E., Ramasco J.J. (2014). Cross-Checking Different Sources of Mobility Information. *PLoS ONE* 9, e105184.
- Lin M., Hsu W.-J. (2014). Mining GPS data for mobility patterns: A survey. *Pervasive and Mobile Computing*, 12, 1–16.
- Lord S., Després M., Kestens Y. (2018). Approcher la complexité de la mobilité et des territoires dans la vieillesse: l'intégration de données et de méthodes mixtes dans l'analyse des relations personne-environnement. *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*, Montréal, ACFAS p 194-207.
- Neatt K., Millward H., Spinney J. (2016, April). Aggregation and spatial analysis of walking activity in an urban area: results from the Halifax space-time activity survey. In *IOP Conference Series: Earth and Environmental Science* (Vol. 34, No. 1, p. 012022). IOP Publishing.
- Servigne S., Lesage N., Libourel Th. (2005). Quality Components, Standards, and Metadata. *Fundamentals of Spatial Data Quality*, ISTE, London, Newport Beach, p. 179-210.
- Shoval N., Kwan M.-P., Reinau K.H., Harder H. (2014). The shoemaker's son always goes barefoot: Implementations of GPS and other tracking technologies for geographic research. *Geoforum* 51, 1–5.
- Stopher P. (2009). Collecting and Processing Data from Mobile Technologies, in: Bonnel, P. (Ed.), *Transport Survey Methods: Keeping up with a Changing World*, United Kingdom, Emerald, p. 361-391.
- Site de l'INED : Saisie, codage, apurement, documentation, <https://www.ined.fr/fr/ressources-methodes/methodologie-enquete/les-choix-methodologiques/saisie-codage-apurement-documentation/> ; consulté en avril 2019.

Enrichissement d'orthophotographie par des données OpenStreetMap pour l'apprentissage machine

Gauthier Fillières-Riveau¹, Vincent Barra¹,
Jean-Marie Favreau¹, Guillaume Touya²

1. Université Clermont Auvergne (UCA) - LIMOS UMR 6158,
Clermont-Ferrand, France

gauthier.fillieres@uca.fr;vincent.barra@uca.fr;j-marie.favreau@uca.fr

2. Univ. Paris-Est, LASTIG GEOVIS, IGN, ENSG, Saint-Mande, France

Guillaume.Touya@ign.fr

RÉSUMÉ. Les données géographiques présentes dans OpenStreetMap (OSM), décrites à la fois par leur géométrie et de la sémantique, constituent une source d'information essentielle à la fabrication de cartes en relief pour déficients visuels. Cependant, à l'échelle du carrefour, les détails géométriques nécessaires à la modélisation de certains éléments ne sont pas disponibles. L'utilisation de l'imagerie aérienne constitue une source d'information complémentaire, mais impose un traitement d'image sophistiqué. Dans cet article, nous proposons une approche basée sur l'apprentissage profond (réseau adverse génératif conditionnel), en enrichissant les informations présentes dans les orthophotographies par des données sémantiques et géométriques issues d'OSM. Afin de mesurer l'influence de cet enrichissement, nous présentons les résultats de deux séries d'apprentissage, avec et sans enrichissement.

ABSTRACT. The geographic data in OpenStreetMap (OSM), described through geometry and semantics, is an essential source of information for relief maps production for the visually impaired. However, at the crossroads scale, the geometric details required for modeling some components are not available. The use of aerial imagery is a complementary source of information, but involves advanced image processing. In this article, we propose an approach based on deep learning (conditional Generative Adversarial Networks), by enhancing the information of orthophotographs with semantic and geometric data from OSM. In order to measure the influence of this enrichment, we present the results of two series of learning, with and without enrichment.

MOTS-CLÉS : apprentissage machine ; enrichissement de données ; information géographique volontaire ; segmentation sémantique

KEYWORDS: machine learning ; data enrichment ; VGI ; semantic segmentation

1. Introduction

1.1. Problématiques

La transcription de cartes pour un public en situation de déficience visuelle (PSDV) est aujourd'hui un processus manuel, réalisé par des professionnels tels que les *adaptateurs-transcripteurs* qui sont chargés de la transcription de document par l'adaptation de texte en braille et la préparation pour la mise en relief des contenus graphiques, et les *instructeurs en locomotion* (IL) dont le rôle est de faciliter les déplacements des PSDV, notamment pour la navigation urbaine. Ce processus de transcription est long, et nécessite une compétence experte importante (Kern, 2016). Dans ce travail, nous nous intéressons à l'adaptation et l'automatisation des processus de conception de cartes destinées aux PSDV.

Parmi les usages spécifiques aux PSDV, l'utilisation de cartes à l'échelle du carrefour est une pratique courante (figure 1 a). Utilisés par les IL, ces plans permettent aux PSDV de s'appropriier l'espace et les aménagements urbains.

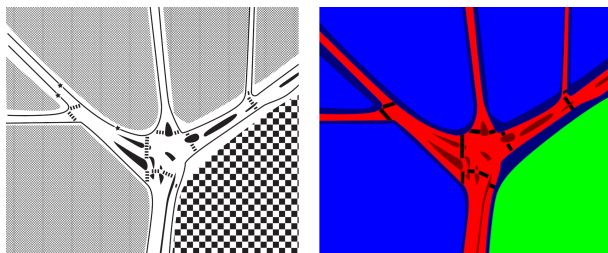


FIGURE 1. a- Une carte de carrefour adaptée par une professionnelle (© CRDV) pour un PSDV, b- la segmentation correspondante dans nos travaux

D'après le travail de collecte de données que nous avons réalisé¹, les cartes fabriquées pour ce besoin spécifique représentent majoritairement les éléments suivants : routes, terre-plein et îlots routiers, trottoirs, parcelles closes et/ou bâtiments, espaces verts, passage piétons, tracé des tramways et des voies ferrées, arrêts de bus. Tous ces éléments sont représentés par leur empreinte réelle au sol, à l'exception des arrêts de bus qui sont représentés par un élément ponctuel, et des voies de tramway et voies ferrées qui sont tracées en linéaire. Construire une telle carte revient donc à réaliser une segmentation sémantique, où chaque élément du territoire est étiqueté par l'un de ces éléments.

1. <https://activmap.limos.fr/Participez/Collecte/> : en juin 2019, 113 cartes collectées auprès d'adaptateurs transcripateurs et d'instructeurs en locomotion

Une part importante de ces informations sont présentes dans les bases de données géographiques telles qu’OpenStreetMap (OSM). Cependant, si OSM est une ressource pertinente pour la production de cartes classiques (Girres, Touya, 2010), notamment dans le cadre de classification par apprentissage machine (J. Chen, Zipf, 2017 ; Kurath *et al.*, 2017), nous identifions facilement des limitations dans le cas de l’application envisagée : des éléments significatifs tels que les trottoirs et les îlots centraux ne sont quasiment pas présents dans les bases de données, et un grand nombre d’informations sont décrites de manière ponctuelle (passages piétons) ou linéaire (routes), leur emprise réelle n’étant pas renseignée. D’autres bases de données contiennent ces informations mais ne proposant pas un modèle ouvert d’accès aux données et/ou de représentation mondiale, nous les avons écartées. En étudiant le processus de fabrication de ces cartes, nous avons constaté que les professionnels de la déficience visuelle s’appuyaient en grande partie sur les orthophotographies des carrefours représentés. Nous nous sommes donc inspiré de cette pratique pour tenter de palier les lacunes précédentes d’OSM.

1.2. *Approches existantes*

Différentes approches permettent d’améliorer la perception de la carte pour les PSDV. Certains travaux se concentrent sur la sélection et la mise en relief de la carte, avec une représentation des données telles que présentes dans la base de référence (Červenka *et al.*, 2016) ou après généralisation (Touya *et al.*, 2018). D’autres se focalisent sur l’augmentation de la représentation par l’ajout d’informations sonores (Josselin *et al.*, 2016 ; Boularouk *et al.*, 2017). Il existe également des travaux combinant ces deux approches, mais la carte est représentée avec un style symbolique des données (Watanabe *et al.*, 2014). Dans le cas d’utilisation que nous envisageons, la représentation se doit d’être fidèle aux réalités du terrain, malgré les informations d’emprise incomplètes ou non présentes dans la base de données de référence. De plus, le premier usage envisagé de ces cartes correspond à une utilisation par des IL accompagnant les PSDV sur le terrain. L’enrichissement par information sonore n’est donc pas nécessaire dans un premier temps.

Nous présentons dans cet article une approche basée sur l’apprentissage profond. L’arrivée progressive des réseaux de neurones à convolution (CNNs) dans de nombreux domaines applicatifs, notamment en imagerie (Guo *et al.*, 2016), a rendu caduque un grand nombre de méthodes classiques de segmentation issues des mathématiques et heuristiques (Shotton *et al.*, 2008). Les réseaux neuronaux permettent également la prise en compte de savoir-faire dans le processus d’entraînement, ce qui conforte l’orientation de nos travaux vers ces techniques. Les CNNs fonctionnent généralement avec un mécanisme de couches de mise en commun (*pooling*), ce qui pose problème pour la réalisation de segmentations sémantiques, ces couches entraînant une perte de l’information de localisation et de contexte initial. Ce constat a conduit à l’émergence d’un nouveau para-

digne : l'utilisation de l'information globale et locale, avec l'arrivée des *Fully Connected Networks* (FCN) (Long *et al.*, 2014).

On trouve dans la littérature deux approches basées sur ce paradigme pour répondre aux problèmes de segmentation : les réseaux de type encodeur-décodeur (Ronneberger *et al.*, 2015), adaptés des réseaux utilisés en classification, et ceux utilisant des convolutions dilatées (Yu, Koltun, 2016 ; L. Chen *et al.*, 2016), basées sur une conception dédiée aux problèmes de segmentation. Ces approches font référence dans le domaine de la segmentation sémantique, cependant les données issues des cartes que nous avons collectées comportent des biais et incertitudes qui après quelques expérimentations se sont révélées incompatibles avec ces méthodes. En effet les passages piétons et terre-pleins sont très mal détectés (<15%) et les faux positifs sont abondants (>70%) avec un U-Net (Ronneberger *et al.*, 2015), contrairement aux résultats obtenus par un réseau de type *Generative Adversarial Networks* (GAN) (Goodfellow *et al.*, 2014) (tableau 2) entraîné en parallèle pour un même jeu de donnée. Ce dernier semble donc plus adapté à nos données pour cette segmentation. Les méthodes de génération d'image par un GAN, et plus particulièrement le GAN conditionnel (cGAN) (Mirza, Osindero, 2014 ; Isola *et al.*, 2016) peuvent apporter des résultats en segmentation, généralement moins probants que ceux des techniques dédiées à la segmentation, cependant nos premières expérimentations montrent des résultats prometteurs avec les cartes collectées.

Nous proposons dans la suite de cet article une approche par apprentissage couplant des données OSM et des orthophotographies en s'appuyant sur un ensemble de cartes collectées, afin de produire une segmentation automatique de carrefours pour la production de cartes en relief. La troisième partie présente l'implémentation et les résultats obtenus sur des carrefours dans la ville de Clermont-Ferrand. Enfin, la dernière partie propose une synthèse et ouvre différentes perspectives de recherche.

2. Méthode

2.1. GAN conditionnel

L'approche que nous proposons s'appuie sur un cGAN (Isola *et al.*, 2016) dont le but est d'apprendre à modéliser la mise en correspondance entre une image d'entrée et une image de sortie. Dans cette approche, le réseau comporte deux composantes principales : le *générateur* et le *discriminateur*. Le générateur utilisé est un encodeur-décodeur U-Net, appliquant des transformations à l'image d'entrée et générant une nouvelle image ayant les mêmes dimensions. Le rôle du discriminateur est de déterminer si l'image de sortie a été produite par le générateur ou si elle est « vraie », en utilisant un discriminateur Markovien spécifique (patchGAN).

L'entraînement du réseau s'effectue en deux étapes, la première pour entraîner le discriminateur, la seconde pour entraîner le générateur (figure 2). Le principe général consiste à améliorer d'abord le discriminateur. Cela engendre l'amélioration du générateur qui cherche à tromper le discriminateur en faisant passer l'image qu'il génère pour une image cible. S'il réussit alors le discriminateur est ajusté, sinon ce sont les poids du générateur qui sont modifiés. Le réseau s'entraîne en bouclant de façon cyclique sur le jeu de données.

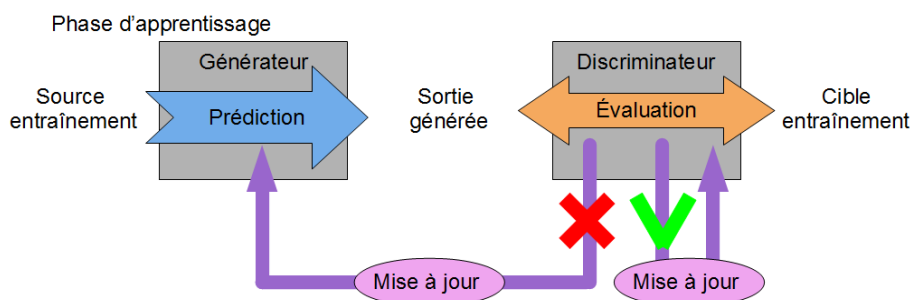


FIGURE 2. Schéma simplifié du processus d'entraînement du GAN conditionnel.

L'utilisation la plus répandue de GAN est la génération d'image, notamment à partir de descriptions textuelles (Zhang *et al.*, 2016) ou de segmentation (Isola *et al.*, 2016 ; Wang *et al.*, 2017).

2.1.1. Données sources

Les données d'entrée du réseau correspondent aux images que l'on cherche à segmenter. Pour adapter l'approche à notre problématique, nous avons choisi des données accessibles et libres. Deux types de données ont ainsi été retenues : les *orthophotographies* (ou orthophotos), qui sont des images issues de prise de vue aériennes, et les *données géographiques vectorielles* d'OSM. Les orthophotos comportent plusieurs spécificités intéressantes qui facilitent leur utilisation en segmentation par apprentissage machine (Kaiser *et al.*, 2017). Elles sont géométriquement rectifiées et recalées dans un système de coordonnées exploitable par un système d'information géographique (SIG). Pour homogénéiser leur traitement, elles sont corrigées radiométriquement (luminosité, saturation et contraste des couleurs unifiées). OSM est une base de données géographique volontaire (VGI) proposant une grande quantité d'informations avec une pertinence et une précision significatives (Haklay, 2010 ; Zielstra, Zipf, 2010). OSM s'appuie sur une représentation des données par géométrie vectorielle, augmentées d'un ensemble d'étiquettes (association clé / valeur), et de relations entre ces objets géométriques. Proposées sous licence libre, OSM est ouvert à la contribution de tout volontaire. Sa représentation géométrie/sémantique permet un enrichissement régulier assuré par des conventions de description adoptées de manière collégiales par la communauté des contributeurs.

Les données issues d'OSM ont été ajoutées par superposition aux orthophotos avec la couleur de la classe correspondante. Nous avons ainsi choisi pour les routes un rouge saturé (RGB 255; 0; 0), très peu présent en photographie naturelle, et complètement absent des données orthophotos que nous avons utilisées. Les tracés vectoriels sont discrétisés au plus fin (largeur d'un pixel) dans le but de minimiser l'impact sur les données orthophotos.

Ces images sont ensuite découpées en vignettes de 256x256 pixels, en utilisant la résolution de l'orthophoto. Ce choix de définition des données d'entrée résulte du compromis entre la quantité d'information à cette dimension pour notre échelle, la résolution des orthophotos utilisées et le temps d'apprentissage.

2.1.2. Données cibles

Les données cibles servent à la phase d'entraînement et de tests du réseau, et correspondent aux segmentations de référence. En puisant dans les cartes issues de notre collecte, nous avons sélectionné 15 cartes adaptées, en retenant celles correspondant à l'échelle du carrefour, et transmises dans un format numérique exploitable. Cette quantité de cartes peut sembler faible pour entraîner un réseau de neurones, cependant, la couverture géographique et le format vectoriel permettent l'export avec une grande résolution en corrélation avec les orthophotos. De fait, la subdivision pour coller à la définition utilisée en entrée permet d'obtenir de nombreux échantillons d'entraînement. Nos données d'entrée et cibles ayant les mêmes définitions, nous pouvons consolider la quantité de données par les techniques d'augmentation (Taylor, Nitschke, 2017).

Une étape de pré-traitement est nécessaire pour préparer ces cartes au format attendu par le cGAN. Les cartes initiales (figure 1 a) en noir et blanc représentent les classes par des textures ou par la matérialisation des frontières. Nous avons transposé les motifs et régions présentes par un code couleur définissant les catégories d'éléments représentés (figure 1 b). Ces cartes non géoréférencées et préparées pour une impression au format A4 doivent être recalées sur les données géographiques.

2.2. Définition de l'erreur et de sa mesure

Pour évaluer un réseau de neurones, on teste les résultats du modèle par comparaison des images de référence attendues pour un jeu de données que l'on a écarté de l'étape d'entraînement.

Afin de mesurer la qualité de la segmentation, nous proposons de quantifier l'erreur de classification pour de chaque pixel, que l'on classe en déterminant son appartenance à un ensemble de classes définies. L'erreur correspond alors à une différence entre la classe détectée et la classe réelle.

En cartographie, les ontologies géographiques et les liens sémantiques sont des domaines grandement étudiés (Couclelis, 2010 ; Dobesova, Brus, 2011), nous proposons de les utiliser pour classer les erreurs de segmentation.

En considérant un niveau d'erreur de classification inversement proportionnel au niveau de lien sémantique entre la catégorie déterminée par la segmentation et celle correspondant à la classe issue de l'image de référence, une mesure d'erreur sémantique est calculée. Dans l'exemple de la figure 3, l'erreur de classification de la classe passage piéton avec celle de la route est moindre qu'avec l'élément terre-plein ou trottoir, et elle est maximale pour l'association avec la classe parcelle.

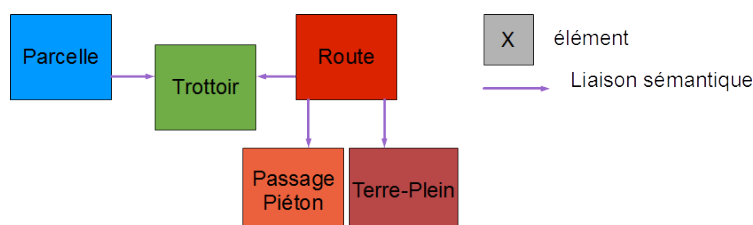


FIGURE 3. représentation non exhaustive de liens sémantiques entre des éléments de notre segmentation.

De plus, nous avons choisi de pondérer l'erreur en fonction des répercussions occasionnées dans l'utilisation. L'objectif étant la navigation piétonne, les classifications erronées d'éléments ne perturbant pas la navigation seront ainsi minorées, tandis que celles entravant les déplacements sont majorées.

Nous proposons donc la classification d'erreur suivante : **faible** : il existe un lien fort et la navigation n'est pas ou peu influencée; **modérée** : il existe un lien et la navigation est influencée, ou, il n'existe pas de lien mais aucune influence sur la navigation; **forte** : il n'existe pas de lien et la navigation est perturbée.

3. Implémentation et résultats

L'implémentation de la méthode est faite en suivant les processus tels que décrit sur le schéma (figure 4), en partant de la préparation des données, puis la phase d'apprentissage suivie du test et des mesures d'erreur du modèle afin de pouvoir analyser nos résultats.

Le cGAN utilisé est basé sur l'implémentation TensorFlow² du modèle proposé par (Isola *et al.*, 2016). La phase d'entraînement a été réalisée sur un ser-

2. <https://github.com/affinelayer/pix2pix-tensorflow>

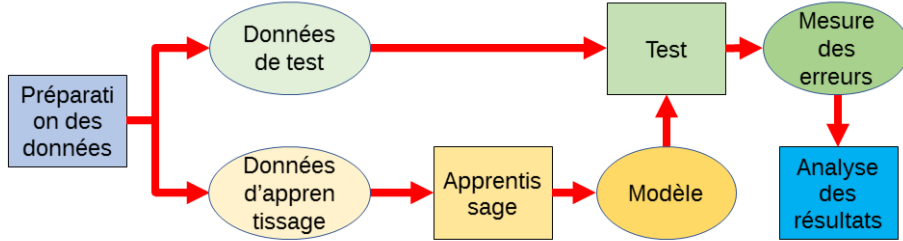


FIGURE 4. schéma global du processus implémenté.

veur de calcul équipé d'un processeur Intel®Xeon®CPU E5-2640 v4 @ 2.40GHz et d'une carte NVIDIA®GP104GL.

La préparation des données nous a permis d'obtenir deux jeux de 375 paires d'images sources/cibles. Les deux jeux sont identiques à cela près que le premier comporte les orthophotos simples alors que le second comprend les orthophotos enrichies des données OSM. Nous avons utilisé 300 de ces paires pour l'entraînement et les 75 restantes pour l'évaluation des modèles.

L'évaluation de l'erreur a été réalisée en suivant la méthode présentée section 2.2, pour produire la pondération illustré par le tableau 1.

TABLE 1. Classification de l'importance des erreurs en fonction de l'association des classes

Classe	Route	Passage piéton	Terre plein	trottoir	parcelle/bâti	espace vert
route	-	faible	modérée	faible	haute	haute
Passage piéton	faible	-	faible	faible	haute	haute
terre plein	modérée	faible	-	haute	haute	haute
trottoir	faible	faible	haute	-	modérée	modérée
parcelle/bâti	haute	haute	haute	modérée	-	modérée
espace vert	haute	haute	haute	modérée	modérée	-

3.1. Données utilisées

Nous avons utilisé une orthophoto haute résolution (ORTHO HR®), co-produite par l'Institut National de l'Information Géographique et Forestière (IGN) et le Centre Régional Auvergne-Rhône-Alpes de l'Information Géographique (CRAIG) pour le Puy-de-Dôme. Avec une résolution de 20 cm à 25 cm par pixel, elle permet de décrire au mieux l'échelle du carrefour qui nous in-

téresse, et est disponible sous licence ouverte³. Les données OSM de la région Auvergne ont été téléchargées courant 2018 depuis le serveur geofabrik⁴.

La préparation des données a été réalisée avec le logiciel libre de conception et d'analyse cartographique QGIS⁵. Un plugin QGIS écrit en Python pour un usage interne à l'équipe a été utilisé pour isoler le tracé des routes OSM : en filtrant les différentes associations de clé/valeur dans les étiquettes OSM, le plugin crée une couche vectorielle correspondante pour chaque classe sémantique avec un style appliqué prédéfini. En agrégeant la couche regroupant le tracé des routes à la couche composée de l'orthophoto, nous obtenons notre enrichissement. Les zones correspondant à nos cartes sont alors exportées avec et sans enrichissement pour générer le jeu de données d'entrée.

Les cartes collectées ont été coloriées grâce à un logiciel de dessin vectoriel⁶ afin de préparer les cartes texturées au format attendu des cartes de segmentation (figure 1). Les déformations et simplifications effectuées par les professionnels lors de l'adaptation rendant difficile l'utilisation des outils de géoréférencement de QGIS, nous avons effectué un appariement manuel de mise en correspondance des données réelles et de la segmentation. L'emprise grossière et des simplifications de géométries de certains éléments apportent cependant des imperfections et débordements pour certaines classes.

Les paramètres du modèle cGAN utilisé sont ceux proposés par défaut, à l'exception du nombre de cycle complet (*epoch*) pour lequel nous avons testé différentes valeurs, et qui donne de bons résultats à partir de 200 epochs avec le jeu de données enrichies utilisé. Les résultats présentés dans cet article sont issus d'entraînement stoppé à 300 epochs.

3.2. Analyse des résultats

Afin de visualiser les améliorations apportées par l'enrichissement de données réalisé, nous avons procédé à une comparaison des résultats obtenus pour un jeu de test identique avec et sans cet enrichissement. Les erreurs sont relevées et classées suivant leur importance entre l'image de référence et celle générée par le modèle.

Les résultats (tableau 2) confirment nettement l'amélioration apportée par enrichissement. Le type d'erreur constaté met en avant une amélioration de la cohérence de la segmentation. Nous remarquons ainsi que le gain de précision et de cohésion proviennent principalement de la diminution du nombre d'erreurs de forte importance. C'est ici majoritairement l'association route et

3. <http://professionnels.ign.fr/orthohr-par-departements/#tab-3>

4. <http://download.geofabrik.de/europe/france/auvergne.html>

5. <https://www.qgis.org/fr/site/>, version 2.18

6. Inkscape : <https://inkscape.org>

TABLE 2. Résultats sur le jeu de test : proportion de bonne classification et erreurs par ordre d'importance pour les données avec et sans OSM.

Mesure	modèle sans OSM	modèle avec OSM
Bonne classification	70.76 %	81.9 %
Erreur faible	13.73 %	10.54 %
Erreur modérée	7.06 %	5.36 %
Erreur forte	8.45 %	2.21 %

parcelles/bâtis qui est en cause : sans enrichissement, certains toits de bâtiments et cours privatives sont identifiés comme route, ce qui ne se produit plus en utilisant les données OSM. Nous constatons également que la segmentation des routes avec enrichissement ne souffre plus des problèmes d'occlusion (arbres par exemple) et de détection dans les zones recouvertes d'ombre.

On peut en grande partie expliquer les erreurs de forte importance restantes par les biais dans nos cartes de référence. Les professionnels se concentrent en effet sur l'information aux abords directs des intersections, en délaissant et en simplifiant les informations mineures en périphérie de carte, telles que des rues peu fréquentées (figure 5). À l'inverse, notre segmentation incluant les données OSM travaille sans faire la distinction entre le centre et la périphérie de la carte.



FIGURE 5. Biais sur la classification des routes entre la carte adaptée manuellement (gauche), l'orthophoto enrichie (centre) et le résultat généré par le réseau entraîné avec les données OSM (droite)

Afin de mesurer l'importance de l'accroche aux données OSM dans l'apprentissage, nous avons testé le modèle appris en incluant des erreurs dans les données enrichies : absence d'un tronçon de tracé de route (figure 6), absence complète de tracé avec l'utilisation de données non enrichies, ou mauvais placement du tracé. À l'exception du cas de l'absence partielle de tracé de route, les performances chutent en deçà de celles sans enrichissement, ce qui montre l'influence des données OSM enrichies sur l'apprentissage par le réseau.

Cela met également en évidence le fait que la méthode tire partie des informations liées à la sémantique, telles que la liaison passage piéton / route (quart inférieur gauche figure 6). En effet, un passage piéton bien détecté en présence du tracé de la route ne l'est plus en retirant ce tracé.



FIGURE 6. Erreur issue d'erreur de données OSM sur la classification. Réf rence (gauche), orthophoto avec donn es OSM incompl tes (centre), r sultat du r seau entra n  avec enrichissement (droite)

4. Conclusion et perspectives

L'utilisation de donn es OSM en superposition d'orthophotos permet bien de guider et am liorer les r sultats d'apprentissage pour notre cas d'utilisation. Nous avons pu constater un gain de pr cision et de coh rence s mantique des segmentations produites. L'utilisation d'autres  l ments OSM, tels que les contours parcellaires ou les points de passages pi tons pour enrichir les donn es d'entr e et guider l'apprentissage est la prochaine  tape logique de nos travaux.

On peut  galement travailler sur certains  l ments sur-repr sent s en terme de surface (parcelle / b ti) de la segmentation pour att nuer la pr pond rance d'une classe sur toutes les autres afin de se concentrer sur l'affinage de l'espace de navigation. Il serait ainsi envisageable de ne plus op rer une segmentation compl te, o  chaque pixel appartient   une classe, mais d'effectuer la classification uniquement sur les pixels d tect s comme des  l ments d'int r ts pour lesquels les donn es issues du VGI sont incompl tes.

Une autre suite envisag e consisterait   apprendre avec des donn es d'autres sources, comme les mod les num riques de surfaces (MNS), qui ne seraient pas utilis es par la suite lors de la phase de segmentation avec le mod le appris. Dans le principe un r seau secondaire apprend   g n rer ces sources annexes, et ensuite ce second mod le est utilis  comme une forme d'extracteur de d'informations compl mentaires sur les donn es d'entr e (Piasco *et al.*, 2018).

Bibliographie

- Boularouk S., Josselin D., Altman E. (2017). Ontology for a Voice Transcription of OpenStreetMap Data: The Case of Space Apprehension by Visually Impaired Persons. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 11, n  5, p. 585–589.
- Chen J., Zipf A. (2017, 04). Deepvgi: Deep learning with volunteered geographic information. In, p. 771-772.
- Chen L., Papandreou G., Kokkinos I., Murphy K., Yuille A. L. (2016). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution,

SAGEO'2019

and fully connected crfs. *CoRR*, vol. abs/1606.00915.

- Couclelis H. (2010, 11). Ontologies of geographic information. *International Journal of Geographical Information Science*, vol. 24, p. 1785-1809.
- Dobesova Z., Brus J. (2011). Coping with cartographical ontology. In *Proceedings of SGEM*, p. 377-384.
- Girres J.-F., Touya G. (2010, aug). Quality assessment of the french openstreetmap dataset. *Transactions in GIS*, vol. 14, n° 4, p. 435-460.
- Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S. *et al.* (2014, Jun). Generative Adversarial Networks. *arXiv e-prints*, p. arXiv:1406.2661.
- Guo Y., Liu Y., Oerlemans A., Lao S., Wu S., Lew M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, vol. 187, p. 27 - 48.
- Haklay M. (2010). How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and Planning B: Planning and Design*, vol. 37, n° 4, p. 682-703.
- Isola P., Zhu J., Zhou T., Efros A. A. (2016). Image-to-image translation with conditional adversarial networks. *CoRR*, vol. abs/1611.07004.
- Josselin D., Roussel D., Boularouk S., Saidi A., Matrouf D., Bonin O. *et al.* (2016, juin). Sonorous cartography for sighted and blind people. In *AGILE'2016 - 19th AGILE International Conference on Geographic Information Science*. Helsinki, Finland.
- Kaiser P., Wegner J. D., Lucchi A., Jaggi M., Hofmann T., Schindler K. (2017). Learning aerial image segmentation from online maps. *CoRR*, vol. abs/1707.06879.
- Kern R. (2016, décembre). Cartographie et malvoyance - du papier au numérique. *Cartes & Géomatique*, vol. 229-230, p. 167-174.
- Kurath S., Das Gupta R., Keller S. (2017, 01). Osmdeepod - object detection on orthophotos with and for vgi. , vol. 1, p. 173-188.
- Long J., Shelhamer E., Darrell T. (2014). Fully convolutional networks for semantic segmentation. *CoRR*, vol. abs/1411.4038.
- Mirza M., Osindero S. (2014). Conditional generative adversarial nets. *CoRR*, vol. abs/1411.1784.
- Piasco N., Sidibé D., Gouet-Brunet V., Demonceaux C. (2018, juin). Apprentissage de modalités auxiliaires pour la localisation basée vision. In *Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP)*. Marne-la-Vallée, France.
- Ronneberger O., Fischer P., Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation. *CoRR*, vol. abs/1505.04597.
- Shotton J., Johnson M., Cipolla R. (2008, June). Semantic texton forests for image categorization and segmentation. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, p. 1-8.
- Taylor L., Nitschke G. (2017). Improving deep learning using generic data augmentation. *CoRR*, vol. abs/1708.06020.

- Touya G., Christophe S., Favreau J.-M., Ben Rhaïem A. (2018). Automatic derivation of on-demand tactile maps for visually impaired people: first experiments and research agenda. *International Journal of Cartography*.
- Wang T., Liu M., Zhu J., Tao A., Kautz J., Catanzaro B. (2017). High-resolution image synthesis and semantic manipulation with conditional gans. *CoRR*, vol. abs/1711.11585.
- Watanabe T., Yamaguchi T., Koda S., Minatani K. (2014, juillet). Tactile Map Automated Creation System Using OpenStreetMap. In *Proceedings of 14th ICCHP*. Paris, France.
- Yu F., Koltun V. (2016). Multi-scale context aggregation by dilated convolutions. *CoRR*, vol. abs/1511.07122.
- Zhang H., Xu T., Li H., Zhang S., Huang X., Wang X. *et al.* (2016). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. *CoRR*, vol. abs/1612.03242.
- Zielstra D., Zipf A. (2010). A comparative study of proprietary geodata and volunteered geographic information for germany. In *13th AGILE international conference on geographic information science*, vol. 2010.
- Červenka P., Břinda K., Hanousková M., Hofman P., Seifert R. (2016). Blind Friendly Maps. In K. Miesenberger, C. Bühler, P. Penaz (Eds.), *Computers Helping People with Special Needs*, vol. 9759, p. 131–138. Springer International Publishing.

Classification d'aires de dispersion à l'aide d'un facteur géographique : application à la dialectologie.

Clément Chagnaud¹, Philippe Garat²,
Paule-Annick Davoine^{3,1}, Guylaine Brun-Trigaud⁴

1. Université Grenoble Alpes, CNRS, Grenoble INP, LIG
clement.chagnaud@univ-grenoble-alpes.fr

2. Université Grenoble Alpes, CNRS, Grenoble INP, LJK
philippe.garat@univ-grenoble-alpes.fr

3. Université Grenoble Alpes, CNRS, Grenoble INP, PACTE
paule-annick.davoine@univ-grenoble-alpes.fr

4. Université Côte d'Azur, CNRS, BCL
guylaine.brun-trigaud@univ-cotedazur.fr

RÉSUMÉ. Nous proposons une procédure d'analyse statistique multidimensionnelle couplant des méthodes de projection et de classification pour identifier des ensembles cohérents au sein d'un corpus d'entités géographiques surfaciques que l'on appelle aires de dispersion. La méthodologie intègre un facteur géographique dans la construction de l'espace de représentation pour la projection des données. En appliquant ces méthodes sur des données géolinguistiques, nous pouvons identifier et expliquer de nouvelles structures spatiales au sein d'un corpus d'aires de dispersion de traits linguistiques.

ABSTRACT. We propose a multidimensional statistical analysis procedure using projection and classification methods, in order to identify coherent clusters into a set of surface entities called dispersion areas. The methodology includes a geographical factor to build the representation space for the projection of the data. By applying this method on geolinguistic data, we are able to identify and explain new spatial patterns among a set of dispersion areas of linguistic features.

MOTS-CLÉS : géolinguistique, géomatique, classification, analyse spatiale, statistiques, humanités numériques

KEYWORDS: geolinguistics, geomatics, clustering, spatial analysis, statistics, digital humanities

1. Contexte et positionnement

La dialectologie s'intéresse à l'étude des traits linguistiques caractéristiques des langues à tradition orale comme les parlers locaux, appelés patois ou dialectes. Ces traits linguistiques peuvent être de nature très différente - phonétique, morphosyntaxique, lexicale, sémantique ou prosodique - et évoluent dans un espace géographique donné, dans le temps et au contact de la société (Chambers, Trudgill, 1998).

La géolinguistique s'intéresse à l'étude des phénomènes linguistiques à travers l'analyse de la distribution spatiale des traits linguistiques et à leur variabilité. Pour cela, les géolinguistes cherchent à identifier les frontières délimitant les aires occupées par les différents traits linguistiques. Ces frontières, appelées *isoglosses*, sont définies à partir de données recueillies lors de vastes campagnes d'enquêtes de terrain décrivant, en un lieu donné, les variations phonétiques ou lexicale d'une *notion* linguistique. Ces données brutes ont servi, notamment à élaborer les atlas linguistiques, dont le plus emblématique est l'*Atlas Linguistique de la France* (Gilliéron, Edmont, 1902-1910). Par exemple, la figure 1, illustre les différentes variations lexicales (*lemmes*) de la notion "chêne" issue de la base de données du THESOC¹.

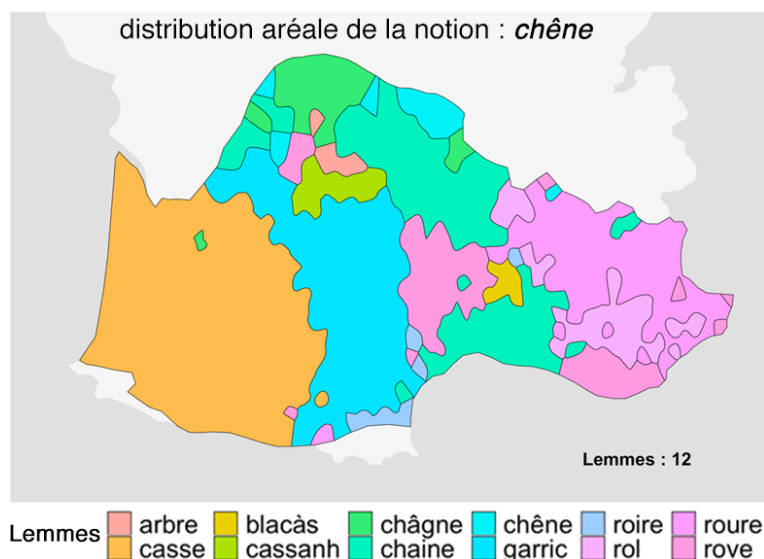


FIGURE 1. Distribution aréale des lemmes de la notion "chêne" sur le domaine occitan. Données issues du THESOC.

1. <http://thesaurus.unice.fr>

Les isoglosses délimitent des *aires de dispersion* qui matérialisent l'occupation d'un trait linguistique dans l'espace. L'étude de ces aires de dispersion s'intéresse à la dimension synchronique des phénomènes, c'est-à-dire à l'analyse de leur distribution géographique à un instant donné (Lafkioui, 2015). À travers l'analyse empirique d'un corpus de cartes, les géolinguistes peuvent identifier des motifs spatiaux récurrents (telle que l'occupation d'une zone géographique singulière) dans les aires de dispersion de certains traits linguistiques (Léonard, 2001) (Brun-Trigaud *et al.*, 2005). Ils réalisent ainsi des analyses typologiques en regroupant des aires de dispersion qui sont circonscrites à des espaces particuliers et qui sont similaires par leurs formes (Brun-Trigaud, 2012) (voir figure 2). Ces analyses permettent aussi de révéler des phénomènes jusqu'alors invisibles (Brun-Trigaud, Malfatto, 2013) qu'on ne retrouve pas dans les structures des aires linguistiques classiques. Ces études s'accompagnent généralement d'une volonté d'expliquer ces phénomènes par des facteurs socio-historiques ou géographiques (Saussure, 1971) (Dalbera, 2013).

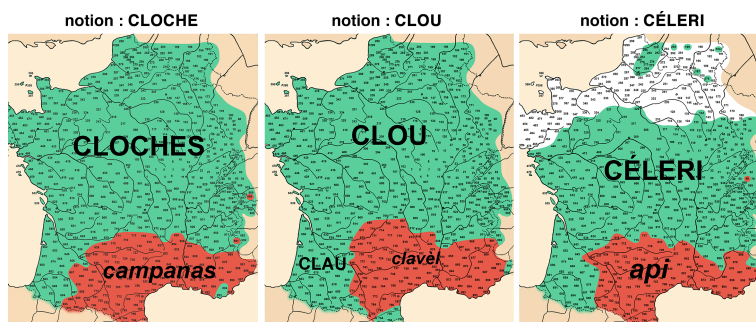


FIGURE 2. Les aires de dispersion des lemmes *campanas*, *clavèl* et *api* (en rouge) constituent un motif spatial récurrent. Cartes interprétées de l'Atlas Linguistiques de la France issues du livre de (Brun-Trigaud *et al.*, 2005)

Si ces analyses typologiques ont d'abord été réalisées de manière intuitive, basées sur des connaissances d'experts, en étudiant la distribution aréale de plusieurs cartes d'un domaine d'étude, elles ont pu être corroborées par des méthodes statistiques de classification appliquées à un plus grand nombre de cartes via une approche dialectométrique (Brun-Trigaud *et al.*, à paraître). À partir de données issues de plusieurs centaines de cartes d'atlas linguistique, les points d'enquête (lieux géographiques visités lors des enquêtes de terrain) sont classifiés en fonction de leurs similarités linguistiques. En utilisant la distance de Levenshtein (Miller *et al.*, 2009) il est possible d'établir une distance linguistique quantitative entre points d'enquête (Heeringa, 2004) ; ils sont ensuite regroupés avec des méthodes classiques de classification (souvent ascendantes hiérarchiques) (Everitt *et al.*, 2011). Il existe des outils (comme Gabmap (Leinonen *et al.*, 2016)) qui utilisent ces approches dialectométriques pour pro-

duire des cartographies représentant des régions de cohérences linguistiques autour des points d'enquête.

Les méthodes statistiques de classification permettent en général de retrouver des structures proches des typologies déterminées empiriquement. Toutefois elles ne traitent pas les aires de dispersion en tant que telles mais plutôt les points d'enquête dont elles sont issues.

Notre proposition de classification consiste à utiliser les aires de dispersion des cartes du corpus comme individus statistiques qui sont classifiées en fonction de leur similarité spatiale. De plus, ni l'analyse typologique empirique des géolinguistes ni l'approche statistique des outils dialectométriques (Nerbonne *et al.*, 2011) n'étudient les liens qui existent entre les caractéristiques linguistiques des structures identifiées et d'autres types de données de contexte (caractéristiques géographiques, géohistoriques, socio-économiques, etc.) susceptibles de constituer un facteur explicatif.

Dans ce contexte, notre objectif est de proposer une procédure d'analyse spatiale adaptée pour traiter un grand corpus d'aires de dispersion. Nous cherchons en particulier à identifier des ensembles cohérents d'aires de dispersion présentant des co-occurrences spatiales. La difficulté réside dans le fait que les aires de dispersion sont très hétérogènes en taille et en forme, autrement dit leurs limites ne sont jamais les mêmes. Pour comprendre les structures identifiées, notre approche s'appuie sur l'intégration d'un *facteur géographique* pour lequel nous pouvons mesurer l'impact sur ces structures. A ce stade, le facteur géographique dont nous parlons est une partition de l'espace en plusieurs entités géographiques. L'intérêt de la méthode est de croiser des données spatiales multi-sources et multi-formes afin d'établir des liens entre des phénomènes linguistiques et des réalités géographiques, historiques, sociologiques, etc.

Les sections suivantes exposent les étapes successives de notre méthode de classification ainsi que son implémentation. Pour finir, nous illustrons la méthode au travers du cas d'étude sur les données géolinguistiques issues du THESOC.

2. Méthode

Nous cherchons à classer un ensemble hétérogène d'aires de dispersion en groupes cohérents en intégrant comme facteur de regroupement un découpage géographique. Nous émettons l'hypothèse que des critères géographiques tels que l'organisation des bassins versants ou les limites d'anciennes provinces peuvent avoir une influence sur la diffusion des traits linguistiques dans l'espace au cours du temps. Les aires de dispersion se présentent sous la forme d'objets géographiques surfaciques représentés initialement dans l'espace à deux dimensions d'une carte.

Classer ces objets nécessite de les comparer entre eux. Pour cela nous devons être en mesure de les caractériser quantitativement selon des critères spatiaux communs. Nous transformons les aires de dispersion en points dans un espace de représentation multi-dimensionnel propre au facteur géographique choisi.

Notre corpus d'aires de dispersion devient alors un ensemble de points dans un espace de représentation qui peut maintenant faire l'objet d'une classification. Nous proposons de nous appuyer sur deux méthodes de classification supervisée : la *Classification Ascendante Hiérarchique* (CAH) et les nuées dynamiques (k-means) (Nakache, Confais, 2004).

Les étapes de notre méthode sont : la création d'un espace de représentation adapté, puis la projection des objets dans cet espace de représentation et enfin leur classification.

2.1. Espaces de représentation

Nous choisissons dans un premier temps de discrétiser l'espace étudié en n unités spatiales de forme hexagonale (maillage plus ou moins fin selon le niveau de discrétisation souhaité); cela constitue un espace de représentation initial E (de dimension n) où toute aire de dispersion est discrétisée sous la forme de n coordonnées surfaciques : la i -ème coordonnée est la part de surface occupée dans la maille numéro i . Sur l'exemple de la figure 4 (a), l'aire *chêne* occupe seulement quelques unités spatiales. La classification dans un tel espace serait peu efficace car les objets seraient trop dispersés (fléau de la dimension).

L'idée est désormais de construire un espace de représentation F de dimension p plus petite. La dimension p devra être raisonnablement faible afin d'assurer une classification efficace par la suite. Nous utilisons le facteur géographique pour construire cet espace F , en procédant comme suit :

1. calcul de la *matrice des aires* A de dimension $n \times p$ en croisant toutes les unités spatiales avec toutes les régions du facteur géographique : la cellule (i,j) de la matrice A contient l'aire de l'intersection entre la $i^{\text{ème}}$ unité spatiale et la $j^{\text{ème}}$ région du facteur géographique (figure 3 (a)). Avec cette matrice des aires A nous pouvons obtenir le profil de chaque unités spatiale en fonction du découpage géographique choisi en calculant les pourcentages lignes de A . Inversement nous pouvons exprimer les régions du découpage en fonction des unités spatiales en calculant les pourcentages colonnes de la matrice A .

2. mise en oeuvre d'une *analyse des correspondances* (CA) (Rencher, 2002) sur la matrice A . Cette méthode d'analyse statistique multidimensionnelle va efficacement représenter les unités spatiales comme une ensemble de n points dans l'espace vectoriel F (figure 3 (b)).

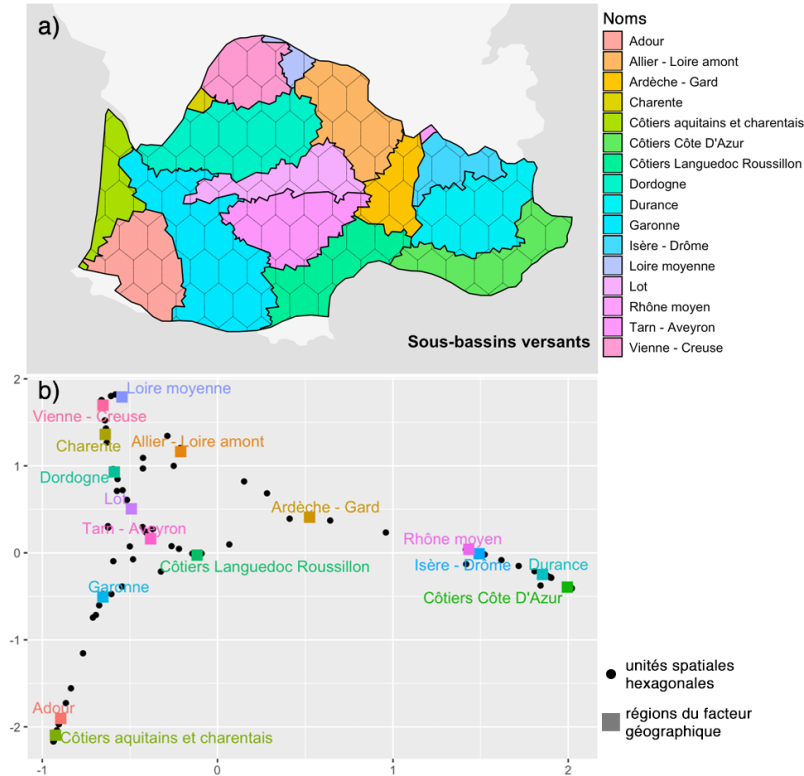


FIGURE 3. Analyse des correspondances : (a) superposition des régions du facteur géographique et du maillage hexagonal (b) représentation simultanée (semi-barycentrique) des p régions du facteur géographique et des n unités spatiales hexagonales, selon les deux premiers axes factoriels.

2.2. Projection barycentrique

Les unités spatiales sont positionnées dans l'espace de représentation F propre au facteur géographique choisi ; nous projetons les aires de dispersion dans ce même espace. De la même façon que pour la première étape décrite dans la section 2.1 nous calculons une matrice des aires A' entre les aires de dispersion et les unités spatiales. La cellule (i,j) de la matrice A' contient l'aire de l'intersection entre la $i^{\text{ème}}$ aire de dispersion et la $j^{\text{ème}}$ unité spatiale. On calcule ensuite les pourcentages lignes de cette matrice afin d'obtenir le profil de chaque aire de dispersion en fonction des unités spatiales. Nous utilisons ces profils pour projeter les aires de dispersion dans l'espace de représentation en tant que barycentres des unités spatiales qui les composent (figure 4 (b)).

Classification d'aires de dispersion

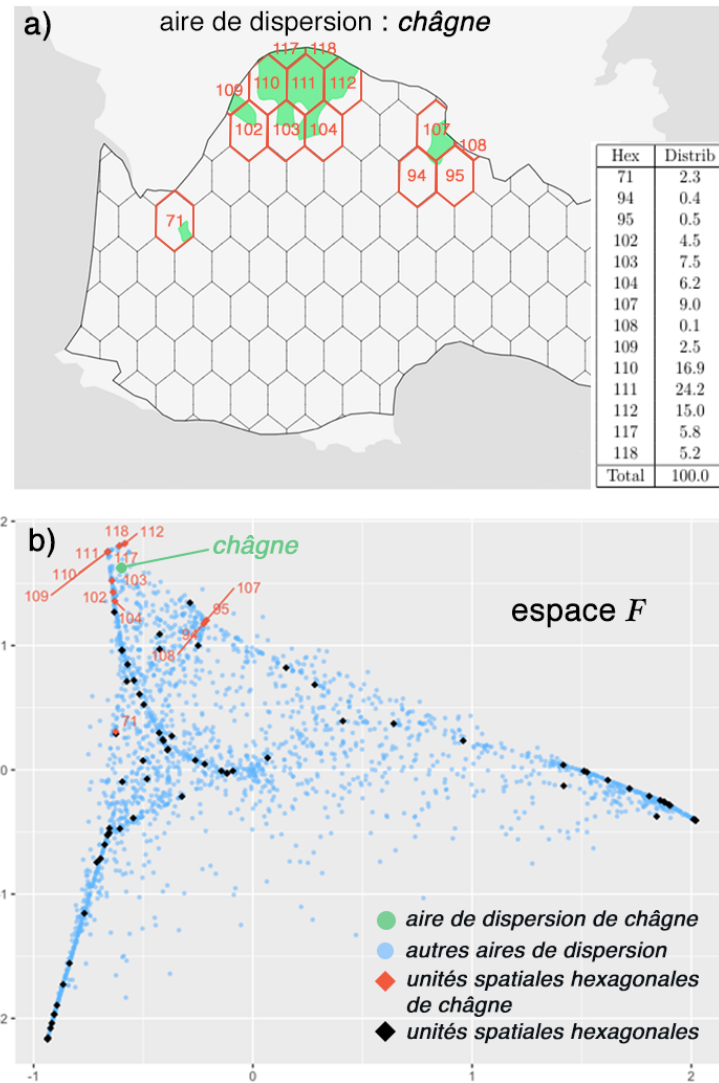


FIGURE 4. Projection barycentrique de l'aire de dispersion du lemme '*chêne*' dans l'espace de représentation F ici visualisé selon ses deux premiers axes factoriels.

2.3. Classification

La dernière étape de notre analyse consiste en une procédure de classification supervisée des points des aires de dispersion dans l'espace de représentation

F . Plusieurs possibilités existent compte tenu du grand nombre de méthodes de classification disponibles dans la littérature (Everitt *et al.*, 2011). A ce stade nous procédons de la manière suivante : tout d'abord nous utilisons un algorithme de Classification Ascendante Hiérarchique (CAH) avec différents critères d'agrégation tels que Ward, saut minimal, saut maximal et distance moyenne. Le dendrogramme produit par la CAH peut être coupé à n'importe quel niveau en fonction du nombre de classes souhaitées. Nous proposons la possibilité de consolider les classes avec un l'algorithme *k-means* dont les points de départ sont initialisés avec les centroïdes des classes déjà établies précédemment par la CAH (Nakache, Confais, 2004).

Afin de mesurer le pouvoir classificatoire du facteur géographique, nous proposons deux indicateurs (R_E^2 et R_F^2) qui représentent la part de l'inertie inter-classes sur l'inertie totale de l'ensemble des données lorsque celles-ci sont représentées respectivement dans l'espace E et dans l'espace F . Plus ces indicateurs sont proches de 1 plus les classes sont compactes ce qui signifie que la classification est efficace. Il est ainsi possible pour les praticiens de comparer différents facteurs géographiques et d'évaluer lesquels sont les plus pertinents pour leur jeu de données.

3. Implémentation de la méthode

La méthode décrite ci-dessus a été implémentée dans un environnement d'analyse exploratoire de données géographiques développé avec le langage R². Il offre à l'utilisateur la possibilité d'explorer le processus de classification en agissant de façon interactive sur les paramètres de classification et de visualiser graphiquement et cartographiquement les résultats. La visualisation et la classification étant liées dynamiquement, l'utilisateur peut explorer chaque classe pour faciliter l'interprétation des regroupements.

3.1. Exploration visuelle

La représentation cartographique a pour objectif de mettre en évidence le profil de concentration de chaque classe constituée. Pour chaque classe, nous identifions les zones recouvertes localement par un nombre n_k d'aires de dispersion de la classe. Un indice (ou score de concentration) n_k/N_{max} , est alors créé, N_{max} étant le plus grand nombre d'aires superposées identifié pour cette classe. Plus la valeur de n_k/N_{max} est proche de 1 plus la concentration en aires sur la k -ème zone est forte. Ainsi on visualise l'épicentre de la classe qui peut être interprété comme l'origine géographique de diffusion d'un phénomène (voir figure 5).

2. <https://cran.r-project.org>

Classification d'aires de dispersion

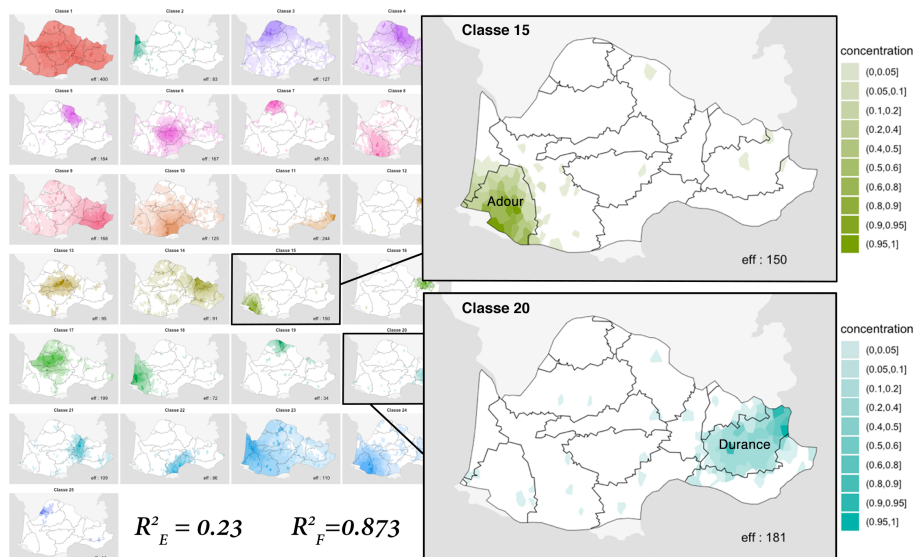


FIGURE 5. Classification en 25 classes à l'aide du facteur géographique des sous-bassins versants. Agrégation avec le critère Ward et consolidée avec *k-means*.

Nous permettons également d'explorer chaque classe, notamment à travers une analyse des caractéristiques géographiques (contour, extension, localisation) et thématiques des aires de dispersion qui la compose. L'analyse géographique vise à dégager le meilleur représentant de la classe (parangon) en relation avec le profil de concentration. Les autres individus de la classe sont triés par ordre décroissant de représentativité.

3.2. Caractérisation des classes

Afin de mieux caractériser chaque classe nous procédons à une analyse thématique à l'aide de variables illustratives : la distribution des thèmes et sous-thème de chaque classe est comparée à celle de l'ensemble du corpus. Sur le plan graphique, les diagrammes *radar* mettent en évidence les écarts de distribution comme l'illustrent les figures 6 et 7.

4. Cas d'étude

Nous avons appliqué cette méthode sur un corpus de 235 cartes géolinguistiques issues du *Thesaurus Occitan* ou THESOC (Dalbera *et al.*, 2012) qui rassemble les atlas linguistiques régionaux du domaine occitan de la France. Pour

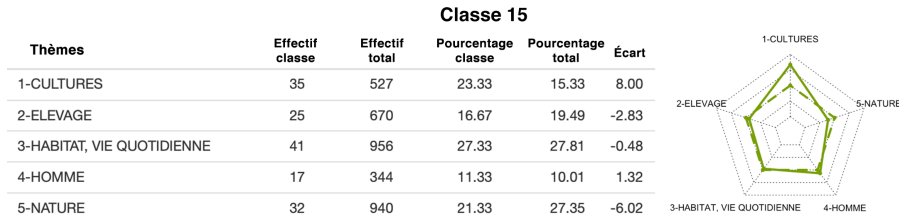


FIGURE 6. Analyse thématique de la classe 15 (Adour) en cinq thèmes : tableau des fréquences et diagramme radar. Sur le diagramme, le trait en pointillé représente le profil moyen, le trait plein représente le profil de la classe.

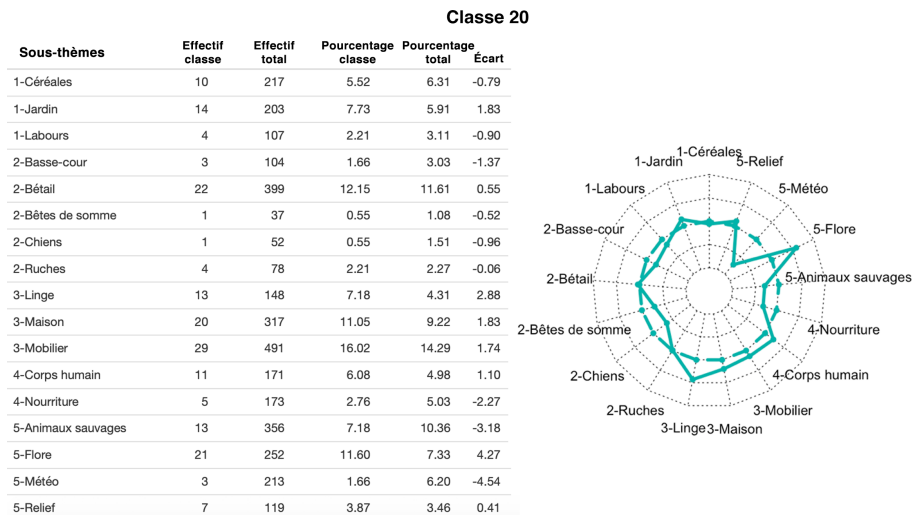


FIGURE 7. Analyse thématique de de la classe 20 (Durance) en dix-sept sous-thèmes : tableau des fréquences et diagramme radar. Sur le diagramme, le trait en pointillé représente le profil moyen, le trait plein représente le profil de la classe.

rappel, chaque carte décrit la distribution spatiale des lemmes utilisés pour désigner une notion linguistique collectées sur 645 points d'enquête répartis dans la région étudiée. Pour chacune des notions linguistiques, les aires de dispersion propres à chaque lemme ont été définies au moyen de méthodes d'interpolation appliquées à des données qualitatives (Chagnaud *et al.*, 2017). Un corpus de 3437 aires lexicales est ainsi créé, à partir duquel, des co-occurrences spatiales

et les facteurs géographiques associés doivent être identifiés. Nous proposons d'appliquer différents types de facteurs géographiques à notre étude de cas: limites environnementales (les sous-bassins versants (figure 3 a) ou les hydro-écorégions³), administratives (les départements) ou historiques (les Généralités en 1789 ou les provinces gauloises en 450)).

En utilisant le facteur géographique des sous-bassins versants, l'analyse montre qu'il est préférable de travailler avec au moins 25 classes pour voir se dégager des regroupements intéressants. On voit immédiatement que certains sous-bassins versants regroupent un grand nombre de lemmes qui sont bien confinés à l'intérieur de leurs limites, c'est le cas par exemple du bassin *Côtiers Côte d'Azur* avec 244 lemmes, de celui de la *Durance* avec 181 lemmes ou encore de l'*Adour* avec 150 lemmes (voir figure 5). Si les cas d'une cohérence parfaite entre une région du facteur géographique et une aire lexicale sont assez rares, on peut néanmoins trouver une bonne adéquation comme par exemple dans le cas de l'aire de *dragon* (lemme de la notion "*faux à blé*") avec le bassin de l'Adour ou celle de l'aire du lemme *estraçaire* (de la notion "*chiffonnier*") sur la Côte d'Azur. Ces phénomènes de répartition trouvent peut-être leurs origines dans les mouvements migratoires de ces métiers saisonniers. Ainsi dans le cas de la classe décrivant le bassin de l'Adour, le diagramme radar de la figure 6 nous montre une prédominance d'aires appartenant au vocabulaire de la culture (auquel appartient le terme *dragon*). Quant au diagramme radar de la figure 7 qui rend compte des lemmes regroupés dans le bassin de la Durance, il indique par la présence des sous-thèmes que les notions rattachées à la flore sont très présentes dans ce groupe. Ces analyses permettent de trouver s'il existe, par exemple, une interprétation par les thèmes qui pourrait unir les aires lexicales regroupées par la classification.

En changeant de facteur géographique de nouveaux regroupements apparaissent. Chaque critère géographique a sa pertinence et génère des regroupements différents, montrant bien que les parlers ne sont pas seulement influencés par le fait d'appartenir à une aire linguistique, mais que d'autres facteurs entrent en jeu et donnent leur propre cohérence. Ils conduisent néanmoins à des classifications plus ou moins efficaces (voir table 1). Cette table nous montre que le découpage administratif des provinces de Gaule en 450 semble être le facteur géographique qui donne les classes les plus compactes alors que le découpage des départements donne une classification moins efficace.

5. Discussion et conclusion

En proposant une méthode de classification d'objets géographiques surfaciques représentant des aires de dispersion de phénomènes géolinguistiques, il devient alors possible d'identifier des co-occurrences spatiales et de créer des

3. <https://www.irstea.fr/fr/les-hydroecoregions-de-france-metropolitaine>

Facteur Géographique	R_E^2	R_F^2
Aucun	NA	0,374
Généralités en 1789	0,233	0,900
Provinces Gaule en 450	0,191	0,950
Sous-bassins versants	0,230	0,873
Départements	0,308	0,722
Hydro-écorégions	0,222	0,897

TABLE 1. Comparaison du pouvoir classificatoire des différents facteurs géographiques à l'aide des indicateurs R_E^2 et R_F^2 . Méthode d'agrégation : Ward. Nombre de classes : 25. Kmeans : oui.

groupes cohérents. Cette méthode implémentée selon une approche d'analyse exploratoire de données géographiques offre l'opportunité aux linguistes d'identifier plus facilement les liens existants entre des phénomènes linguistiques et des facteurs géographiques (ceux-ci étant jusqu'à maintenant essentiellement basés sur leur connaissance du territoire). La force de ces liens est quantifiée grâce aux indicateurs de pouvoir classificatoire (R_E^2 et R_F^2). Cependant de futurs travaux devront ajuster ces indicateurs pour tenir compte du nombre de régions des facteurs géographiques car ils influencent les valeurs des indicateurs. L'environnement d'analyse interactif permet d'étudier facilement plusieurs classifications en faisant varier les critères d'agrégation, il offre également la possibilité d'analyser les classes ainsi établies en proposant aux linguistes des clés d'explication. Il faut garder à l'esprit que notre méthode, au travers d'un facteur géographique, engendre des classes spatialement compactes, mais que ces classes ne sont pas toutes pertinentes d'un point de vue linguistique. Déterminer la validité des classes reste du ressort de l'expert.

La validation complète de la méthode nécessite de l'appliquer à une diversité de jeux données géolinguistiques mais aussi dans le cadre d'autres thématiques mobilisant des données similaires.

Remerciements

Ces travaux ont été réalisés dans le cadre du projet ANR ECLATS soutenu par l'Agence Nationale de Recherche ANR-15-CE-380002.

Bibliographie

Brun-Trigaud G. (2012). Essai de typologie des aires lexicales dans l'Atlas Linguistique du Centre. *Annales de Normandie*, vol. 62, n° 2, p. 77-93.

Classification d'aires de dispersion

- Brun-Trigaud G., Le Dù J., Le Berre Y. (2005). *Lectures de l'Atlas Linguistique de la France de J. Gilléron et E. Edmont : du temps dans l'espace*. CTHS.
- Brun-Trigaud G., Malfatto A. (2013). Limites dialectales vs limites lexicales dans le domaine occitan : un impossible accord? In *Current Approaches to Limits and Areas in Dialectology*, p. 293-310. Cambridge Scholars Publishing.
- Brun-Trigaud G., Malfatto A., Sauzet M. (à paraître). Essai de typologie des aires lexicales occitanes : regards dialectométriques. In *Fidélités et dissidences. 12e Congrès de l'Association Internationale d'Etudes Occitanes*.
- Chagnaud C., Garat P., Davoine P.-A., Carpitelli E., Vincent A. (2017). Shinydialect: A cartographic tool for spatial interpolation of geolinguistic data. In *Proceedings of the 1st ACM SIGSPATIAL workshop on Geospatial Humanities*, p. 23-30. ACM.
- Chambers J. K., Trudgill P. (1998). *Dialectology* (2^e éd.). Cambridge Univ. Press.
- Dalbera J.-P. (2013). La trajectoire de la dialectologie au sein des sciences du langage. De la reconstruction des systèmes dialectaux à la sémantique lexicale et à l'étymologie. *Corpus*, vol. Dialectologie : corpus, atlas, analyses, n° 12, p. 173-200.
- Dalbera J.-P., Ranucci J.-C., Olivieri M., Brun-Trigaud G. (2012). La base de données linguistique occitane Thesoc. Trésor patrimonial et instrument de recherche scientifique. *Estudis Romànics* 34, p. 367-387.
- Everitt B. S., Landau S., Leese M., Stahl D. (2011). *Cluster Analysis, 5th ed.* United Kingdom, John Wiley.
- Gillieron J., Edmont E. (1902-1910). *Atlas Linguistique de la France*. Champion.
- Heeringa W. (2004). *Measuring dialect pronunciation differences using levenshtein distance*. Thèse de doctorat non publiée, University of Groningen.
- Lafkioui M. B. (2015). Méthodologie de recherche en géolinguistique. *Corpus*.
- Leinonen T., Çoltekin c., Nerbonne J. (2016). Using Gabmap. *Lingua*.
- Léonard J.-L. (2001). Aréologie dialectale et modularité des réseaux dialectaux : étalement spatial et structural des processus (morpho)phonologiques dans le réseau dialectal basque. In *Actes du XV^e Congrès international de l'Académie basque, 17-19 septembre 2001*, p. 141-168.
- Miller F. P., Vandome A. F., McBrewster J. (2009). *Levenshtein distance: Information theory, computer science, string (computer science), string metric, damerau-levenshtein distance, spell checker, hamming distance*. Alpha Press.
- Nakache J., Confais J. (2004). *Approche pragmatique de la classification: arbres hiérarchiques, partitionnements*. Technip.
- Nerbonne J., Colen R., Gooskens C., Kleiweg P., Leinonen T. (2011, 01). Gabmap - a web application for dialectology. *Dialectologia*.
- Rencher A. C. (2002). *Methods of multivariate analysis* (second edition éd.). Hoboken, NJ, USA, Wiley-Interscience.
- Saussure F. de. (1971). *Cours de linguistique générale*. Paris, Payot.

Mesurer les “vides alimentaires” dans un contexte urbain hétérogène, réflexion méthodologique et application à Lyon-Saint-Etienne

Luc Merchez¹, Hélène Mathian¹, Julie Le Gall¹

1. UMR 5600 Environnement Ville Société
Ecole Normale Supérieure de Lyon, 15 parvis René Descartes - BP 7000 69342
Lyon Cedex 07 - FRANCE

luc.merchez@ens-lyon.fr, helene.mathian@ens-lyon.fr, julie.legall@ens-lyon.fr

RESUME. La question de l'alimentation et de la caractérisation des environnements alimentaires a déjà fait l'objet de nombreuses études et développements méthodologiques pour rendre compte des différentiels d'accessibilité. Aux Etats-Unis, essentiellement à l'aune de questions sur la santé, ces études ont conduit à identifier des “désert alimentaires”. Cette question éminemment spatiale, qui repose sur des notions d'accessibilités, est souvent abordée par des enquêtes et entretiens, ou des approches quantitatives basées sur des calculs d'accessibilités géographiques. Nous nous situons dans ces dernières et proposons d'explorer la transférabilité de cette notion de “désert” à un espace métropolitain français. La démarche nécessite de revenir sur les données disponibles, et de justifier l'enchaînement méthodologique mis en oeuvre au regard des spécificités du terrain métropolitain étudié. Cette première étape de modélisation constitue les fondements d'un outil d'exploration de différentes dimensions de la notion de “désert alimentaire”.

ABSTRACT. The issue of food and characterization of food environments has been the subject of many studies and methodological developments to report and account for accessibility differentials. In the United States most studies are interested in this question with regard to health issues and have led to identify "food deserts". As part of a project for which the main goal is to explore the transferability of this notion to other spaces, with a broader perspective on spatial justice, we propose a generic framework to shed light on this highly spatial question, generally addressed by qualitative (surveys, interviews) or quantitative methods (GIS-based models). This study requires a review of the available data and justification of the methodological sequence according to the characteristics of a large metropolitan area. This first modelling step provides the basis for a tool for exploring different dimensions of the concept of "food desert".

MOTS-CLES : ENVIRONNEMENT ALIMENTAIRE, COMMERCES, MODELES D'ACCESSIBILITE, ANALYSE SPATIALE, OUTIL D'EVALUATION

SAGEO'2019 – Clermont-Ferrand, 13-15 novembre 2019

Introduction

La question de la mesure de l'environnement alimentaire n'est pas nouvelle, comme le prouve le nombre de revues méthodologiques que l'on peut lire depuis 2010 (Chaire & al, 2010, Kelly & al, 2011, Caspi & al, 2012, Black & al, 2014, McKinnon & al. 2009). La notion commune qui sous-tend la caractérisation de ces espaces dans la majorité des articles, est la notion d'« *accessibility to healthy and affordable food* » (Beaulac & al, 2009) qui a d'une part donné lieu à différentes définitions et interprétations de la notion d'environnement alimentaire (*food environment*) et d'autre part de la notion d'accessibilité. Après être revenus sur ces notions (§1), nous explorons dans cet article les données disponibles (§2) et les besoins méthodologiques associés (§3) pour construire un outil d'évaluation paramétrable des différenciations d'environnements alimentaires, et explorer la transférabilité de la notion de "désert alimentaire" à une aire métropolitaine française, ici celle de Lyon/Saint-Etienne.

1- La question de l'évaluation des environnements alimentaires

1.1 Des environnements alimentaires à leurs mesures

Au cœur de la réflexion sur les environnements alimentaires et les injustices alimentaires, la superposition de cartes présentant à l'échelle américaine les déserts alimentaires (*food deserts*) parallèlement au taux d'obésité, au taux de prévalence des maladies cardio-vasculaires, au taux de pauvreté ou à la composition raciale de la population (Walker et al, 2010) semble offrir en quelques regards une démonstration sans faille de la relation entre composition de l'offre alimentaire et insécurité alimentaire. Plus globalement, les approches établissant des liens entre précarité, alimentation et santé se sont largement diffusées, y compris au sein des communautés scientifiques et acteurs publics, notamment ceux de l'aire métropolitaine de Lyon/Saint-Etienne à laquelle nous nous intéresserons ici (Agences d'urbanisme de Lyon et de Saint-Etienne, 2016).

L'ensemble des études partent du postulat que de multiples facteurs influent sur le comportement alimentaire, qu'il s'agisse de caractéristiques individuelles ou des caractéristiques de l'environnement (Glanz et al, 2005). Plusieurs approches cohabitent sur ces questions : les approches plus qualitatives, qui partent de l'individu, de ses caractéristiques, de ses pratiques, de ses perceptions, et celles, plus statistiques, qui vont se situer au niveau de l'espace résidentiel et de "l'environnement alimentaire", et construire des mesures décrivant les contraintes et potentialités de cet environnement. Dans cette deuxième perspective, les données sur le fonctionnement -pratiques d'achat, pratiques de consommation, ...- sont difficiles à acquérir (enquêtes essentiellement) et le plus souvent insuffisamment spatialisées à niveau fin.

Cet article ouvre la boîte noire des questions qui se posent quand on cherche à appréhender depuis un contexte urbain hétérogène français (européen) les logiques d’organisation qui conduisent à des différences d’environnements alimentaires (ici réduits aux commerces alimentaires).

1.2 Déserts ou vides alimentaires ?

Les multiples insatisfactions autour de la transposition de la notion de « désert alimentaire » nous ont conduit à poser la question : son usage est-il possible et même souhaitable en contexte français ?

Les études anglo-saxonnes ciblent d’emblée les enjeux de santé, et s’attachent de ce fait à isoler des commerces proposant une offre alimentaire diversifiée et potentiellement saine (*healthy food stores*) ou alors clairement identifiée comme malsaine (*non-healthy food stores*). Pour diverses raisons, cette simplification binaire ne nous apparaît pas transposable au cas français. En premier lieu, le paysage commercial se révèle selon nous plus diversifié qu’en Amérique du Nord, témoignant de modes de consommation (et pratiques d’achats associés) spécifiques à la société française, et plus largement à la sphère euro-méditerranéenne. Ce refus de jugement de valeur quant à la qualité (saine/malsaine) des commerces nous a poussés à considérer autant que possible l’ensemble des commerces de détail en alimentaire, et aussi à nous écarter de la notion de « désert alimentaire » pour proposer et mobiliser celle de « vide alimentaire ». Cette notion se veut ainsi plus neutre, sans discriminer a priori certains segments de populations ou de commerces.

Nous proposons d’interroger la transférabilité des concepts et méthodes dans d’autres zones géographiques et culturelles, en nous inspirant de travaux déjà effectués dans le courant des études basées sur l’accessibilité géographique dites “*GIS-based*” (cf Chaire & al, 2010) qui décrivent les lieux de résidence en fonction de l’accès aux commerces environnants. La variété des méthodes documentées sont le plus souvent limitées - (1) à la prise en compte d’une seule voire deux dimensions d’accessibilité (densité, proximité, disponibilité...) - (2) à des types spécifiques de commerces simplifiant ainsi la question de comportements associés différenciés. Cette variété ne facilite pas la comparaison entre études et leur transfert à des espaces différents. Nous explorons ici les potentialités d’une méthode:

- multi-factorielle: prenant en compte simultanément de plusieurs dimensions de l’accessibilité géographique
- multi-échelle: prenant en compte les différentiels de densités et donc de pratiques spatiales.

2- L’environnement alimentaire au prisme de l’offre commerciale

2.1 Approche empirique de l’offre alimentaire commerciale

La mesure de l’environnement alimentaire au prisme des commerces présents sur un territoire a longtemps buté sur l’accès aux données de commerces avec une

géolocalisation fine. Si l'ouverture et la mise à disposition des données publiques lève des verrous successifs dans l'accès aux données, il n'en reste pas moins que leurs qualités sont mises à l'épreuve, tant du point de vue de la géolocalisation que de la catégorisation et la description des commerces. Kelly & al. (2011) soulignaient déjà qu'en raison du renouvellement rapide des commerces (taux de rotation), les répertoires officiels d'entreprises s'avèrent limités en termes de complétude, tributaires de la survivance de données obsolètes (faux actifs) et porteurs de possibles erreurs dans la géolocalisation et la catégorisation des commerces.

Dans un premier temps, nous ciblons uniquement les commerces sédentaires (commerces de détail), homogènes dans leur temps d'ouverture¹.

On distingue deux familles de données, celle relevant d'inventaire(s) officiel(s) et reposant généralement sur une nomenclature précise, et celles pouvant relever d'une compilation de données collectées via des WebServices de cartographie (Google Maps, OpenStreetMap, Pages jaunes, ...). Le choix d'une base en particulier relève des besoins identifiés en termes d'exhaustivité, de reproductibilité, de suivi temporel, ou d'enrichissement (croisement avec d'autres variables), ce qui est l'enjeu de la méthodologie que nous développons ici.

La Table 1 présentée ci-après fournit une rapide synthèse comparative, en ciblant essentiellement les bases de la première famille (inventaires officiels), ainsi qu'OpenStreetMap, de plus en plus souvent mobilisé dans ce genre d'études.

TABLE 1: Comparaison des bases de données de commerces alimentaires (hors restauration et débits de boissons)

	Base SIRENE https://www.data.gouv.fr	Base Perm. des Equip. (BPE xy) https://www.insee.fr	Base CCI https://www.lyon-metropole.cci.fr	Open StreetMap https://www.openstreetmap.org/
Couverture géo. potentielle	France entière	France entière	Lyon /St-Etienne/ Roanne	Monde entier
Fréq. de mise à jour	en continu	annuelle	annuelle	en continu
Granularité thématique	15 catég. de commerces	8 catég. de commerces	15 catég. de commerces	#17 catég. de commerces
Potentialité de la source: variables supplémentaires	83 dont 23 relatives à l'étab.	Aucune	3 (nom, année création, tranche effectifs)	2 (nom, horaires d'ouverture)
Nb d'étab. recensés: Métrop Lyon/ Lyon	3189/ 1543	2234/1018	2332/1179	1260/ 680

¹ Les commerces non sédentaires (dont les marchés) ne représentant par ailleurs qu'une faible part de la dépense alimentaire des ménages 7,3% pour la région lyonnaise en 2017 par exemple (CCI Lyon métropole, 2017)

Le travail comparatif réalisé sur ces différentes bases nous a amenés à choisir la base SIRENE. Nous l’avons privilégié pour sa mise à jour en continu, sa nomenclature homogène, son exhaustivité (même si SIRENE surestime le nombre de commerces, cet excédent étant variable selon les types de commerces), et la description des commerces par des variables complémentaires.

2.2 L’offre alimentaire dans la zone métropolitaine Lyon/Saint-Etienne

Un des enjeux de notre exploration est de tester la transférabilité de mesures des environnements alimentaires dans un espace hétérogène. En effet les études se limitent généralement aux zones les plus denses des espaces urbains/suburbains (Chaire et al, 2010), alors qu’un espace hétérogène engendre des pratiques d’achats différentes. Le territoire d’étude (Figure 1) correspond à la réunion des deux unités urbaines de Lyon et Saint-Etienne (190 communes, 2 millions d’habitants pour 1800 km²). Ici les densités communales varient de moins de 100 hab/km² à plus de 10 000 hab/km², allant de la ville à la campagne, et ce territoire bipolaire présente également l’intérêt de proposer des contextes socio-économiques très différents entre pôles (le revenu fiscal médian s’élevant en 2010 à 17000 euros pour Saint-Etienne métropole, 20000 pour la métropole de Lyon), mais aussi au sein de chaque unité urbaine et à des niveaux très fins.

Sur les 15 catégories de commerces alimentaires sélectionnées à partir de la nomenclature SIRENE, nous avons opéré certains regroupements sur la base des types de produits commercialisés et de leur plus ou moins grande spécialisation. Au final neuf types ont été retenus, pour 4643 commerces (Figure 1).

Pour approcher les pratiques de la population, la zone d’étude a été carroyée selon une maille de 200m sur 200m qui permet de différencier l’espace à toutes les échelles et de caractériser les environnements alimentaires au niveau d’organisation de l’hétérogénéité sociale (Figure 2).

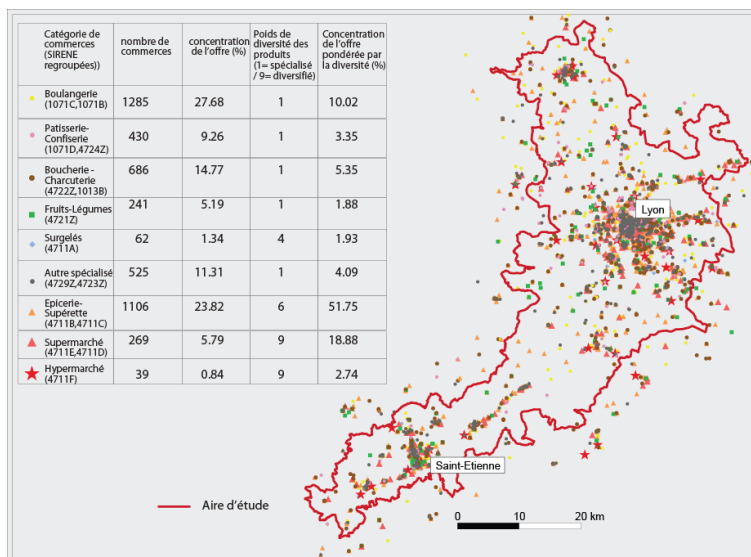


FIGURE 1. Distribution des commerces sur la zone d'étude

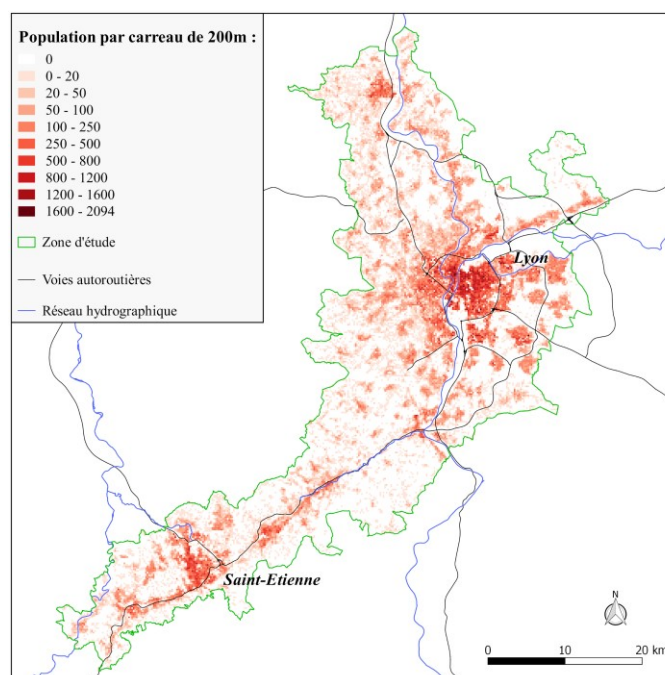


FIGURE 2. La zone métropolitaine Lyon-Saint-Etienne : des densités de population contrastées

3. Evaluer l’environnement alimentaire

L’objectif est de construire une méthodologie pour un outil d’évaluation ouvert qui permette l’identification de zones les plus défavorisées et de discuter l’existence de “vides alimentaires”. La méthode que nous présentons ici est une première étape devant permettre d’évaluer la manière dont les commerces encadrent le territoire compte tenu des pratiques des populations, et les hétérogénéités spatiales des environnements alimentaires des populations. Nous présentons l’ensemble des paramètres du modèle afin de pouvoir les discuter, les faire évoluer ou varier pour être utilisés dans différents scénarios. Dans cette première étape un certain nombre d’hypothèses simplificatrices ont été faites, les deux majeures étant que l’on considère que les achats sont faits à partir des lieux de résidences, et que les distances euclidiennes approchent les distances pratiquées. Ceci pour nous focaliser sur un objectif, à savoir intégrer parmi les variables la taille des environnements alimentaires, de manière à rendre compte des pratiques différenciées.

3.1 Description multicritère des environnements alimentaires commerciaux

L’espace est appréhendé à partir de la grille présentée ci-dessus, décrite par la population résidente. L’espace commercial considéré est constitué de 4643 commerces décrits par l’une des 9 catégories d’appartenance retenues (Figure 1). L’originalité de la méthode est de considérer simultanément les trois mesures d’accessibilité les plus fréquemment utilisées (Chaire & al., 2010), à savoir la *densité*, la *proximité* et la *diversité* commerciale dans ces environnements. Nous proposons ainsi de construire pour chaque maille une description multi-critère des environnements, sans occulter la relativité de la notion d’accessibilité, dépendante des capacités d’accès au sens large (en fonction du moyen de transport utilisé, de la congestion, ...).

La table 2 présente les différentes formes des indicateurs retenus. Les 2 premiers indicateurs, *densité* et *diversité*, sont calculés dans un environnement délimité² : un cercle dont le rayon correspond à une portée moyenne de déplacement pour effectuer ses achats alimentaires. Cette portée a été conçue pour être évolutive dans l’espace: la présentation de la fonction d’évolution fait l’objet du développement du paragraphe suivant (§3.2). L’indicateur de *densité* est le simple calcul du nombre de commerces accessibles dans cet environnement, rapporté à sa surface (Dens), et permet d’appréhender globalement l’offre quelle que soit la catégorie. Elle est complétée par la mesure de la *diversité* de cette offre. Classiquement la *diversité* peut être mesurée à plusieurs niveaux, au niveau de l’établissement, selon le nombre de catégories de commerces présentes (Div1), ou au niveau de l’offre de produits, au sens de la pondération présentée dans la figure 1 (Div2). A diversité de catégories égale (Div1), une zone sera privilégiée en terme de diversité de l’offre (Div2) si les catégories représentées ont des poids supérieurs à la première. Enfin le dernier indicateur s’abstrait de toute délimitation d’environnement, et approche la *proximité*

² Une zone tampon de 5 kilomètres autour de la zone d’étude permet de prendre en compte les commerces extérieurs à la zone, et ainsi éviter les effets de bords.

sur la base de la distance à trois catégories de commerce (Dist). Pour être comparables la mesure est rapportée à la taille de l'environnement (Distp).

TABLE 2: Description multivariée des environnements

Notion	mesure brute	mesure relative	illustration
Densité	nombre total de commerces	nombre total de commerces rapporté à la superficie du voisinage ((Dens)	
Diversité	Nombre de catégories différentes de commerces alimentaires dans un voisinage donné (Div1)	Somme des catégories représentées pondérées par la note de diversité de leur offre (Div2)	
Proximité	distance cumulée à 3 types de commerces «essentiels» (grande surface, épicerie-supérette, boulangerie) (Dist)	Même mesure rapportée au rayon du voisinage de la maille (Distp)	

3.2 – Différencier l'espace : définir des tailles d'environnements reflétant les pratiques sur un territoire aux densités diversifiées

La deuxième originalité de notre travail est de rendre variable la taille des environnements. L'environnement servant de base aux différents calculs précédents doit refléter l'espace des pratiques d'achats alimentaires. On définit cet environnement comme un champ possible d'achats à partir de chacune des mailles. Pour l'instant, l'inscription spatiale de ce champ est circulaire : l'espace est considéré anisotrope et euclidien. Tout l'enjeu de la question est la définition des portées (rayon) de ces environnements que l'on ne peut considérer comme fixe sur l'ensemble de l'espace. En effet, cet espace très hétérogène recèle des comportements variés qu'il s'agisse du mode de transport utilisé pour les achats alimentaires ou qu'il s'agisse de la localisation/polarisation de ces achats. A l'évidence la densité de commerces varie avec la densité de population, comme l'accessibilité à ces commerces varie aussi avec le type de transport utilisé. Un bon proxy de la variabilité des distances parcourues pour les achats alimentaires serait donc celui associé à la densité de population. Mais actuellement rares sont les données permettant d'établir la forme du lien entre le niveau de densité et le comportement en termes de déplacement pour les achats alimentaires à un niveau géographique très fin.

La question est donc de définir un modèle théorique de variabilité de la portée de ces déplacements sur l'espace considéré en fonction de la densité de population. Ce modèle permet de définir la taille de l'environnement alimentaire associé à chaque maille selon sa densité (Figure 2). Nous décrivons dans ses grands traits la méthodologie.

Mesurer les “vides alimentaires” dans un contexte urbain hétérogène

Pour modéliser des comportements moyens relatifs au déplacement, nous nous sommes appuyés sur les données de cadrage fournies par les résultats des enquêtes “consommation des ménages” réalisées tous les 5 ans pour la métropole lyonnaise (CCI, 2009, CCI, 2017) diffusées à un niveau agrégé (44 secteurs³), ainsi que sur des paramètres globaux issus d’enquêtes de déplacements et de leurs motifs (INSEE, 2015; SYTRAL, Agence d’Urbanisme aire métropolitaine lyonnaise 2016). Les paramètres et données mobilisées sont :

- (1) A un niveau macro (national), le temps moyen de déplacement pour les achats. Le choix a été fait de considérer un temps moyen de $T=10mn$ pour faire ses achats alimentaires (INSEE, 2015; SYTRAL, Agence d’Urbanisme aire métropolitaine lyonnaise, 2016).
- (2) A un niveau méso, celui de l’échelon des secteurs de l’enquête CCI (2009), la répartition par types d’achats et par mode de déplacement. Sur la métropole de Lyon, les dépenses liées aux achats alimentaires se font par exemple à 55% en voiture contre 20% à pied.

Les premières données (1) concernent les vitesses $v(k)$ par mode k et permettent de calculer la distance parcourue associée à un mode de déplacement pour un temps de trajet $T=10mn$. Pour l’instant nous ne considérons que les modes de transport “voiture” et “à pied”, ceux-ci constituant la majeure partie des modes déclarés (environ 85%); On peut alors calculer la distance de déplacement $d(k) = v(k)*T$. On considère une vitesse de 5km/h pour les déplacements à pied et de 15km/h en voiture (vitesse en milieu urbain dense). On obtient ainsi respectivement des portées de 800m et 2.5km. Les secondes données (2) donnent pour chaque secteur (s) la part moyenne des dépenses réalisée avec le mode de transport k . Soit $p(s,k)$ cette part.

A partir de (1) et (2) on calcule pour chaque secteur la distance moyenne parcourue pour un achat soit $d(s) = \sum_k p(s,k)*d(k)$

A l’échelon des secteurs de l’enquête CCI (2009) on calcule la densité bâtie⁴, ce qui permet de prendre en compte la composante résidentielle et l’activité humaine. Si l’on croise cette densité bâtie avec la distance moyenne parcourue pour les achats effectués en voiture, on observe une courbe qui peut être modélisée, ici par une fonction logarithmique $f : d(s)=f(dens_batie(s))$ (Figure 3a)

C’est donc sur la base de ce modèle empirique qu’ont été calculées les portées des voisinages de calcul de chaque maille sur l’ensemble de l’espace à partir de leur densité bâtie⁵ (Figure 3b)

³ Les secteurs font en moyenne 11,5 km² en moyenne (médiane : 3, min : 0.4, max : 57.7)

⁴ La densité bâtie d’une maille est le rapport entre le coefficient d’emprise au sol des bâtiments multiplié par le nombre de niveaux, rapportée à la surface de la maille (Source: BD TOPO-IGN).

⁵ Pour éliminer des variabilités trop fortes de portée de proche en proche, les densités bâties ont été lissées avec un rayon de 300m

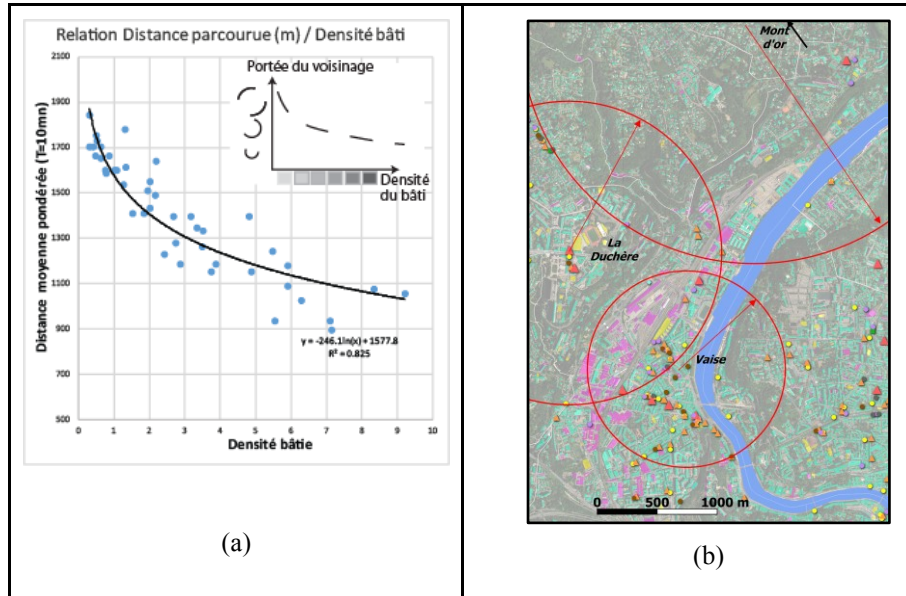


FIGURE 3. (a) Calibrage du modèle de variabilité de la portée des déplacements
(b) Illustration de délimitation variable des environnements alimentaires (9^{ème} arr. de Lyon)

4 Résultats et perspectives

4.1 Premiers résultats

Les premiers traitements montrent l'intérêt de cette approche multifactorielle et multiscale. Tout d'abord l'analyse conjointe des différents indicateurs montre que l'on approche des dimensions dé-corrélées, dont le croisement permet bien de différencier l'espace selon différentes configurations thématiques et spatiales. La figure 4 illustre ce résultat (par ailleurs mesuré systématiquement par une ACP). Elle présente les valeurs des trois indicateurs choisis pour une portion très dense de notre espace d'étude, à savoir le centre-est de l'agglomération de Lyon.

Chaque lieu est donc caractérisé par différentes formes d'inégalités d'accès, ces inégalités pouvant se compenser ou se cumuler selon les cas. A titre d'exemple, un même lieu peut se révéler assez peu favorisé en terme de distance aux commerces (point A, figure 4), mais en revanche présenter des valeurs élevées pour l'indicateur de diversité pondérée (point B, figure 4), ici en raison de la relative proximité d'hypermarchés et autres commerces "à portée de voiture". La combinaison de ces 3 indicateurs peut se révéler utile si l'on souhaite disposer d'un indice composite unique d'accessibilité alimentaire (définie ici comme l'accessibilité de l'environnement alimentaire commercial).

Mesurer les “vides alimentaires” dans un contexte urbain hétérogène

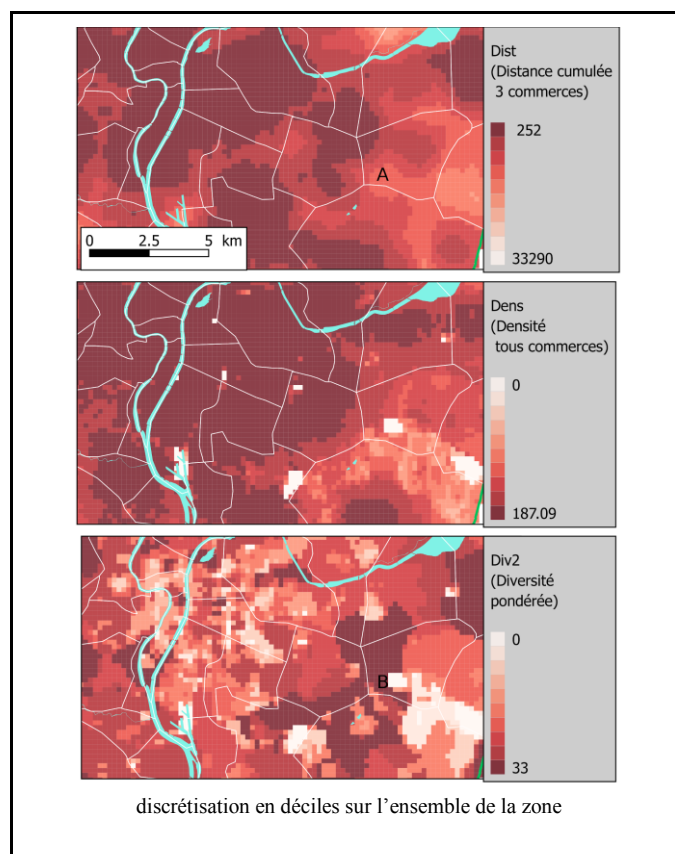


FIGURE 4. Cartographie des indicateurs - Zoom sur l'agglomération de Lyon

L'étape suivante pour finaliser la méthodologie est de tester la capacité de la combinaison des trois indicateurs à identifier les vides alimentaires, puis de distinguer pour ces vides les profils des populations résidentes sur certaines caractéristiques qui peuvent être aggravantes ou au contraire, venir contrebalancer la faible accessibilité alimentaire. L'identification d'un vide alimentaire (via les indicateurs précédents) n'aura en effet pas la même signification s'il concerne une zone de forte précarité sociale ou s'il concerne une zone globalement favorisée, au sein de laquelle les habitants sont actifs, mobiles, et aptes à accéder à une offre alimentaire plus éloignée.

4.2 Ajustements possibles et perspectives

L'intérêt de la chaîne méthodologique qui a été mise en place est qu'elle est reproductible. La Table 3 résume les différentes entrées et paramètres du modèle, ce

qui permet dans un premier temps d'en faire la critique et de les améliorer : par exemple les données servant à l'ajustement du modèle de variabilité des portées, ou la forme du modèle lui-même qui doit pouvoir être représentatif de l'ensemble de l'espace considéré⁶.

TABLE 3 : Entrées et paramètres de la chaîne méthodologique

	Entrées et Paramètres du modèle	Valeur
Données	Taille mailles d'observation	200m
	Densité et rayon de lissage	densité bâtie et 300m
	Commerces et catégories agrégées	table SIRENE + corresp. 9 categ.
Paramètres	Temps moyen de déplacement pour les achats	10mn
	Mode de transport et vitesses moyennes	à pied (5km/h) / voiture -TC (15km/h)
	Distribution de la part des achats par mode de transport et pour différents lieux	table enquête consommations des ménages (2009)
	Fonction de variabilité de la portée	fonction logarithmique

On a ici une base à améliorer. On peut tester l'amélioration des résultats à partir de certains choix opératoires, comme l'utilisation de distances réseaux, ou dans le calcul des indicateurs la prise en compte des configurations spatiales des commerces pour le calcul de la proximité (commerces dispersés ou concentrés). Par ailleurs pour l'instant, les indicateurs d'accessibilité n'intègrent pas la dimension temporelle, qui peut aussi être un facteur de variabilité, notamment si l'on souhaite introduire les modes alternatifs de distribution comme les marchés ou les AMAP.

L'originalité de ce travail repose en particulier sur la prise en compte des déplacements autrement qu'à pied, alors que la plupart des études sont fondées sur la marchabilité. Sur cette question la voiture est à la fois un levier sur la question de l'émission des gaz à effet de serre, mais aussi un critère de discrimination quant à l'équité alimentaire.

L'enjeu est ensuite de construire autour de cette méthodologie un outil ouvert (transférable à d'autres territoires) et interactif, du choix des paramètres à la

⁶ En particulier nous attendons les récents résultats des enquêtes consommation des ménages (10ème enquête, 2016) sur Lyon-Saint-Etienne (désormais regroupées), qui renseignent également sur le lieu de travail des individus sondés, ce qui peut permettre de dépasser la seule association lieu de résidence / lieux d'achats, forcément réductrice.

visualisation des résultats, ainsi qu’un protocole d’évaluation de ces résultats. L’intérêt est de rendre possible l’élaboration de scénarios de fonctionnement différents en jouant sur les paramètres qui influenceront par exemple sur les portées des voisinages : ce pourrait être un scénario “100% des achats alimentaires de proximité à pied”, ou encore l’introduction d’une part croissante d’achats réalisés en mode doux, comme cela semble se profiler dans certaines villes. Cet outil permettra de travailler en différentiel par rapport à des scénarios de référence, et stimuler la réflexion sur des cas de territoires plus vulnérables que d’autres.

Bibliographie non numérotée et références

Agences d’urbanisme de Lyon et de Saint-Etienne (2016), *Le système alimentaire de l’aire métropolitaine lyonnaise*, 88 p.

Beaulac J., Kristjansson E., Cummins S. (2009). *A Systematic Review of Food Deserts, 1966-2007*. Preventing Chronic Disease 6(3), 105-115

Black C., Moon G., Baird J. (2014). *Dietary inequalities: What is the evidence for the effect of the neighbourhood food environment?*. Health & Place 27, 229–242.

Caspi C.E., Sorensen G., Subramanian S.V., Kawachi I. (2012). *The local food environment and diet : a systematic review*. Health & Place 18(5), 1172–1187.

CCI Lyon Métropole (2009). *Schéma Directeur d’Urbanisme Commercial (SDUC) de la métropole lyonnaise 2009 - 2015*, 266 p.

CCI Lyon Métropole (2017). *10ème enquête consommateurs, région, lyonnaise, principaux résultats*, 32 p.

Charreire H., Casey R., Salze P., Simon C., Chaix B., Banos A., Badariotti D., Weber C., Oppert J.M. (2010). *Measuring the food environment using geographical information systems : a methodological review*. Public Health Nutrition 13(11), 1773–1785.

Glanz K., Sallis J.F., Saelens B.E., Frank L.D. (2005). *Healthy nutrition environments : concepts and measures*. American Journal of Health Promotion 19(5), 330–333

INSEE (2015). *Le temps des courses depuis 1974*, INSEE Première n° 1533, 4 p.

Kelly B., Flood V.M., Yeatman H. (2011). *Measuring local food environments : an overview of available methods and measures*. Health & Place 17(6), 1284–1293.

McKinnon R.A., Reedy J., Morrisette M.A., Lytle L.A., Yaroch A.L. (2009). *Measures of the food environment : a compilation of the literature, 1990–2007*. American Journal of Preventive Medicine 36(4), 124–S133

SYTRAL, Agence d’Urbanisme aire métropolitaine lyonnaise (2016), *Enquête déplacements 2015 de l’aire métropolitaine lyonnaise, Principaux résultats*, 44 p.

Walker E., Keane C.R., Burke J.G. (2010). *Disparities and Access to Healthy Food in the United States: A Review of Food Deserts Literature*. Health & Place 16, 876–884.

L'Ontologie d'Alerte Choucas : de la modélisation des connaissances à un outil support d'un raisonnement géovisuel

Application à la recherche de victime en haute-montagne

Matthieu Viry, Marlène Villanova-Oliver

*Univ. Grenoble Alpes, CNRS, Grenoble INP (Institute of Engineering Univ. Grenoble Alpes), LIG
38000 Grenoble, France
{matthieu.viry; marlene.villanova-oliver}@univ-grenoble-alpes.fr*

RÉSUMÉ. Cet article présente l'Ontologie d'Alerte Choucas, développée dans le cadre du projet ANR éponyme, et son exploitation dans un outil support d'un raisonnement géovisuel. L'ontologie formalise le domaine d'une alerte et de son traitement par un secouriste dans le cas particulier de la recherche de victime en haute-montagne. Elle structure ainsi les concepts exploités par le secouriste élabore des hypothèses de localisation probable de la victime à partir d'informations fournies par le requérant (la personne qui appelle les secours). L'ontologie est exploitée par un prototype de géovisualisation dont les fonctionnalités visent à accompagner, faciliter et optimiser le raisonnement du secouriste. L'outil permet de saisir des indices collectés (souvent complexes et imparfaits), d'en fournir une restitution cartographique, de construire et d'affiner peu à peu des zones de localisation.

ABSTRACT. This article presents the Choucas Alert Ontology, developed in the framework of the ANR project of the same name, and its exploitation in a support tool of a geovisual reasoning. The ontology formalizes the field of an alert and its treatment by a rescuer in the particular case of the victim search in high mountains. It thus structures the concepts used by the rescuer and elaborates hypotheses of the probable location of the victim based on information provided by the person who calls for help. The ontology is exploited by a geovisualization prototype whose functionalities aim to support, facilitate and optimize the reasoning of the rescuer. The tool allows to capture collected clues (often complex and imperfect), to provide a cartographic restitution, to build and to refine localization areas.

MOTS-CLÉS : Ontologie, raisonnement spatial, géovisualisation, secours.

KEYWORDS: Ontology, spatial reasoning, geovisualization, search and rescue.

1. Introduction

Lors de certaines interventions, quand la localisation par GPS n'est pas possible, les secouristes du PGHM (Peloton de Gendarmerie de Haute-Montagne) doivent recouper différentes informations données par téléphone par la personne ayant déclenché l'alerte (le requérant) afin de localiser la victime. Ces informations sont de natures diverses : lieu de départ, itinéraire emprunté, description de l'environnement actuel, et font souvent appel à des termes ou expressions ambigus ou flous. Le secouriste doit exploiter ces éléments et procède à une analyse géovisuelle exploitant différents supports (cartes, couches métiers, etc.) et mobilisant ses propres connaissances. Dans ce contexte, le projet CHOUCAS¹ se propose de « formaliser, modéliser et implémenter les stratégies convoquant les méthodes de visualisation les plus adaptées au contexte d'utilisation » (Olteanu-Raimond, 2017). Cet article porte sur la modélisation des connaissances, qui sous-tendent le raisonnement du secouriste, sous la forme d'une ontologie, appelée Ontologie d'Alerte Choucas. Celle-ci résulte de l'analyse de cas réels d'alertes, de l'étude de scénarios d'utilisation, de l'identification des besoins en termes d'acquisition et d'exploitation des informations, ainsi que de la mise en évidence de concepts sur lesquels s'appuie le raisonnement des secouristes. Plusieurs objectifs ont guidé ce travail : modéliser les étapes du raisonnement et les connaissances mobilisées par les secouristes lors de la résolution de ces cas ; favoriser les échanges entre les partenaires du projet et garantir l'interopérabilité des développements ; faciliter le prototypage rapide d'interfaces permettant d'opérationnaliser et de tester les différents concepts modélisés. L'ontologie est exploitée dans un outil visant à assister, faciliter voire optimiser le raisonnement géovisuel. Nous travaillons en particulier sur l'interface pour proposer des composants de collecte des informations fournies et des modes de restitution des résultats de localisation calculés aussi efficaces et intuitifs que possible. Nous suivons ici Adrienko et Adrienko, (2007) qui donnent des prérequis à une visualisation intelligente pour la gestion de crises : la réalisation d'une ontologie du domaine, l'identification des acteurs impliqués et leurs besoins en termes d'informations et enfin des techniques et des méthodes spécifiques et adaptées aux tâches à effectuer pour manipuler, organiser et présenter les résultats. Tout comme pour la définition de l'ontologie, notre démarche est ici également « centrée utilisateurs » dans la réalisation de composants logiciels, conçus et évalués avec les secouristes. L'évaluation des composants de géovisualisation proposés ouvre parfois sur de nouveaux modes de raisonnement non envisagés jusqu'ici, ce qui a pu avoir des répercussions sur la formalisation des connaissances portées par l'ontologie. La méthodologie employée est de nature itérative et incrémentale, portée par un cycle prenant en compte l'utilisateur, ses retours, les besoins émergeant des nouveaux composants de restitution et d'interaction proposés. Celle-ci s'inscrit dans une logique comparable aux travaux de Robinson *et al.*, (2005) qui adoptent une démarche centrée humain pour le développement d'outils de géovisualisation pour l'épidémiologie. Ils ont recours à un processus de co-création et à différentes méthodes d'évaluation qualitative (verbalisation, *focus*

¹ANR Choucas : Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne. ANR-16-CE23-0018.

group) mises en œuvre pour valider les propositions. Ces méthodes sont proches de celles que nous mobilisons pour tester la validité de nos propositions et l'ergonomie de notre interface auprès des utilisateurs finaux comme déjà évoqué dans (Viry *et al.*, 2019) lors d'un premier prototypage. Ces préoccupations s'inscrivent dans le cadre de défis, persistants dans le domaine de la géovisualisation, tels que la prise en compte des utilisateurs finaux, la compréhension des facteurs (perceptuels et cognitifs) impliqués dans l'analyse géovisuelle ou encore les réflexions sur le choix d'une géovisualisation adaptée au type de tâche à réaliser (Çöltekin *et al.*, 2017).

Cet article est organisé de la façon suivante. Dans la section 2, nous présentons plusieurs travaux sur la modélisation des connaissances pour l'analyse géovisuelle et de prise de décision. La section 3 introduit l'Ontologie d'Alerte Choucas qui formalise le domaine d'une alerte et de son traitement par un secouriste lors de la recherche de victime en montagne. Dans la section 4, nous montrons comment certains concepts décrits sont matérialisés dans le prototype d'aide à la localisation. Les orientations de nos prochains travaux sont données en section 5.

2. Positionnement par rapport à l'état de l'art

2.1 Formalisation des connaissances et analyse (géo) visuelle

Abi-Zeid *et al.* (2010) décrivent un système permettant d'aider à la localisation, suite à l'envoi d'un signal de détresse, d'aéronefs disparus. Les auteurs utilisent des approches d'ingénierie des connaissances pour acquérir, structurer et modéliser le raisonnement des secouristes et le restituer dans une interface de recherche interactive. Cet article présente des points communs avec nos méthodes (ingénierie de la connaissance, utilisation de cas réels) et avec notre problématique mais n'intègre pas la complexité des indices manipulés dans le cas de la recherche de victimes en montagne comme abordés dans nos travaux (voir section 3).

Konstantopoulos *et al.* (2006) décrivent une approche à base d'ontologies pour l'organisation des opérations des services de lutte contre les incendies. Les modèles proposés se concentrent sur les aspects opérationnels des secours que nous ne traitons pas dans le projet Choucas. Si les auteurs identifient le besoin d'étendre le spectre de leurs ontologies pour supporter plus efficacement le raisonnement spatio-temporel des secouristes, ils ne font toutefois pas de proposition en ce sens à ce jour.

La question de l'analyse d'un support géovisuel dans le cadre des secours est abordée par Löffler *et al.* (2007). Ils soulignent l'apport de la visualisation 3D dans les cas de gestion des secours (pour traduire l'orientation, la hauteur des objets, etc.) ainsi que l'intérêt de systèmes qui permettent de répondre à des questions complexes nécessitant de décoder des requêtes en langage naturel pour les transformer en requêtes sur des objets présents dans le système d'information utilisé. Ceci est également un axe étudié dans le projet Choucas (Bunel *et al.*, 2019) et qui fait l'objet de certaines de nos propositions (voir section 4).

2.2 Informations imprécises et prise de décision

L'information fournie par le requérant est souvent porteuse d'imperfections de différentes natures (imprécision, incertitude, etc). Lorsque l'incertitude n'est pas éliminable du domaine de connaissance modélisé, Caglioni et Fusco (2014), proposent de faire évoluer le cadre clair et déterministe des ontologies formelles vers des approches probabilistes, floues, possibilistes ou basées sur les évidences. Ils identifient en particulier un fort potentiel d'application à des ontologies prenant en compte l'incertitude lorsqu'il s'agit de modéliser des connaissances géographiques.

Kinkeldey *et al.* (2017) s'intéressent aux effets de la représentation de l'incertitude de données géospatiales lors de la prise de décision. Ils comparent un panel de 87 expérimentations et synthétisent les éléments concernant l'acceptation de l'incertitude, les effets de sa représentation, les techniques utilisées pour représenter l'incertitude, l'impact du niveau d'expertise. Les effets observés révèlent ses différences dans les décisions prises, dans la justesse de celles-ci, dans le type d'erreur commis, dans le temps nécessaire à la prise de décision, dans la confiance accordée à une décision, ou encore dans la charge de travail / cognitive liée à la prise de décision. Des techniques pour le rendu de l'incertitude qui s'avèrent plus efficaces (e.g. représentation graphique *versus* affichage d'une valeur numérique) sont mises en avant. Les auteurs évoquent également le concept de « *info-gap decision theory* » (voir Ben-Haim, 2006, pour plus de précisions à ce sujet) pour expliquer le niveau d'incertitude toléré avant qu'une décision prise ne soit modifiée. Ces travaux sont pris en compte dans nos démarches.

Enfin Joussemme *et al.* (2018) s'intéressent à la capacité des experts à raisonner avec des informations venant de sources différentes et de qualité variable. Dans le cadre de leurs expérimentations, les auteurs étudient notamment « le besoin d'information » (informations que l'utilisateur pense être utiles) et « la pertinence effective de l'information » (quelles sont les informations qui aident réellement à résoudre le problème donné). Ils mettent en avant que les informations exprimées avec une forte certitude, mais pas nécessairement avec une forte précision, sont celles qui ont le plus d'impact sur le fait de croire que la solution proposée est la bonne. La représentation de la confiance accordée aux indices par les secouristes et son impact sur les décisions prises font également parties de nos préoccupations.

3. L'Ontologie d'Alerte Choucas

L'Ontologie d'Alerte Choucas formalise les concepts impliqués dans le traitement d'une alerte par un secouriste du PGHM. Ces concepts sont ceux sur lesquels se base le raisonnement du secouriste, opéré *dans un environnement de géovisualisation*. La formalisation intègre ainsi des concepts liés à la restitution d'informations et aux interactions proposées dans l'outil puisqu'ils sont des vecteurs de ce raisonnement.

L'Ontologie d'Alerte Choucas, développée en OWL 2, structure les différents éléments recueillis ou élaborés par le secouriste et qui sont utiles pour la localisation

trouve. Elle correspond à la traduction d'un premier indice (*cf.* ci-après) : il peut par exemple s'agir de l'ensemble d'un massif si la victime est « perdue en Chartreuse », ou d'un rayon de plusieurs kilomètres autour de son dernier point connu.

Un **Indice** correspond à un fragment du discours du requérant (qui dialogue avec les secours) et décrit la position actuelle, ou une position passée, de la victime. Ce fragment de discours constitue un élément de localisation incluant des informations spatiales mais aussi temporelles. Un indice peut être décrit par un **instant** ou une **durée** de validité. Des propriétés d'ordonnement (indice suivant et/ou précédent) permettent de décrire la chronologie entre éléments de localisation, notamment utile quand un marquage temporel n'est pas possible. Les données temporelles sont également exploitées pour la reconstitution d'*Itinéraire* (*cf.* plus loin). Le crédit accordé à un **Indice** est également modélisé sous la forme d'une valeur de **Confiance** parmi 3 modalités : faible, moyenne ou fort. Le secouriste peut ainsi moduler le poids d'un indice si besoin (cela répond à des situations où ce que dit le requérant paraît par exemple peu plausible au secouriste).

Un indice implique une **Relation de Localisation** définie entre deux **Entités Spatiales**, respectivement une **Cible** et un **Site**. Cinq sous-classes de **Relations de Localisation** sont définies (non représentées ici) pour porter différentes sémantiques. Ces concepts sont issus d'une Ontologie de Relations de Localisation en cours de développement au sein du projet Choucas. Inspirée de Bateman et al., (2010), elle qualifie et matérialise finement les relations spatiales entre la victime et/ou les éléments du territoire (**Objets** ou **Catégories d'Objets**). À titre d'exemple (Figure 2), une **Relation de Localisation Binaire** exprime qu'une **Cible** (la victime) est liée à un **Site** par un **type de relation** (tel que voir, entendre, être passé par, être à proximité immédiate de, etc.). Dans une **Relation de Localisation Emboîtée**, l'indice est composé d'une **cible** (la victime), d'un type de relation spatiale et d'un **site**, mais ce dernier a la particularité d'être lui-même issu d'une relation de localisation dont la **cible** et le **site** sont des **Objets** ou **Catégories d'Objets du Territoire** (voir Figure 1).

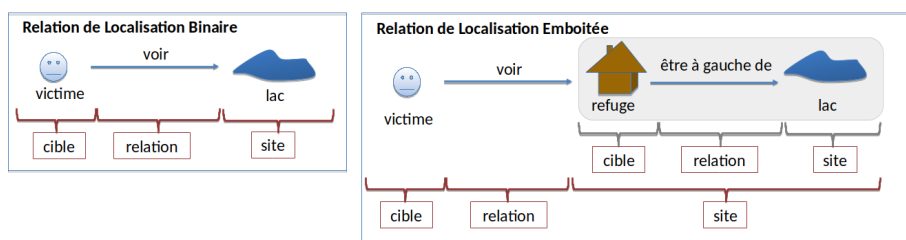


Figure 2. Relations de localisation simple: binaire et emboîtée.

Ces relations de localisation, binaires ou emboîtées, peuvent également être qualifiées de **Relation Spatiale Simple** ou de **Relation Spatiale Étendue** (Bunel et al., 2018). Cette dernière permet d'exprimer la finesse de la description d'une relation de localisation en langage naturel (« je suis *juste à côté* d'un pont », « je vois un lac à *environ 500m* ») en incluant des modificateurs qui accentuent/atténuent tout ou partie de l'expression (Bunel et al., 2019).

Une *Entité Spatiale (ES)* décrit un objet précis (*ES Repère*, e.g. « le lac Achard ») ou un ensemble d'objets (*ES Candidate*, e.g. « les lacs ») par rapport auxquels la victime peut être localisée, *via* une relation spatiale décrite précédemment. La distinction effectuée entre *ES Repère* et *ES Candidate(s)* permet de considérer les informations imprécises (quand le requérant ne sait pas nommer l'ES). L'objectif est de raffiner ces informations imprécises, notamment afin de faire émerger, d'un ensemble d'*ES Candidates* (« les lacs ») l'*ES Repère* qui est décrite par le requérant (« le lac Achard » par exemple). Les éléments du territoire pouvant être mobilisés dans le cadre d'une alerte en tant qu'*ES* sont donnés par une Ontologie des Objets de Référence en cours de construction par le projet Choucas et mobilisant des concepts relatifs à la topographie, l'hydrographie, aux constructions, aux formes du relief, etc. (voir Mustière *et al.*, 2011, ainsi que Sinha *et al.*, 2018).

Un *Indice* a vocation à être traduit en une *Zone de Localisation (ZL)* lui correspondant. Il s'agit de passer d'une représentation conceptuelle d'informations de localisation à la représentation graphique de ce qu'elles expriment. Ces zones peuvent être de deux types selon la nature de l'indice : une *ZL Compatible* correspond à la matérialisation d'une relation de localisation par l'affirmative (« être à côté d'un refuge ») alors qu'une *ZL Incompatible* correspond à la matérialisation d'une relation par la négative (« ne pas être sur un chemin »).

Un *Itinéraire* correspond à un ensemble de points de passage, à des jalons de temps distincts, et éventuellement une destination (elle-même décrite par une *ES Repère* ou par un ensemble d'*ES Candidates*). La création d'un *Itinéraire* peut mobiliser des *Indices* précédemment saisis (tels que « être parti de la mairie de Chamrousse il y a 3h », « avoir été au bord du lac Achard il y a 2h » et « avoir aperçu le grand pic de Belledonne il y a 1h ») comme points de passage. Comme pour les *Indices*, l'*Itinéraire* de la victime peut être traduit en une *ZL Compatible*. À cet effet, des méthodes de reconstitution d'itinéraires sont étudiées dans le projet Choucas, notamment en comparant les indices disponibles avec les données de plateformes communautaires de partage *via* l'appariement de textes de randonnées et de traces associées (voir Medad *et al.*, 2018).

Enfin, notre ontologie intègre des éléments permettant la transformation des *Indices* en *Zone de Localisation* par le biais de différents *Services* offerts à l'utilisateur. L'objectif est de respecter son raisonnement en lui permettant de matérialiser un indice selon la forme géovisuelle qui lui paraît la plus adaptée, et en offrant un compromis entre précision des calculs effectués et temps de calcul nécessaire. Par exemple, la possibilité est offerte de matérialiser l'indice « je vois le sommet du Néron » en utilisant le service de calcul de bassin de visibilité (précis mais coûteux) ou en utilisant le service de création de zone tampon pour créer une large zone autour de l'objet. L'ontologie référence donc pour un *Type d'Indice* les *Services* de calcul et de rendu graphiques répertoriés dans un catalogue définis pour le projet Choucas (Halilali *et al.*, 2018). Les paramètres de ces Services sont exprimés comme des classes ou des propriétés des ontologies définies dans le projet Choucas (e.g. la propriété *hauteur* des Entités Spatiales *cible* et *site* d'une Relation de Localisation pour invoquer un service de calcul de bassin de visibilité).

Disposant d'un ensemble de *Zone de Localisation* créées à partir d'*Indices*, il est alors possible, par appel à un Service de fusion de ces zones, d'obtenir une ***Zone de Localisation Probable (ZLP)*** de la victime. Par fusion, on entend la détermination de la portion du territoire dans laquelle, par recoupement des différents *Indices*, la victime doit se trouver. Cette *ZLP* peut comporter une ou plusieurs portions disjointes de territoires qu'il faut restituer au secouriste. La section suivante aborde la matérialisation des concepts de l'Ontologie d'Alerte Choucas dans un prototype de géovisualisation que nous développons.

4. Matérialisation des concepts dans le prototype de géovisualisation

Matérialiser les concepts de l'ontologie dans un prototype de géovisualisation qui supporte le raisonnement du secouriste nécessite de s'intéresser *i)* au développement de composants pour la saisie des informations qui viennent peupler les concepts de l'ontologie (composants de collecte des indices notamment) et *ii)* à la retranscription dans la carte des instances de concepts (stockées, comme les *Indices*, ou calculées comme les *Zones de Localisation*). Nous présentons dans ce qui suit des propositions pour ces deux aspects (saisie et retranscription visuelle) à travers différents exemples qui prolongent nos premiers travaux (Viry *et al.*, 2019). Différents *Services* ont été instanciés (pour la création d'une zone tampon, la création d'une zone à la main levée, le calcul d'inter-visibilité et le calcul de l'ombre portée – calculs respectivement réalisés avec les fonctionnalités *r.viewshed* et *r.sun* de la bibliothèque GRASS) et permettent de créer des *Zones de Localisation* matérialisées par des géométries vectorielles polygonales à partir d'*Indices* décrivant des *Relations de Localisation Simple*. Les *Entités Spatiales* mobilisables proviennent de la base de données OpenStreetMap et les données d'altitude proviennent du programme *Copernicus* de l'Agence Européenne de l'Environnement.

4.1 Zone Initiale de Recherche

Définie par le secouriste, la Zone Initiale de Recherche (ZIR) correspond à la définition à gros grain d'une zone dans laquelle mener la recherche sur le support cartographique. Cette zone est suffisamment vaste pour qu'il soit certain que la victime s'y trouve. Ainsi, pour le PGHM de Grenoble, peuvent être définies plusieurs ZIR ayant un sens par rapport à sa zone de compétence (le département de l'Isère) et au domaine montagnard (la ZIR Chaîne de Belledonne, la ZIR du Massif de la Chartreuse, etc.). Ces présélections de ZIR sont instanciées dans notre ontologie en utilisant la couche « massifs » produite par le PGHM de Grenoble.

Une ZIR est matérialisée par un masque de couleur blanc-gris disposant d'une opacité de 50 %. Il s'agit d'une boîte englobante contenant le massif ou la portion du territoire sélectionnée par l'utilisateur. Cette matérialisation permet de faire un focus sur la zone de recherche sans priver le secouriste du contexte environnant, pouvant éventuellement l'inciter à annuler ou élargir cette zone si nécessaire.

4.2 Informations relatives à la victime et à sa pratique

Un panneau permet de saisir les différentes informations relatives à la victime : son état de santé (notamment de sa capacité éventuelle à se déplacer) et le fait que la victime est également ou non le requérant. Ces deux éléments sont utiles pour questionner à bon escient le requérant afin d'obtenir des indices sur l'environnement immédiat (ou plus lointain, ce qui suppose un déplacement) de la victime.

La ou les activités pratiquées peuvent être sélectionnées parmi plusieurs propositions (instances de la classe *Activité*, voir Figure 1). Ceci permet d'activer rapidement des couches d'informations propres à l'activité en question non nécessairement présentes sur le fond de carte initial (entrées de grottes pour la *spéléologie*, sentiers pour l'activité *randonnée pédestre*, etc.).

4.3 Indices

La création d'*Indices* peut se faire de plusieurs manières dont certaines ont été présentées dans (Viry *et al.*, 2019). La diversité des solutions vise à ne pas contraindre l'utilisateur à un seul type de raisonnement et à supporter des cas d'utilisations variés (variabilité du nombre d'indices recueillis simultanément et de la taille de la zone initiale de recherche ; préférence de l'utilisateur pour la lecture de la carte, pour la recherche d'une entité nommée dans une liste, pour la saisie de plusieurs indices à la suite, etc.). Nous illustrons la saisie et la matérialisation de deux indices dans l'interface à partir de deux fragments de textes, issus de cas réels :

- « *Je suis juste à côté d'un sentier de randonnée* » peut être saisi dans le panneau de création des indices (Figure 3-A) par la sélection de l'ensemble des *sentiers de randonnées (Catégories d'Objets du Territoire)* de la *Zone Initiale de Recherche* et le choix de *proximité immédiate* comme type de relation de localisation. Ce type de relation permet de définir deux valeurs pour traduire au mieux l'expression : la distance par rapport à l'objet de référence (le *Site*) et un rayon correspondant à l'imprécision de la localisation par rapport à celui-ci. Ici, des valeurs de, respectivement, 0m et 50m sont indiquées pour décrire la présence de la victime dans une zone de 50 mètres maximum autour du sentier de randonnée (cette valeur pour modéliser la présence *juste à côté* peut paraître large mais elle offre un aperçu initial de la zone compatible et peut être réduite par exemple après confirmation d'une distance moindre). Le curseur relatif à la confiance est positionné sur « *fort* », correspondant à une prise en compte totale de l'indice dans le calcul de la *Zone de Localisation Probable*, et le curseur temporel est positionné sur « *maintenant* ». Une fois la création de l'indice validée, la zone correspondante s'affiche en surimposition au fond de carte (Figure 3-B).

Comme alternative à la saisie, nous introduisons ici un composant Bloc-Note illustré dans la Figure 4 spécifiquement développé pour le projet. Ce composant propose une zone de saisie de texte sous forme d'items (signalés par des puces). Chaque item correspond à un élément de localisation, saisi selon les termes du requérant. Au delà de simplement consigner les informations, ce composant permet via un menu contextuel d'activer la complétion automatiquement du panneau de création d'indices à partir du texte saisi. La Figure 4 illustre la traduction du texte « *est juste à côté* » en une *Relation Spatiale* de *proximité immédiate* et du texte

« sentier de rando » en une référence aux *Catégories d'Objets du Territoire Sentier de Randonnées* qui deviennent des *Entités Candidates de la Relation Spatiale*).

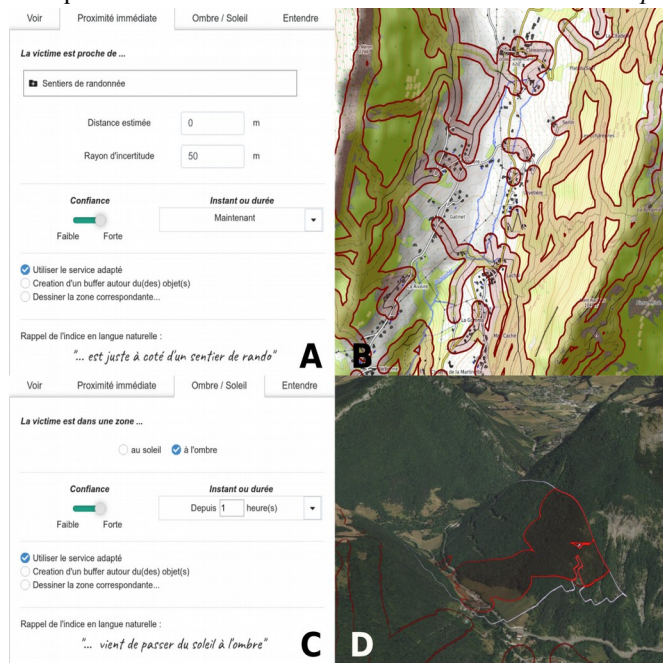


Figure 3. Saisie (A et C) et exemple de restitution (B et D) des indices « je suis juste à côté d'un sentier de randonnée » et « je viens de passer du soleil à l'ombre ».

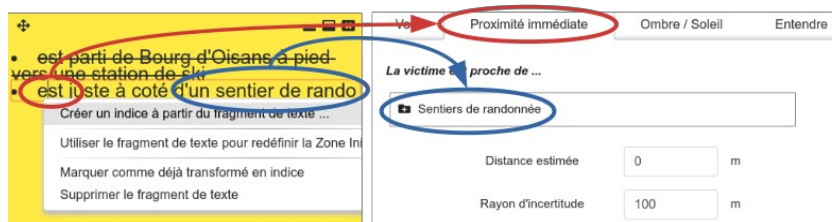


Figure 4. Utilisation d'une zone de notes (texte libre) pour la création d'indices. Le fragment de texte rayé correspond à un indice déjà traité.

- « Je viens de passer du soleil à l'ombre » peut être créé en utilisant le même panneau de création d'indices que précédemment (Figure 3-C). La sélection de la relation de localisation « ombre / soleil » et du service approprié permet d'afficher les paramètres adéquats : il s'agit ici des modalités « à l'ombre » et « au soleil » qui décrivent la position de la victime. Cette fois, le curseur temporel est positionné de

façon à sélectionner la durée « depuis 1 heure ». Ces paramètres permettent d'obtenir une zone correspondant à la différence entre les zones d'ombre il y a une heure et les zones d'ombre actuelles. Un affichage en trois dimensions utilisant un modèle numérique de terrain et une orthophotographie (Figure 3-D) peut être utilisé conjointement ou en remplacement d'un fond de carte traditionnel (Figure 3-A).

5. Conclusion et perspectives

Nous avons présenté l'Ontologie d'Alerte Choucas qui permet de modéliser la phase d'alerte du processus de recherche de victime en montagne et plus particulièrement les concepts d'indices et de zones de localisation probable. Un prototype d'interface² mobilisant l'essentiel des concepts de l'ontologie est développé pour supporter plus efficacement le processus de raisonnement du secouriste. Ces propositions bénéficient d'autres réalisations du projet Choucas, notamment en garantissant l'interopérabilité avec deux autres ontologies (Ontologies des Objets de Référence de Repères et des Relations de Localisation) et différents services de calculs.

Dans les suites immédiates données à ce travail, un premier accent est mis sur l'exploitation de l'ontologie pour guider la suggestion de nouveaux indices qu'il serait pertinent d'obtenir en fonction de ce qui est déjà connu et du fait de leur pouvoir discriminant pour éliminer ou réduire des ZLP. Un second travail porte sur la prise en compte plus fine de la confiance dans les indices, notamment lors de l'étape de fusion de ceux-ci. Nous travaillons également à la matérialisation tant en termes de composant de saisie qu'en termes de restitution géovisuelle du concept de relation de localisation emboîtée introduit en partie 3.2 (« voir un refuge à coté d'un lac ») ainsi qu'à la matérialisation d'autres types de relations de localisation nécessitant également de relever certains défis comme lorsqu'il s'agit de transcrire des informations comme « voir un groupe de 5 bâtiments en bas d'une colline ».

Bibliographie

- Andrienko N., Andrienko G. (2007). Intelligent Visualisation and Information Presentation for Civil Crisis Management. *Transactions in GIS*, vol. 11, n° 6, décembre 2007, p. 889-909, <https://doi.org/10.1111/j.1467-9671.2007.01078.x>
- Bateman J.A., Hois J., Ross R., Tenbrink T. (2010). A linguistic ontology of space for natural language processing. *Artificial Intelligence*, 174, p. 1027-1071, <https://doi.org/10.1016/j.artint.2010.05.008>
- Ben-Haim Y. (2006). *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*, 2nd edition, Academic Press, London, ISBN 0-12-373552-1.
- Bunel M., Olteanu-Raimond A.-M., Duchêne C. (2018). Référencement spatial indirect : modélisation à base de relations et d'objets spatiaux vagues. *SAGEO 2018, 6-9 Novembre 2018*, Montpellier.

²Deux vidéos de démonstration ainsi qu'un lien vers le code source sont disponibles sur http://steamer.imag.fr/?page_id=792.

- Bunel M., Duchêne C., Olteanu-Raimond A.-M., Villanova-Oliver M., Bonhoure G., Jouan T. (2019). An approach for identifying and analysing reference features and spatial relations used in mountain emergency calls. *ICC 2019, 15-20 Juillet 2019*, Tokyo, Japon.
- Cagliani M., Fusco G. (2014). Formal Ontologies and Uncertainty. In *Geographical Knowledge. Tema. Journal of Land Use, Mobility and Environment, 2014: INPUT 2014 - Smart City: Planning for Energy, Transportation and Sustainability of the Urban System*, <https://doi.org/10.6092/1970-9870/2530>
- Çöltekin A., Bleisch S., Andrienko G., Dykes J. (2017). Persistent Challenges in Geovisualization – a Community Perspective. *International Journal of Cartography*, vol. 3, n° sup, p. 115-39, <https://doi.org/10.1080/23729333.2017.1302910>
- Halilali M. S., Gouardères E., Devin F., Gaio M. (2018) Plateforme logicielle pour l'intégration et la composition de services géospatiaux. *SAGEO 2018, 6-9 Novembre 2018*, Montpellier.
- Jousselman A.-L., Pallotta G., Locke J. (2018). Risk Game: Capturing impact of information quality on human belief assessment and decision making. *International Journal of Serious Games*, vol. 5, n° 4, p. 23-44, <https://doi.org/10.17083/ijsg.v5i4.258>
- Kinkeldey C., MacEachren A. M., Riveiro M., Schiewe J. (2017). Evaluating the Effect of Visually Represented Geodata Uncertainty on Decision-Making: Systematic Review, Lessons Learned, and Recommendations. *Cartography and Geographic Information Science*, vol. 44, n° 1, p. 1-21, <https://doi.org/10.1080/15230406.2015.1089792>
- Konstantopoulos S., Paliouras G., Chatzinotas S. (2006). SHARE-ODS: An Ontology Data Service for Search and Rescue Operations. *SETN*, 2006.
- Löffler J., Ernst V., Schon J., Pottebaum J., Koch R. (2007). Intelligent Use of Geospatial Information for Emergency Operation Management. *Proceedings of the fourth international conference on information systems for crisis management, ISCRAM, 2007*.
- Medad A., Gaio M., Le Nir Y., Mustière S. (2018). Appariement automatique de données hétérogènes: textes, traces GPS et ressources géographiques. *SAGEO 2018, 6-9 Novembre 2018*, Montpellier.
- Mustière S., Abadie N., Aussenac- Gilles N., Bessagnet M.-N., Kamel M., et al. (2011). Analyses linguistiques et techniques d'alignement pour créer et enrichir une ontologie topographique. *Revue Internationale de Géomatique*, 2011, Lavoisier, vol. 21, n°2, p. 155-180.
- Olteanu-Raimond A.-M., Davoine P.-A., Gaio M., Gouardères E., Van Damme M.-D., Villanova-Oliver M., Brasebin M., et al. (2017). Projet CHOUCAS: Intégration de données hétérogènes et raisonnement spatial pour l'aide à la localisation des victimes en montagne. *SAGEO 2017, Novembre 2017*, Rouen.
- Sinha G., Arundel S. T., Hahmann T., Usery E. L., Stewart K., Mark D. M. (2018). The Landform Reference Ontology (LFRO): A Foundation for Exploring Linguistic and Geospatial Conceptualization of Landforms (Short Paper). *10th International Conference on Geographic Information Science (GIScience 2018)*, Leibniz International Proceedings in Informatics (LIPIcs).
- Viry M., Villanova-Oliver M., Gautier J., Sreeves M., Davoine P.-A. (2019). Improving the search for victims in mountain with geovisualization and competing hypotheses management. *ICC 2019, 15-20 Juillet 2019*, Tokyo, Japon.

Intensification, periurbanization and specialization as significant short-term land system dynamics in the Mediterranean basin

Fusco J.¹, Marraccini E.² Debolini M.³

1. UMR CNRS IMBE. Aix-Marseille Université - Campus Aix Technopôle de l'environnement Arbois Méditerranée Avenue Louis Philibert Bât Villemin - BP 80 F-13545 Aix en Provence, johannafusco@hotmail.fr

2. UP INTERACT 2018.C102, UniLaSalle, Beauvais, France. elisa.marraccini@unilasalle.fr

3. UMR EMMAH INRA/UAPV, Avignon, France. marta.debolini@inra.fr

RESUME. Les systèmes territoriaux méditerranéens connaissent d'importants changements en termes d'usage du sol et de pratiques agricoles. Ces changements sont mis en évidence par un certain nombre d'études locales ou régionales, mais il manque une vue d'ensemble des dynamiques au niveau du bassin méditerranéen. Cet article identifie les principales dynamiques des systèmes territoriaux sur une courte période (2005-2015) et leur localisation. Nous avons relevé quatre types de changements importants : (1) de l'agriculture mixte aux vergers spécialisés ; (2) des zones agricoles aux zones urbaines et/ou périurbaines ; (3) de l'agroforesterie aux systèmes arables, et (4) des sols avec végétation éparse aux zones agricoles. Ces dynamiques sont des signes avant-coureurs d'intensification, périurbanisation et spécialisation de l'agriculture.

ABSTRACT. Mediterranean land systems are undergoing significant changes in terms of agricultural land use and practices. In the literature these changes are highlighted by a number of local or regional studies, however, an overview of the land system changes at the Mediterranean basin level is lacking. This paper identifies the main land system dynamics over a short time period (2005-2015) and their location, in order to highlight the trajectories that are also taking place in other parts of the world. We identified four significant types of change: (1) from mixed agriculture to specialized fruit groves; (2) from agricultural areas to urban and/or periurban areas; (3) from agroforestry to arable systems, and (4) from predominantly bare soils to agricultural areas. These ongoing dynamics can be characterized as intensification, periurbanization and specialization of agriculture.

MOTS-CLES : Systèmes territoriaux méditerranéens, classification expert, analyse des clusters non-supervisée, systèmes agricoles, trajectoires, machine learning.

KEYWORDS: Mediterranean land system, expert-based classification, unsupervised cluster analysis, agricultural systems, trajectories, machine learning

SAGEO'2019 – Clermont Ferrand, 13-15 novembre 2019

1. Introduction

The Mediterranean basin has a long history in terms of production of food and fiber. Current agricultural systems and the complementarity between crops are complex and derive from a long history of agricultural and Mediterranean traditions (Zeder, 2008). Many typical land uses and landscapes are due to the reshaping of the environment (e.g. terraces and irrigation canals, hedgerows) (Pinto-Correia and Vos, 2004; Blondel, 2006) made it necessary by the harshness of natural conditions for agriculture (poor and shallow soils, steep slopes and dry summers) and the limited availability of resources, particularly water (Caraveli, 2000). These traditional agricultural practices, including agroforestry and agro-silvo-pastoralism, are known to provide food production while improving the good functioning of ecosystems. However, if most of these agricultural practices were common in the past, they decreased significantly since the beginning of the 20th century in favour of more economically profitable ones (Pinto-Correia and Vos, 2004) often accompanied by unsustainable practices.

The spatial patterns of Mediterranean agricultural systems, based on mixed crops and polycultural systems, are however difficult to identify and map if only a land cover and land use approach is used (Malek and Verburg, 2017). Following the increasing availability and quality of land use data, the development of more sophisticated methodologies and the increasing interest in the management of different land uses, land use and land cover studies have tended to combine the concepts of land use and management through “land systems” (Verburg et al., 2013). In this case, the geographical unit for modelling and predicting future changes is not directly the land cover or the land use, but a more complex class that combines land use and management through indicators such as population density, irrigation systems, and agricultural inputs. The rationale behind this systemic approach is to consider the world as a complex system where natural/environmental and human/socio-economic components are fully integrated and considered at the same time (GLP, 2005). A land system approach is thus well adapted to tackling Mediterranean agricultural complexity, in order to anticipate and manage the impacts of future changes on ecosystem services.

This paper aims to describe and to map current Mediterranean land systems and their ongoing dynamics with a focus on agricultural areas. For this purpose, we exploit the systematic, data-driven and exploratory aspects of cluster analysis, supported for technical and thematic reasons by expert knowledge. This work is part of the ARIMNET2-DIVERCROP project, which aim to analyse the possible impact of land system change at Mediterranean level on biodiversity and food security.

2. Materials and Methods

2.1. Study area

We defined the spatial extent of the Mediterranean region based on the combination of two references. 1) The northern part (from Portugal to Anatolia) corresponds to the “Mediterranean Mountains”, “Mediterranean North” and “Mediterranean South” environmental zones defined by Metzger et al (2005) in their climatic environmental stratification of Europe. We also decided to include the “Alpine South” zones that are spatially adjacent to the two previous ones (located in the Pyrenees, the Massif Central, the Alps, and the Dinaric Mountains) in order to capture characteristic agricultural abandonment dynamics. 2) We retained as the southern and oriental parts of the Mediterranean basin the spatial extent proposed by Malek and Verburg (2017) in their Mediterranean land system analysis.

2.2. Data

We chose to perform our land system classification at a 2km spatial resolution. This choice is based on two main reasons: 1) a 10km resolution was tested at first, and was found to be not fine enough to capture the spatial patterns of several types of fine-scale changes, such as urban sprawl. 2) This resolution is as an optimal compromise between the original high data resolution (table 1) and our intention to capture and to synthesize global patterns of human-environment interactions and changes in the studied area. The sources and the main characteristics (table 1) of the original data used to describe human-environmental interaction on the Mediterranean basin, are described below. These data have been chosen as they represent the most relevant aspects of Mediterranean land systems and they describe a similar time span. In terms of diachronic analysis, these was possible through the LU and the population change, as the others data were static or available in just one data. The data had considered firstly as the same weight on the classification, but then we considered on the cluster analysis just them which better described the class separation.

2.3. Land system classification

2.3.1. Expert-based classification

A preliminary expert-based classification was needed before performing any unsupervised classification, for two main reasons: (1) the heaviness of the dataset (716105 lines x 88 columns) prevented any classification algorithm; (2) some typical Mediterranean agricultural systems according to the local did not emerge sufficiently from the unsupervised classification.

TABLE 1. Sources for the spatial database used to model Mediterranean land systems and Information provided.

Information	Source	Reference	Variables	Unit	Temporal resolution	Original spatial
Land Cover / Land Use	ESA/CCI Land Cover Project. https://mans.elie.ucl.ac.be/CCI/viewer/	ESA Climate Change Initiative - Land Cover	15 classes of Land Cover	Percentage of each class in	2005; 2015	300m
Topography	GMTED 2010	Danielson & Gesch 2011.	Elevation	Average in 2km pixel	/	1 km
			Average slope	Average in 2km pixel		
			Soil nutrient availability			
Soil quality	Harmonized World Soil Database v 1.2	Fischer et al. 2008	Rooting conditions	Average of 0 to 7 indicator	2006	1 km
			Soil workability			
Bioclimatic data	WorldClim v2	Fick & Hijmans 2017	Temperature	Average in 2km pixel (°C)	Average for years 1970 to 2000	1 km
			Rainfall			
Population	Landscan ¹	https://landscan.ornl.gov/	Population	Amount in 2km pixel	2005; 2015	1 km
Livestock	Gridded Livestock of the World (GLW3)	FAO 2007	Cattle, goats & sheep	Amount in 2km pixel	2006	1 km
Crop physical	MAPSpam	You et al. 2014	9 classes of crops	Number of ha in 2km pixel	2005	10 km

¹ This product was made utilizing the LandScan (2005)TM High Resolution global Population Data Set copyrighted by UT-Battelle, LLC, operator of Oak Ridge National Laboratory under Contract No. DE-AC05-00OR22725 with the United States Department of Energy.

These specific agricultural systems, notably the multifunctional ones such as agroforestry, highlight the methodological challenge of classifying Mediterranean land systems systematically at a global scale. Statistical classification methods require very precise, perhaps qualitative micro-scale information on agricultural practices, which are impossible to consider at this scale or extent, and which are neither homogenous nor consistent. Global data with homogenous and consistent coverage can also cause classification problems. In fact, when representing arboriculture, the “rainfed croplands with tree or shrub cover” land cover class was sufficient to enable us to identify this class in the northern part of the Mediterranean basin. However, it seemed totally absent from the southern part of the Mediterranean basin, if only land use and land cover variables were considered, whereas it is, in fact, present. We thus had to mobilize expert knowledge and crop surface information from the MAPSPAM database to represent arboriculture. The classification method is summarized on Fig.1.

2.3.2 Unsupervised cluster analysis

After the expert based classification, we performed the unsupervised classification on the remaining pixels. Considering the large number of pixels and variables to classify, even after the expert based classification, we needed to apply a classification method developed for big data analysis. Although several methods of unsupervised cluster analysis suited for big data have been used yet for Land Systems characterization (e.g. Letourneau et al. 2012; Václavík et al 2013; van der Zanden et al. 2016), we chose the CLARA algorithm – Clustering LARge Applications- created by Kaufman and Rousseeuw (1986,1990), available in the R package cluster (Maechler et al. 2018) and extensively described in Kassambara, 2017. This data mining and clustering algorithm is an extension of the k-medoids methods and deals with large datasets by using a sampling approach (Kaufman and Rousseeuw 1990). The quality of the medoids resulting from the sample is measured by the average dissimilarity between every object in the entire data set and the medoid of its cluster, defined as the cost function. The final clustering results correspond to the set of medoids with the minimal cost. Not all the variables in the database were used for all the cluster analysis. The choice of the variable was based on an iterative method: we started by including the largest set of variables and then noted which indicators differentiated the most among classes and then repeated the clustering using a smaller set of indicators.

2.3.3 Map validation and land system dynamics assessment

Our land system classification was based on an iterative approach; at each step we identified the main bias of the preliminary results and then returned to the methodology to correct them. One of the main tools was local expert validation. First of all, we discussed with the local partners the most relevant variables to include in the analysis, during the first DIVERCROP project meeting in Tunis (December 2017). Then, our initial classification was sent to each local partner in

the project with a zoom of the map on the country of origin (Portugal, Spain, Italy, France, Tunisia, Malta and Algeria). For each country, the classification was analysed by at least three experts. The questions for the partners were “Does the proposed classification relevantly represent your country’s Land System diversity?”, “Are the various Land System classes correctly located across your country?” and “Please share any additional opinion / information on Land Systems characterization across your country”. Their answers enabled us to correct for any bias, such as the underestimation of the arboriculture on the southern Mediterranean or the misclassification of agroforestry areas.

Concerning the dynamics assessment, we first compared the surface of each land system classes for the two years. We then developed a transition matrix in order to evaluate the main land system trajectories.

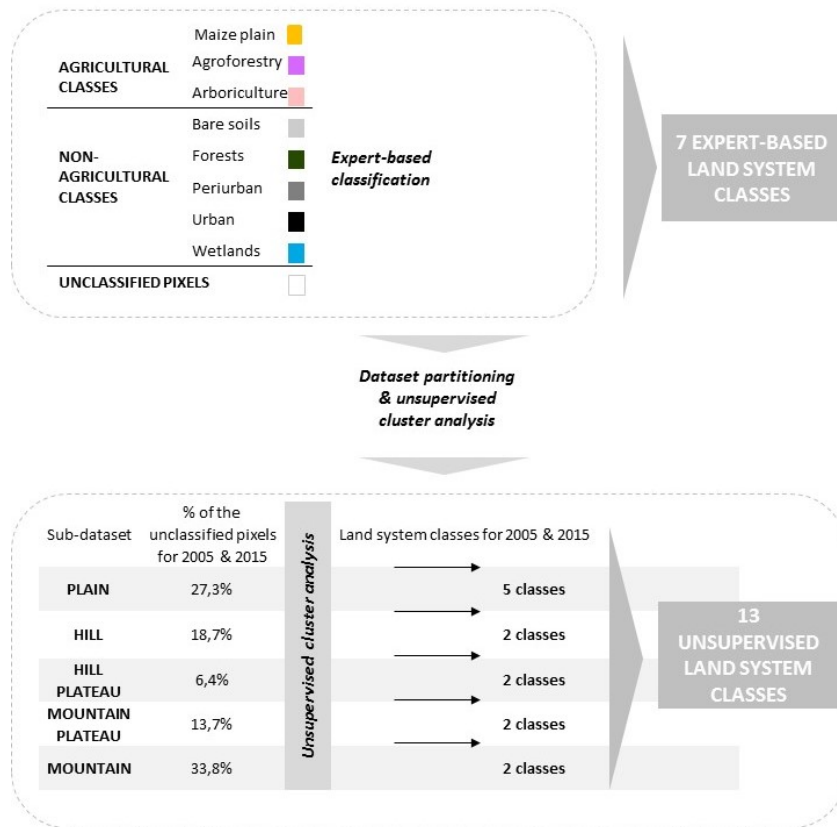


FIGURE 1. Overall scheme of the methodology.

3. Results

3.1 Land system classification

We obtained 20 land system classes. Table 2 presents the main characteristics of each class, Fig. 2 shows the two maps obtained for 2005 and 2015.

TABLE 2. Detailed description of the 20 Mediterranean land systems

Topographic class	n°	Land system	Detailed description of the land system
Irrigated or flooded plain	1	Predominance of arable crops (maize, cereals and oil crops) alternated with vegetables, livestock	Most of the surface is dedicated to irrigated croplands (73%), in particular oil crops and cereals (corn). Vegetables are also well represented. Temperate climate, high population density.
	2	High population density and intensive cattle breeding	Most of the surface is dedicated to irrigated croplands (92%) alternated with urban areas (6%). The population density is very high and cattle breeding is predominant.
	3	Predominance of corn associated with vegetable and cattle breeding	
Rainfed plain	4	Highly productive plains with a predominance of cereal crops combined with oil crops, fruit groves and livestock	Arable crops are predominant (45% of the land use), in particular wheat (29ha), oil crops (26ha) and corn (4ha) associated with temperate fruits (7ha) and vegetables (2ha). Intensive livestock.
	5	Semiarid plains with sparse vegetation and bare soils, pulses	Very low amount of rainfall (150mm/y) and 13% of land use occupied by bare soils. Sparse vegetation is also well represented (15%). Among the crops, pulses are important, associated with cereals.
Hill plateau	6	Arable crops, mainly cereals associated with permanent pastures and sheep / goat breeding	Wheat is the predominant crop (average 67ha/4km ²) associated with pulses and vegetables. Grasslands represent 5% of the land use. High presence of sheep and goat breeding.
	7	Fruit groves associated with arable crops	High presence of temperate fruits (around 6ha) associated with oil crops and cereals (half of the surface than in the previous class). 20% of the surface dedicated to sparse herbaceous cover.
Hill	8	High slopes characterized by olive groves and vineyards associated with natural vegetation (trees and shrubs)	Fruit groves associated with natural vegetation (forest is 25% of the surface and sparse tree and shrublands 13%). Average slope around 8%.

	9	Sparse crops, mainly cereals associated with livestock, mostly sheep / goats	Arable crops (31% of the land use), in particular wheat and other cereals associated with oil crops. Average slope 5%. Intensive sheep and goat breeding.
Mountain plateau	10	Sparse natural vegetation and bare soils, residual agriculture, mainly cereals	Low rainfall (around 400mm/y), 10% of the surface occupied by bare soils or sparse vegetation (18%). Cereals are the main crops (75ha) and the average altitude is around 1000 m.
	11	Arable crops (cereals and oil crops) with high presence of sheep / goat breeding	Intensive cereal and arable crops (84ha of cereals and 12ha of oil crops). Low rainfall amount (460mm/y).
Mountain	12	High mountain permanent pastures (sheep / goats) alternated with sparse natural vegetation	The average altitude is around 1700 m a.s.l. with very high slopes (15%). Grasslands occupy 17% of the land use associated with natural sparse vegetation. Extensive livestock systems.
	13	Grasslands alternated with arable crops (cereals and pulses). Sheep and cattle breeding	The average altitude is around 1000 m a.s.l. with high slopes (10%). Grasslands are alternated with cereals and pulses. Extensive livestock systems.
Other agricultural classes	14	Agroforestry	25% or more of the cell's LU/LC are mosaic croplands associated with natural vegetation. Sheep breeding is predominant.
	15	Fruit groves	50% or more of the cell's LU/LC are rainfed croplands with tree or shrub cover and temperate fruits are around 24 ha
Non-agricultural classes	16	Bare soils	95% or more of the cell's LU/LC are bare soils
	17	Forest	95% or more of the cell's LU/LC are forests
	18	Periurban areas	Around 34% of the cell's LU/LC are urban areas. Cereals are the most common crop, associated with temperate fruits and vegetables.
	19	Urban areas	Around 92% of the cell's LU/LC are urban areas
	20	Wetlands	>60% of the land use is wetlands

The most represented class is mountain areas characterized by the presence of both cereals and pulses associated with sheep and cattle breeding (around 16% of the total Mediterranean area). This class is relatively equal in the north and south of the basin. The highly productive rainfed plains, which represent around 12% of the total surface, are mainly located along coastal areas, but also in some internal regions, such as the south of Portugal (Alentejo), the north of Morocco (Saiss), the south of France (Garonne and Rhone) and the north of Italy (Po valley) where they are associated with irrigated surfaces of wheat, rice and corn. The hilly areas are characterized by permanent crops such as vineyards and olive groves which

Short-term land system dynamics in the Mediterranean basin

represent around 10% of the Mediterranean basin as well as the high mountain areas with permanent pastures. Vineyards and olive groves are located mainly in the internal hilly areas of the northern side of the Mediterranean, but also in the coastal hilly areas of Morocco and Algeria. High mountain areas with permanent pastures represent the typical land system of Alpine areas, including the Pyrenees, Balkans, Atlas Mountains and internal Turkish mountains.

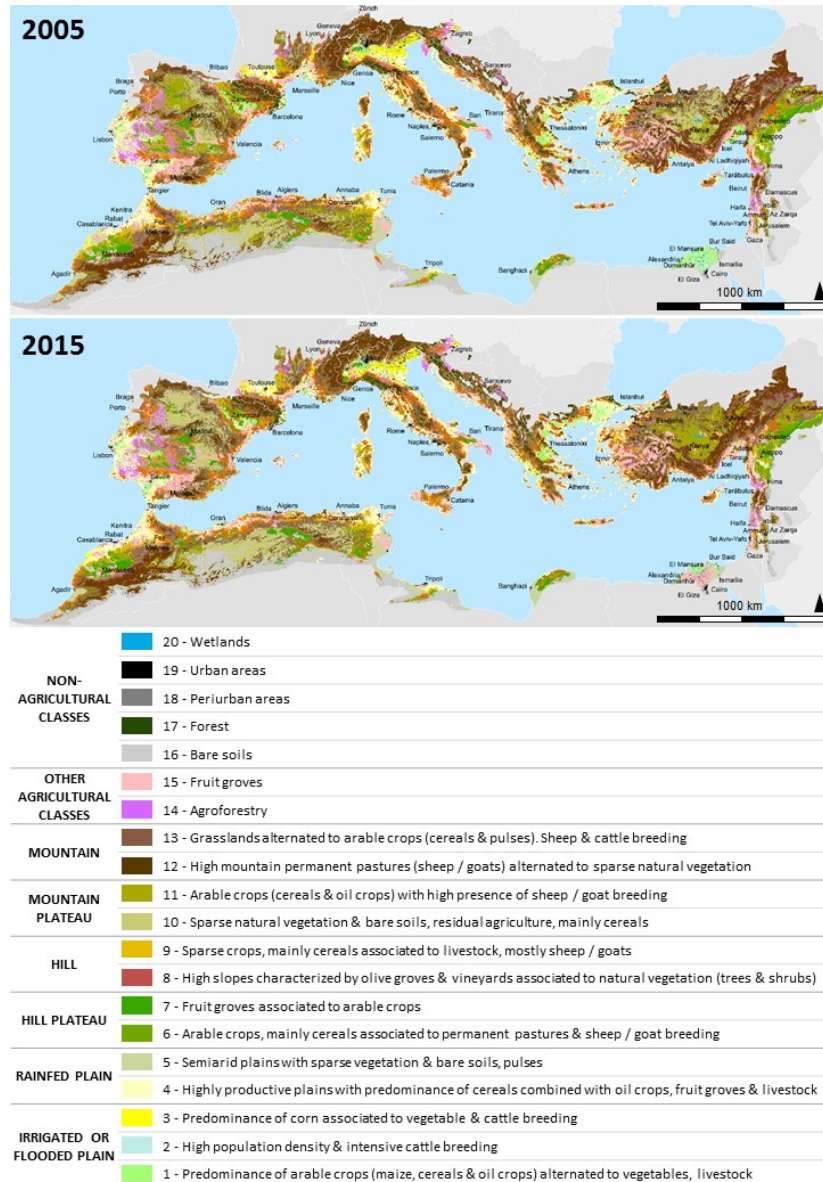


FIGURE 2. Maps of Mediterranean land systems for 2005 and 2015

3.2 Land system dynamics

As shown in Fig. 3 and Fig. 4 the main changes are: a decrease in irrigated plains where arable crops and vegetables are grown; a decrease in agroforestry; an increase in fruit groves and periurban areas.

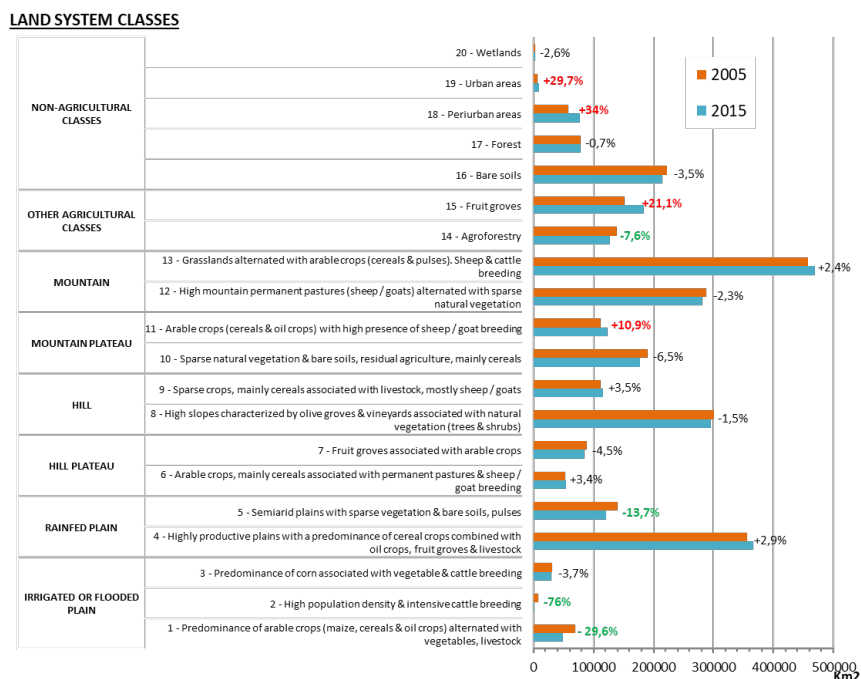


FIGURE 3- Changes in Mediterranean land system areas between 2005 and 2015. **Values in red:** 4 highest change rates. **Values in green:** 4 lowest change rates.

Four main trajectories were identified in all parts of the Mediterranean basin:

1) From mixed agriculture to specialized fruit groves. This corresponds to a change from different types of land systems including arable crops, vegetables and livestock, to specialized fruit groves. In 2005 the areas involved were by: (1) irrigated plains with a predominance of arable crops and vegetables (54%), (2) irrigated plains with a high population density and intensive cattle breeding (13%), (3) rainfed highly productive plains with a predominance of cereal crops associated with fruit groves and breeding (5%), (4) high slope hilly areas characterized by olive groves and vineyards associated with natural vegetation (5%), (5) mountain plateau areas with sparse natural vegetation and bare soils (8%), and other systems. The first three cases would seem to be the result of specialization, whereas the hilly and mountain areas are the result of intensification. These dynamics take place in different types of topography, from plains to mountains. We observed these dynamics in the Nile plain, in some coastal areas of Turkey, but also in Albania, southern Italy (Apulia), Valencia (Spain) and in some coastal areas of the Maghreb.

Specialization is more frequent in the coastal areas of the southern and northern Mediterranean, whereas intensification is mainly present in hilly and mountainous parts of the basin located in the most internal zones.

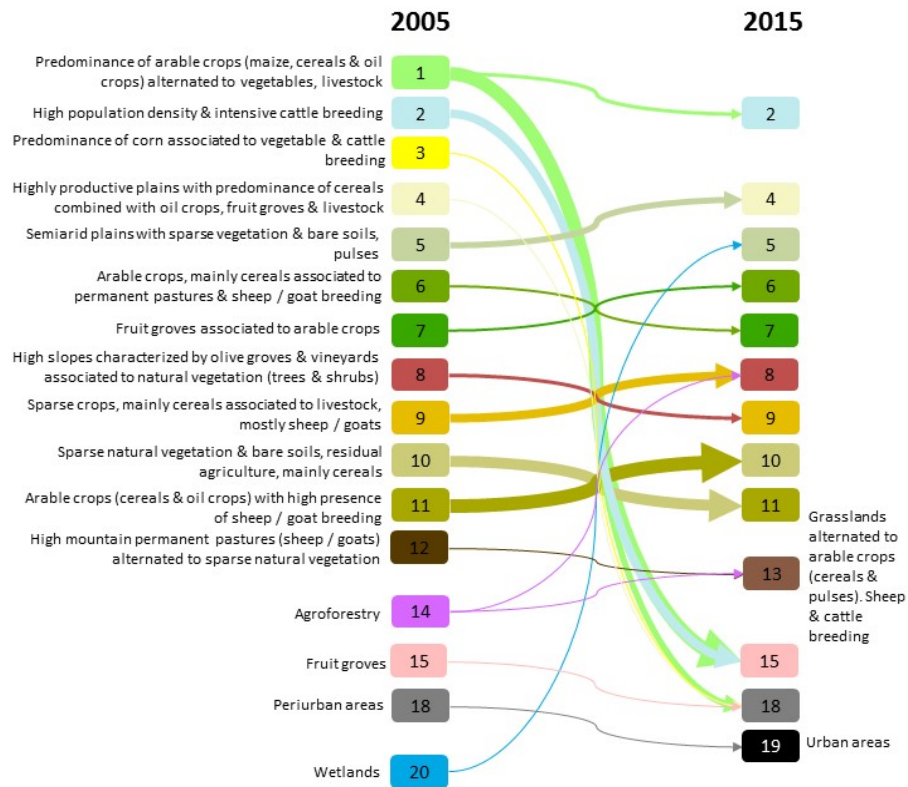


FIGURE 4 - Transition plot displaying the main land system conversions between 2005 and 2015. The arrows' width traduces the intensity of change

2) From agricultural areas to urban and/or periurban areas. The highly productive plains with a predominance of cereal crops or hill plateau areas with arable crops or fruit groves are the most affected by this kind of change (36% and 6%, respectively) together with arboriculture areas (10%). This means that Mediterranean areas are losing their most productive agricultural lands to urbanization. These areas are widespread around the Mediterranean basin, mainly in the north-western and south-eastern parts.

3) From agroforestry to arable systems. This highlights the simplification of agroforestry systems through the loss of trees and the increase in arable lands, until the complete disappearance of agroforestry and the transition to arable crops and vegetables. The main dynamics are from agroforestry to: (1) mountains with arable

crops, sheep and cattle breeding (29%); (2) high slope hilly areas with fruit groves (25%); (3) rainfed highly productive plains with a predominance of cereal crops (15%); and (4) mountainous permanent pastures (10%). All these cases involve an intensification process associated with the simplification of the agricultural system. The areas most affected by these changes are in the internal parts of Portugal, some coastal areas of Algeria, and the internal hilly parts of the Balkans and Turkey.

4) From predominantly bare soils to agricultural areas. There has been a change from predominantly bare soils to extensive agricultural areas, such as mountain plateaus with sparse natural vegetation (50%), mountain areas with permanent pastures (27%), and mountain areas with arable crops (12%). It has especially affected the southern Mediterranean, particularly Morocco and Algeria, near the borders between agricultural and desert areas, or in the southern Atlas Mountains.

Acknowledgements: DIVERCROP is funded through the ARIMNet2 2016 Call by the following funding agencies: ANR (France), IRESA (Tunisia), INIA (Spain), FCT (Portugal), ATRSNV (Algeria), MIPAAF (Italy) and MCST (Malta). ARIMNet2 (ERA-NET) received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n° 618127. Special thanks to all of the local partners of the DIVERCROP project for discussing the method and validating the results.

References

- Blondel, (2006). The “Design” of Mediterranean Landscapes: A Millennial Story of Humans and Ecological Systems during the Historic Period. *Human Ecology* 34(5):713–29.
- Caraveli, H. (2000). A comparative analysis on intensification and extensification in mediterranean agriculture: dilemmas for LFAs policy. *Journal of Rural Studies* 16(2),231–42.
- Danielson, J.J., & Gesch, D.B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)*: U.S. Geological Survey Open-File Report 2011–1073, 26 p.
- FAO. 2007. *Gridded livestock of the world 2007*, by G.R.W. Wint and T.P. Robinson. Rome, pp 131
- Fick, S.E. and R.J. Hijmans, 2017. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*.
- Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuisen, L. Verelst, D. Wiberg, 2008. *Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008)*. IIASA, Laxenburg, Austria and FAO, Rome, Italy.
- GLP (2005). *Science plan and implement strategy* (IGPB Report N° 53/IHDP - Report N° 19, 54 pp. Stockholm IGPP secretariat.
- Kassambara A. (2017). Practical guide to cluster analysis in R. Unsupervised machine learning. In *Multivariate analysis I*, edited by sthda.com, 186pp.

Short-term land system dynamics in the Mediterranean basin

- Kaufman, L., and Rousseeuw, P. J. (1986), Clustering large data sets (with discussion), in *Pattern Recognition in Practice II*, edited by E. S. Gelsema and L. N. Kanal, Elsevier/North-Holland, Amsterdam, pp. 425-437.
- Kaufman, L., Rousseeuw, P. (1990). Finding Groups in *Data: An Introduction to Cluster Analysis*. Wiley, New York. <http://www.sthda.com/english/articles/27-partitioning-clustering-essentials/89-clara-clustering-large-applications/>
- Letourneau A, Verburg PH, Stehfest E (2012) A land-use systems approach to represent land-use dynamics at continental and global scales. *Environmental Modelling & Software*, 33, 61–79.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K. (2018). *Cluster: Cluster Analysis Basics and Extensions*. R package version 2.0.7-1.
- Malek, Ž., & Verburg, P. (2017). Mediterranean land systems: Representing diversity and intensity of complex land systems in a dynamic region. *Landscape and Urban Planning*, 165, 102-116.
- Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Múcher, C. A., Watkins, J. W. (2005). A climatic stratification of the environment of Europe. *Global Ecology and Biogeography* 14, 549–563.
- Pinto-Correia, T. and Vos, W. (2004). Multifunctionality in Mediterranean landscapes – past and future. In: Jongman, R.H.G (Ed.): *The new dimensions of European landscapes*. Wageningen UR Frontis Series, Springer, 135-164.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Václavík, T., Lautenbach, S., Kuemmerle, T., Seppelt, R. (2013). Mapping global land system archetypes. *Global Environmental Change* 23(6): 1637-1647.
- Verburg, P.H, Erb, K.H, Mertz,O. & Espindola, G. (2013). Land System Science: between global challenges and local realities. Editorial overview. *Current Opinion in Environmental Sustainability* 5: 433-437
- You, L., U. Wood-Sichra, S. Fritz, Z. Guo, L. See, and J. Koo. 2014. *Spatial Production Allocation Model (SPAM) 2005 v2.0* [2017]. Available from <http://mapspam.info>
- Van der Zanden, E. H., Levers, C., Verburg, P. H., & Kuemmerle, T. (2016). Representing composition, spatial structure and management intensity of European agricultural landscapes: A new typology. *Landscape and Urban Planning*, 150, 36–49. <http://dx.doi.org/10.1016/j.landurbplan.2016.02.005>.
- Zeder M.A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *PNAS* 105:11597-604.

eSTIMe : une approche visuelle, interactive et modulable pour l'analyse multi-points de vue des mobilités quotidiennes

Aline Menin¹, Sonia Chardonnel²,
Paule-Annick Davoine^{1,2}, Michael Ortega¹,
Etienne Dublé¹, Luciana Nedel³

1. Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
pre.nom@univ-grenoble-alpes.fr

2. Univ. Grenoble Alpes, CNRS, Science Po Grenoble, PACTE, 38000
Grenoble, France
sonia.chardonnel@univ-grenoble-alpes.fr

3. Federal University of Rio Grande do Sul, Institute of Informatics, Porto
Alegre, Brazil
nedel@inf.ufrgs.br

RÉSUMÉ. Nous proposons un environnement de géovisualisation qui vise à accompagner l'analyse de la mobilité quotidienne selon trois points de vues : les déplacements et les flux, les trajectoires individuelles, et la dynamique de peuplement du territoire. Chacun de ces points de vue s'appuie sur la définition d'indicateurs spécifiques qui font l'objet de représentations visuelles innovantes. L'approche choisie pour favoriser l'analyse visuelle repose sur le concept de tableaux de bord modulables et mobilise des techniques issues du domaine de la visualisation et de l'interaction homme-machine.

ABSTRACT. We propose a geovisualization environment to assist the analysis of daily mobility data focused on three standpoints: the flows and displacements, the individual trajectories, and the territory settlement dynamics. We derive specific indicators from the data to study these aspects, which we explore through synchronized novel visual representations displayed on analytic flexible dashboards. Our approach combines concepts and techniques from visualization and human-computer interaction domains.

MOTS-CLÉS : mobilité quotidienne, géovisualisation, analyse geovisuelle, tableau de bord analytique, données spatio-temporelles

KEYWORDS: daily mobility, geovisualization, spatio-temporal data, geovisual analysis, analytical dashboard

1. Introduction

L'intérêt pour l'analyse de données décrivant la mobilité quotidienne des populations est aujourd'hui partagé par de nombreux domaines au sein des politiques urbaines : offre de transport, accessibilité aux services urbains, environnement (qualité de l'air), santé publique et bien-être (modes actifs, environnements adaptés, sécurisés). De ce fait, beaucoup d'experts (chargés d'études et diagnostics, gestionnaire des collectivités territoriales) ont besoin de manipuler des données de mobilité - sans être des spécialistes du transport - pour extraire des connaissances synthétiques et facilement exploitables. Les jeux de données décrivant les déplacements quotidiens proviennent aujourd'hui de sources diversifiées telles que des enquêtes en population (enquêtes ménages déplacements, enquêtes assistées par GPS) ou des corpus de «traces» d'activités géolocalisées extraits des *Call Detail Record* (télécommunications) ou des réseaux sociaux. La plupart du temps, l'approche commune utilisée pour formaliser la description des déplacements d'un individu à partir de ces données consiste à générer une chronique (par exemple de 24 heures) d'activités et de déplacements spatialisés et enrichis sémantiquement (type d'activité, modes de déplacements). En ressort une structure classique (Peuquet, 1994) permettant d'explorer des entités d'intérêts dans le temps et dans l'espace à différents niveaux d'agrégation. Pour traiter ces données spatio-temporelles, les environnements de géovisualisation et d'analyse exploratoire dédiés à l'étude des mobilités quotidiennes se sont multipliés. L'apport majeur de ces outils est notamment de mieux donner à voir la dynamique du phénomène en proposant des indicateurs et des méthodes de visualisations intégrant le temps. Dans le domaine de l'expertise des politiques urbaines les exploitations des données privilégient l'entrée par les déplacements et les flux, même si les méthodes d'analyses et de représentations des dynamiques du territoire sont aussi exploitées pour l'action (Armoogum *et al.*, 2015; Hurez, Pélata, 2016). Plus généralement, beaucoup d'environnements existants sont conçus avant tout pour résoudre un problème spécifique, comme la détection d'embouteillages, l'exploration de l'utilisation de systèmes de vélos partagés (Yan *et al.*, 2018), l'étude de la ségrégation sociale dans une région métropolitaine (Le Roux *et al.*, 2017) ou encore l'analyse de trajectoires de vie ou de déplacements individuels (Cochey, Tabaka, 2007; Thudt *et al.*, 2013; Otten *et al.*, 2015). Par ailleurs, la plupart de ces environnements privilégient une seule forme de visualisation centrée soit sur la représentation des dynamiques du territoire (avec des cartes dynamiques par exemple (L. Shi *et al.*, 2017)), soit sur celle des individus en mouvement (avec des vues en 3D de trajectoires dans un cube spatio-temporel (Tominski *et al.*, 2012)). A notre connaissance, il n'existe pas d'outil donnant la possibilité aux experts d'appréhender dans un même environnement à la fois des indicateurs synthétiques résumant les caractéristiques des déplacements et des flux, mais aussi des analyses et visualisations représentant le caractère dynamique du phénomène de mobilité à l'échelle des individus et des territoires.

Dans cet article, nous présentons l'approche *eSTIMe* (Spatio-Temporal Exploration of Individual Mobility data) qui vise à accompagner l'analyse de la mobilité quotidienne selon trois points de vues : les déplacements et les flux, les trajectoires individuelles, et la dynamique de peuplement du territoire. Nous décrivons dans un premier temps l'approche qui sous-tend notre proposition, à savoir les attendus des utilisateurs et les principes de visualisation et d'interaction. Dans une seconde partie, nous décrivons, à partir d'un jeu de données implémenté, les différents indicateurs de mobilité et les visualisations associées, avant de présenter l'architecture de l'environnement. Dans une dernière partie sont discutés les résultats des expérimentations menées auprès de plusieurs groupes d'utilisateurs.

2. Approche conceptuelle

2.1. Attendus

Les analyses de la mobilité quotidienne gagneraient à être développées selon des points de vues complémentaires permettant d'analyser dans un même environnement de travail des déplacements, des trajectoires individuelles et des dynamiques de peuplement. Les déplacements résumés par des indicateurs synthétiques (taux de mobilité, parts modales...) sont une première approche des tendances générales des pratiques de mobilité à l'échelle d'un territoire. Les flux, construits par agrégation des déplacements entre deux zones, révèlent la structure des liens entre les sous-espaces du territoire. L'analyse des trajectoires individuelles aide à comprendre comment les individus ordonnent leurs activités et déplacements au fil du temps dans le contexte spatial du territoire. Des typologies de trajectoires (Robette, 2011) peuvent par exemple être élaborées pour dresser des profils d'emplois du temps. Enfin, l'analyse de l'évolution (au cours de la période d'observation) de la densité de présence des individus dans chaque sous-espace (tenant compte de leur mobilité) permet de comprendre les rythmes du territoire et de sa population.

Dans ce contexte, les spécifications qui ont guidé la conception et la réalisation de notre environnement de visualisation sont les suivantes : (1) décrire la mobilité quotidienne selon plusieurs points de vue à travers l'analyse et la comparaison de différents indicateurs au niveau individuel et global qui se complètent mutuellement ; (2) explorer les indicateurs de mobilité dans le temps afin d'identifier les rythmes quotidiens de la ville et de sa population ; (3) donner à l'utilisateur la possibilité de composer «à façon» son environnement visuel en choisissant les indicateurs, les ensembles spatiaux et les périodes temporelles qu'il souhaite analyser et comparer. Partant de ces trois premières spécifications, nous fondons notre proposition sur les principes de l'analyse visuelle, définie comme la capacité à traiter visuellement des données au moyen d'un environnement logiciel et d'en extraire de l'information (Cook, Thomas, 2005). Disposer d'un environnement support à l'analyse visuelle qui intègre différents points de vue implique de s'interroger sur (1) la façon dont les diffé-

rents indicateurs de la mobilité peuvent être représentés visuellement ; (2) les processus d'interactions entre visualisation et utilisateur.

2.2. Principes et proposition

Nous proposons d'utiliser une approche par tableau de bord reconnue pour favoriser l'analyse visuelle des données quantitatives en les rendant interprétables au moyen de représentations visuelles (Kitchin *et al.*, 2015). Cette approche repose sur le principe de la co-visualisation au sein d'une interface commune, de représentations graphiques dynamiques, interactives et interconnectées entre elles. Chaque graphique représente un indicateur synthétique issu des traitements statistiques appliqués aux jeux de données brutes. L'utilisateur interroge et visualise les indicateurs aux moyens des représentations graphiques en fonction de ses questionnements. Les tableaux de bord analytiques offrent la possibilité de mettre en relation les indicateurs, de les comparer, d'explorer différentes combinaisons au gré de son analyse pour en extraire des patterns et des tendances (Rivard, Cogswell, 2004). Appliqués à la gestion de données urbaines, ils constituent des outils très puissants car ils fournissent très rapidement et efficacement des informations sur les différents aspects d'un système urbain et de ses changements (Kitchin *et al.*, 2015). Ces tableaux de bord sont généralement déployés sur plusieurs écrans en raison de la multiplicité des indicateurs à visualiser (Langner *et al.*, 2019). Par ailleurs, (Pagno *et al.*, 2015) recommandent l'utilisation d'appareils mobiles pour interagir avec les écrans et (Kister *et al.*, 2017) proposent des techniques interactives combinant écran mural et appareils mobiles pour faciliter l'analyse visuelle. Partant de ces recommandations, nous pensons qu'en déportant les processus d'interaction entre l'utilisateur et les représentations graphiques (effectués généralement avec la souris) sur une tablette interactive et connectée à de multiples tableaux de bord *modulables*, il est possible de (1) bénéficier d'un plus large espace d'affichage ; (2) renforcer les capacités interactives de l'application et donc accroître le pouvoir exploratoire de l'environnement.

Une des problématiques récurrentes concerne la visualisation des dynamiques spatio-temporelles. Celles-ci sont généralement représentées au moyen de visualisations associant des cartes animées et des graphiques temporels (timelines ou timewheels) eux-mêmes interactifs et/ou animés (Edsall, Peuquet, 1997). Néanmoins, "leur perception est confrontée à des limites cognitives, d'inhibition rétroactive ou de concurrence entre stimuli"(Saint-Marc, 2017). Dans la plupart des cas, ces diagrammes temporels jouent essentiellement un rôle de support pour interroger et filtrer les données. Ils sont basés sur l'existence de liens dynamiques entre cartes et diagrammes, et la technique du brushing (Kaddouri *et al.*, 2014). Nous proposons de déporter les processus de filtrage et de visualisation spatio-temporelle sur le dispositif mobile interactif, et d'y intégrer une technique d'animation basée sur le mouvement de la tablette, lui-même dirigé par l'utilisateur. Les angles d'inclinaison de la tablette correspondent alors aux périodes temporelles et l'animation cartographique associée

est contrôlée via le mouvement effectué par l'utilisateur (Menin *et al.*, 2018). L'interaction corporelle effectuée lors de l'oscillation de la tablette, associant la position et l'orientation des poignets lors de la recherche d'information, tend à favoriser la mémorisation de l'information (Arvola, Holm, 2014).

3. L'environnement eSTIME

3.1. Les données mobilisées

L'environnement eSTIME est conçu de manière générique pour accueillir différentes sources de données de mobilité. Pour l'heure nous avons implémenté les données de l'édition 2010 de l'enquête ménages et déplacements (EMD) réalisée auprès d'un échantillon représentatif de la population de la région métropolitaine grenobloise (354 communes) découpée en 97 secteurs de tirage (CEREMA, 2010). Tous les membres de plus de cinq ans des ménages interrogés décrivent l'ensemble de leurs déplacements effectués pendant les 24 heures précédant le jour de l'enquête (de 4 h la veille à 4 h le jour même), tous motifs et tous modes confondus. L'enquête a été menée auprès de 7 600 ménages, soit 16 000 personnes interrogées ayant effectué plus de 62 000 déplacements en une journée. Des coefficients de redressement, basés sur la commune de résidence, la taille et la motorisation des ménages, sont ensuite appliqués à l'échantillon qui représente ainsi 800 000 personnes générant un peu plus de 2,8 millions de déplacements par jour ouvré. Nous utilisons les données redressées, définies sur trois niveaux d'agrégations spatiales: 97, 39 et 12 secteurs.

3.2. Les indicateurs et représentations associées

Comme énoncé auparavant, l'environnement eSTIME permet aux utilisateurs d'analyser la mobilité selon des points de vue complémentaires : déplacements et flux; dynamiques de peuplement du territoire; trajectoires individuelles. Dans les trois cas, les données peuvent être explorées par l'intermédiaire de visualisations représentant les indicateurs définis à différents niveaux d'agrégations spatiales et temporelles, et selon deux entrées thématiques : les activités appelées « motifs » dans l'EMD (5 modalités : domicile, travail, études, courses et loisirs) et les modes de transport (4 modalités : automobile, vélo, marche à pied et transport en commun). Des données décrivent aussi les profils sociaux des individus : elles sont utilisées pour l'analyse des trajectoires individuelles.

– **La carte** : les dynamiques de peuplement sont représentées à travers des indicateurs qui traduisent la présence au sein du territoire de la population selon différentes périodes de la journée : la densité de présence, qui décrit le rapport entre le nombre de personnes présentes dans chaque secteur et sa superficie ; le taux de migration, qui identifie les secteurs qui ont perdu ou gagné de la population au cours de la journée ; l'indice d'attractivité, qui détermine la densité réelle de chaque secteur par rapport à la densité de ses résidents (André-Poyaud *et al.*, 2008) ; le taux d'activités, qui décrit le volume de personnes présentes

dans chaque secteur pour effectuer chaque activité. La visualisation de ces indicateurs, s'effectue de façon classique au moyen de cartes choroplèthes (voir Figure 1a(A)) ou basées sur des symboles proportionnels. Par exemple, la figure A (taux de migration sur 24 heures) permet de distinguer deux pôles "attractifs" (Grenoble et Voiron) gagnant de la population, des secteurs périphériques en perdant beaucoup et enfin des secteurs intermédiaires tel que le Grésivaudan, espace périurbain proche de Grenoble pour lequel nous décrirons par la suite comment sa population organise ses activités et ses déplacements quotidiens.

– **Le chronogramme** : Cette visualisation permet d'explorer les patterns des activités réalisées par l'ensemble de la population sur le territoire étudié pendant la journée. Il se présente sous la forme d'un graphique à bandes empilées, pour lequel l'axe des abscisses représente le temps et l'axe des ordonnées représente la proportion de personnes par activité (Pistre *et al.*, 2015). La figure 1a (B) illustre le rythme global de la population du Grésivaudan qui pour près de 70% est sortie de son domicile à 9h00 du matin afin de travailler, étudier et pour une plus faible part (10%) faire du shopping et des loisirs, avant de progressivement (entre 16h00 et 19h00) arrêter le travail au profit pour une part des activités de loisirs jusqu'à 22h00, heure où l'ensemble la population est de retour au domicile.

– **La roue de mobilité** : les pratiques des personnes mobiles pendant la journée peuvent être explorées à travers la roue de mobilité. Sa conception a été inspirée par les roues temporelles (ou timewheels), souvent utilisées pour visualiser des séries temporelles cycliques (Zhao *et al.*, 2008). Elle consiste d'un diagramme circulaire qui simule une horloge de 24 heures (voir Figure 1a(C)). Il comporte deux anneaux formés par 24 rectangles représentant des périodes d'une heure. Le cercle extérieur de la roue affiche le rapport entre le nombre de personnes en mouvement et la population du secteur sur la période donnée ; le cercle intérieur affiche la répartition des personnes mobiles sur la période donnée selon le motif ou le mode de transport. Enfin, le *pie chart* central présente la distribution temporelle des motifs ou modes de transport. Suivant le rythme des activités de la population du Grésivaudan, les moments de déplacements (figure 1aC) les plus intenses sont entre 8h00 et 9h00, puis entre 16h00 et 19h00 (avec un petit pic sur la pause méridienne). Quant aux modes, l'usage de l'automobile est majoritaire à toutes les heures de la journée.

– **Le diagramme de flux** : il s'agit ici de représenter non seulement les flux de déplacement quotidiens des individus en termes d'origine - destination, mais aussi d'y associer les motifs et les modes de transport, à l'échelle de la journée, ou par plage horaire. Certes les cartes de flux sont une technique largement utilisée pour visualiser les déplacements des individus, mais elles présentent des contraintes qui affectent leur lisibilité (Bahoken, 2016). Nous représentons donc les flux entre les secteurs au moyen d'un diagramme de cordes, de manière similaire à ce qui est fait par (X. Shi *et al.*, 2017). Les arcs représentent les secteurs géographiques qui sont reliés entre eux par des rubans dont l'épaisseur décrit le volume d'échanges de flux entre deux secteurs. Cette visualisation

peut être générée de façon globale pour l'ensemble des flux, ou de façon plus spécifique pour les flux selon leurs motifs ou des modes de transport. Dans ce cas, la couleur des rubans et des arcs définit le motif de déplacement ou le mode de transport. Une opération de filtrage peut être réalisée de façon interactive pour afficher seulement les rubans reliés à l'arc d'un secteur choisi, ainsi que des détails sur les flux (entrants ou sortants) reliés à ce secteur. Globalement, les flux du Grésivaudan sont soit internes, soit reliés aux secteurs centraux de la ville de Grenoble et des communes de son agglomération (Figure 1b (gauche)). Si l'on distingue les motifs de déplacements, on remarque que les flux vers le travail partent massivement du Grésivaudan pour rejoindre les communes de l'agglomération hors Grenoble alors que les flux des loisirs sont plus souvent internes au Grésivaudan mais vont aussi vers les secteurs centraux de la ville de Grenoble.

– **Les actogrammes** : la représentation en actogramme montre l'ordre temporel des activités d'une personne au cours d'une journée (Chardonnel *et al.*, 2010). Chaque individu de l'enquête est représenté par une ligne et chaque ligne est composée de segments de couleurs différentes. La couleur correspondant à l'activité réalisée. L'abscisse représentant le déroulement temporel des trajectoires montre le moment et la durée de l'activité. La superposition des individus en ligne, autrement appelé tapis, permet de comparer visuellement les séquences entre elles. L'utilisateur peut aussi visualiser les résultats d'une analyse de similarités de séquences (basées sur l'*Dynamic Hamming Distance*) en affichant le tapis de chaque groupe de la typologie. Ainsi, la figure 1c(gauche) montre le résultat d'une classification réalisée sur un échantillon de la population de l'enquête caractérisée par son utilisation quotidienne de la voiture. Les actogrammes des deux groupes montrent clairement deux structures de trajectoires distinctes : la première est ordonnée par l'activité travail qui débute entre 8h00 et 9h00 et se termine vers 17h00 offrant des possibilités d'activités plus variées avant un retour massif aux domiciles vers 20h00 ; la seconde dont le pattern est moins "net" illustre des emplois du temps plus diffus dans le temps avec soit des activités d'études ou de travail, soit des loisirs.

– **Le cube spatio-temporel (STC)** : pour compléter l'approche par les actogrammes qui montrent uniquement la dimension temporelle des séquences individuelles, nous utilisons la représentation 3D classique du cube spatio-temporel (Kraak, 2003). Ceci permet d'explorer les formes et l'étendue des trajectoires individuelles dans l'espace et dans le temps. Nous intégrons dans le cube la possibilité de caractériser les trajectoires au moyen de la couleur pour visualiser les activités et les modes de transport qui les composent. L'utilisateur peut explorer un ensemble de trajectoires dans le cube en les sélectionnant en fonction des caractéristiques socio-démographiques des individus. Il peut en outre valoriser le résultat des typologies de séquences (cf. ci-dessus) en affichant les trajectoires spatio-temporelles de chaque groupe. La figure 1c(droite) projette dans l'espace-temps du cube les trajectoires du premier groupe de la typologie de l'échantillon d'automobilistes. Les domiciles sont dispersés sur l'en-

semble du périmètre de l'enquête, alors que l'on observe une concentration des trajectoires sur les secteurs plus centraux, notamment les 2 pôles repérés dans l'actogramme (Grenoble et Voiron) pendant la période consacrée au travail.

3.3. Les interactions

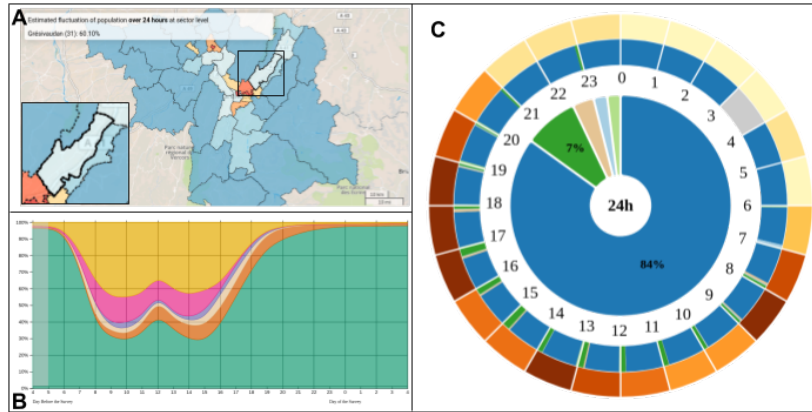
L'application repose sur l'utilisation des techniques d'interactions couramment utilisées dans le domaine de l'analyse visuelle interactive: «pan», zoom, filtrage, sélection spatio-temporelle et attributaires à la demande. Chaque représentation visuelle définie ci-dessus correspond à une vue. Les vues sont interactives et liées entre elles selon la technique *brushing and linking*, de sorte que les modifications réalisées sur une visualisation sont automatiquement répercutées sur les autres visualisations. Nous proposons également de gérer le temps ou l'espace sur les vues de façon indépendante, afin de permettre la comparaison des indicateurs selon différentes granularités spatiales et temporelles.

La carte et le diagramme temporel sont intégrés dans l'interface mobile et jouent le rôle de «contrôleur». Ils prennent en charge les interactions nécessaires à l'exploration des autres vues, en permettant la sélection des zones spatiales et des périodes temporelles selon différentes granularités spatio-temporelles. Rappelons que le diagramme temporel permet la gestion de l'animation cartographique pour la représentation des dynamiques territoriales via le mouvement d'inclinaison de la tablette. Puisque la mobilité quotidienne des individus tend à former des patterns qui se répètent selon un rythme de 24 heures, le diagramme temporel intégré dans le contrôleur peut être de type cyclique (timewheel) ou linéaire (timeline). Du point de vue des interactions, ces deux représentations se différencient par le niveau et la forme d'inclinaison de la tablette: la timewheel requière plus de degrés de liberté en termes de manipulation que la timeline. D'un point de vue fonctionnel, il est possible de déployer les visualisations proposées sur plusieurs tableaux de bord (écrans) connectés au contrôleur. Ceci permet de multiplier les possibilités de visualisation et d'exploration des données. Chaque tableau est composé de quatre fenêtres, initialement vierges, et dans lesquelles l'utilisateur y affiche les indicateurs de son choix, personnalisant ainsi son tableau de bord en fonction de ses besoins d'analyse.

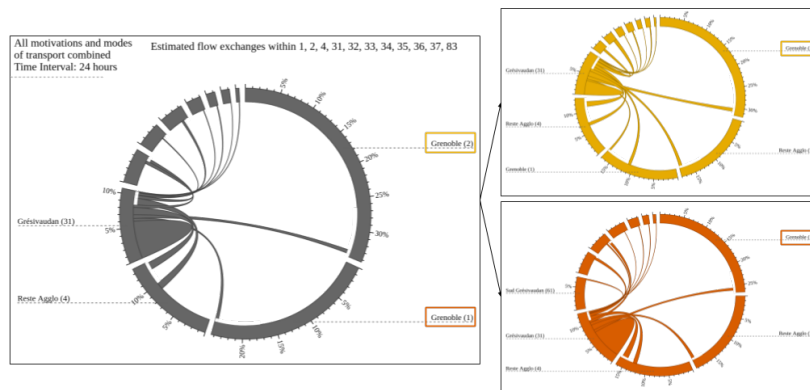
4. Implémentation

L'environnement eSTIME est basé sur l'utilisation d'au moins deux écrans : le premier s'affiche sur une tablette et contient l'interface interactive (appelée le *Contrôleur*) qui intègre les fonctionnalités classiques de navigation et d'exploration des données. Le second écran prend la forme d'un tableau de bord, sur lequel s'affichent et s'organisent les représentations sélectionnées. L'utilisateur peut avoir plusieurs tableaux de bord (et plusieurs écrans) connectés au Contrôleur. L'architecture d'eSTIME (voir Figure 2) est basée sur les technologies Web. Le coté serveur s'appuie sur la technologie WebSocket et est structuré en deux modules : le premier développé en Java gère la communication entre les mul-

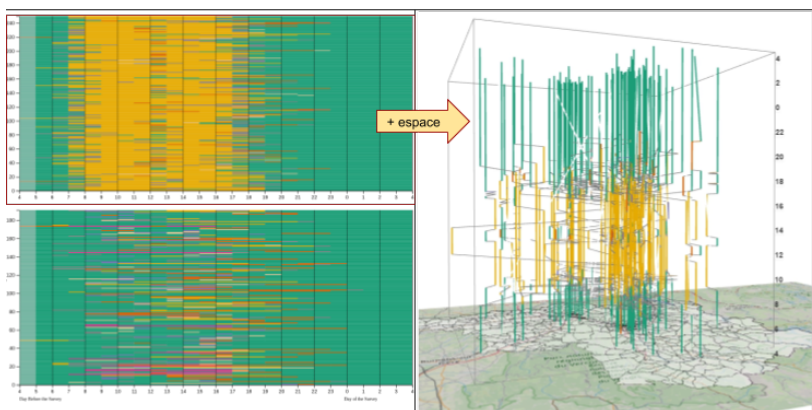
L'analyse visuelle des mobilités quotidiennes



(a) La dynamique de peuplement du territoire est représenté au moyen de cartes choroplèthe (A) et des symboles proportionnels; des chronogrammes (B) ; et via la roue de mobilité (C).



(b) Les flux sont représentés au moyen d'un diagramme de cordes. Les flux arrivant à chaque secteur sont différenciés de ceux qui partent par une lacune blanche entre la corde et l'arc.



(c) Les résultats de la classification des trajectoires sur le temps sont explorés à l'aide des actogrammes (A) et projetés sur l'espace à l'aide du STC (B).

FIGURE (1) Chaque représentation peut supporter les indicateurs caractérisés par des activités/motifs (dans cette figure : domicile, loisirs, courses, études, travail, autres) ou modes de transport (voiture, vélo transport en commun, marche à pied.)

tiples écrans ; le second développé en Python, utilise la bibliothèque OpenGL pour générer le graphique 3D du cube spatio-temporel. Le côté client est développé en Javascript et utilise les bibliothèques D3 (Data-Driven Documents) et Leaflet, pour générer, respectivement, les graphiques et les représentations cartographiques en 2D. Les données sont stockées localement sous la forme de fichiers *geojson* pour les informations spatiales, *json* pour les données de flux, et *csv* pour les autres. Les données sont pré-traitées par des scripts R qui utilisent les packages *Flows* et *TraMineR*. Ce dernier sert à créer les typologies de séquences grâce à l'algorithme *Dynamic Hamming Distance*.

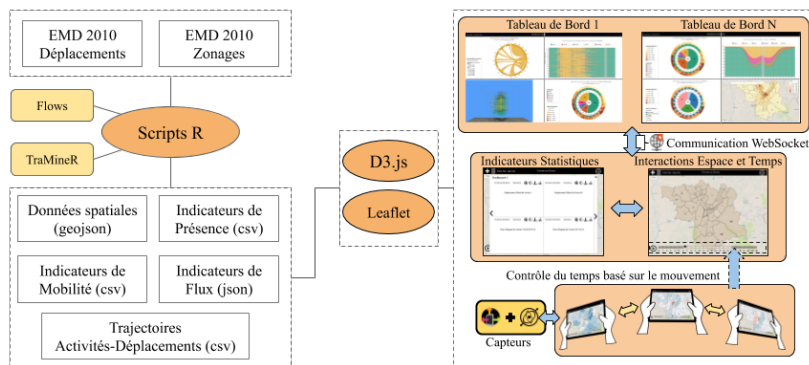


FIGURE (2) Architecture fonctionnelle de l'environnement eSTIME.

5. Résultats des expérimentations

Tout au long du développement d'eSTIME, des expériences ont été menées pour évaluer son ergonomie et son efficacité en termes d'analyse. Les expérimentations ont fait l'objet d'une analyse quantitative et qualitative¹. Nous synthétisons ici les résultats issus de trois expérimentations :

- la première s'appuie sur une analyse qualitative de retours d'expérience effectuée auprès de 13 sujets. Elle a porté sur l'appropriation générale de l'environnement et sur l'utilisation du mouvement pour contrôler l'animation spatio-temporelle, à travers la timeline ou la timewheel. D'une façon générale, l'ergonomie de l'interface a été considérée comme «moyenne» par les sujets, avec une charge cognitive relativement importante. En revanche, les sujets ont répondu correctement aux questions analytiques (précision de 80 pourcent). Lors de cette expérimentation, le contrôle du temps par la timeline a été jugé plus facile qu'avec la timewheel.

- la seconde, réalisée auprès de 17 sujets a consisté à évaluer l'usage du mouvement pour interagir avec le temps (via la tablette). Au cours de cette expérimentation, nous avons comparé quatre techniques d'animation du temps :

1. (Menin *et al.*, 2019) présente la procédure et les résultats en détails.

l'animation interactive traditionnelle contrôlée à la fois par la souris et à la fois par des touches sur l'écran; l'interaction via l'inclinaison de la tablette, à travers le mouvement soit cyclique, soit linéaire. Même si le mouvement cyclique a eu des effets négatifs sur l'exactitude des réponses, l'animation par le mouvement linéaire a été aussi efficace que l'animation interactive classique et les utilisateurs l'ont appréciée.

– la troisième a fait évalué l'efficacité de l'outil à des fins d'analyse de la mobilité par 21 sujets experts du domaine. Cette expérience est basée sur une série de questions analytiques auxquelles les sujets devaient répondre après exploration des données. Les réponses aux questions posées ont été satisfaisantes (avec un score de 80 pourcent de précision). Les utilisateurs ont souligné que l'utilisation de la tablette comme deuxième écran rendait la tâche plus facile. Dans la mesure où il est possible de visualiser les indicateurs sur un écran et interagir sur un autre (en l'occurrence ici la tablette), les utilisateurs ont fait valoir que ce dispositif contribuait positivement à l'analyse visuelle et qu'il pouvait éventuellement réduire les efforts physiques requis autrement pour manipuler la souris.

Globalement, notre approche a été jugée positivement car elle offre une diversité d'analyses visuelles des données de mobilité : les participants (y compris des non-experts) ont été capables de répondre aux questions analytiques simples (ex.: quels sont les secteurs qui attirent de la population?) et complexes (ex.: pour les secteurs les plus attractifs, d'où viennent les personnes sur les périodes de début de journée, de la pause méridienne et de fin de journée?).

6. Conclusion

Nous avons présenté l'approche conceptuelle eSTIME et son implémentation à partir des données d'enquête ménages déplacements. Notre proposition est basée sur le concept de tableau de bord modulable déployé sur un dispositif multi-écrans connecté à une tablette interactive sur laquelle sont déportés les processus d'animation cartographique pour interagir avec la dimension spatio-temporelle de la mobilité. Notre proposition répond à la volonté de développer les capacités d'analyse exploratoire multi-points de vue des environnements de géovisualisation. L'originalité de notre proposition se situe aussi dans l'intégration au sein d'un même environnement de représentations permettant l'analyse à la fois au niveau agrégé (cartes, roue de la mobilité, diagramme de flux, chronogramme) et au niveau individuel (actogramme, cube spatio-temporel). Certaines de ces représentations sont issues d'une adaptation de visualisations existantes tels que le diagramme de flux et le cube spatio-temporel; d'autres sont innovantes, telle que la roue de la mobilité. Enfin, l'ergonomie et l'efficacité de l'outil ont fait l'objet d'expérimentations auprès d'utilisateurs montrant le potentiel de l'approche eSTIME.

7. Remerciement

Ces travaux ont été bénéficié d'un financement doctoral de la Région Auvergne Rhône-Alpes.

Bibliographie

- André-Poyaud I. I., Chardonnel S., Charleux L. L., Tabaka K. (2008). La mobilité au cœur des emplois du temps des citoyens. In F. P. Yves Chalas (Ed.), *La mobilité qui fait la ville*, p. 67-95. CERTU.
- Armoogum J., Guilloux T., Richer C. (2015). *Mobilité en transitions : Connaître, comprendre et représenter*. CEREMA.
- Arvola M., Holm A. (2014). Device-orientation is more engaging than drag (at least in mobile computing). In *Proceedings of the 8th nordic conference on human-computer interaction: Fun, fast, foundational*, p. 939-942.
- Bahoken F. (2016). *Mapping flow matrices, a contribution*. Theses, Université Paris Diderot (Paris 7).
- CEREMA A.-C. (2010). *Enquête Ménages Déplacements (EMD), Grenoble, Grande région grenobloise*.
- Chardonnel S., Charleux L. L., Thibault P. (2010). Analyser les routines dans les emplois du temps par la mesure des concordances d'actogrammes. In *Mobilités urbaines et risques des transports : approches géographiques*, p. 23-50.
- Cochey E., Tabaka K. (2007). Modes de représentation des trajectoires quotidiennes des habitants. In *Huitièmes Rencontres de Théo Quant*.
- Cook K. A., Thomas J. J. (2005). *Illuminating the path: The research and development agenda for visual analytics*. Rapport technique. Pacific Northwest National Lab.(PNNL), Richland, WA (United States).
- Edsall R., 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*, p. 182-189.
- Hurez C., Pélatà J. (2016). *À la recherche du territoire «pratiqué»*. <http://mappemonde.mgm.fr/118gv7/>. (Accédé le 9 mai 2019)
- Kaddouri L., Blaise J.-Y., Davoine P.-A., Mathian H., Saint-Marc C. (2014). *État des lieux des représentations dynamiques des temporalités des territoires*. Research Report.
- Kister U., Klamka K., Tominski C., Dachselt R. (2017). Grasp: Combining spatially-aware mobile devices and a display wall for graph visualization and interaction. *Computer Graphics Forum*, vol. 36, n° 3, p. 503-514.
- Kitchin R., Lauriault T. P., McArdle G. (2015). Smart cities and the politics of urban data. In *Smart urbanism*, p. 32-49.
- Kraak M.-J. (2003). The space-time cube revisited from a geovisualization perspective. In *The 21st International Cartographic Conference*.

- Langner R., Kister U., Dachselt R. (2019). Multiple coordinated views at large displays for multiple users: Empirical findings on user behavior, movements, and distances. *IEEE Trans. on Visualization and Computer Graphics*, p. 608-618.
- Le Roux G., Vallée J., Commenges H. (2017). Social segregation around the clock in the Paris region (France). *Journal of Transport Geography*, vol. 59, p. 134-145.
- Menin A., Chardonnel S., Davoine P.-A., Nedel L. (2018). Exploring Shifting Densities through a Movement-based Cartographic Interface (Short Paper). In *10th International Conference on Geographic Information Science (GIScience'18)*.
- Menin A., Chardonnel S., Davoine P.-A., Nedel L. (2019). eSTIME: Towards an All-in-One Geovisualization Environment for Daily Mobility Analysis. In *Proceedings of the 32nd Conference on Graphics, Patterns and Images (SIBGRAPI'19)*.
- Otten H., Marian D., Boris M. (2015). Are there networks in maps? An experimental visualization of personal movement data. In *Personal Visualization: Exploring Data in Everyday Life*, p. 5-7.
- Pagno B., Costa D., Guedes L., Freitas C. D. S., Nedel L. (2015). Guidelines for designing dynamic applications with second screen. In *2015 XVII Symposium on Virtual and Augmented Reality*, p. 42-51.
- 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*, vol. 84, n° 3, p. 441-461.
- Pistre P., Commenges H., Guerrero D., Proulhac L. (2015). Operational Definitions of Time for Longitudinal Data Analysis: Illustration in the Field of Spatial Mobilities. *Nouvelles Perspectives en Sciences Sociales*, vol. 10, n° 2, p. 199-236.
- Rivard K., Cogswell D. (2004). Are you drowning in BI reports? Using analytical dashboards to cut through the clutter. *Information Management*, p. 26.
- Robette N. (2011). *Explorer et décrire les parcours de vie: les typologies de trajectoires*. CEPED.
- Saint-Marc C. (2017). *Formalization and geovisualization of historical natural risk events to understand spatial dynamics*. Theses, Université Grenoble-Alpes.
- Shi L., Jiang T., Zhao Y., Zhang X., Lu Y. (2017). UrbanFACET: Visually Profiling Cities from Mobile Device Recorded Movement Data of Millions of City Residents.
- Shi X., Yu Z., Chen J., Xu H., Lin F. (2017). The visual analysis of flow pattern for public bicycle system. *Journal of Visual Language and Computing*, p. 1-9.
- Thudt A., Baur D., Carpendale S. (2013). Visits : A Spatiotemporal Visualization of Location Histories. *EuroVis*, p. 2312.
- Tominski C., Schumann H., Andrienko G., Andrienko N. (2012). Stacking-based visualization of trajectory attribute data. *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, n° 12, p. 2565-2574.
- Yan Y., Tao Y., Xu J., Ren S., Lin H. (2018). Visual analytics of bike-sharing data based on tensor factorization. *Journal of Visualization*, vol. 21, n° 3, p. 495-509.
- Zhao J., Forer P., Harvey A. S. (2008). Activities, ringmaps and geovisualization of large human movement fields. *Information Visualization*, p. 198-209.

Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière

Comprendre, cartographier et explorer le passé pour maîtriser son environnement

Meddy Escuriet¹

1. UMR Territoires, Université Clermont-Auvergne,
49 boulevard François Mitterrand, Clermont-Ferrand, France
escuriet.meddy@ladapt.net

RESUME. L'idée d'organiser des semaines thématiques autour de la géographie pour des personnes en situation de handicap usagères d'un Etablissement et Service d'Aide par le Travail (ESAT) est née d'une rencontre entre deux univers professionnels différents : celui de la recherche en géographie et celui de l'accompagnement médico-social.

Se trouver dans une position de chercheur en géographie en Convention Industrielle de Formation par la Recherche (CIFRe) dans une structure médico-sociale (association LADAPT) nécessite de faire ses preuves vis-à-vis de collègues spécialistes du handicap et des pathologies représentées afin de construire sa légitimité à intervenir auprès de personnes considérées comme fragiles et vulnérables.

Une fois cette étape dépassée, la position de chercheur en immersion comporte de nombreux avantages. D'un point de vue organisationnel, la présence sur site rend par exemple simple l'accès au terrain en facilitant la mise en place des enquêtes. L'intégration à une équipe pluridisciplinaire de professionnels œuvrant au quotidien pour l'accompagnement de la personne permet de tisser de véritables relations de confiance avec les usagers, d'accéder à une connaissance du fonctionnement médico-social depuis l'intérieur. La prise de conscience des problématiques qui animent la structure et qui touchent les personnes en situation de handicap accompagnées stimule la volonté d'action et le désir de proposer des innovations, d'être utile au fonctionnement quotidien de l'association.

Cet article se propose de revenir sur la mise en place d'un ensemble d'ateliers thématiques autour de la géographie organisés pour les usagers de l'ESAT Hors-Murs de LADAPT Puy-de-Dôme.

Ces séances ont été conçues pour être aussi bien outils de recueil de données de recherche qu'outil d'accompagnement médico-social. Organisées suite au déménagement de la structure dans de nouveaux locaux, elles poursuivent l'objectif de stimuler leur connaissance et leur appropriation de ce nouvel environnement, rendue difficile par les problématiques cognitives auxquelles elles sont soumises.

SAGEO'2019 – Clermont-Ferrand, 13-15 novembre 2019

SAGEO'2019

Dans cet article, il sera plus particulièrement question de la troisième semaine d'activité, élaborée autour d'une approche géo-historique. En questionnant le passé et les évolutions de la ville de Clermont-Ferrand et du quartier du parc technologique la Pardieu dans lequel est implantée LADAPT, l'histoire a été mobilisée d'une part comme clé de lecture de l'espace et d'autre part comme vecteur de connaissance et d'appropriation du nouvel environnement

ABSTRACT. *The idea of holding thematic weeks about geography for people with disabilities working in an ESAT (a centre providing care through employment) was born from an encounter between two different professional worlds: geography research and medicosocial support.*

Being a researcher in geography integrated into a medicosocial structure in the LADAPT association requires to measure up to colleagues specialised in disability and different pathologies, in order to build legitimacy to intervene with vulnerable people

After this first stage, the immersion for the researcher has many advantages. The location on site facilitate field surveys. The integration into a multidisciplinary team of professionals supporting people daily enables to build relationships of trust with users but also to access knowledge of medicosocial functioning from within. The awareness of the issues animating the structure and affecting the persons being cared stimulates the will to action and the desire to propose innovations, to be useful to the daily functioning of the association.

This article aims to review the implementation of thematic workshops about geography for the users of LADAPT.

These sessions were designed to be both research data collection tools and medico-social support tools. Organised after the relocation of the structure, the purpose is to stimulate the knowledge and the appropriation of a new environment for the people suffering with cognitive problems.

In this article, we will focus on the third week of activity, based on a geo-historical approach. By questioning the past and the evolutions of the city Clermont-Ferrand and the technology park La Pardieu, where LADAPT is located. Firstly, History has been used as a key to read space and secondly as a vector of knowledge and appropriation with the new environment.

MOTS-CLES. *Cartographie, Approche géo-historique, Accompagnement médico-social*

KEYWORDS. *Cartography, geo-historical approach, medical and social support*

1. Introduction : des activités géographiques pour connaître et s'approprier le parc technologique

L'association LADAPT accompagne vers l'insertion sociale et professionnelle des personnes victimes de lésions cérébrales acquises par l'intermédiaire de plusieurs services médico-sociaux et notamment d'un *Établissement et Service d'Aide par le Travail (ESAT)*. Une lésion cérébrale est une destruction des tissus nerveux du

Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière

cerveau, on la qualifie d'« acquise » quand elle n'est pas liée à une affection héréditaire, congénitale ou dégénérative et qu'elle n'est pas causée par la grossesse ou l'accouchement². La survenue d'une lésion cérébrale acquise dans la vie d'une personne peut avoir des causes multiples, souvent brutales. Elle peut se produire, par exemple, à la suite d'un traumatisme crânien provoqué par un accident de la route, ou d'une embolie cérébrale débouchant sur un accident vasculaire cérébral (AVC). (Escuriet, 2018).

Afin de permettre l'insertion sociale et professionnelle l'ESAT de LADAPT Puy-de-Dôme, à Clermont-Ferrand propose un parcours personnalisé en milieu « ordinaire » aux usagers par l'intermédiaire de stages et d'expériences de travail en immersion, dans des entreprises partenaires. Comme ils sont en contrat à l'extérieur, il n'y a pas d'ateliers de production sur le site de l'association, c'est pourquoi le modèle est qualifié « d'hors-murs ». Les usagers se rendent dans les locaux pour bénéficier d'un accompagnement médico-social assuré par les divers professionnels accompagnants (ergothérapeutes, assistant social, monitrice d'atelier, neuropsychologue, chargé d'insertion en entreprise...) en vue de leur intégration en entreprise.

Les personnes ayant subi une lésion cérébrale, éprouvent des difficultés à organiser mentalement les images qu'elles ont de leur environnement et à les structurer en un système cohérent (Lynch, 1969). Ces difficultés compliquent leur orientation et leur prise de repères dans l'espace, elles interviennent surtout lors des phases de planification, d'organisation et de réalisation de la mobilité et affectent, de ce fait, leurs rapports spatiaux. Guy Di Méo définit les rapports spatiaux comme étant les « liens affectifs, fonctionnels et économiques, politiques et juridiques ou purement imaginaires que les individus ou les groupes tissent avec les espaces géographiques où ils vivent, qu'ils parcourent ou qu'ils se représentent » (Di Méo, 2008, p. 1-2).

Les rapports spatiaux des usagers de l'ESAT ont été chamboulées en septembre 2018 suite au déménagement de l'association dans de nouveaux locaux, situés dans un autre quartier de Clermont-Ferrand. De cet événement imprévu, impliquant une réorganisation de leurs habitudes spatiales, est née l'opportunité de mettre en place des ateliers hybrides construits autour de la géographie, aussi bien utiles au recueil de données scientifiques qu'à l'accompagnement médico-social.

La réalisation d'une thèse CIFRe immersive nécessite une vigilance de la part du doctorant qui se situe « dans un double contexte académique et professionnel qui entraîne autant de responsabilités que de conflits » (Rouchi, 2018). Cet article permettra d'illustrer une tentative de mise en dialogue du double contexte qui incombe au doctorant intégré à une structure d'accueil. Nous verrons que la position du chercheur et son immersion au cœur d'une structure médico-sociale n'est ici pas

² Définition issue de la Brain Injury Association of America. Causes of Acquired Brain Injury. (en anglais) Disponible à : <http://www.biausa.org>.

envisagée comme un outil de recueil à part entière. Cependant, le partage, l'échange entre les univers scientifiques et professionnels ont permis la genèse d'une forme d'outil hybride, aussi bien outil de recueil de données qu'outil d'accompagnement médico-social. L'hypothèse qui a conduit à la mise en place de ces ateliers était la suivante : le déménagement de l'association dans un nouveau quartier nécessite un temps d'adaptation et d'appropriation pour des usagers ayant des difficultés pour organiser et structurer leur image mentale de l'espace. Il est possible de raccourcir ce temps et de stimuler leur rapport à l'espace en mettant en place des ateliers thématiques autour de la connaissance du quartier, tout en étant utiles à leur accompagnement médico-social.

Après avoir recueilli et évalué lors de la première semaine d'activités, leurs représentations collectives du quartier, différents ateliers ont été organisés afin de renforcer d'une part leur rapport à l'espace, leur connaissance du quartier et d'autre part leur capacité de travail en entreprise. Conçue comme une introduction aux outils SIG adaptée à des personnes en situation de handicap cognitif, la troisième semaine d'activité fera l'objet d'une présentation plus détaillée. Nous verrons que l'approche par les changements géographiques et paysagers du quartier, induits par son histoire ont constitué en quelque sorte un prétexte pour pousser les usagers de l'ESAT à se déplacer dans leur nouvel environnement, afin qu'ils l'intègrent mentalement, plus rapidement.

2. Le déménagement de l'association dans de nouveaux locaux : un nouvel environnement à connaître et à s'approprier

Depuis 2009, date de sa création, LADAPT Puy-de-Dôme était établi avenue de la Margeride, au sud de Clermont-Ferrand. Le déménagement de l'association dans les locaux actuels du Parc technologique La Pardieu en septembre 2018 a modifié les habitudes de vie et les habitudes spatiales des personnes accompagnées devant se rendre régulièrement dans ce nouveau lieu. Même si les locaux actuels sont situés à quelques kilomètres des anciens, s'y rendre implique une organisation différente. L'itinéraire à emprunter depuis le domicile de ces personnes habitant pour la plupart l'agglomération clermontoise est désormais différent. Aller dans les locaux de l'association peut nécessiter d'avoir recours à un autre type de moyen de transport ou d'emprunter une autre ligne de bus desservant le quartier. Pour une personne non touchée par les problématiques cognitives, réorganiser ses déplacements habituels ne pose pas forcément de soucis. Pour celles qui ont subi une lésion cérébrale cela nécessite une attention toute particulière, se repérer et maîtriser un nouveau trajet ne va pas de soi et peut être difficile.

L'architecte américain Kevin Lynch dans une annexe de son travail sur l'image de la cité (Lynch, 1969) illustre l'impact des problématiques cognitives des personnes cérébrolésées sur la perception de l'espace. Lynch nous expose le « cas d'hommes qui, du fait d'une blessure au cerveau, sont incapables d'organiser leur environnement » (Lynch, 1969, p 147). Pour lui « ils peuvent se montrer capables de parler et de penser de manière rationnelle et même de reconnaître sans difficultés les objets, mais ils ne peuvent pas structurer les images en un système cohérent ». De ce fait, il peut être très difficile pour une personne ayant subi une lésion cérébrale de se

Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière

rendre en autonomie dans un quartier qu'elle ne connaît pas. De plus, un déplacement non maîtrisé, le sentiment d'être perdu, de ne pas savoir où aller, une image mal définie et mal construite entraîne souvent des sentiments de stress et d'angoisse très prégnants, Lynch expliquant qu'une image claire est à la base du développement individuel et que « celui qui possède une bonne image de son environnement en tire une bonne sécurité émotionnelle » (Lynch, 1969, p 5).

Les différents entretiens³ réalisés individuellement avec les usagers de l'ESAT ont fait émerger le fait que la majorité d'entre eux ne se sont pas rendus seuls la première fois dans les nouveaux locaux de l'association. Une demi-journée de visite a, en effet, été prévue par LADAPT avant l'emménagement.

Pour se rendre dans des lieux qu'ils n'ont jamais fréquentés auparavant, ils se doivent d'être très attentif à la planification de ces déplacements inédits, pouvant générer beaucoup de stress.

Une personne a par exemple expliqué que, depuis son accident elle est contrainte de préparer scrupuleusement tous les déplacements dans les lieux où elle n'est jamais allée : « *juste après mon accident, il a fallu que je me mette à préparer tous les déplacements et à noter plein d'éléments pour pouvoir me repérer* ».

Un autre usager a exprimé le fait que ses problèmes de mémoire l'obligeaient : « *à avoir une stratégie, si on me dit rend toi avenue X, si je ne connais pas l'avenue X à ce moment-là je me dit où est-ce-que-c'est ? le stress commence à faire son apparition* ».

Au contraire d'une personne n'ayant pas de difficultés cognitives, le fait de se rendre seulement une fois dans un lieu inconnu auparavant, n'entraîne pas une connaissance et une maîtrise pleine et entière du déplacement pour un individu ayant subi une lésion cérébrale acquise. En effet, les problèmes de structuration et d'organisation des images mentales soulevées par Lynch font, que souvent, seule la répétition des déplacements favorise la prise de points de repères spatiaux et la mise en contexte spatial de ces points de repères les uns par rapport aux autres. En répétant plusieurs fois au cours du temps un déplacement, le sentiment d'être perdu, le stress et l'angoisse va progressivement s'atténuer : « *Au début j'utilisais un plan, ou bien c'est déjà arrivé que mes parents m'amènent [...] maintenant je n'ai pas spécialement besoin de préparer car je connais le trajet par cœur ça se fait tout seul* ».

Tous ces éléments, nous laissent à penser que le déménagement de l'association dans de nouveaux locaux a redéfini le rapport à l'espace des usagers, aussi bien dans sa dimension affective que fonctionnelle. Pour tenter de confirmer cette hypothèse, une évaluation de leur représentation collective du quartier a été menée, en préalable

³ Au cours de la thèse, 16 usagers de l'ESAT Hors-Murs de LADAPT ont été interrogés dans le cadre des entretiens cartographiques, outil d'objectivation des rapports spatiaux (fréquentations spatiales, mobilité)

à la mise en place des ateliers, cette phase de diagnostic a donné lieu à un Jeu de Reconstruction Spatiale.

3. Les activités géographiques : outil de recueil de données, outil d'accompagnement

Dans le but d'interroger leur représentation d'un espace proche (Pontagnier et Rouker, 2016) il a été proposé aux usagers de créer collectivement une carte du quartier du Parc Technologique La Pardieu. Un Jeu de Reconstruction Spatiale (JRS) a été ainsi organisé.

Sept personnes ont réalisé l'exercice. Divisés en deux groupes (un groupe de trois et un groupe de quatre) ils avaient pour mission de créer une carte du quartier en disposant des carrés de mosaïque (éléments surfaciques et ponctuels) et des fils de laine (éléments linaires) sur un support vierge.

L'objectif scientifique était ici d'accéder à leur représentation collective du quartier pour comprendre sa structure et identifier les points de repères spatiaux structurant leur représentation. Alors que le Jeu de Reconstruction Spatiale tel qu'il a été expérimenté (Ramadier et Bronner, 2006 ; Ramadier et Depeau, 2006 ; Guy et Depeau, 2014 ; Tratnjek, 2015) est un exercice individuel, le parti pris a été de l'organiser en collectif, afin de le mettre au service de l'accompagnement vers l'emploi, dans le but de stimuler leur capacité à travailler en équipe.

Pour cela, des consignes précises guidaient la réalisation cartographique. L'obligation de se concerter au sein du groupe pour que la nature et la localisation des éléments sur la carte aient fait l'objet d'un consensus était imposée.

A l'issue de la réalisation des cartes, une analyse des éléments géographiques présents et non présents, de leur localisation, de leur disposition ou encore de l'échelle a été menée, en collaboration avec les usagers qui en ont profité pour livrer leurs impressions sur l'exercice et leurs commentaires sur les cartes.

Les deux représentations cartographiques (groupe 1 et groupe 2) faisaient état d'une connaissance du quartier géographiquement disparate

Ce sont globalement les secteurs périphériques du parc technologique (patinoire, centre commercial, gare ...) qui ont fait l'objet du plus grand nombre d'éléments indiqués et nommés. Les entretiens ont fait émerger que les lieux en bordure des axes de communications étaient connus et fréquentés par les usagers avant même que l'association ne déménage. En effet, toutes les personnes qui se sont prêtées à cet exercice vivent dans l'agglomération de Clermont-Ferrand depuis au moins 5 ans (une seule personne sur les sept n'est pas originaire de la région). A l'inverse, au cœur du parc technologique, peu d'éléments sont identifiés, s'ils le sont, peu sont nommés. Seul le lycée Lafayette avec son architecture circulaire, véritable point de repère spatial aux dires des usagers et la structure de l'avenue et des rues principales du parc ont été matérialisées et nommés, quel que soit le groupe. Ces constats vont dans le sens de l'hypothèse énoncée au départ : les usagers ont une représentation

Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière

collective du cœur du quartier qui est lacunaire, un temps d'adaptation et d'appropriation est nécessaire.

Plusieurs facteurs permettent d'expliquer cette méconnaissance. D'une part, le quartier, au contraire de sa périphérie n'était pas un quartier fréquenté pour la plupart des usagers avant que l'association ne s'y implante. D'autre part, tous les bâtiments sont des locaux de bureaux identiques, d'un ou deux étages construits à la fin des années 1980. Au cœur du quartier, aucun élément architectural ne se détache et ne fournit de point de repères sur lequel s'appuyer (Lynch, 1969).

Toujours en s'intéressant aux points de repères spatiaux, il s'agissait, lors de la seconde semaine d'ateliers, de se pencher sur la manière dont les sens entrent dans le processus de construction de l'image mentale. Elle a été organisée autour du thème de la perception, la fonction selon laquelle l'esprit, le sujet, se représente, pose devant lui les objets (Dupré, 2006). A l'issue de différents trajets imposés au sein du quartier, les usagers avaient pour mission d'indiquer sur un fond de carte individuel tout ce qu'ils avaient « senti » lors du déplacement. Comme nous l'explique Denis Martouzet, « sentir renvoie d'abord, spécifiquement à l'odorat mais aussi, plus largement, à la sensorialité, la capacité qu'ont nos sens de nous approcher du réel qui nous entoure, et à la sensibilité » (Martouzet, 2013, p 7).

L'objectif géographique ici, était toujours de s'intéresser aux points de repères spatiaux pris par les usagers tout en encourageant par la déambulation dans le quartier, la prise de nouveaux points de repères. En leur demandant d'être attentif à tout ce que leur sens leurs renvoyaient, l'accent n'a pas seulement été mis sur les points de repères visuels présents dans le parc technologique. Ainsi, certaines cartes ont mis en évidence le bruit du vent dans les arbres qui jalonnent le quartier tandis que d'autres ont fait apparaître la sensation éprouvée quand on marche dans les pelouses bordant les bâtiments.

La restitution individuelle des cartes réalisés par chacun, a permis de mettre l'accent sur le fait que chaque personne est différente, que chacun est porteur de sa propre sensibilité et qu'il est nécessaire d'en tenir compte quand on travaille en équipe. Tout comme pour le jeu de reconstruction spatiale, cette forme de cartographie sensible (Olmedo, 2011 ; Olmedo, 2015a ; Mekdjian et Olmedo, 2016) visant à faire émerger la subjectivité individuelle, s'est mise au service de l'accompagnement médico-social.

4. L'histoire mobilisée comme clé de lecture de l'espace : comprendre le passé de l'espace pour le connaître et se l'approprier

4.1 Des photographies aériennes comme moyen d'accès à une vision diachronique du quartier de la Pardieu

Après deux semaines d'ateliers sur l'état des lieux de la connaissance du quartier, la prise de repères spatiaux et la perception d'une manière générale, la troisième

semaine d'activité d'accompagnement s'articulait autour des changements géographiques et paysagers induits par le temps qui passe.

Par le biais de l'histoire du quartier, replacée dans le contexte global de l'histoire de Clermont-Ferrand, il s'agissait de retracer la genèse du parc technologique la pardieu pour en clarifier la structure et la physionomie actuelle. A l'image des ateliers précédents, ces activités ont induit des déplacements dans le quartier toujours dans l'objectif de stimuler sa connaissance.

D'un point de vue de l'accompagnement médico-social, il s'agissait ici de stimuler le sens de l'observation des usagers de l'ESAT, de les inciter à s'exprimer avec un vocabulaire géographique et de leur proposer une introduction à l'utilisation des SIG.

En préambule de l'atelier, plusieurs photographies aériennes du quartier ont été distribuées. Extraites du site *remonter le temps* de l'IGN⁴, elles s'échelonnent sur une amplitude allant de 66 ans⁵ (de 1947 à 2013). Il a été demandé aux usagers de classer les neuf photographies aériennes de la plus ancienne à la plus récente.

La photographie de 1947 est structurée par trois éléments linéaires identifiables : la rivière Artière qui la traverse du sud vers le nord, la voie ferrée au nord et un axe qui croise la rivière parallèlement à la voie ferrée (il s'agit de l'actuel boulevard Robert Schuman et l'avenue Ernest Cristal). Au niveau du bâti, quelques exploitations agricoles sont disséminées un peu partout, la très faible taille des parcelles maillant le paysage nous donne une idée du morcellement du foncier. On peut noter la présence en bordure ouest de l'Artière (rive gauche) de petites constructions s'apparentant à des cabanes de jardin qui témoignent du passé maraîcher de la banlieue clermontoise. Sur les photographies de 1956 et de 1964, les saisons durant lesquelles les clichés ont été pris nous permettent de constater clairement la ripisylve⁶ de l'Artière. Les formations arborées alignées sur les parcelles nous confirment alors la présence de vergers en bordure de la rivière. La photographie de 1972, témoigne de l'avancée vers l'est qu'a connu la ville à cette époque. La patinoire a fait son apparition, si on change d'échelle on remarque que des entrepôts jalonnent l'actuelle avenue Gustave Flaubert, un des grands axes commerciaux de la ville aujourd'hui.

En 1974, l'entrepôt des transports urbains clermontois (T2c) est en construction, en 1985 on constate que des bus remplissent son parking. En 1987, même si subsistent les arbres qui bordaient son lit, l'Artière est désormais canalisée et passe sous le Parc technologique en construction. L'actuelle avenue Léonard de Vinci, où est implantée LADAPT, le principal axe du parc, prend place sur les anciens vergers,

⁴ <https://remonterletemps.ign.fr/>

⁵ Les neuf photographies distribuées ont été prises respectivement en 1947, 1956, 1964, 1972, 1974, 1985, 1987, 1999, 2013

⁶ « La ripisylve se définit comme un ensemble de groupements végétaux, la plupart étant dominés par une strate arborée, localisée sur la marge des cours d'eau et inféodée à des milieux régis par l'eau superficielle et souterraine » (Piégay, 1997).

Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière

les premières entreprises viennent de s'installer. En 1999, le contraste est saisissant, le quartier totalement urbanisé a pris sa physionomie actuelle, le lycée Lafayette trône à l'entrée sud et la flotte des bus rouges de la T2c remplace pour partie les anciens véhicules blancs présents sur la photo de 1985. En 2013, peu de changements mis à part la ligne de tramway qui a fait son apparition au nord du Parc technologique.

La seconde étape de cette expérimentation à partir des photographies aériennes consistait à placer sur chaque cliché l'emplacement des locaux de LADAPT. Cela a nécessité, de la part du groupe, d'approfondir la comparaison des images. Les locaux ont été matérialisés sans difficulté, d'abord sur les photos les plus récentes car le bâtiment où l'association est implantée existait. Sur les plus anciennes, la disparition des points de repères, du bâti, du réseau viaire a conduit les usagers à se baser sur les éléments structurants identifiés auparavant (la voie ferrée, la route (actuel boulevard Robert Schuman / avenue Ernest Cristal) ou encore l'Artière). L'emplacement des locaux a été localisé compte tenu de leur position entre la voie ferrée et le boulevard Robert Schuman/ avenue Ernest Cristal. La rivière, même si elle a été canalisée dans les années 1980, a permis au groupe d'affiner le positionnement de l'emplacement des locaux sur toutes les photographies les plus anciennes. En effet, les usagers ont remarqué que, même si la rivière n'existait plus ; certains arbres de sa ripisylve ont traversé les années. Le repérage du positionnement des locaux de l'association par rapport aux structures du paysage et aux résidus de la ripisylve sur les photos les plus récentes ont indicé sa transposition dans le passé, en plein cœur des vergers, au bord de l'Artière. En effet, on remarque dans le quartier, qu'ici ou là, au bord des parkings, le long des clôtures on rencontre encore des cerisiers, des noyers ou autres arbres fruitiers assez vieux au regard de la taille des troncs, trop vieux pour avoir été plantés dans les années 1980 en même temps que le quartier s'urbanisait. A ce stade de l'activité, le groupe a fait part de son étonnement, un usager a par exemple manifesté sa surprise que le parc technologique ait été construit sur d'anciens vergers :

« C'est impressionnant de voir à quel point tout a changé si vite, il y a peine 50 ans, nous aurions été au bord de l'eau au milieu des vergers, quand on voit la croissance de la ville ça fait réfléchir, on se demande s'il y aura encore de la nature dans 50 ans »

Le classement des photographies et la localisation de l'emplacement des locaux a donné aux usagers des clés de lecture du quartier dans le présent. En effet, ils ont pu retracer et comprendre les différentes étapes de son urbanisation, de la canalisation de l'Artière, à la construction des différents bâtiments, jusqu'à l'arrivée du tramway. Au-delà de ça, l'exercice a généré des discussions et des débats au sein du groupe. Ces interactions illustrent bien l'essence même de ces ateliers : outils de recueil de données scientifique au service de l'accompagnement médicosocial. Tout en apprenant de l'histoire du quartier, chacun a pu en améliorer sa connaissance,

comprendre sa genèse, sa structure, enrichir sa représentation de l'espace et travailler sa capacité à s'exprimer face au groupe, à défendre son point de vue, à s'écouter.

En invitant les usagers à se projeter dans le passé pour s'approprier le présent, un défi leur a été lancé en conclusion de la troisième semaine d'atelier : « Retrouver dans le quartier des indices permettant d'identifier précisément le tracé que prenait l'Artière avant que le quartier ne soit construit ».

4.2 En quête de l'Artière sauvage : se projeter dans le passé, initiation aux outils SIG

L'Artière est une rivière affluente de l'Allier. Prenant sa source sur la commune de Saint-Genès-Champanelle, elle sillonne sur environ 30 kilomètres le sud et l'est de l'agglomération de Clermont-Ferrand pour se jeter dans l'Allier aux Martres d'Artière.

Dans l'objectif de retrouver dans le présent et dans le quartier, sous sa physionomie actuelle, le chemin que prenait jadis la rivière sauvage, deux photos aériennes ont été sélectionnées et soumises aux usagers de l'ESAT : celles de 1985 et de 2013. Ces deux clichés n'ont pas été pris au hasard. Celui de 1985, en infrarouge couleur (IRC) technique photographique particulièrement adaptée au repérage de la couverture végétale, est le dernier sur lequel on peut voir l'Artière et sa ripisylve avant la construction des bâtiments. En effet, on voit clairement apparaître les travaux de canalisation de la rivière qui ne vont pas tarder à démarrer. Celui de 2013, le plus récent, nous donne une image du quartier très proche de la physionomie que le Parc technologique a actuellement.

Un exercice de transposition des éléments du passé dans le présent a tout d'abord été proposé aux usagers de l'ESAT : « Imaginez les bâtiments du quartier présents sur la photo de 2013 dans le paysage du quartier en 1985. Est-ce que selon vous, en 2013, il reste, dans le paysage du quartier, des arbres de la ripisylve de l'Artière ? »

Sur la base des éléments structurants communs aux différentes photographies mais également en s'aidant de l'emplacement des entrepôts de la T2c toujours présent aujourd'hui, les usagers ont identifié plusieurs groupes d'arbres qui pourraient appartenir à l'ancienne ripisylve de l'Artière. Avant d'aller sur le terrain vérifier ces hypothèses et suivre l'ancien cours de la rivière, une transposition des bâtiments du quartier sur la photographie de 1985 était nécessaire.

Sur un calque imprimé du cliché pris en 1985, l'empreinte des voies et du bâti du quartier actuel a été réalisée afin d'identifier la trajectoire prise par la ripisylve et le passage de l'Artière entre les bâtiments du présent.

Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière

Cette opération de transposition cartographique aurait pu être réalisée via des outils SIG. En effet, les photographies aériennes numériques utilisables via les SIG sont superposables à une carte si on leur applique une orthorectification qui consiste à recalculer la photographie, « pour la replacer dans la géométrie de projection choisie et la rendre ainsi superposable en tous points à une carte » (Muraz et al, 1999, p 40).

Pour que l'activité proposée soit accessible aux usagers de l'ESAT, ces principes de base de l'exploitation des photographies aériennes en SIG ont été exposés oralement.

Dans le croisement des informations géographiques de 1985 et de 2013, nous nous sommes limités à « une interprétation visuelle des photographies » couplée à un « report manuel sur un fond cartographique » (Muraz et al, 1999, p 39). Ce choix a permis une sensibilisation aux SIG sans rentrer directement dans la complexité des techniques cartographiques utilisées. Le recours au calque, réalisé manuellement, comportait un objectif pédagogique : celui d'illustrer concrètement le principe de superposition des couches que l'on rencontre dans les logiciels de traitement de l'information cartographique.

Pour permettre aux usagers de l'ESAT de consolider l'identification du tracé pris par l'Artière sauvage, un document supplémentaire leur a été donné : le plan cadastral du quartier. Il a été utile car le tracé de certaines parcelles où sont implantés les bâtiments du quartier épousent l'ancien lit de la rivière. La superposition du calque au cadastre a ainsi permis de tester la précision du tracé réalisé, et de l'affiner.

Lors de la déambulation réalisée dans le quartier, au fil de l'Artière oubliée, une dizaine d'arbres ont été identifiés comme pouvant potentiellement avoir été arrosés il y a quelques années par les eaux de cette petite rivière désormais canalisée.

5. Conclusion

Les semaines autour de la géographie ont été appréciées par les usagers : « *C'était très intéressant, ça nous a sorti de notre quotidien et permis de mieux se repérer dans le quartier, maintenant j'y ai mes repères, je peux m'y déplacer facilement, ce qui n'était pas le cas avant* ».

Même si cette personne souligne les effets bénéfiques de ces différentes activités géographiques sur sa capacité à se repérer dans le quartier, il n'est pas possible à l'heure actuelle de livrer un bilan complet de l'efficacité de ces ateliers. En effet, il est prévu de mettre en place un autre jeu de reconstruction spatiale afin d'interroger l'état des représentations collectives du quartier des usagers après qu'ils se soient livrés à ce type d'expérience. Seule la comparaison avec les cartes réalisées au départ nous permettra ou non de valider l'hypothèse initiale en constatant ou non d'avantage d'éléments localisés et nommés au cœur du quartier. Le Jeu de Reconstruction Spatiale n'a pas été reproduit directement à l'issue de la dernière

semaine d'atelier. En effet, compte tenu des difficultés de mémoire plus ou moins fortes en fonction des usagers, le choix a été fait d'attendre plusieurs mois avant de le reconduire afin de voir si les différentes capacités spatiales acquises sont ancrées mentalement par les usagers sur le long terme. A l'issue de cet ultime Jeu de Reconstruction Spatiale, des entretiens seront menés afin de comprendre dans quelle mesure ce sont les ateliers ou au contraire la pratique quotidienne du quartier qui a permis d'améliorer leur rapport à l'espace et leur représentations spatiales.

Imagines au départ comme un véritable outil de recueil de données scientifique sur la construction du rapport à l'espace et d'aide à la connaissance et à l'appropriation d'un nouvel environnement, les ateliers se sont avérés utiles au développement de compétences à faire valoir sur le chemin de l'insertion professionnelle.

Bibliographie

Di Méo G. (2008). La géographie culturelle : quelle approche sociale ? *Annales de géographie*, n°660-661, p. 47-66.

Dupré S. (2006). Perceptions et représentations géographiques : un outil pour aménager les forêts touristifiées ? *Téoros*. 53-61.

Escuriet M. (2018). Approche longitudinale et émotionnelle des déplacements, retour d'expérience sur une méthodologie appliquée aux limitations fonctionnelles et aux situations de handicap, *Communication au colloque MSFS Mobilités spatiales, méthodologies de collecte, d'analyse et de traitement*, 8 et 9 novembre 2018, UFR de droit économie et sciences sociales, Tours (37). URL : <https://hal.archives-ouvertes.fr/hal-01922839>

Gervais T. (2001). Un basculement du regard, *Études photographiques*. URL : <http://journals.openedition.org/etudesphotographiques/916>.

Guy F., Depeau S. (2014). Carte à la une : la carte mentale par le jeu pour comprendre l'espace vécu par des adolescents, *Géocofluences*. URL : <http://geoconfluences.ens-lyon.fr/informations-scientifiques/a-la-une/carte-a-la-une/carte-a-la-une-jrs>

Lynch K. (1969). *L'image de la cité*, Paris, Dunod, 222 p.

Martouzet D. (2013). Sentir et Ressentir la ville [Numéro thématique]. *Noroi*. URL : <https://journals.openedition.org/noroi/4633>

Muraz J., Durrieu S., Labbe S., Andreassian V., Tangara M. (1999). Comment valoriser les photos aériennes dans les SIG ? *Ingénieries - E A T, IRSTEA*, p. 39 - 58.

Un parc technologique, des entreprises, des bureaux ... et au milieu coulait une rivière

Olmedo E. (2011). Cartographie sensible, émotions et imaginaire, *Visions cartographiques*. URL : <http://blog.monediplo.net/2011-09-19-Cartographie-sensible-emotions-et-imaginaire>, article republié le 15 mars 2015 : <http://visionscarto.net/cartographie-sensible>

Olmedo E. (2015a). *Cartographie sensible. Tracer une géographie du vécu par la recherche-crédation*, Thèse de géographie sous la direction de Jean-Marc Besse, Paris, Université Paris I Panthéon Sorbonne, 516 p.

Olmedo E. (2015b). Représenter l'espace vécu. Les enjeux d'une cartographie émotionnelle en géographie, *séminaire géographie des émotions*, ENS-Ulm, 15 janvier 2015, compte-rendu par Célia Innocenti et Aurore Léon. URL : http://geographie.ens.fr/IMG/file/Géo%20des%20émotions/Geoemotions_CR%2015%20janvier.pdf

Olmedo E. (2016). Pour une cartographie émotionnelle des récits des femmes de Sidi Youssef Ben Ali, Marrakech, Maroc. In Fournier M (Dir.), *Cartographier les récits*, Clermont-Ferrand, Presses Universitaires Blaise Pascal, CERAMAC 35, Coll. « À la croisée des SHS », p 179-196.

Pontagnier F., Rouker A., (2016). Territoires Métropolitains : un projet éducation artistique et cartographique, *Comité Français de Cartographie*, n°229-230, Sept 2016, p 199-205.

Ramadier T., Bronner A-C. (2006). Knowledge of the environmental and spatial cognition: JRS as a technique for improving comparisons between social groups, *Environment and Planning B: Planning and Design*, n°33, pp. 285-299. URL : <http://www.envplan.com/abstract.cgi?id=b3248>

Ramadier T., Depeau S. (2006). Approche méthodologique (JRS) et développementale de la représentation de l'espace urbain quotidien de l'enfant, Actes du colloque *Les enfants et les jeunes dans les espaces du quotidien*, 16-17 novembre 2006, Rennes, 24 p.

Rouchi C. (2018). Une thèse CIFRE en collectivité territoriale : concilier la recherche et l'action ? *Carnets de géographes [En ligne]*.

Tratnjek, B. (2015). Représenter l'espace : de l'enfance et la jeunesse à la recherche, *Aggiornamento hist-géo*. URL : <https://aggiornamento.hypotheses.org/2611>

Contribution de la géomatique à l'observation des nids de tortues marines de Tetiaroa.

Benoît Stoll¹, Mehani Hart², Cécile Gaspar³

1. Laboratoire GEPASUD, Université de la Polynésie française

BP6570, 98702, Faa'a – Tahiti, Polynésie française

Benoit.stoll@upf.pf

2. hanihart@gmail.com

3. Association Te Mana O Te Moana, Moorea, Polynésie française

cecile.gaspar.moorea@gmail.com

RESUME. L'atoll de Tetiaroa est un haut lieu de la ponte des tortues vertes (Chelonia mydas) en Polynésie française, l'association Te Mana O Te Moana y effectue la surveillance des nids de tortues marines depuis 2007.

Onze années d'observations ont été intégrées dans un système d'information géographique (SIG) afin de mieux diffuser et analyser ces relevés.

Des traitements géomatiques ont permis de mettre en lumière les sites de pontes préférentiels et d'analyser l'influence de la distance des nids au récif sur le choix des sites de pontes.

Ce travail a aussi permis de proposer des pistes d'amélioration du protocole de collecte des informations sur le terrain afin d'augmenter la qualité des données recueillies et permettre de futurs traitements géomatiques pour une meilleure compréhension du comportement de nidification des tortues marines.

ABSTRACT. The Tetiaroa atoll is a high spot of marine turtle nesting in French Polynesia, Te Mana O Te Moana association is monitoring the turtles nests since 2007.

Eleven years of data collection has been included in a geographic information system in order to spread and analyze these data.

Some geomatic processings allowed enlightening the preferential nesting sites and analysing the influence of the distance between the nests locations and the reef barrier on the nesting site choice.

This work allowed advocating some data collection improvements in order to increase the data accuracy and allow new processings for a better nesting behavior understanding

MOTS-CLES : Atoll de Tetiaroa, Tortues marines, Nidification des tortues, Système d'information géographique, Géomatique.

KEYWORDS: Tetiaroa atoll, Marine turtles, Turtle nesting, Geographical information system, Geomatic.

1. Introduction

1.1. Contexte géographique et historique de l'atoll de Tetiaroa

L'île de Tetiaroa est le seul atoll des îles du vent dans l'archipel de la Société en Polynésie française. Cet atoll fermé qui possède 12 « *motu* » (îlots coralliens) est situé à 50km au nord de l'île de Tahiti par 17°00' Sud et 149°34' Ouest (Figure 1).

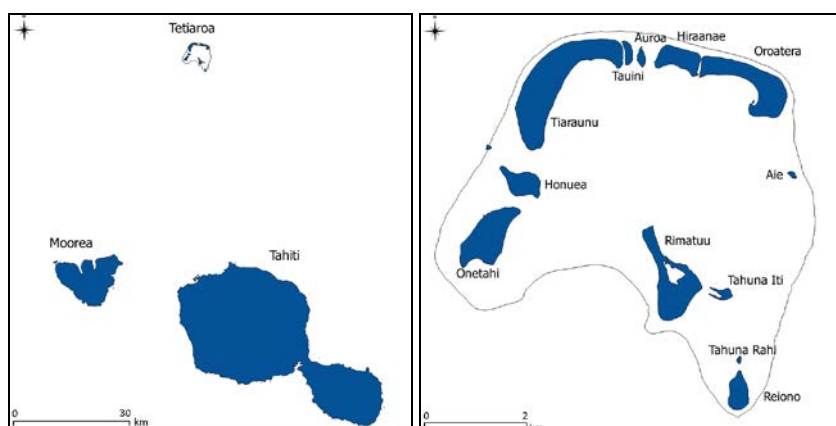


FIGURE 1. Atoll de Tetiaroa – Archipel de la société – Polynésie française

L'acteur Marlon Brando a découvert l'atoll en 1960 lors du tournage du film « La Bounty » et l'a acheté en 1967. Depuis et jusque tout récemment, l'activité anthropique sur l'atoll s'est réduite à l'exploitation d'un petit hôtel sur le *motu* Onetahi et à la présence de quelques pêcheurs venus de Tahiti, ce qui en a assuré la préservation.

En 2014, un nouvel hôtel de grand luxe « The Brando » a ouvert ses portes avec l'accord de Marlon Brando qui souhaitait que l'île devienne un modèle écologique, non seulement pour l'exploitation touristique mais aussi pour les générations futures, en mettant en avant la préservation de l'environnement, la recherche scientifique et l'éducation. La Tetiaroa Society (Tetiaroa_Society, 2016), organisation à but non lucratif est chargée de concrétiser ce vœu, l'éco-station sert de lieu d'accueil pour les équipes de recherche internationales.

1.2. Contexte scientifique

Le programme IDEA (Davies *et al.*, 2016) a pour mission de modéliser l'environnement des îles de Moorea et Tetiaroa. Dans ce cadre et en collaboration avec la Tetiaroa Society, l'université de la Polynésie française a commencé la mise en place d'un observatoire de données scientifiques et environnementales sur Tetiaroa (Stoll and Longine, 2019).

Plusieurs programmes scientifiques ont déjà commencé à produire des données qui sont, ou vont être, incluses dans cette base de données : citons en particulier le programme d'élimination des moustiques (Bossin *et al.*, 2016a, b; Stoll and Bossin, 2019), la cartographie exhaustive des sites archéologiques de Tetiaroa (Molle *et al.*, 2019), l'observation des tortues marines par l'association Te Mana O Te Moana (Petit *et al.*, 2016; Touron *et al.*, 2018) ainsi que l'analyse diachronique de la géomorphologie des motus de Tetiaroa (Stoll and Longine, 2019).

Nous expliciterons ici le programme scientifique de suivi de la ponte des tortues vertes de Tetiaroa ainsi que la constitution d'un fichier vectoriel spatialisé incluant, en données attributaires, toutes les observations sur les 11 saisons de pontes depuis 2007. Cette couche vectorielle a permis les premiers traitements géomatiques mettant en lumière les sites de pontes préférentiels ainsi que l'analyse de l'influence de la distance au récif sur le choix du site de ponte des tortues. Nous proposerons des améliorations sur le protocole terrain afin de permettre de nouveaux traitements géomatiques. Ceux-ci permettront la collecte optimale de nouvelles données et l'optimisation du travail de terrain des bénévoles de l'association. Nous terminerons cet article par des conclusions et perspectives.

2. Programme scientifique

2.1. Suivi des pontes des tortues marines

2.1.1 Collecte des données et cartographie

L'association Te Mana O Te Moana est chargée, entre autres, d'un programme de recensement, d'aménagement et de protection des sites de pontes de tortues marines sur l'atoll de Tetiaroa. Elle y mène tous les ans depuis 2007 des campagnes d'observation des pontes selon un protocole scientifique de relevé d'informations (Gaspar *et al.*, 2008; Petit, 2009; Petit *et al.*, 2013; Petit *et al.*, 2016).

Pendant la saison de pontes des tortues, de septembre à avril, les équipes de l'association arpentent les plages de Tetiaroa de jour comme de nuit à la recherche de traces de tortues (Figure 2). Lorsqu'une trace est découverte plusieurs informations sont relevées : les caractéristiques générales de l'observation (date, relevé GPS, observateur, ...), les caractéristiques de la trace et des creusages, les caractéristiques environnementales, le référencement des nids ainsi que des relevés biométriques sur la femelle lorsqu'elle est toujours présente (Touron *et al.*, 2017).



FIGURE 2. *Trace de montée de tortue (gauche) et nid de tortue (droite)*

Ces informations sont collectées dans un fichier Excel centralisant toutes les observations scientifiques associées à chaque trace, ou nid de tortue, découverts pendant l'année en cours.

La banque de données de l'association est donc constituée de 11 fichiers Excel correspondants aux 11 premières saisons de pontes suivies. Cette banque de données, même si elle permet d'effectuer un bilan annuel des pontes des tortues, ne permet ni la représentation cartographique, ni l'analyse multi-annuelle du comportement de ponte.

Puisqu'un relevé GPS a été effectué sur chaque trace ou nid de tortue, l'utilisation d'un outil SIG (système d'information géographique) s'est imposée. En effet, ceux-ci permettent de manipuler conjointement des données spatialisées de type raster (typiquement des fonds de carte d'images satellites), et des données vectorielles. Ces dernières associent des informations relatives à la géométrie (emplacement d'un point, d'une ligne ou d'un polygone) à d'autres types d'informations, dites données attributaires. De plus, les SIG permettent d'opérer des traitements géomatiques à des fins de représentation (cartes thématiques) ou d'analyse. Le choix d'un SIG libre (Qgis) s'est dès lors imposé. Outre sa gratuité, critère important pour une association, il est aussi d'accès relativement intuitif et aisé pour des membres d'associations souvent non-spécialiste des SIG.

L'ensemble de la banque de données Excel a été centralisée en un seul fichier, et nettoyée de toute erreur (mauvais report de localisation GPS, redondance ou manque d'information, incohérences flagrantes, ...). Ce fichier unique a été importé dans le SIG sous la forme d'un fichier vectoriel de type point, représentant l'emplacement des traces et des nids de tortues de 2007 à 2018. Chaque point étant associé, dans la table attributaire, à toutes les observations scientifiques collectées sur le terrain.

La couche vectorielle ainsi réalisée à partir du suivi de 2007 à 2018 permet d'obtenir la carte Figure 3 où la couleur des points représente la saison de découverte du nid. On y remarque que sur Tetiaroa, les tortues favorisent certaines plages plutôt que d'autres pour y creuser leur nid.

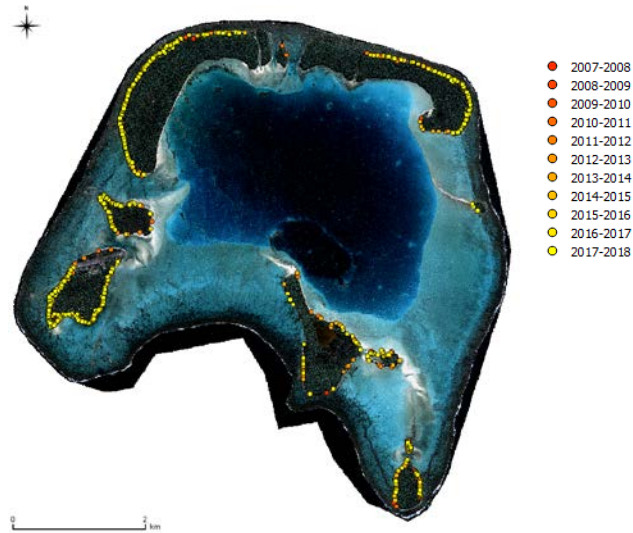


FIGURE 3. *Emplacement des nids de tortues de 2007 à 2018*

Même s'il a été montré que les femelles choisissent leur lieu de naissance comme site de ponte (Bass, 1999) et que la tortue verte marque une préférence pour les grandes étendues de sable nu (Whitmore and Dutton, 1985), et malgré de nombreuses études sur le sujet, les critères de choix des tortues demeurent néanmoins encore mal connus.

Les tortues marines ont un cycle de vie, une physiologie et des traits comportementaux fortement liés à la température de l'environnement (Petit *et al.*, 2013; Touron *et al.*, 2017). La sélection du site de ponte, soumise à différents paramètres (facteurs biotiques, abiotiques, pressions anthropiques), va par conséquent revêtir une importance significative sur le succès des pontes.

Dans la littérature, plusieurs critères environnementaux sont classiquement pris en compte, en voici une liste non-exhaustive : la distance du nid à l'eau (Ackerman, 1997), la nature du substrat (Mortimer, 1980; Cathy, 1994; Kikukawa *et al.*, 1999; Godley *et al.*, 2001; Karavas *et al.*, 2005), le couvert végétal (Janzen, 1994; Karavas *et al.*, 2005; Kamel, 2013), la pente de la plage (Horrocks and McA. Scott, 1991; Kikukawa *et al.*, 1999; Wood and Bjorndal, 2000), la présence d'obstacles (Witherington *et al.*, 2011b, a), la présence de lumières artificielles (Wood and Bjorndal, 2000).

Bénéficiant d'une base de données SIG sur une période relativement longue, nous pouvons envisager l'utilisation de différentes fonctions ou extensions du logiciel Qgis pour analyser la distribution spatiale des nids, et éventuellement calculer certains des paramètres environnementaux précédemment cités. Finalement,

le problème se résume à analyser un nuage de points (la position des nids) : (i) entre eux (densités, distances, ...), (ii) par rapport à des vecteurs (position du récif, de l'eau, de la végétation, ...) ou (iii) par rapport à des données raster (pente, ...). Nous développerons ici deux problématiques, la densité locale des nids et la distance au récif.

2.1.2 Densité locale des nids de tortues

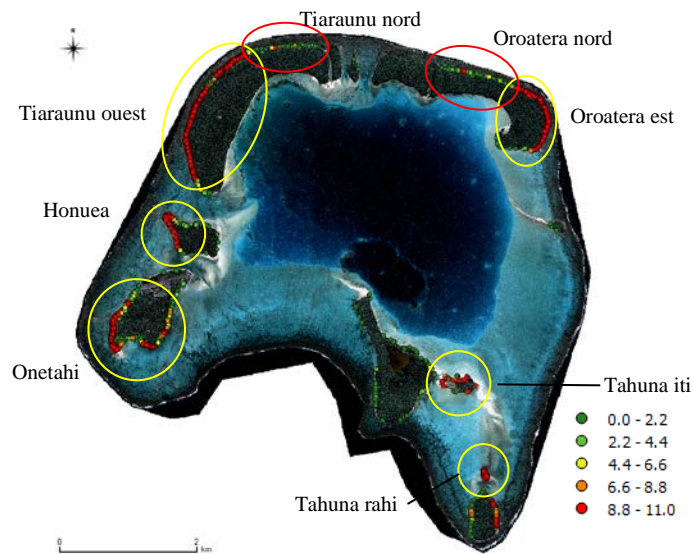


FIGURE 4. Densité locale à 30m des nids de 2007 à 2018 et zones préférentielles

Afin de connaître la position des sites de pontes préférentiels sur Tetiaroa, la densité locale des nids a été calculée dans un rayon de 30 mètres autour de chaque nid. La valeur du rayon a été déterminée suite au calcul de la matrice de distances entre l'ensemble des points et à une analyse de la distance moyenne entre les nids.

Les densités à 30 mètres vont de 0 à 11 nids si l'on considère la période 2007-2018 (Figure 4). Les sites préférentiels de ponte sur Tetiaroa sont situés sur les plages de Tiaraunu ouest, Honuea ouest, Onetahi sud, Oroatera est, Tahuna iti et Tahuna rahi (cercles jaune). Les résultats obtenus année par année n'ont montré aucun changement significatif sur la position des sites de pontes préférentiels, ceux-ci sont stables depuis 2007.

2.1.3 Distance des nids au récif

A la lecture de cette carte, on remarque que les sites affichant la plus forte densité semblent être situés près du récif. On peut donc se demander si la proximité de la plage par rapport au récif constitue un élément décisif dans le choix du site de

Contribution de la géomatique à l'observation des nids de tortues marines de Tetiaroa

ponte. Une étude portant sur les 11 saisons a été menée pour mettre en lumière cette éventuelle corrélation.

L'opération revient à calculer la distance minimale entre chaque point d'une couche vectorielle (l'emplacement de chaque nid) et l'ensemble des nœuds d'un vecteur ligne (le récif) (Stoll and Longine, 2019). Cela peut être effectué par l'utilisation d'extensions disponibles sur Qgis ou par le biais d'une matrice de distance entre les points représentant la position des nids et les nœuds extraits du vecteur ligne représentant le récif, la matrice est travaillée sous Excel pour associer à chaque nid la distance minimum au récif, et re-importé dans Qgis en effectuant une jointure attributaire.

La distance au récif de chaque nid trouvé entre 2007 et 2018 est ainsi comprise entre 76 et 1168 mètres (Figure 5).

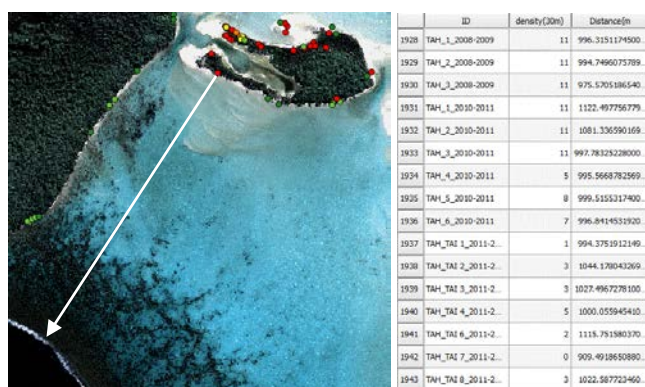


FIGURE 5. Distance au récif le plus proche et table attributaire (Tahuna Iti)

La Figure 6 représente la distance moyenne au récif de tous les nids dans les intervalles de densité à 30 mètres de [0-1], [2-3], [4-5], [6-7], [8-9] et [10-11].

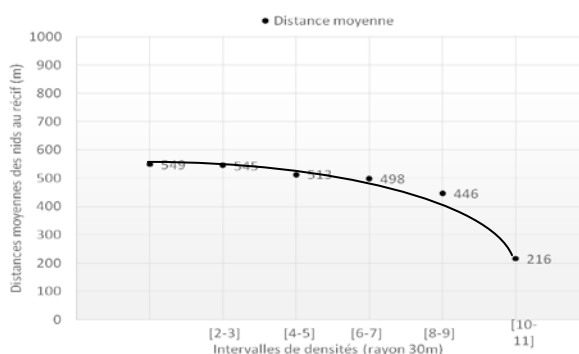


FIGURE 6. Densité locale à 30m vs. distance au récif le plus proche

Cette courbe semble montrer que les zones de nidification les plus denses (densités de 10 à 11) sont relativement proches du rivage (de l'ordre de 200 mètres), alors que les zones les moins denses (densité entre 0 et 9) sont situées au-delà de 400 mètres du récif. Cependant, plusieurs cas sur la Figure 4 viennent contredire cette analyse et montrent que d'autres facteurs sont à prendre en compte.

Ainsi le *motu* Tahuna iti (en jaune sur la Figure 4), bien qu'étant très éloigné du récif (de l'ordre du kilomètre), possède une des densités de nids la plus importante de l'atoll.

A contrario, les plages du nord des *motu* Oroatera et Tiaraunu (en rouge sur la Figure 4), très proches du récif, ont une très faible densité de nids. Cela pourrait s'expliquer par la présence d'obstacles rocheux dans le lagon et sur les plages ; l'aragonite à ces endroits forme des escarpements importants et étendus (Figure 8). Cela peut non seulement constituer un obstacle pour le trajet des femelles, plusieurs cas de tortues coincées ont été rapportés (Touron *et al.*, 2018), mais aussi être encore plus critique pour les jeunes tortues émergeantes qui peuvent rester coincées, à la merci des oiseaux, crabes et autres prédateurs lors de leur parcours vers le lagon.



FIGURE 8. L'aragonite peut être un obstacle pour les tortues juvéniles

Dans l'hypothèse où les femelles tortues reviennent nicher sur la plage où elles sont nées (Bass, 1999), il semble logique de penser que, générations après générations, les plages les moins favorables à l'accès des femelles adultes en quête de sites de ponte et/ou au retour des jeunes tortues au lagon seraient délaissées, alors que les plages les plus favorables verraient leur attractivité augmenter au fil des saisons.

2.2 Améliorations possibles

2.2.1. Critères géomorphologiques des nids

Lors de la découverte d'un nid, les équipes de bénévoles notent, entre autres, des paramètres géomorphologiques sur l'emplacement du nid, comme la distance à l'eau, l'altitude, la distance au récif, la pente etc. Cependant, ces informations souffrent d'un biais inévitable lié à leur caractère estimatif. Automatiser le calcul de ces paramètres, lorsque cela est possible, permettrait d'éviter ce biais.

Contribution de la géomatique à l'observation des nids de tortues marines de Tetiaroa

La connaissance de l'emplacement précis des nids de tortues associée à un modèle numérique 3D des plages de Tetiaroa permettrait de calculer automatiquement certains paramètres géomorphologiques. Le programme IDEA ayant fourni plusieurs modèles 3D des plages de Tetiaroa à partir des données Pléiades tri-stéréo 2014 et Lidar 2017, nous avons accès au MNS (modèle numérique de surface) et au MNT (modèle numérique de terrain) de Tetiaroa.

Des tests ont été effectués sur ces données afin de vérifier la faisabilité de l'extraction automatique des paramètres de pente, hauteur, distance à l'eau, distance au récif. Si cela ne pose pas de problème géomatique, deux contraintes doivent toutefois être respectées pour que les données extraites soient fiables et précises :

D'une part il est indispensable de travailler sur un MNT concomitant à la découverte du nid. Or la morphologie des plages de sable change à chaque épisode de forte houle ou de tempête importante. Ainsi, il faudrait obtenir le modèle 3D des plages de Tetiaroa en début de saison de ponte (vers le mois de septembre) et après chaque événement climatique significatif susceptible de modifier la morphologie de la plage.

Cela peut être envisagé en utilisant la reconstruction 3D par drone car les caméras et algorithmes existent. Le seul point délicat est le recalage de ces images, qui nécessite de disposer de points d'amer sur les plages au moment du passage du drone. Puisque cette opération doit être effectuée plusieurs fois par an, il est envisagé de sceller dans le béton des points d'amer permanents, dont on relèverait la position GPS très précise lors de l'installation. Ces points de calage pouvant être réutilisés moyennant un simple nettoyage avant le passage du drone et également servir au recalage des images satellites à venir.

D'autre part il est aussi nécessaire de disposer d'un protocole de localisation GPS avec une précision au moins inférieure à la précision spatiale du modèle 3D. Concrètement, le programme IDEA a fourni un MNS Pléiades 2014 à 70cm et un MNT Lidar 2017 à 30cm. La technologie drone devrait avoir une précision spatiale équivalente. Une localisation GPS centimétrique, entre 2 et 20cm est donc nécessaire.

L'association a longtemps utilisé des GPS grand public avec une précision de l'ordre de la dizaine de mètres, ce qui s'avère insuffisant pour ce genre d'application. Depuis 2018, elle possède un système GPS différentiel RTK (Stoll and Longine, 2019) qui permet d'obtenir une précision centimétrique. Cependant son utilisation est trop contraignante pour une utilisation au jour le jour par du personnel non-spécialiste. En attendant d'obtenir du matériel GPS centimétrique plus adapté aux missions terrain, décision a été prise pour la saison 2018-2019 de localiser uniquement les nids ayant éclos, en une fois en fin de saison de ponte.

2.2.2. Problématique de la « red line »

Les techniques géomatiques ont aussi été envisagées pour servir d'aide à la décision pour les équipes de bénévoles de l'association. En effet, lorsqu'un nid de tortue est submergé, il n'est plus viable. Pour éviter cela, les bénévoles ont 24 heures pour déplacer un nid sans incidence sur les éclosions ; au-delà, le risque est trop grand sur la viabilité des juvéniles.

A chaque découverte d'un nid, une décision doit être prise de déplacer ou non le nid s'il est en aval de la limite de la zone inondable de la plage (appelée « red line »). Pour estimer cette « red line », les bénévoles utilisent la laisse de mer : celle-ci dessine sur le sable une bande, constituée de débris flottants, à la limite supérieure des vagues. Classiquement, dans de nombreux pays, la laisse de basse mer ou de haute mer est utilisée pour cartographier la limite entre la terre et la mer.



FIGURE 9. Relevé de la laisse de mer comme proxy de la zone inondable

La laisse de mer tend à s'effacer avec le temps, mais elle peut être relevée aisément avec un GPS différentiel (Figure 9), et être intégrée comme repère vectoriel dans le GPS servant à la localisation des nids. Pour être pertinent, ce relevé doit être fait en début de saison de pontes et après chaque événement climatique significatif : forte houle, tempête, ...

Une fois le protocole de cartographie 3D des plages de Tetiaroa mis en place, il sera possible de mettre au point une méthodologie de calcul de la « red line » à partir de ces données 3D et des données issues d'un marégraphe. Cet instrument permet de mesurer le niveau de la mer à un endroit donné, or l'intérieur du lagon de Tetiaroa en a été équipé en 2018. Le calcul de la « red line » devra être effectué après chaque mise à jour du modèle 3D des plages.

3. Conclusions et perspectives

Après un nettoyage de la banque de données Excel des nids de tortues entre 2007 et 2018, un fichier vectoriel unique de l'emplacement des nids de tortues a été créé, incluant dans sa table d'attributs toutes les observations relevées sur le terrain par les membres de l'association Te Mana O Te Moana. L'analyse géomatique sous Qgis de cette donnée vectorielle a permis de repérer les sites de pontes prioritaires en calculant la densité locale de nids à 30 mètres. Un calcul de la distance de chaque

nid au récif a montré que même si les tortues semblent préférer les plages les plus proches du récif pour pondre, ce facteur ne peut pas être le seul retenu.

L'utilisation de fonctions géomatique a été étudiée pour extraire des facteurs géomorphologiques à partir de la seule localisation précise des nids et d'un modèle numérique 3D des plages à jour, permettant d'éviter les biais dans l'estimation des mesures.

Pour cela, des évolutions matérielles et méthodologiques sont nécessaires : il faut équiper l'association de tablettes durcies avec localisation GPS différentielle afin de pouvoir alimenter la base de donnée directement en limitant les erreurs de recopie. Un réseau de points d'amer permanents devra être installé sur les plages de pontes. Un protocole de modélisation 3D par image drone devra être mis au point afin de relever le modèle des plages en début de saison de ponte et après chaque évènement climatique. Enfin, à l'aide des données du marégraphe et de ces modèles 3D, il faut étudier un protocole pour déterminer l'emplacement de la « red line », zone de la plage où les nids n'ont que peu de risque d'être submergés.

Ces évolutions méthodologiques permettront de faciliter la collecte des données sur le terrain, d'en améliorer la précision et par là même de permettre des analyses plus fines pour répondre aux nombreuses questions qui se posent encore sur la nidification des tortues marines. Ces données alimenteront le futur observatoire de données scientifiques et environnementales en cours de mise en place sur Tetiaroa.

Remerciements :

Merci à la Tetiaroa Society, au programme IDEA, Armin Gruhn, Tao Guo et Serkan Ural de l'ETH Zurich pour les données Pléiades2014 et Lidar2017, et à tous les membres et bénévoles de l'association Te Mana O Te Moana et en particulier Quentin et Margot pour leur aide scientifique et logistique.

Bibliographie :

Ackerman, R.A., 1997. The nest environment and the embryonic development of sea turtles. *The biology of sea turtles* 1, 83-106.

Bass, A., 1999. Genetic analysis to elucidate the natural history and behaviour of hawksbill turtles (*Eretmochelys imbricata*) in the wider Caibbean: a review and re-analysis. *Chelonian Conserv Biol* 3, 195-199.

Bossin, H., Petit, H., Marie, J., Stoll, B., Cheong Sang, M., 2016a. Releases of *Wolbachia* mosquito males to sterilize and suppress a population of *Aedes Polynesiensis* on Tetiaroa Atoll, French Polynesia. *International Congress for Tropical Medicine and Malaria, Brisbane - Australia.*

Bossin, H., Petit, H., Marie, J., Stoll, B., Cheong Sang, M., 2016b. Suppression of an isolated population of the mosquito vector *Aedes polynesiensis* on the atoll of Tetiaroa, French Polynesia, by sustained release of *Wolbachia*-incompatible male mosquitoes. *ICE - XXV International Congress of Entomology, Orlando, Florida, USA.*

Cathy, R.R., 1994. Loggerhead nest morphology: effects of female body size, clutch size and nesting medium on nest chamber size., *proceedings of the 14th annual sea turtle*

symposium on sea turtle biology and conservation. NOAA Technical Memorandum NMFS SEFSC, pp. 25-28.

Davies, N., Field, D., Gavaghan, D., Holbrook, S.J., Planes, S., Troyer, M., Bonsall, M., Claudet, J., Roderick, G., Schmitt, R.J., Amaral Zettler, L., Berteaux, V., Bossin, H., Cabasse, C., Collin, A., Deck, J., Dell, T., Dunne, J., Gates, R., Harfoot, M., Hench, J.L., Hopuare, M., Kirch, P., Kotoulas, G., Kosenkov, A., Leichter, J.J., Lenihan, H., Magoulas, A., Martinez, N., Meyer, C., Stoll, B., Swalla, B., Tartakovsky, D.M., Teavai Murphy, H., Turyshev, S., Valdivinos, F., Williams, R., Wood, S., Consortium, I., 2016. Simulating social-ecological systems: the Island Digital Ecosystem Avatars (IDEA) consortium Gigascience - Biomed Central. The Open Publisher.

Gaspar, C., Petit, M., Leclerc, N., 2008. Rapport final relatif au suivi des sites de pontes sur l'atoll de Tetiaroa (saison 2007-2008). Association Te Mana O Te Moana, Direction à l'environnement de Polynésie française, p. 104.

Godley, B.J., Broderick, A.C., Hays, G.C., 2001. Nesting of green turtles (*Chelonia mydas*) at Ascension island, south atlantic. *Biological Conservation* 97, 151-158.

Horrocks, J.A., McA. Scott, N., 1991. Nest site location and nest success in the hawksbill turtle *Eretmochelys imbricata* in Barbados, West Indies. *Marine Ecology Progress Series* 69, 1-8.

Janzen, F.J., 1994. Climate change and temperature-dependent sex determination in reptiles. *Pop Biol* 91, 7487-7490.

Kamel, S.J., 2013. Vegetation cover predicts temperature in nests of the hawksbill sea turtle: implications for beach management and offspring sex ratios. *Endangered Species Research* 20, 41-48.

Karavas, N., Georghiou, K., Arianoutsou, M., Dimopoulos, D., 2005. Vegetation and sand characteristics influencing nesting activity of *Caretta caretta* on Sekania beach. *Biological Conservation* 121, 177-188.

Kikukawa, A., Kamezaki, N., Ota, H., 1999. Factors affecting nesting beach selection by loggerhead turtles (*Caretta caretta*): A multiple regression approach. *Journal of Zoology* 249, 447-454.

Molle, G., Hermann, A., Lagarde, L., Stoll, B., 2019. The long-term history of Teti'aroa (Society Islands, French Polynesia): New archaeological and ethnohistorical investigations. *Journal of Pacific Archaeology*

Mortimer, J.A., 1980. The influence of beach sand characteristics on the nesting behavior and clutch survival of green turtles (*Chelonia mydas*). *Copeia*, 802-817.

Petit, M., 2009. Rapport final relatif au suivi des sites de pontes de tortues sur l'atoll de Tetiaroa (octobre 2008-juillet 2009). Te Mana O Te Moana.

Petit, M., Bignon, F., Besson, M., Gaspar, C., 2013. Suivi des pontes de tortues vertes sur l'atoll de Tetiaroa (Polynésie française) durant la saison 2012-2013. Te Mana O Te Moana.

Petit, M., Gaspar, C., Leport, G., Esposito, C., Stabile, V., 2016. Saisons de ponte 2014-2015 et 2015-2016 de la tortue verte (*Chelonia mydas*) sur l'atoll de Tetiaroa. Te Mana O Te Moana.

Stoll, B., Bossin, H., 2019. Observatoire de données de Tetiaroa : Contribution à la modélisation de l'environnement des moustiques., SAGEO, Clermont Ferrand.

Contribution de la géomatique à l'observation des nids de tortues marines de Tetiaroa

Stoll, B., Longine, M., 2019. Géomorphologie à Tetiaroa : Prémices d'un observatoire de données scientifiques et environnementales. SAGEO, Clermont Ferrand.

Tetiaroa_Society, 2016. TETIAROA CONSERVATION AND SUSTAINABLE USE PLAN (CASUP).

Touron, M., Gaspar, C., Carpentier, A., Van Der Helm, M., Stabile, V., 2017. Bilan de la saison de ponte 2016-2017 de la tortue verte (*Chelonia mydas*) sur l'atoll de Tetiaroa, en Polynésie française. Te Mana O Te Moana, p. 53.

Touron, M., Genet, Q., Gaspar, C., 2018. Rapport final de la saison de ponte 2017-2018 des tortues vertes (*Chelonia mydas*) de l'atoll de Tetiaroa - Polynésie française. Te Mana O Te Moana.

Whitmore, C.P., Dutton, P.H., 1985. Infertility, embryonic mortality and nest-site selection in leatherback and green turtles in Suriname. *Biol Cons* 34, 251-272.

Witherington, B.E., Hirama, S., Mosier, A., 2011a. Barriers to sea turtle nesting on Florida (United States) Beaches: Linear extent and changes following storms. *Journal of Coastal Research* 27, 450-458.

Witherington, B.E., Hirama, S., Mosier, A., 2011b. Sea turtles responses to barriers on their nesting beach. *Journal of Experimental Marine Biology and Ecology* 401, 1-6.

Wood, D., Bjorndal, K.A., 2000. Relation of temperature, moisture, salinity and slope to nest site selection in loggerhead sea turtles. *Copeia*, 119-128.

La programmation logique pour les entrepôts de données spatiales

François Pinet

Université Clermont Auvergne, Irstea, UR TSCF, Centre de Clermont-Ferrand, 9 avenue Blaise Pascal CS 20085, F-63178 Aubière, France.

RESUME. Nous assistons à une augmentation de la mémoire vive des ordinateurs et à une utilisation de plus en plus importante des bases de données en mémoire. La programmation logique offre aujourd'hui des fonctionnalités intéressantes pour raisonner sur des données en mémoire. Partant de ce constat, nous proposons dans cet article, une approche pour représenter des entrepôts de données spatiales et leurs requêtes en Prolog. Nous nous concentrons sur une fonction principale des entrepôts, à savoir l'agrégation des données. Nous mettons en lumière une manière simple, exploitant directement les fonctionnalités existantes de Prolog, pour réaliser ce type de requêtes, en prenant en compte des critères spatiaux.

ABSTRACT. There is an increase in computer RAM size and a growing use of in-memory databases. Logical programming now offers interesting features to reason on data. Starting from this observation, we propose in this communication, an approach to represent spatial data warehouses and their queries in Prolog. We focus on the main function of data warehouses, i.e., the data aggregation. We emphasize a simple method, directly exploiting the existing functionalities of Prolog, to represent this type of queries, taking into account spatial criteria.

MOTS-CLES : PROGRAMMATION LOGIQUE, ENTREPÔT DE DONNEES

KEYWORDS: LOGIC PROGRAMMING, DATA WAREHOUSES

1. Introduction

Les propositions de (Robinson, 1965) et (Kowalski, 1974) ont établi des bases opérationnelles pour raisonner à partir de formules de la logique du 1^{ère} ordre. Ces travaux ont favorisé l'avènement de la programmation logique et de son langage le plus emblématique, à savoir Prolog (Colmerauer et Roussel, 1992). Les systèmes de raisonnement à base de règles restent jusqu'à aujourd'hui importants dans le domaine de l'intelligence artificielle.

En Prolog, les données utilisées pour le raisonnement sont généralement toutes montées en mémoire. Le couplage entre Prolog et les bases de données a été très étudié, l'objectif étant de permettre d'utiliser les possibilités offertes par Prolog avec de grands volumes de données stockés sur disques. En effet, Prolog intègre des fonctionnalités spécifiques qui peuvent être intéressantes pour traiter les données (requêtes récursives, fonctions pour les graphes, solveurs de contraintes, traitement de données textuelles, etc.). Dans les années 80, Datalog est apparu avec pour objectif de proposer des méthodes spécifiques pour les bases de données (Ceri et al., 1989 ; Bidoit, 1991 ; Abiteboul et al., 2000). On parle alors du concept de base de données déductive. Prolog propose une évaluation descendante des règles (de la tête des règles vers leur corps), alors que Datalog préconise une évaluation ascendante (du corps des règles vers leur tête), cette dernière méthode étant considérée comme plus adaptée aux traitements par lots des données. En Datalog, le processus de raisonnement peut être optimisé grâce à des méthodes dynamiques de réécritures des règles en fonction de la requête (voir les méthodes des ensembles magiques (Bancilhon, 1986)).

Sur le plan de la formalisation, les opérations de base de l'algèbre relationnelle peuvent être exprimées facilement sous la forme d'une requête conjonctive, c'est-à-dire une règle. Prenons la règle, $r1(X,Z) \leftarrow r2(X,Y) \wedge r3(Y,Z,c)$, avec X,Y,Z des variables et c une constante. Cette règle est exprimable directement en Prolog. En fait, cette règle correspond à : (1) une opération d'équi-jointure entre des relations $r2$ et $r3$, puisque $r2(X,Y)$ et $r3(Y,Z,c)$ partagent une variable commune, à savoir Y ; (2) une opération de sélection (le dernier attribut de la relation $r3$ doit être égal à la constante c) ; (3) une opération de projection sur les attributs X et Z , si on voit $r1$ comme la relation résultante de la requête.

Aujourd'hui, avec l'augmentation de la mémoire vive des ordinateurs et la progression des bases de données en mémoire, Prolog redevient un bon candidat pour raisonner sur des bases de données. Partant de ce constat, nous proposons dans cet article, une approche pour requêter un entrepôt de données spatiales décrit en Prolog. Nous nous focalisons sur une fonction principale des entrepôts, à savoir l'agrégation des données (Bimonte, 2007 ; Vaisman et Zimanyi, 2014). Notre objectif est de proposer une manière simple, exploitant directement les fonctionnalités existantes de Prolog, pour réaliser ce type de requêtes. Cet aspect n'a pas vraiment été traité dans la littérature. L'utilisation de Prolog pour agréger des données a été brièvement évoquée dans un exemple présenté dans (SWI Prolog, 2019). L'exemple donné dans (SWI Prolog, 2019) reste succinct et ne traite pas tous

les cas d'agrégation. Nous proposons donc un canevas plus complet pour représenter les agrégations des entrepôts de données, et montrons des applications pour le spatial. Nous utilisons le SWI Prolog qui est aujourd'hui la version de référence pour ce langage (Wielemaker, 2012).

2. Cas d'étude d'entrepôt de données spatiales

Nous allons illustrer nos propositions sur l'exemple suivant, portant sur des objectifs de ventes de commerciaux. Considérons le schéma de données suivant :

```
salesperson(SalespersonID:integer, SalespersonName:string, #StoreID:integer).  
store(StoreID:integer, StoreLocation:point).  
fact(#SalespersonID:integer, SalesGoal:numeric, ProductID:integer).
```

La relation *salesperson* représente les commerciaux avec leur identifiant, leur nom et le magasin auquel ils appartiennent. La relation *store* correspond aux magasins avec leur localisation. Les faits sont représentés dans *fact*. Chaque instance de *fact* correspond à un objectif de vente (*SalesGoal*) pour un commercial et un produit. Pour simplifier l'exemple, nous n'avons pas intégré de dimension temporelle – on considère que toutes les données de l'entrepôt portent sur la même période. Nous donnons ci-dessous des instances de données qui seront utilisées dans le reste de l'article. Les instances sont représentées sous la forme de faits Prolog.

```
salesperson('idsp0001', 'Peter', 'ser0001').  
salesperson('idsp0002', 'James', 'ser0002').  
salesperson('idsp0003', 'Bill', 'ser0001').  
salesperson('idsp0004', 'John', 'ser0003').  
  
store('ser0001', point(475310.5,2101000.3)).  
store('ser0002', point(555310.5,2000007.2)).  
store('ser0003', point(574444.7,2111333.4)).  
  
fact('idsp0001', 100000, 'idprdtP1').  
fact('idsp0001', 550000, 'idprdtP3').  
fact('idsp0001', 500000, 'idprdtP5').  
fact('idsp0002', 300000, 'idprdtP2').  
fact('idsp0003', 500000, 'idprdtP2').  
fact('idsp0004', 100000, 'idprdtP4').  
fact('idsp0004', 350000, 'idprdtP5').
```

Ainsi, par exemple, le vendeur Peter ayant comme identifiant idsp0001 a un objectif de ventes de 100000 pour le produit idprdtP1.

A noter que nous avons volontairement limité le nombre de dimensions afin d'avoir un exemple simple. Il n'y a pas de dimension spatiale – il est question d'objectifs de ventes concernant une unique période. Nous n'avons pas développé la dimension produit.

3. Représentation des agrégations de base

Afin de représenter une agrégation des données en Prolog, nous proposons d'utiliser l'opérateur *aggregate* de Prolog. Il sert à calculer des agrégats à partir de prédicats logiques. Afin de l'utiliser spécifiquement pour réaliser des requêtes d'agrégations dans un entrepôt de données, il faut notamment pouvoir préciser les jointures entre les relations. Pour ce faire, nous exploitons le lien vu en introduction entre algèbre relationnelle et requêtes conjonctives. Nous proposons le canevas de requête suivant, exprimé en Prolog :

```
aggregate(
  (aggregation_function_1,...,aggregation_function_N),
  Attribut_1^..Attribut_M^(relation_1,...,relation_O,...,condition_1,...,condition_P),
  (aggregation_result_1,...,aggregation_result_N)).
```

aggregation_function i et *aggregation_result i* sont respectivement (1) une fonction d'agrégation (par ex. count, sum, etc.) et (2) la variable allant stocker le résultat de l'agrégation, obtenu en utilisant cette fonction. Il est ainsi possible d'utiliser plusieurs fonction d'agrégation dans la même requête. *relation_1,...,relation_O* sont les relations nécessaires à l'agrégation. Ici les jointures entre les relations sont représentées comme dans les requêtes conjonctives traditionnelles (*relation_1,...,relation_O* correspond ici à *relation_1* \wedge ... \wedge *relation_O*). *Attribut_1,...,Attribut_M* sont les attributs inclus dans *relation_1,...,relation_O* qui ne sont pas utilisés pour l'agrégation. Alors que dans les requêtes SQL, on précise dans la clause GROUP BY les attributs de regroupement, en Prolog, à l'inverse, on précise les attributs qui ne sont pas utilisés pour le regroupement. En effet, par défaut, l'opérateur *aggregate* de Prolog regroupera tous les attributs présents dans *relation_1,...,relation_O*. La notation *Attribut i* \wedge ... permet d'exclure certains attributs du regroupement. *condition i* permet de préciser des conditions (telles que des sélections au sens de l'algèbre relationnelle).

Ainsi, la représentation en Prolog suivante va permettre de sommer *SalesGoal* pour toutes les instances de relation *fact* :

```
aggregate(sum(SalesGoal),
  SalespersonID^ProductID^(fact(SalespersonID,SalesGoal,ProductID)),
  SalesGoalSum).
```

La programmation logique pour les entrepôts de données spatiales

Le résultat est : SalesGoalSum = 2400000.

Passer à une agrégation par commerciaux ne demande qu'une légère modification de la requête. Par rapport à la requête jute au-dessous, il suffit de n'exclure que l'attribut Product de l'agrégation :

```
aggregate(sum(SalesGoal),
ProductID^(fact(SalespersonID,SalesGoal,ProductID)), SalesGoalSum).
```

Le résultat est :

```
SalespersonID = idsp0001, SalesGoalSum = 1150000 ;
SalespersonID = idsp0002, SalesGoalSum = 300000 ;
SalespersonID = idsp0003, SalesGoalSum = 500000 ;
SalespersonID = idsp0004, SalesGoalSum = 450000.
```

Il est également possible de ne sélectionner que les agrégations respectant un critère donné (ci-dessous > 350000) :

```
aggregate(sum(SalesGoal),
ProductID^(fact(SalespersonID,SalesGoal,ProductID)), SalesGoalSum),
SalesGoalSum > 350000.
```

Voici un exemple de requête illustrant une équi-jointures relationnelles :

```
aggregate(sum(SalesGoal),
SalespersonID^ProductID^StoreID^(
fact(SalespersonID,SalesGoal,ProductID),salesperson(SalespersonID,SalespersonName,StoreID)),
SalesGoalSumByPersonName).
```

L'expression "fact(SalespersonID, SalesGoal, ProductID), salesperson(SalespersonID, SalespersonName, StoreID)" correspond à la représentation de l'équi-jointure entre fact et salesperson. Le résultat est le même que la requête précédente sauf que les noms des commerciaux sont affichés plutôt que leur ID.

4. Représentation des agrégations en prenant en compte un critère spatial

Dans les entrepôts, les requêtes peuvent bien sûr porter sur la partie spatiale des données. Comme indiqué dans notre canevas de requête, il est possible de préciser des conditions dans l'agrégation, et celles-ci peuvent porter sur les objets spatiaux de l'entrepôt - dans l'exemple, la localisation des magasins (des points). Supposons que l'on souhaite sommer *SalesGoal* par magasin, mais uniquement pour les magasins qui sont dans une zone géographique donnée. Nous utilisons dans la requête un prédicat *inside(P,R)* qui réussit si le point *P* est inclus spatialement dans le rectangle *R*, et qui échoue sinon. La formulation de la requête sera celle-ci :

```
aggregate(sum(SalesGoal),
SalespersonID^ProductID^SalespersonName^(fact(SalespersonID,
```

SAGEO'2019

```
SalesGoal,ProductID), salesperson(SalespersonID, SalespersonName, StoreID),
store(StoreID, StoreLocation),
inside(StoreLocation,rectangle(500000,2000000,600000,2200000))), SalesGoalSumByStore).
```

Les résultats seront :

```
StoreID = ser0002, StoreLocation = point(555310.5, 2000007.2), SalesGoalSumByStore = 300000 ;
StoreID = ser0003, StoreLocation = point(574444.7, 2111333.4), SalesGoalSumByStore = 450000.
```

En effet, le magasin *ser0001* qui a comme coordonnée (475310.5,2101000.3) n'est pas dans le rectangle (500000, 2000000, 600000, 2200000).

A titre de comparaison, une représentation en pseudo-langage SQL spatial de cette requête serait :

```
SELECT store.StoreID, store.StoreLocation
FROM fact, salesperson, store
WHERE fact.SalespersonID=Salesperson.salespersonID and salesperson.StoreID=store.StoreID and
inside(store.StoreLocation, rectangle(500000,2000000,600000,2200000))
GROUP BY store.StoreID;
```

Une légère modification de la requête Prolog précédente permettra de calculer une seule somme globale (de *SalesGoal*) pour tous les magasins inclus dans la zone géographique indiquée. Il suffit d'exclure *StoreID* et *StoreLocation* du regroupement.

```
aggregate(sum(SalesGoal),
SalespersonID^ProductID^SalespersonName^StoreID^ StoreLocation^(fact(SalespersonID,
SalesGoal,ProductID), salesperson(SalespersonID, SalespersonName, StoreID),
store(StoreID, StoreLocation),
inside(StoreLocation,rectangle(500000,2000000,600000,2200000))), SalesGoalSum).
```

Le résultat sera :

```
SalesGoalSumByStore = 750000.
```

Il serait possible de préciser, assez directement, des jointures spatiales dans la conjonction de relations en utilisant des opérations spatiales adaptées. A noter que des fonctions spatiales, implémentées dans un Space Package, ont été définies pour Prolog, avec des passerelles vers les formats GML, KML et WKT. Plus de détails peuvent être trouvés dans (van Hage, 2010).

5. Conclusion

Dans cet article, nous avons proposé un canevas général et simple pour définir un entrepôt de données et ses agrégations en Prolog. Nous avons montré un cas d'étude avec des requêtes portant sur des critères spatiaux. Notre objectif dans cet

article était de montrer que l'on peut exprimer simplement, en Prolog, les requêtes typiques des entrepôts de données.

Il reste à illustrer l'utilisation, au sein des requêtes, de fonctionnalités propres à Prolog. Prolog est un langage de programmation et il peut offrir un éventail important de fonctions, par exemple : des expressions récursives, des traitements sur les graphes, sur des données textuelles, des solveurs de contraintes, etc. Par ailleurs, des contraintes d'intégrité (Pinet, 2010) peuvent aussi nativement être définies en Prolog sous la forme de règles, afin de tester le respect de conditions complexes sur les données. Les données elles-mêmes peuvent être complexes, ce qui est particulièrement adapté aux données spatiales. Effectivement, les attributs d'une base de données modélisée en Prolog peuvent être de structures complexes, par exemple prendre la forme de formule logique.

L'aspect performance est enfin à mesurer, en fonction de différentes volumétries de données. Il faudrait comparer les performances de Prolog avec les systèmes de bases de données en mémoire existants. Néanmoins, quel que soit le résultat de cette comparaison, l'intérêt principal de l'utilisation de Prolog pour les entrepôts réside, selon nous, principalement dans le couplage direct entre programmation logique et base de données.

Bibliographie

- Abiteboul S., Hull R., Vianu V. (2000). *Fondements des bases de données*, Addison-Wesley
- Bancilhon F., Maier D., Sagiv Y., Ullman J. (1986). Magic sets and other strange ways to implement logic programs, PODS '86 Proceedings of the fifth ACM SIGACT-SIGMOD symposium on Principles of database systems, p. 1-15.
- Bidoit N. (1991). Bases de données déductives.
- Bimonte S. (2007). Intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation. Mémoire de thèse, INSA de Lyon, 219 pages.
- Ceri S., Gottlob G., Tanca L. (1989). What you Always Wanted to Know About Datalog (And Never Dared to Ask). IEEE Transactions on Knowledge and Data Engineering vol. 1(1), p. 146 - 166.
- Colmerauer A., Roussel P. (1992). *La naissance de Prolog*, 31 p.
- Kowalski R. (1974). *Predicate Logic as a Programming Language*, Department of Artificial Intelligence, Edinburgh University. 1973. Also in Proceedings IFIP Congress, Stockholm, North Holland Publishing Co., 1974, p. 569-574.
- Pinet F. (2010). Modélisation des contraintes d'intégrité dans les systèmes d'information environnementaux. Mémoire d'HDR, 1er avril 2010, Université Blaise Pascal, 110 pages.
- Robinson J. A. (1965). *A Machine-Oriented Logic Based on the Resolution Principle*, J. Assoc. Comput. Mach. vol. 12, p. 23-41.
- SWI Prolog - Database. (2019) <http://www.swi-prolog.org/howto/database.html>

SAGEO'2019

- Vaisman, A., Zimanyi E. (2014). Data Warehouse Systems: Design and Implementation, Springer*
- van Hage W.R., Wielemaker J. Schreiber G. (2010). *The Space Package: Tight Integration between Space and Semantics*. Trans. In GIS, vol. 14(2), p.131-146.
- Wielemaker J., Schrijvers T., Triska M., Lager T. (2012) *SWI-Prolog*. Theory and Practice of Logic Programming vol. 12(1-2), p.67-96.

Diagrammes temporels mixtes et analyse harmonique pour l'identification de cycles

Jacques Gautier¹, Paule-Annick Davoine², Claire Cuntty³

1. Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
jacques.gautier@univ-grenoble-alpes.fr

2. Univ. Grenoble Alpes, CNRS, Science Po Grenoble, PACTE, 38000 Grenoble, France
paule-annick.davoine@univ-grenoble-alpes.fr

3. Laboratoire EVS, Université Lyon II, Lyon, France
claire.cuntty@univ-lyon2.fr

RÉSUMÉ. Les environnements de géovisualisation dédiés à l'analyse de phénomènes spatio-temporels intègrent depuis longtemps des représentations graphiques permettant d'apprécier l'aspect cyclique du temps. En revanche, ils autorisent rarement une recherche visuelle de récurrences cycliques au cours d'une analyse exploratoire. La recherche visuelle de cycles est cependant possible au travers de diagrammes intégrant les aspects linéaires et cycliques du temps, lors d'une modification graduelle de l'échelle cyclique représentée. Couplés à des outils d'analyse harmonique suggérant aux utilisateurs des pistes de recherche, ces diagrammes représentent une opportunité pour l'analyse exploratoire de cycles dans les phénomènes spatio-temporels.

ABSTRACT. Geovisualization environments that allow the analysis of spatio-temporal phenomena, have used for a long time graphical representations that allow to estimate the cyclic aspect of time. However, they rarely enable a visual search for cyclic recurrences during an exploratory analysis. The visual search for cycles is yet possible through the use of diagrams that integrate both linear and cyclic aspects of time, during a gradual modification of the represented cyclic scale. When they are associated with harmonic analysis tools which suggest searching paths to users, these diagrams give an opportunity for an exploratory analysis of cycles in spatio-temporal phenomena.

MOTS-CLÉS : Diagramme mixte, Analyse harmonique, Analyse exploratoire, Récurrences cycliques, Visualisation

KEYWORDS: Mixt diagram, Harmonic analysis, Exploratory analysis, Cyclic recurrences, Visualization

1. Introduction

La représentation de l'aspect cyclique du temps est étudiée depuis longtemps en visualisation (Carlis, Konstan, 1998 ; Weber *et al.*, 2001 ; Müller, Schumann, 2003), et beaucoup d'environnements de géovisualisation intègrent des représentations permettant d'apprécier visuellement la cyclicité de certains phénomènes (*Time Wave* (Li, 2010), *PerSE* (Swedberg, Peuquet, 2017)). En revanche, la dimension cyclique y est majoritairement représentée au travers d'échelles correspondant à des rythmes de la nature ou à des activités humaines (Hornsby, Egenhofer, 2002). Ces environnements permettent d'explorer des données temporelles à travers une liste finie d'échelles cycliques usuelles, mais peu d'entre eux offrent, dans une démarche exploratoire, la possibilité d'une recherche visuelle de récurrences cycliques dont la période de réapparition peut présenter n'importe quelle valeur. Ce que nous désignons ici par le terme de « récurrences cycliques » fait référence à la réapparition d'un phénomène, aussi bien qu'à une évolution de son intensité, respectant une certaine périodicité. Proposer des représentations visuelles permettant l'identification de ces récurrences cycliques, au sein d'environnements de géovisualisation, présente alors un enjeu de recherche pour l'analyse visuelle des phénomènes spatio-temporels. L'analyse de récurrences cycliques fait l'objet de recherche dans d'autres domaines tels que le traitement du signal : des outils provenant de ce domaine, comme l'analyse harmonique (Gasquet, Witomski, 2000 ; Bony, 2001 ; Katznelson, 2004), permettent de déceler ces comportements cycliques dans les données temporelles. Nous émettons l'hypothèse que l'intégration de tels outils dans les environnements de visualisation, suggérant à l'utilisateur des pistes de recherche, pourrait améliorer l'efficacité d'une analyse exploratoire de structures cycliques dans les données temporelles.

Nous présentons à travers ce travail une approche de visualisation associant représentations graphiques et outils d'analyse harmonique, dans l'objectif de faciliter l'identification des récurrences cycliques lors de l'analyse exploratoire de séries d'événements spatio-temporels. Dans un premier temps, nous exposons les possibilités offertes par les diagrammes temporels et par l'analyse harmonique pour identifier ces cycles dans des données temporelles. Nous présentons ensuite comment l'approche visuelle offerte par les diagrammes temporels peut être associée à l'analyse harmonique pour permettre la recherche de cycles au cours d'une analyse exploratoire. Enfin, nous présentons comment cette proposition, intégrée au sein d'un environnement de visualisation, nous permet d'identifier des cycles au sein de jeux de données majoritairement aléatoires.

2. Visualisation de récurrences cycliques

2.1. Identification de cycles dans les représentations graphiques du temps

La majorité des diagrammes temporels intégrés au sein d'environnements de

géovisualisation sont basés sur les modèles de la *timeline*, représentant le temps de manière linéaire, et de la *timewheel*, représentant le temps au travers d'une échelle cyclique (R. M. Edsall *et al.*, 1997 ; R. Edsall, Peuquet, 1997). L'identification d'une récurrence cyclique dans les données temporelles au travers de ces outils pose toutefois certaines limites :

- la *timeline* n'intégrant pas de représentation de la structure cyclique du temps, l'utilisateur doit identifier une réapparition de patterns graphiques similaires séparés par une distance graphique régulière (Figure 1.A);
- la *timewheel* étant la représentation, soit d'une instance unique d'une échelle cyclique considérée, soit d'une projection de différentes instances de cette échelle sur un même espace graphique, elle permet difficilement l'identification de la réapparition d'un pattern dans plusieurs instances (Figure 1.B) : l'observation d'un événement, apparaissant le 3 mars, dans une *timewheel* représentant une échelle d'un an, ne permet pas de dire si cet événement n'apparaît qu'une seule année, ou s'il se répète chaque année le 3 mars.

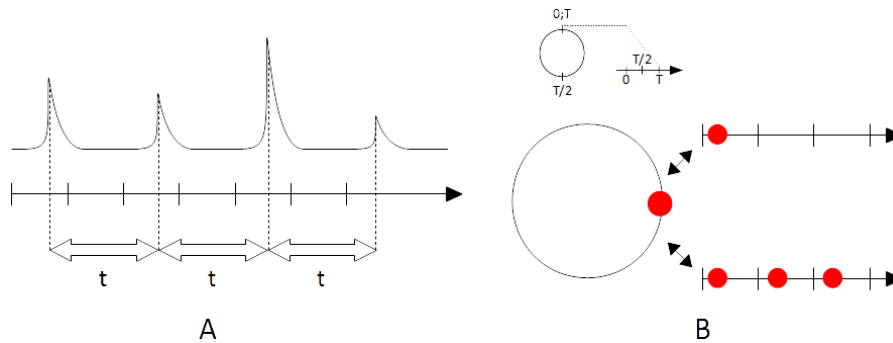


FIGURE 1. Difficulté d'identification d'un cycle dans les timelines et timewheels

D'autres diagrammes proposent d'allier les avantages de la *timeline* et de la *timewheel*, en intégrant dans une même représentation les aspects linéaires et cycliques du temps (R. Edsall, Peuquet, 1997 ; Van Wijk, Van Selow, 1999 ; Weber *et al.*, 2001 ; Li, 2010). Ces diagrammes, que l'on peut appeler diagrammes mixtes, représentent le temps de manière linéaire, à travers une succession d'instances d'une échelle cyclique considérée. À titre d'exemple, nous pouvons citer :

- le diagramme *Spiral Graph* (Weber *et al.*, 2001) (Figure 2.A), basé sur une version linéaire de la timewheel, se présentant sous la forme d'une spirale dont les anneaux représentent les instances successives d'une échelle cyclique;
- les *tiles maps* (Van Wijk, Van Selow, 1999) (Figure 2.B), représentant le temps au travers d'une échelle linéaire et d'une échelle cyclique sur deux axes orthogonaux, selon deux granularités différentes, créant une matrice dont

chaque colonne représente une instance de l'échelle cyclique (ex: une journée), et chaque ligne un même moment dans ces différentes instances (ex: 15h);

– les diagrammes *Time Coil* (R. Edsall, Peuquet, 1997) (Figure 2.C.1) et *Time Wave* (Li, 2010) (Figure 2.C.2), basés sur la combinaison des structures de la *timeline* et de la *timewheel*, et composés d'une série de motifs graphiques, représentant chacun une instance d'une échelle temporelle cyclique définie, se succédant le long d'un axe représentant la dimension linéaire du temps.

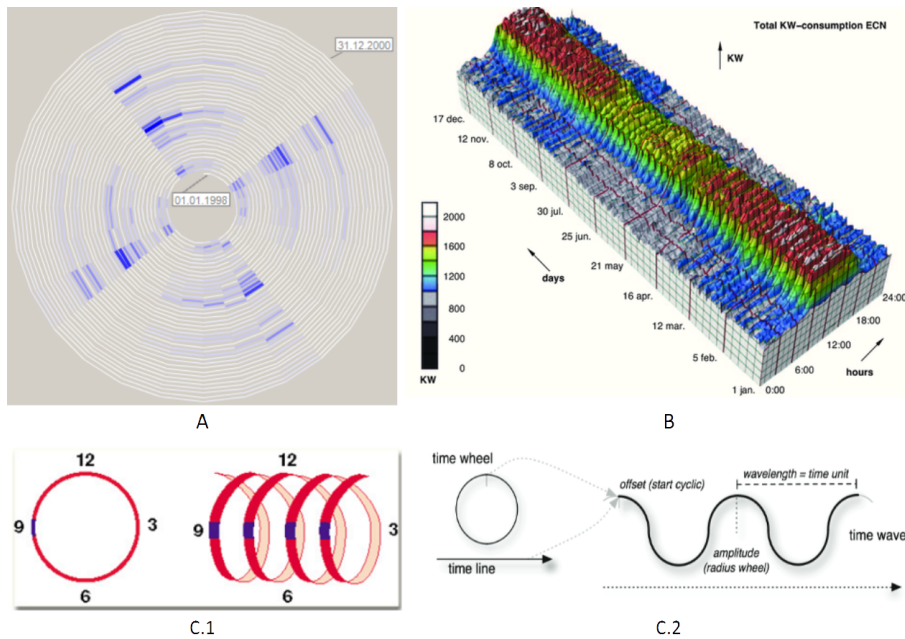


FIGURE 2. Exemples de diagrammes temporels mixtes (R. Edsall, Peuquet, 1997, Van Wijk, Van Selow, 1999, Weber et al., 2001, Li, 2010)

Ces diagrammes permettent d'identifier visuellement une récurrence cyclique dans des données temporelles, à travers l'observation d'un alignement de patterns graphiques similaires le long d'un axe, lorsque la période du cycle approche la durée de l'échelle cyclique représentée. La recherche d'une récurrence cyclique se base ainsi sur le même procédé : l'utilisateur doit chercher pour quelle échelle cyclique un alignement graphique est observable.

L'objectif de nombreux environnements de visualisation intégrant de tels diagrammes n'est cependant pas de déterminer si un phénomène est cyclique ou non, en identifiant les possibles récurrences cycliques de ce dernier. Pour beaucoup, l'objectif est de permettre l'analyse, à travers une représentation

combinée du temps linéaire et d'une (ou de plusieurs) échelle(s) cyclique(s), de la structure temporelle de phénomènes, souvent déjà considérés comme cycliques (tels que ceux liés au cycle jour/nuit). Cet objectif entraîne la limitation du nombre d'échelles cycliques paramétrables dans ces environnements.

Selon Hornsby et Egenhofer (Hornsby, Egenhofer, 2002), les échelles 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 paramétrables dans les environnements de visualisation intégrant des diagrammes mixtes sont ainsi souvent circonscrites à ces périodes usuelles, ce qui limite d'emblée les périodes pour lesquelles il est possible d'identifier une récurrence cyclique des événements. Par ailleurs, le caractère discontinu de la série d'échelles cycliques paramétrables (le passage d'une échelle cyclique d'un mois à une échelle cyclique d'un an s'effectue généralement sans transition) empêche l'observation de la formation d'alignements graphiques, caractéristiques d'une récurrence cyclique, lors d'une modification graduelle de l'échelle cyclique du diagramme. Or, l'observation de la formation d'un alignement permet à l'utilisateur de savoir s'il approche d'une échelle pour laquelle une récurrence est identifiable. Il est donc nécessaire, pour rendre possible une telle recherche, de permettre une modification dynamique et continue de la durée des échelles cycliques représentées (Weber *et al.*, 2001).

2.2. La recherche de cycles par l'analyse harmonique

La recherche de récurrences cycliques au moyen de diagrammes temporels mixtes est exploratoire : c'est l'identification visuelle d'alignements graphiques par l'utilisateur, lors d'une modification de l'échelle cyclique, qui lui permet de détecter la présence de récurrences dans les données temporelles. Il existe d'autres méthodes, faisant notamment appel à des outils d'analyse mathématiques, qui permettent d'identifier des récurrences cycliques et donc d'aider l'utilisateur dans sa démarche exploratoire.

L'analyse harmonique est une branche des mathématiques étudiant la représentation des signaux comme des superpositions de fonctions sinusoidales élémentaires, de fréquence et d'amplitude différentes, appelées harmoniques (Bony, 2001 ; Katznelson, 2004). Les analyses harmoniques ont connu de nombreuses applications en physique sous le nom d'analyse spectrale, et font appel à des outils mathématiques, tels que les transformées de Fourier, permettant de décomposer une fonction mathématique en une somme de fonctions périodiques de différentes fréquences (Gasquet, Witomski, 2000). Lorsqu'une fonction mathématique représente un phénomène physique, cette fonction est appelée signal, et sa transformée de Fourier est appelée spectre. Ce spectre permet de recenser les fréquences, fondamentales et harmoniques, de toutes les fonctions sinusoidales permettant de recomposer le signal originel. L'identification des

fonctions sinusoidales recomposant au mieux un signal, représentant un phénomène physique, permet alors d'identifier les fréquences de possibles récurrences cycliques d'un phénomène physique.

Nous émettons donc l'hypothèse qu'une analyse harmonique effectuée sur des données temporelles, en arrière-plan d'une représentation de ces dernières dans un diagramme mixte, permettrait d'identifier de possibles récurrences cycliques dans le jeu de données étudié. Les résultats trouvés pourraient alors être retournés de manière graphique à l'utilisateur en guise de pistes d'exploration, sous la forme d'une suggestion de valeurs à affecter à la durée de l'échelle temporelle cyclique représentée.

3. Diagrammes mixtes et transformées de Fourier pour la recherche de cycles

Pour permettre la recherche de récurrences cycliques dans les données retranscrivant des phénomènes spatio-temporels, au cours d'une analyse exploratoire, nous proposons un diagramme temporel mixte combiné à des outils d'analyse harmonique utilisant la transformée de Fourier.

3.1. Diagramme temporel mixte en dents de scie

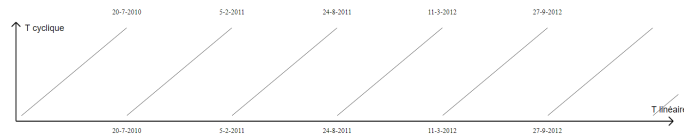


FIGURE 3. Diagramme temporel en dents de scie

Notre proposition est en partie inspirée des *tiles map* (Van Wijk, Van Selow, 1999), et prend la forme d'une fonction périodique en dents de scie (Figure 3). Le temps est représenté sur deux axes selon la même granularité temporelle, l'axe des abscisses représentant le temps de manière linéaire, l'axe des ordonnées représentant le temps selon une échelle cyclique. Une instance de l'échelle cyclique est représentée par un axe linéaire oblique, la représentation d'une suite d'instances de l'échelle cyclique prenant ainsi la forme d'une suite de lignes brisées.

Les données représentées prennent la forme de séries événementielles ponctuelles, pouvant correspondre à des événements ponctuels observables dans le temps et dans l'espace, ou à des instants pour lesquels une variable modélisée

par une série temporelle dépasse une valeur seuil (Figure 4.A). Ces événements sont représentés sur le diagramme par des entités graphiques ponctuelles. Afin de faciliter l'analyse de nombreuses données, les événements temporellement proches sont agrégés, et ces agrégats sont représentés par des entités graphiques, dont la taille est proportionnelle au nombre d'événements agrégés. Les entités graphiques sont placées le long de la courbe du diagramme en fonction du moment d'apparition de l'événement représenté, ou en fonction du « barycentre » des moments d'apparition des événements agrégés (Figure 4.B).

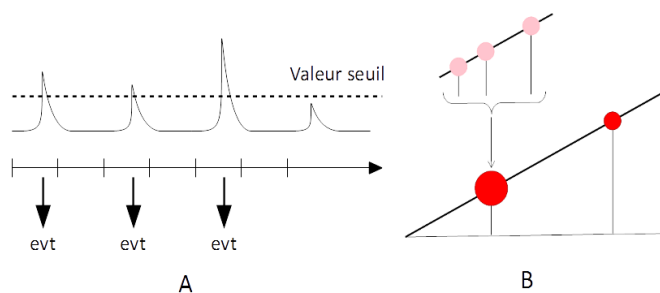


FIGURE 4. Représentation des données dans le diagramme temporel mixte

La durée T de l'échelle cyclique représentée par le diagramme peut être modifiée de manière dynamique et graduelle. Lors d'une modification de l'échelle cyclique (Figure 5), la longueur graphique utilisée pour représenter T reste inchangée sur l'axe des ordonnées, mais est modifiée sur l'axe des abscisses. Ainsi, lorsque T diminue, le nombre de lignes brisées du diagramme augmente et celles-ci apparaissent de plus en plus rapprochées et de plus en plus verticales. À l'inverse, lorsque T augmente, le nombre de lignes brisées du diagramme diminue et celles-ci apparaissent de plus en plus espacées de plus en plus horizontales. La recherche visuelle de récurrences cycliques s'effectue en modifiant graduellement la durée T de l'échelle temporelle cyclique, au moyen d'un curseur via une interface de contrôle, dans le but d'observer la formation d'un alignement d'entités graphiques sur le diagramme.

La formation de cet alignement est observable lorsque la valeur de T approche celle de la période P d'une récurrence cyclique. 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 de durée $T1$, l'orientation de ces droites indique si la période P est inférieure ou supérieure à $T1$:

- si $P < T1$, la ou les droites formées par l'alignement des entités graphiques seront descendantes.
- si $P > T1$, la ou les droites formées par l'alignement des entités graphiques seront ascendantes.

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. Les entités sont alors alignées, parallèlement à cet axe, lorsque $T = P$ (Figure 5).

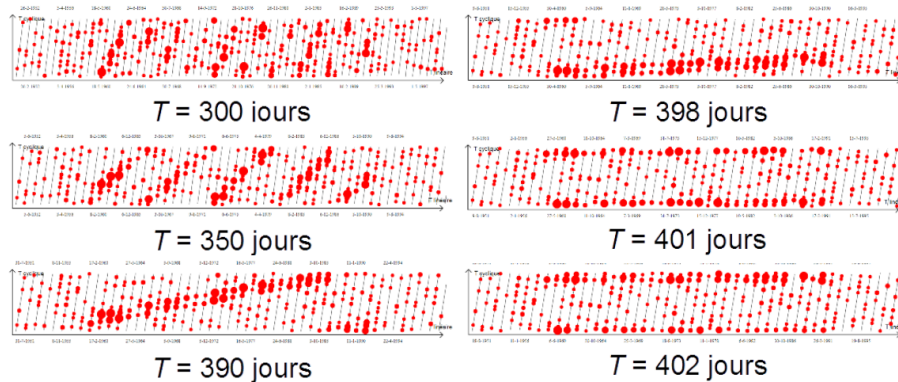


FIGURE 5. Recherche visuelle d'une récurrence cyclique par modification de la durée de l'échelle cyclique

3.2. Améliorer la recherche visuelle par l'analyse harmonique

En arrière-plan de la représentation graphique, une analyse harmonique par transformée de Fourier est effectuée sur les données étudiées, dans le but d'y détecter de possibles comportements cycliques. Afin d'effectuer cette analyse, nous recréons dans un premier temps un signal échantillonné à partir des séries événementielles que nous étudions. Pour cela, nous discrétisons l'étendue temporelle linéaire de notre analyse selon un intervalle d'échantillonnage (paramétrable par l'utilisateur), et nous calculons, pour chaque intervalle temporel, le nombre d'événements y apparaissant. La représentation graphique du signal obtenu prend la forme d'un diagramme en bâtons (Figure 6.A). Nous appliquons sur le signal échantillonné obtenu une transformée de Fourier discrète. Nous sélectionnons alors les fréquences correspondant aux pics les plus significatifs du spectre obtenu (Figure 6.B).

La durée des périodes correspondantes est alors représentée de manière graphique, en étant projetée sur le curseur de modification de la durée de l'échelle cyclique, dans l'interface de contrôle (Figure 6.C). Ces périodes constituent alors une suggestion d'échelles temporelles cycliques "remarquables" que l'utilisateur peut "inspecter". Ce dernier peut faire évoluer la durée de l'échelle cyclique vers ces valeurs, afin de vérifier si la formation d'alignements graphiques est observable sur le diagramme mixte (Figure 6.D).

Diagrammes mixtes et analyse harmonique

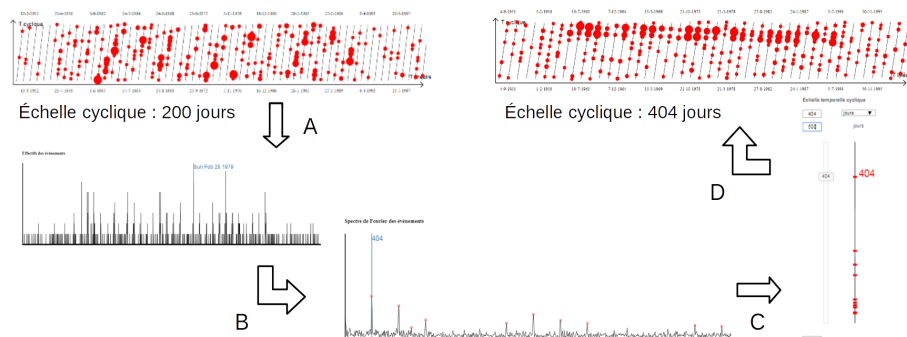


FIGURE 6. Utilisation de l'analyse harmonique pour assister la recherche de récurrences cycliques

4. Validation de la proposition

Notre proposition de diagramme mixte associé à une analyse harmonique est intégré dans un environnement de géovisualisation nommé GRAPHIST¹ (Gautier, 2018). Nous avons pu, au moyen de cet environnement, appliquer notre proposition à l'analyse d'une série de huit jeux de données créés de manière procédurale, chacun composé :

- d'une série d'événements cycliques localisés dans l'espace, mais dont la période du cycle et la zone d'apparition sont définies aléatoirement ;
- d'une série d'événements dont le lieu et le moment d'apparition sont générés de manière aléatoire. Parmi les jeux de données créés, quatre comportent 80% d'événements aléatoires (jeux A), les jeux restants en comportant 90% (jeux B). Nous émettons l'hypothèse que si l'identification visuelle de récurrences cycliques dans de tels jeux de données est possible au moyen de notre interface, notre approche peut être utilisée afin d'identifier des structures analogues dans des jeux de données réels.

Les paramètres de création de chaque jeu sont choisis de manière aléatoire, puis stockés dans un fichier annexe, de manière à être accessibles a posteriori, mais inconnus lors de l'expérimentation. Il s'agit donc d'identifier la composante cyclique de chaque jeu de données, d'estimer les paramètres de cette composante cyclique, puis de comparer les résultats estimés avec les paramètres stockés. Plus précisément, la tâche consiste à estimer (Figure 7) :

- la durée P de la période du cycle, identifiable lors de l'observation d'un alignement graphique parallèle à l'axe des abscisses du diagramme ;

1. GRAPHIST est disponible en ligne https://steamer.imag.fr/?page_id=735

- l'étendue temporelle linéaire E couverte par le cycle, estimée en observant l'étalement de l'alignement formé sur l'axe des abscisses du diagramme ;
- la dispersion temporelle D du cycle, définissant l'intervalle cyclique durant lequel les événements apparaissent : « des événements apparaissent tous les P jours, plus ou moins k jours », où $k = P * D$. D est estimée en observant l'étalement de l'alignement formé sur l'axe des ordonnées du diagramme.

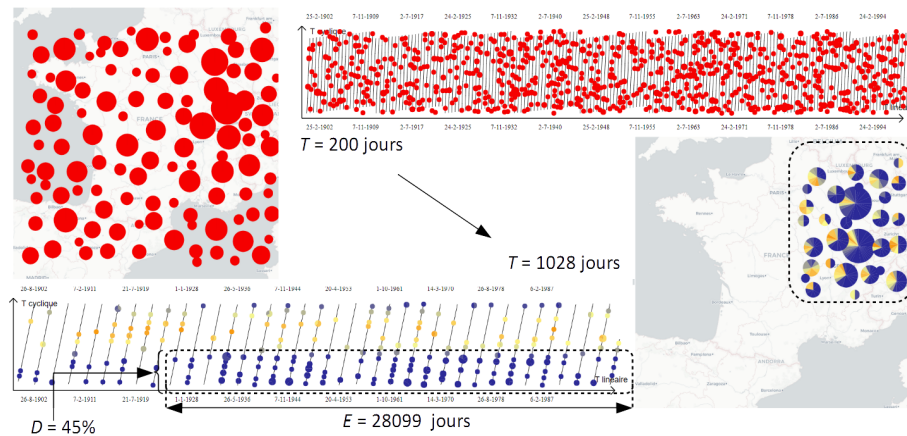


FIGURE 7. Identification d'événements cycliques (jeu B6) (Gautier, 2018)

En combinaison avec le diagramme mixte, GrAPHiST représente dans une carte l'ensemble des événements présents sur le diagramme temporel. Ces événements sont classés selon des périodes d'apparition relatives à l'échelle cyclique représentée, et une couleur est affectée à chaque classe. Les événements spatialement proches sont agrégés, et sont représentés par des entités graphiques dont la taille est fonction du nombre d'événements agrégés. Ces entités prennent la forme de *pie charts*, dont les sections représentent la proportion d'événements de chaque classe. Une concentration spatiale d'événements cycliques (et donc apparaissant durant la même période cyclique) sera observable par des *pie charts* présentant une couleur majoritaire (Figure 7).

La Table 1 présente les résultats des analyses effectuées sur les huit jeux de données. Pour chacun de ces derniers, l'analyse harmonique a permis de suggérer, au travers de l'interface graphique, une échelle cyclique pour laquelle un alignement graphique, légèrement décalé par rapport à l'axe horizontal du diagramme, est observable. Une légère modification de l'échelle cyclique a permis de placer cet alignement parallèlement à l'axe des abscisses, afin d'estimer la valeur de P . L'analyse exploratoire effectuée a permis d'estimer précisément la période du cycle de chaque jeu de données, hormis pour le jeu A2, où la période estimée correspond à un harmonique du cycle (moins de 2% de marge

d'erreur ; 3e colonne de la Table 1). En revanche, l'estimation de l'étendue et de la dispersion temporelle de chaque cycle reste peu précise (respectivement jusqu'à 26% et 15% de marge d'erreur ; 4e et 5e colonnes de la Table 1).

En dépit du manque de précision sur ces derniers points, ces résultats sont satisfaisants et valident notre proposition: une analyse exploratoire, au moyen du diagramme en dents de scie couplé à une analyse harmonique, permet de révéler l'existence de cycles au sein de séries d'événements majoritairement aléatoires, et d'estimer précisément leur période de réapparition.

TABLE 1. Résultats de l'expérimentation (Gautier, 2018)

Jeu	Proportion d'événements aléatoires	Taux d'imprécision de P estimée	Taux d'imprécision de E estimée	Taux d'imprécision de D estimée
A1	80 %	1.9%	0.2%	0.5%
A2	80 %	X	11.7%	15.0%
A3	80 %	0.1%	0.1%	1.2%
A4	80 %	0.6%	6.8%	3.0%
B1	90 %	0.2%	14.3%	6.0%
B2	90 %	0.4%	26.0%	12.0%
B3	90 %	0 %	4.9%	3.0%
B4	90 %	1.6%	13.8%	3.0%

5. Conclusion

Intégrée dans un environnement de visualisation, notre proposition peut être utilisée pour identifier des comportements cycliques isolés dans l'espace, ou pour identifier les différentes périodicités d'un phénomène selon l'espace, afin d'aider la compréhension, voire la prévision de ces phénomènes. Différents cas d'application peuvent être imaginés, comme l'analyse de données météorologiques, marégraphiques ou sismiques, mais également de données sociales.

Notre démarche ayant été de nous concentrer en priorité sur l'analyse d'événements ponctuels, nos propositions de représentations graphiques ne permettent pas de prendre en compte des données décrivant des événements étendus dans le temps. De plus, notre approche n'est pas adaptée à l'identification de comportements cycliques dont la période de réapparition évolue, apparaissant de manière plus ou moins fréquente au cours du temps. Pour répondre à ces limites, de nouvelles représentations graphiques doivent être élaborées, et notre proposition de diagramme mixte associé à une analyse harmonique constitue un support pour la conception de ces nouvelles représentations. De futurs axes de travail peuvent être l'extension de notre approche afin de permettre l'analyse d'événements étendus dans le temps, et l'intégration de représentations graphiques permettant d'identifier, et d'estimer, une évolution temporelle de la période de réapparition d'un comportement cyclique.

Remerciements

Cette recherche a été soutenue par la Région Auvergne-Rhône-Alpes.

Bibliographie

- Bony J.-M. (2001). *Cours d'analyse: théorie des distributions et analyse de fourier*. Editions Ecole Polytechnique.
- Carlis J. V., Konstan J. A. (1998). Interactive visualization of serial periodic data. In *Acm symposium on user interface software and technology*, p. 29–38.
- Edsall R., 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*, p. 182–189.
- Edsall R. M., Kraak M.-J., MacEachren A. M., Peuquet D. J. (1997). Assessing the effectiveness of temporal legends in environmental visualization. In *Gis/lis*, vol. 97, p. 28–30.
- Gasquet C., Witomski P. (2000). *Analyse de fourier et applications: filtrage, calcul numérique, ondelettes*. Dunod.
- Gautier J. (2018). *GrAPHiST: An exploratory analysis approach for the identification of dynamics of spatio-temporal phenomena*. Theses, Université Grenoble Alpes (France). Consulté sur <https://hal.archives-ouvertes.fr/tel-02067484>
- Hornsby K., Egenhofer M. J. (2002). Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence*, vol. 36, n° 1, p. 177–194.
- Katznelson Y. (2004). *An introduction to harmonic analysis*. Cambridge University Press.
- Li X. (2010). The time wave in time space: a visual exploration environment for spatio-temporal data.
- Müller W., 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*, p. 737–745.
- Swedberg B., Peuquet D. (2017). Perse: visual analytics for calendar related spatiotemporal periodicity detection and analysis. *GeoInformatica*, vol. 21, n° 3, p. 577–597.
- Van Wijk J. J., 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*, p. 4–9.
- Weber M., Alexa M., Müller W. (2001). Visualizing time-series on spirals. In *Infovis*, vol. 1, p. 7–14.

Qui a peur du changement climatique?

Lucile Sautot^{1,4}, Eric Chraïbi^{1,2}, Jacques Fize^{1,3},
Sébastien Peillet^{1,3}, Ludovic Journaux⁶,
Flavie Cernesson^{1,4}

1. UMR TETIS, AgroParisTech, CIRAD, CNRS, IRSTEA Maison de la
Téledétection 500 rue J.F. Breton 34093 Montpellier Cedex 5

2. *prenom.nom@irstea.fr*

3. *prenom.nom@cirad.fr*

4. *prenom.nom@agroparistech.fr*

5. Agrosup Dijon, Laboratoire d'Informatique de Bourgogne, 26 Bd Petitjean
21000 Dijon

prenom.nom@agrosupdijon.fr

RÉSUMÉ. Le Grand Débat National a été un événement politique important de l'année 2019. En utilisant les propositions en ligne récoltées lors de cette consultation, nous proposons une représentation du sentiment d'impact du changement climatique parmi les contributeurs du Grand Débat National en France Métropolitaine. Nous analysons les causes de ce sentiment, grâce à une analyse supervisée des contributions (via la Grande Annotation) et nous montrons l'intérêt complémentaire d'une analyse non-supervisée (par extraction de mots clés). Nous montrons la richesse du jeu de données que constitue le Grand Débat National, et les enjeux analytiques autour de ces données. Nous soulevons également quelques limites importantes à prendre en compte dont la représentativité du jeu de données.

ABSTRACT. The French "Grand Débat National" (Great National Debate) was an important political event in 2019. Using the online proposals collected during this consultation, we propose a representation of the feeling of the impact of climate change among the contributors to the "Grand Débat National" in Metropolitan France. We analyze the causes of this feeling, through a supervised analysis of the contributions (via the Great Annotation) and we show the complementary interest of an unsupervised analysis (by extracting keywords). We show the richness of the data set that

SAGEO'2019 - Clermont-Ferrand, 13-15 novembre 2019

SAGEO'2019

constitutes the "Grand Débat National", and the analytical stakes around these data. We also point out some important limitations to be taken into account, including the representativeness of the dataset.

MOTS-CLÉS : Grand Débat National, Changement climatique

KEYWORDS: Grand Débat National, Global warming

Préambule

Cet article est dédié à Monsieur Edouard Philippe, en remerciement du discours prononcé le 1er février 2019 à l'occasion des 80 ans du CNRS, où il indiquait que, nous citons, "*il serait mortifère, et probablement criminel, pour la démocratie que les scientifiques ne participent pas [au Grand Débat National], qu'ils n'apportent pas leur contribution, leur capacité à mettre en perspective, leurs avis dans ce grand débat national*". Dont acte.

1. Introduction

Le Grand Débat National a été un événement politique important de l'année 2019. Il a permis, via une consultation en ligne et l'organisation de plusieurs débats à travers le pays, de collecter l'opinion et les propositions des habitants sur plusieurs sujets de société considérés comme importants, plus particulièrement sur le domaine de la transition écologique.

Les contributions recueillies lors de ce débat sont actuellement proposées en open-data et révèlent une richesse thématique intéressante. Vu le nombre de contributeurs (134177 pour le questionnaire en ligne relatif à la transition écologique) et le nombre de réponses permettant une expression libre (12 questions sur 16) sur le sujet, un travail d'analyse important est alors nécessaire pour permettre de synthétiser les multiples réponses et en tirer des conclusions pertinentes.

Dans cet article, nous proposons de donner un premier aperçu des contributions recueillies dans le cadre de la thématique du "changement climatique". Outre la réponse directe sur ce que pensent les gens de l'impact du changement climatique sur leur vie quotidienne, le jeu de données constitue à lui seul un verrou scientifique du point de vue du traitement informatique. Notre objectif est donc ici de proposer quelques pistes permettant de décrire le sentiment des contributeurs vis à vis de l'impact du changement climatique à travers les territoires de la France métropolitaine.

Nous commencerons d'une part par décrire les données utilisées ainsi que les précautions de manipulation à prendre pour leur analyse (Section 2). D'autre part, dans le but d'illustrer nos propos, nous présenterons quelques résultats

intéressants dont la localisation des contributeurs et la répartition spatiale du sentiment d'impact du changement climatique. Enfin, avant de conclure (Section 4), nous proposerons quelques hypothèses explicatives quant à la peur du changement climatique vue par les contributeurs (Section 3).

2. Données utilisées

2.1. Sources des données

Les données extraites pour cette analyse proviennent du site du Grand Débat National¹. Nous nous sommes essentiellement concentrés sur les données concernant la transition écologique. Sur l'ensemble de ces données seules ont été considérées celles faisant référence aux réponses aux propositions et pas aux questionnaires rapides.

Nous avons également utilisé les données mises à disposition par la Grande Annotation². Ce projet a pour objectif de faire annoter les contributions enregistrées dans le cadre du Grand Débat National par des annotateurs. En effet, l'analyse directe des questions ouvertes n'est pas possible car l'intelligence artificielle seule ne permet pas de restituer fidèlement les idées, les opinions et les sentiments exprimés par les participants. Pour palier à ce problème, il a été décidé de faire appel aux internautes à transcrire les réponses ouvertes en réponses synthétiques et compréhensibles par les outils informatiques.

2.2. Description des données sélectionnées

Pour réaliser notre analyse, quelques prétraitements et sélections de données sont nécessaires. Cette préparation est réalisée en 4 points:

1. Tout d'abord, nous utilisons les codes postaux renseignés par les contributeurs dans leurs propositions. Parmi les 134177 personnes ayant répondu au questionnaire "Transition écologique", 131472 ont donné un code postal exploitable permettant de les localiser. Les données issues du Grand Débat ont téléchargées le 29/03/2019.

2. Ensuite, nous utilisons les réponses à deux questions issues des propositions relatives à la transition écologique :

- a) La question 146 : Diriez-vous que votre vie quotidienne est aujourd'hui touchée par le changement climatique ? (réponses possibles : oui/non/absence de réponse)

- b) La question 147 : Si oui, de quelle manière votre vie quotidienne est-elle touchée par le changement climatique ? (réponse ouverte)

1. <https://granddebat.fr/pages/donnees-ouvertes>

2. <https://grandeannotation.fr/data>

Parmi les 134177 personnes localisables, nous disposons de 128129 réponses non nulles à la question 146 et de 86544 réponses non nulles à la question 147.

3. Ensuite, les réponses de 4502 personnes ont été annotées via le site de la Grande Annotation. Ces annotations correspondent à celles téléchargées le 15/05/2019, et sont des annotations convergentes. Chaque réponse est annotée selon les catégories présentées dans le Tableau 1.

4. Enfin, nous utilisons l'ensemble des propositions relatives à la transition écologique pour réaliser une extraction de mots clés.

2.3. Précautions vis-à-vis des données

Les contributions du Grand Débat National ne sont pas exemptes de biais. Ainsi, les résultats que nous présentons dans cet article sont le reflet de l'opinion des contributeurs, mais ne peuvent en aucun cas être considérés comme un reflet fidèle de l'opinion de la population. En effet, la méthode d'enquête de ce débat le rend plus proche de la production participative (crowdsourcing) ou du réseau social, que du sondage. La différence entre les statistiques classiques et l'analyse des réseaux sociaux (et plus généralement l'analyse de contenu généré par les utilisateurs) est bien connue, mais son importance a néanmoins été reconnue même en présence de différents types de biais (Baeza-Yates, Sáez-Trumper, 2016 ; Morstatter, Liu, 2017 ; Tufekci, 2014).

En effet, aucune donnée concernant la sociologie des contributeurs (âge, sexe, catégorie socio-professionnelle, ...) n'est présente dans le jeu de données. Il n'est donc pas possible de s'assurer que les contributeurs soient un échantillon représentatif de la population. La seule donnée aisément accessible est le code postal renseigné par le contributeur. (Bennani *et al.*, 2019).

Concernant les données de la Grande Annotation, il n'y a pas d'information sur le choix des questionnaires qui ont été annotés (environ 4612 parmi les 92795 questionnaires non vides pour les réponses traitées). La méthode de sélection des contributions annotées n'est pas explicitement décrite.

3. Résultats

3.1. Localisation des contributeurs

Nous proposons, sur les Figures 1, 2 et 3, une représentation cartographique du nombre de contributeurs. Pour représenter spatialement ces données, nous avons utilisé les codes postaux déclarés par les utilisateurs. Environ 6503 personnes (soit environ 5% des contributeurs) n'ont pas déclaré de code postal exploitable : leurs réponses ne seront pas prises en compte.

La Figure 1 montre le nombre de contributeurs répartis dans une maille hexagonale, où le centre de chaque maille est distant de dix kilomètres de

Qui a peur du changement climatique?

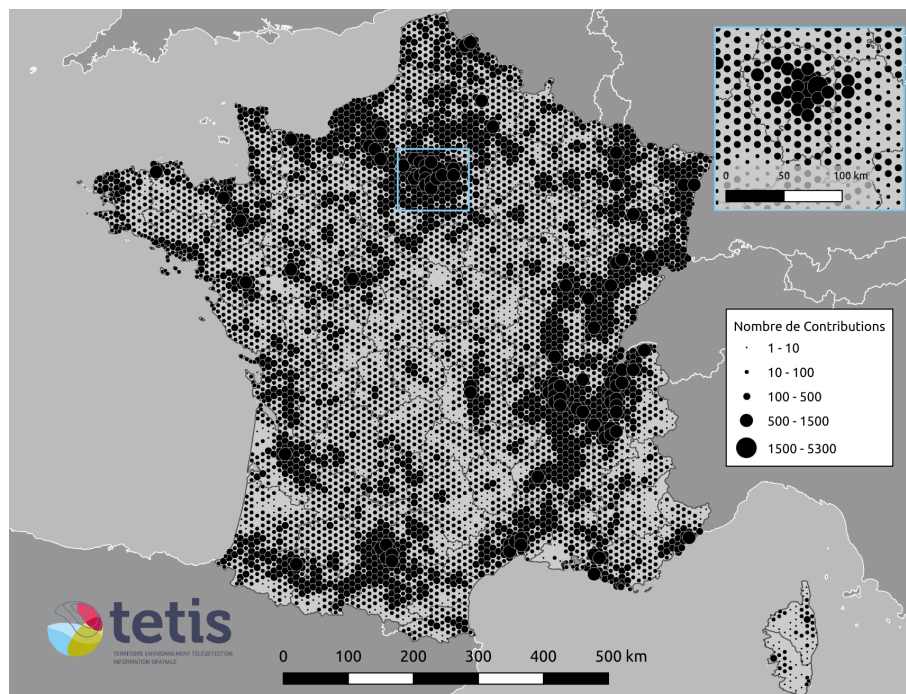


FIGURE 1. Localisation des contributeurs en France métropolitaine

ses voisins. Un code postal est attribué à une maille quand le barycentre de la zone couverte par le code postal est inclus dans la maille. Sur 6300 codes postaux possibles, 5911 sont déclarés par des contributeurs. Les contributeurs sont donc présents sur une grande partie du territoire national. Les zones où les contributeurs sont les plus nombreux sont : la région parisienne, la vallée de la Seine, la vallée du Rhône, la vallée de la Saône, la région toulousaine. Certaines zones représentées sans contributions (les landes, le delta du Rhône, la frontière italienne) correspondent à des zones où les communes sont très peu densément peuplées.

La Figure 2 montre une vue agrégée par département des données de la Figure 1. La Figure 3 montre le nombre de contributeurs pour 10000 habitants, agrégé par département. Sur cette Figure, on constate que la participation a été moins importante en région parisienne alors qu'elle a été importante dans le Lot, en Haute-Saône, dans le Jura, dans la Meuse. Ainsi, le nombre de contributeurs ne semble pas directement lié au nombre d'habitants.

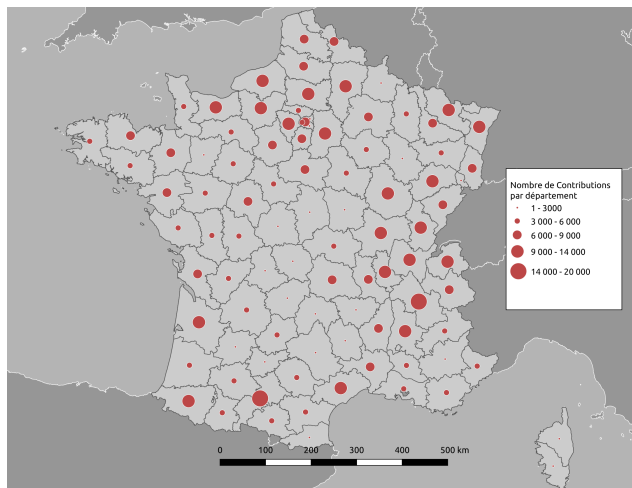
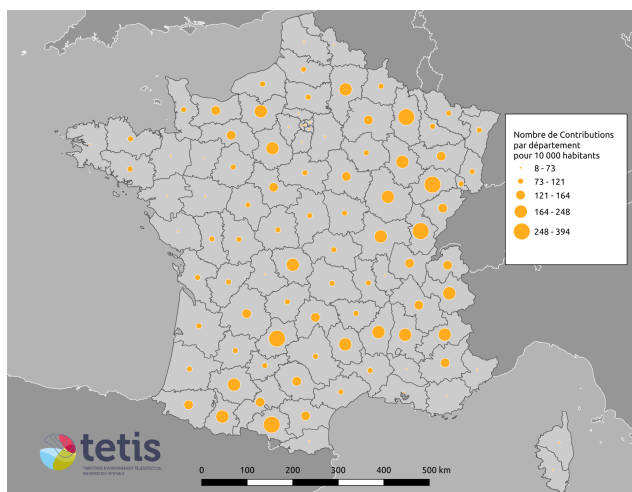


FIGURE 2. Nombre de contributeurs par département en France métropolitaine.



La représentation non conventionnelle d'un ratio par des symboles proportionnels a été choisie avec pour unique but de faciliter la comparaison avec la figure 2

FIGURE 3. Nombre de contributeurs pour 10000 habitants par département en France métropolitaine.

3.2. Analyse des réponses aux questions 146 et 147

Concernant la question 146 (*Diriez-vous que votre vie quotidienne est aujourd'hui touchée par le changement climatique ?*), 64% des contributeurs ont répondu oui, 29% ont répondu non et 7% n'ont pas répondu.

Qui a peur du changement climatique?

Pour représenter le sentiment d'impact du changement climatique, nous exprimons ce dernier par un indice de sentiment estimé par la relation suivante :

$$\text{sentiment} = \frac{\text{Nombre de non} - \text{Nombre de oui}}{\text{Nombre de non} + \text{Nombre de oui}} \quad (1)$$

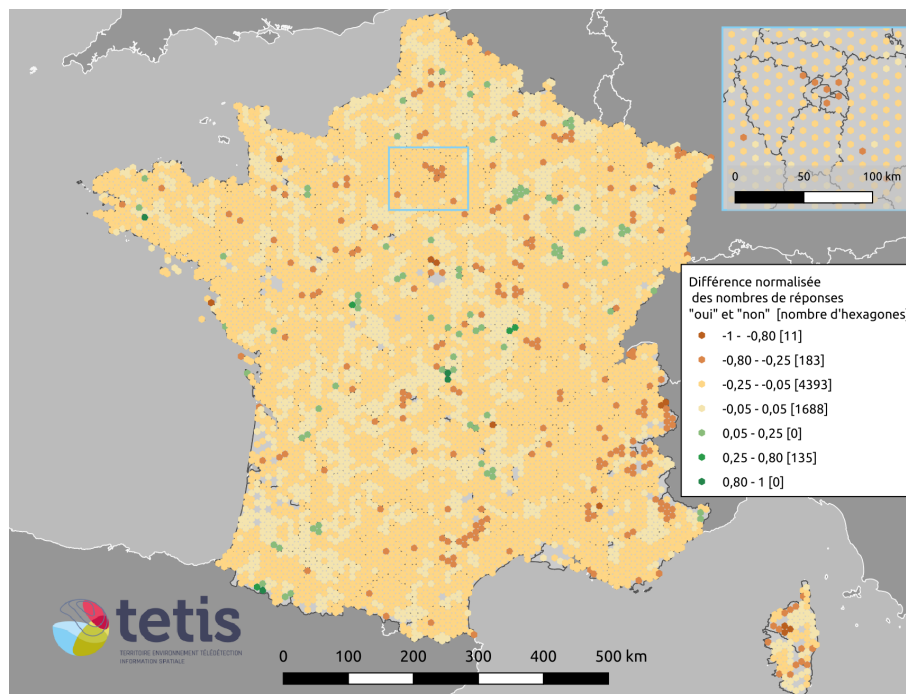
Ainsi, plus l'indice est négatif, plus un nombre important de personnes pense que le changement climatique impacte leur vie quotidienne. Au vu de la grande disparité dans le nombre de contributeurs par point (voir la Figure 1), la division par le nombre de réponses non nulles permet de pondérer l'indice et comparer les différentes zones géographiques. Cet indice a été conçu pour permettre la représentation graphique des réponses binaires (oui/non).

La Figure 4 présente la répartition géographique de l'indice présenté ci-dessus. Nous avons conçu notre échelle de sorte que la classe "neutre" (une proportion quasiment égale de oui et de non) soit la plus petite possible. Sur une majeure partie du territoire national, les contributeurs qui considèrent que le changement climatique présente un impact sur leur vie quotidienne sont plus nombreux que ceux considérant qu'il n'a pas d'impact. On remarque cependant que dans certaines zones, il y a plus de contributeurs qui ne constatent pas d'impact du changement climatique sur leur vie quotidienne. Ainsi, au Nord de Moulins, entre Poitiers et Tours, au Nord de Clermont-Ferrand, au Pays-Basque à la frontière espagnole, et à l'Est de Quimper, on trouve quelques communes où les contributeurs sont plus nombreux à considérer que le changement climatique n'a pas d'impact sur leur vie quotidienne.

Il s'agit à présent de proposer des hypothèses permettant d'interpréter cette répartition spatiale.

Après analyse, seules quelques communes présentent une majorité de contributeurs pensant que le changement climatique n'a pas d'impact sur la vie quotidienne. Cette observation nous oriente sur une première hypothèse sur les différences sociales entre ces territoires et les politiques publiques qui y ont été menées, conduisant à une perception du sujet différente (Dumont, 2019).

Une seconde hypothèse pourrait être la confrontation des contributeurs à des événements climatiques particuliers. Par exemple, la contribution suivante semble s'appuyer sur une expérience personnelle : "*Canicules et sécheresse à répétition à Grenoble. A cela s'ajoute une pollution de l'air fréquente: des habitants déménagent pour aller habiter dans la montagne à cause des problèmes d'air (asthme, etc.), notamment de jeunes parents. La boucle est bouclée: ils doivent ensuite prendre la voiture pour se rendre à leur lieu de travail, et contribuent de plus belle à la pollution locale. Heureusement des solutions alternatives se développent localement (bus propre, vélo, etc.)!*". Des travaux ont tentés de montrer le lien entre événements climatiques et perception du changement climatique (Akerlof *et al.*, 2013 ; Borick, Rabe, 2014 ; Egan, Mullin, 2012) sans réussir à montrer une corrélation immédiate.



Dans cette carte, la taille des points n'est pas proportionnelle au nombre de contributeurs.

FIGURE 4. *Le sentiment des contributeurs de l'impact du changement climatique sur leur vie quotidienne*

La question 147 (*Si oui, de quelle manière votre vie quotidienne est-elle touchée par le changement climatique ?*) apparaît comme pouvant apporter des éléments de réponses à la répartition géographique observée sur la Figure 4. Son traitement présente deux écueils importants : premièrement, il s'agit d'une réponse en texte libre, ne pouvant donc être facilement analysée sur un grand corpus; deuxièmement, elle ne concerne que les gens sentant un impact du changement climatique, car aucune question n'interroge les personnes ne sentant pas d'impact du changement climatique.

En première approche, nous proposons de nous appuyer sur les données de la Grande Annotation, qui ont attribué une catégorie à 4612 réponses à cette question 147. Notez qu'une réponse peut appartenir à plusieurs catégories.

Le Tableau 1 présente les résultats obtenus grâce à la Grande Annotation pour les réponses à la question 147, en termes de thématiques et de nombre de contributeurs. La pollution semble être la préoccupation majeure des contributeurs. Malgré les résultats obtenus par (Akerlof *et al.*, 2013 ; Borick, Rabe,

Qui a peur du changement climatique?

2014; Egan, Mullin, 2012), on peut noter que les événements climatiques extrêmes représentent la deuxième catégorie la plus attribuée.

TABLE 1. Nombre de contributeurs dans chaque catégorie annotée par la Grande Annotation pour la question 147

Catégorie	Nb de contributeurs
Pollution (air, eau, sol)	1411
Événements extrêmes (incendie, inondation, sécheresse, tempêtes)	1097
Saisons modifiées, dérégulées	975
Canicule, nécessité de la clim	785
Maladies, santé	622
Modification de l'environnement (biodiversité, paysage)	569
Impact financier	141
Manque d'eau	92
Impact psychologique	76
Effort écolo	59
Arrivée de migrants climatiques	46
Hausse du niveau des mers	34
Chauffage l'hiver	13
Idee marquée	2
Sans réponse / Hors sujet / Inclassable	81
Autres	80
Je ne suis pas touché	83

3.3. Extraction de mots clés

Pour compléter ces résultats obtenus via une annotation, nous avons réalisé une extraction de mots clés non supervisée de l'ensemble des propositions. Nous avons ensuite recherché les verbes les plus fréquemment associés à chaque mot-clé. Plus précisément, Les mot-clés sont extraits de manière automatique à l'aide de l'approche proposée dans (Lossio-Ventura *et al.*, 2014). Cette approche a été proposée pour des données textuelles dans le domaine biomédical mais elle a fait ses preuves dans d'autres domaines comme l'agronomie (Roche *et al.*, 2015).

Enfin, les relations de cooccurrence entre mot-clé et adjectif sont extraites à l'aide d'une fenêtre glissante, c-a-d si un adjectif apparaît dans une fenêtre autour d'un mot-clé.

Enfin, nous avons représenté ces relations sous forme de graphe ³ (voir Figure 5), en conservant pour chaque mot clé les cinq verbes les plus fréquemment associés. Cette représentation permet de préciser l'opinion des contributeurs, par rapport aux catégories de la Grande Annotation, en offrant une richesses sémantique bien plus importante.

3. Le graphe est visible à cette adresse : <http://jacquesfize.com/gdn/index.html>

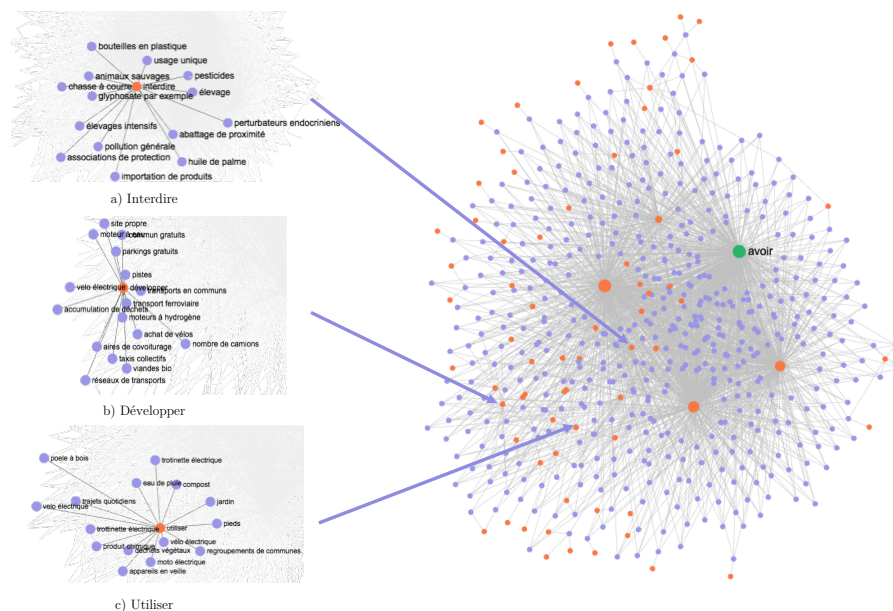


FIGURE 5. Aperçu du graphe des mots-clés identifiés dans la totalité des réponses aux questions concernant la transition écologique

4. Conclusion

En conclusion, nous avons proposé une première représentation du sentiment d'impact du changement climatique parmi les contributeurs du Grand Débat National en France Métropolitaine. Nous avons également amorcé une analyse plus fine, sur les causes de ce sentiment, grâce à une analyse supervisée des contributions (via la Grande Annotation) et une analyse non-supervisée (par extraction de mots clés) qui est à exploiter plus en détails. Nous avons, dans cet article, montré la richesse du jeu de données que constitue le Grand Débat National, et les enjeux analytiques autour de ces données. Nous avons également soulevé quelques limites importantes à prendre en compte dont la représentativité du jeu de données.

Nos travaux futurs se concentreront sur la poursuite de nos analyses des contributions textuelles, et des localisations des contributeurs, afin de mieux cerner les raisons du sentiment d'impact du changement climatique. La question reste donc ouverte : qui, en France métropolitaine, a peur du changement climatique ?

Bibliographie

- Akerlof K., Maibach E. W., Fitzgerald D., Ceden A. Y., Neuman A. (2013). Do people “personally experience” global warming, and if so how, and does it matter? *Global Environmental Change*, vol. 23, n° 1, p. 81–91.
- Baeza-Yates R. A., Sáez-Trumper D. (2016). Wisdom of the crowd or wisdom of a few? an analysis of users’ content generation. *CoRR*, vol. abs/1604.03044. Consulté sur <http://arxiv.org/abs/1604.03044>
- Bennani H., Gandré P., Monnery B. (2019). Les déterminants locaux de la participation numérique au grand débat national : une analyse économétrique. Consulté sur https://economix.fr/pdf/dt/2019/WP_EcoX_2019-7.pdf?1.0
- Borick C. P., Rabe B. G. (2014). Weather or not? examining the impact of meteorological conditions on public opinion regarding global warming. *Weather, Climate, and Society*, vol. 6, n° 3, p. 413–424.
- Dumont G.-F. (2019). Si les gilets jaunes m’étaient contés. . . . *Population & Avenir*, n° 742, p. 3-3. Consulté sur <https://www.cairn.info/revue-population-et-avenir-2019-2-page-3.html>
- Egan P. J., Mullin M. (2012). Turning personal experience into political attitudes: The effect of local weather on americans’ perceptions about global warming. *The Journal of Politics*, vol. 74, n° 3, p. 796–809.
- Lossio-Ventura J. A., Jonquet C., Roche M., Teisseire M. (2014). Biomedical Terminology Extraction: A new combination of Statistical and Web Mining Approaches. , p. 13.
- Morstatter F., Liu H. (2017, 1^{er} 6). Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media*, vol. 1, p. 1–13.
- Roche M., Fortuno S., Lossio-Ventura J. A., Akli A., Belkebir S., Lounis T. *et al.* (2015, septembre). Automatic extraction of keywords from scientific publications for indexing and open data in agronomy. *Cahiers Agricultures*, n° 5, p. 313–320. Consulté sur <http://revues.cirad.fr/index.php/cahiers-agricultures/article/view/31146>
- Tufekci Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. *CoRR*, vol. abs/1403.7400. Consulté sur <http://arxiv.org/abs/1403.7400>

Combiner cartes à dire d'acteurs et lectures de paysage pour analyser les circuits agricoles et alimentaires de la Vallée de la Levrière

Un itinéraire méthodologique innovant

Sylvie Lardon¹, Elisa Marraccini², Davide Rizzo², Marc Benoit³

1. UMR Territoires, INRA & AgroParisTech

9, Avenue Blaise pascal, CS 20085, 63178 Aubière Cedex, France

sylvie.lardon@agroparistech.fr

2. UP INTERACT 2018.C102 UniLaSalle, Beauvais, France

3. UR ASTER, INRA 662 avenue Louis Buffet.F-88500 Mirecourt, France

marc.benoit@inra.fr

RESUME. Pour aborder la complexité des systèmes agricoles et alimentaires dans les territoires, une approche systémique est requise, tant les dynamiques sont inter-reliées et nécessitent des regards croisés. Nous proposons une approche de diagnostic prospectif participatif valorisant la transversalité et l'articulation des échelles. L'itinéraire méthodologique suivi vise à faire ressortir les perceptions et motivations quant à ce territoire de vie, et ce, à travers des cartes à dire d'acteurs et des lectures de paysage in situ ainsi que l'organisation d'un jeu de territoire. Les diagnostics thématiques retranscrivent des informations obtenues lors des entretiens individuels menés à travers le territoire, et se voient enrichis des discussions et débats qui se sont tenus lors d'une après-midi de rencontre entre acteurs. L'étape des scénarii et des actions reprend les idées et les volontés évoquées durant cette journée à propos de l'avenir de la Vallée et des moyens à mettre en œuvre pour le développement de ce territoire. Cette approche empirique, qui part du terrain et des objectifs pédagogiques de formation des étudiants, pour répondre aux objectifs de l'action et réinterroger les concepts et méthodes de la recherche, donne un fil directeur pour penser la genericité de l'itinéraire méthodologique et un cadre d'analyse pour les transitions territoriales. Ainsi, tout au long de l'itinéraire méthodologique, il pourrait y avoir des traitements géomatiques complémentaires pour consolider les analyses, extraire les connaissances et faciliter la participation.

MOTS-CLES : DIAGNOSTIC DE TERRITOIRE, PROSPECTIVE, PARTICIPATION,

SAGEO'2019 – Clermont-Ferrand, 13-15 novembre 2019

SAGEO'2019

ABSTRACT: *To address the complexity of farming and food systems at the territorial level, a systemic approach is required, as the dynamics are interrelated and require cross-referencing. We propose a prospective participatory diagnostic approach that emphasizes transversality and the articulation of scales. The methodological itinerary followed aims to highlight the perceptions and motivations for this territory of life, through actors' maps and readings of landscape in situ and the organization of a territory game. The thematic diagnosis transcribes information obtained during individual interviews conducted throughout the territory, and are enriched by discussions and debates held during an afternoon of meetings between actors. The stage of the scenarios and actions takes again the ideas and the wills evoked during this day about the future of the Valley and the means to be implemented for the development of this territory. This empirical approach, which starts from the field and the educational objectives of student training, to answer the objectives of the action and to re-examine the concepts and methods of the research, gives a guiding thread to think the genericity of the methodological itinerary and a framework of analysis for territorial transitions. Thus, all along the methodological itinerary, there could be complementary geomatics treatments to consolidate the analyzes, extract the knowledge and facilitate the participation.*

KEYWORDS: *TERRITORIAL DIAGNOSIS, PROSPECTIVE, PARTICIPATION*

1. Introduction

Les territoires sont en transition et voient l'accélération et le foisonnement d'initiatives et de procédures, émanant tant des initiatives locales que des incitations globales. La territorialisation des politiques publiques et la coordination des actions collectives sont au cœur des dynamiques actuelles de changement dans les territoires (Amblard et al., 2017). Dans le champ de l'agriculture et de l'alimentation, les enjeux touchent aux modalités d'articulation et d'intégration des différents acteurs, activités et espaces dans les territoires (Benoît et al., 2006). De nombreux diagnostics de territoire sont réalisés, pour des enjeux de plus en plus transversaux, par exemple sur les questions de transitions alimentaires (Lardon et al., 2016).

Pour aborder la complexité des territoires, paysages, systèmes agricoles et alimentaires, une approche systémique est requise, tant les dynamiques sont inter-reliées et nécessitent des regards croisés (Caron et al., 2017). L'identification et la valorisation des ressources sont à prendre en compte en termes de potentiels pour l'avenir, ce qui engage nos capacités d'anticipation et de construction du futur et fait appel aux démarches de prospective territoriale (Lardon et Noucher, 2016). Comment rendre plus fluides les interconnexions nécessaires entre acteurs et chercheurs ? Comment adapter la diversité des méthodes à la diversité des points de vue et des enjeux portés par les acteurs des territoires ? Quels changements opérer dans nos représentations et nos façons de travailler pour aborder la complexité du monde de demain ?

Nous proposons une approche de diagnostic prospectif participatif valorisant la transversalité et l'articulation des échelles (Lardon et Piveteau, 2005). Nous partons d'une démarche réalisée, avec une entrée délibérément partenariale et pédagogique. Le fil directeur de notre analyse est un travail co-construit avec des étudiants du master DEBATs¹ que nous avons encadrés en décembre 2018, en partenariat avec le CAUE de l'Eure² et l'Association de la vallée de la Levrière (AVL). Dans le cadre du programme « Paysages en transitions territoriales », les partenaires de l'Eure, avec leur projet « Vallées Habitées »³ s'interrogeaient sur la redynamisation des vallées par la relocalisation des circuits alimentaires. Nous avons mené une démarche de recherche-formation-action (Lardon et al., 2015), où la formation des étudiants est le moteur à la fois d'une mobilisation des acteurs et des chercheurs pour produire des connaissances scientifiques et pour l'action.

¹ Chaire Unesco DEBATs : Développement durable, biodiversité et aménagement des territoires (<https://www.chaire-unesco-debats.org/master-debats/>)

² Conseil d'Architecture, d'Urbanisme et d'Environnement de l'Eure (<https://caue27.fr/>)

³ Projet Vallées Habitées porté par le CAU27 et AVL (<https://vallees-habitees.fr/>)

L'itinéraire méthodologique suivi vise à faire ressortir les perceptions et motivations quant à ce territoire de vie, et ce, à travers des cartes à dire d'acteurs et des lectures de paysage *in situ* ainsi que l'organisation d'un jeu de territoire. Les diagnostics thématiques retranscrivent des informations obtenues lors des entretiens individuels menés à travers le territoire, et se voient enrichis des discussions et débats qui se sont tenus lors d'une après-midi de rencontre entre acteurs sur l'avenir de la Vallée et les moyens à mettre en œuvre pour le développement de ce territoire.

2. Matériels et Méthodes

2.1. La Vallée de la Levrière

La Levrière est une rivière Normande située entre le Vexin Normand et le Pays de Bray, à l'extrémité Est de l'Eure. Elle prend sa source en lisière forestière à Bézu-la-Forêt et rejoint l'Epte à Gisors, 24 km plus au Sud-Ouest. La vallée se trouve à proximité de Beauvais, entre Paris et Rouen. Très faiblement urbanisée, la vallée est ponctuée de petits centres-bourgs (dont Bézu-la-Forêt, Martagny, Mainneville, Mesnil-sous-Vienne, Hébecourt, Sancourt, Saint-Denis-le-Ferment qui sont adhérent à l'AVL avec la commune d'Amécourt, située dans la vallée de l'Epte).

2.2. L'itinéraire méthodologique des circuits alimentaires de la vallée de la Levrière

Pour répondre aux préoccupations de l'AVL et du CAUE27, à savoir une connaissance des atouts et contraintes de la vallée de la Levrière, de la vision de ses acteurs, des actions concrètes pouvant avoir un vrai impact socio-économique, les étudiants ont co-construits collectivement la question suivante : « Quels circuits pour la vallée de la Levrière d'ici à 2040 ? ». Ils ont suivi l'itinéraire méthodologique enseigné par Benoît et al. (2006).

2.2.1. Etape préparatoire

Le travail a commencé par une présentation des objectifs de l'exercice par l'équipe pédagogique, à la suite de quoi l'ensemble de la promotion a pu discuter à propos de la commande afin de mieux l'appréhender. Cette phase de réflexion et d'appropriation s'est prolongée lors de la reformulation de la question commune à tous les acteurs mobilisés sur le territoire de la Levrière. L'étape suivante correspond à l'élaboration du fond de carte, qui fera office de support par la suite pour l'ensemble des entretiens. Il fournit quelques points de repère mais évite de livrer trop d'informations, pour ne pas influencer le dire des acteurs. Enfin, la dernière fut la prise de contact par téléphone avec des acteurs du territoire de la vallée de la Levrière, afin de les interviewer. Le choix des acteurs a été fait à partir de la liste fournie par les commanditaires (principalement élus et agriculteurs, ainsi que les membres de l'association), complétée par des contacts directs à l'initiative des étudiants (commerçants, autres professionnels) et au hasard des rencontres sur le terrain.

2.2.2. Etape de déambulation in situ

La prise de contact avec la vallée de la Levrière a commencé par un parcours sur le territoire pour découvrir le patrimoine naturel, architectural, historique et paysager de la vallée. Cette étape aura permis de relever certains éléments saillants faisant la particularité du territoire en mobilisant des lectures de paysages en petits groupes. C'est aussi une façon d'entrer dans le paysage par une déambulation individuelle et collective et laisser libre court aux premières impressions.

La première rencontre avec les commanditaires (CAUE27 et AVL) visait à ré-expliciter les attentes de ces derniers et à s'accorder sur des objectifs partagés. Ainsi, si la question présentée lors de l'étape préparatoire (portant sur le développement de "circuits courts") semblait avant tout porter sur les systèmes alimentaires, le groupe s'est accordé sur une vision plus large des circuits, pour l'ouvrir à tout type de circuits (agricoles, alimentaires, énergétiques, circulations, touristiques, etc.).

Par groupes, il y a eu la rencontre avec des résidents de la vallée aux profils variés (élus, agriculteurs, enseignants, commerçants... et des territoires limitrophes (vallée de l'Epte, plateaux avoisinants, et même jusqu'à Gisors). La méthode des cartes à dire d'acteurs a été mobilisée pour ces rencontres, consistant à utiliser le fond de carte simplifié élaboré au préalable et à recueillir les dires des personnes interrogées (données chaudes), ces dernières spatialisant leurs propos sur la carte. Cette mise en territoire de leurs discours permet une combinaison des divers dires d'acteurs qui sont ainsi spatialement comparables. Ces cartes à dire d'acteurs récoltées par les différents groupes ont ensuite été mutualisées. De grands thèmes récurrents et structurants ont été dégagés et les informations s'y rapportant ont été reportées sur des cartes thématiques par des chorèmes. Ces cartes ont été enrichies au fur et à mesure que de nouvelles informations étaient recueillies.

2.2.3. Jeu de territoire

Après avoir pris connaissance des règles du jeu de territoire (Lardon, 2013), le groupe s'est attelé à sa préparation : (1) Mise au propre des cartes chorématiques issues des cartes à dire d'acteurs et élaboration d'un narratif correspondant afin de servir de base à l'étape "diagnostic" du jeu de territoire ; (2) Elaboration de fiches actions sur base des pistes évoquées par les acteurs rencontrés et sur base d'observation des étudiants ; (3) Répartition des participants par table en veillant à respecter un certain équilibre entre les types d'acteurs ; (4) Répartition des rôles à chacun des étudiants pour l'animation de l'atelier.

Lors du diagnostic, les étudiants ont présenté les cartes chorématiques réalisées par eux, qui ont été soumises ensuite à discussion avec les participants, pour une appropriation collective du diagnostic. Ces cartes affichées au mur, ont été enrichies par ces échanges. Pendant la phase de scénarii et d'actions, les participants ont répondu à la question suivante : "Quels circuits pour la vallée en 2040 ?" dans un exercice de réflexion prospective. Au cours de cette étape, les participants ont

spatialisé leurs ambitions et rêves pour la Vallée sur le fond de carte commun. Chaque scénario a ensuite été nommé par leurs auteurs de manière à faire ressortir l'idée centrale. A l'issue d'une étape de scénario où tout était possible, s'est ouverte une étape visant à identifier les actions concrètes permettant de réaliser ces scénarios souhaités : Quoi ? Qui ? Où ? Quels partenariats ? Comment ? Avec quels financements ? Il était attendu que ces actions soient spatialisées sur le fond de carte. Enfin lors de la mise en commun et du débat, les participants ont présenté leurs scénarii respectifs par table, exposant aux autres groupes leur démarche et leurs idées. Puis les actions identifiées par les différents groupes ont été présentées et débattues et ont été identifiées les idées communes et divergentes entre les groupes.

3. Le diagnostic à partir de l'analyse du paysage et des cartes à dire d'acteurs

Deux méthodes spécifiques ont été mobilisées pour que les étudiants s'imprègnent des enjeux du territoire : l'analyse de paysage (Benoit et al, 2006) et les cartes à dire d'acteurs (Debolini *et al.*, 2012).

3.1. Perceptions paysagères

Le paysage a été pris dans une approche cohérente avec la définition qu'en donne la Convention Européenne : « une partie de territoire telle que perçue par les populations » (art. 1a). Afin de permettre aux étudiants d'interagir avec les habitants, le groupe a conduit une analyse de paysage, dans laquelle une attention particulière est portée sur la formalisation des premières impressions (Brooks, 1998). L'analyse de la vallée s'est faite à partir des 4 communes : Bézu-la-Forêt, Martigny, Mainneville et Hébecourt situées dans la vallée de la Levrière. Celles-ci sont structurées autour de petits centres bourgs, densément urbanisés et ce, sur un espace restreint. La Levrière, est peu visible et peine à s'affirmer vis-à-vis du bâti. Les habitations s'étendent généralement sur les plateaux. Le paysage de la vallée est structuré par un alignement de végétaux et un relief offrant des dénivelés de part et d'autre. En outre, il présente des courbures qui incitent à découvrir le site. On trouve parfois des percées qui permettent de dégager une vue sur l'étendue de ses plateaux, stratifiés par la présence de différents éléments structurants : bâtis, bosquets d'arbres, haies, grillages, etc. L'élevage est l'activité principale. Les contraintes existantes sont la privatisation des espaces de la vallée. La présence de clôtures, de murs et de grillages délimitent des parcelles plus ou moins grandes et empêchent l'accès aux promeneurs, ce qui tend à appuyer le manque d'espace de rencontre.

3.2 Diagnostic

Le diagnostic a été élaboré à partir des impressions des étudiants et des expressions des acteurs et synthétisé sous forme de chorèmes qui donnent à voir les principales dynamiques du territoire : pôles d'attraction externes qui vident le territoire ; manque de cohésion sociale ; présence d'un patrimoine naturel riche mais fragile ; le bâti touristique comme un atout historique ; des productions agricoles qui sont en large majorité exportée en dehors du territoire.

Selon les dires d'acteurs, ce territoire est un lieu "vidé" par des pôles d'attraction externes et éparpillés. Les principaux axes routiers se trouvent en dehors de cette zone « en creux » et les transports ci sont quasi-inexistants (deux transports scolaires et absence de transports en commun). La plupart des services sont situés en dehors de la vallée. Certaines activités économiques (ex. touristiques) au sein de la vallée sont alimentées par une demande extérieure, notamment de Paris. Cela soulève le problème d'une très faible consommation des produits et services locaux par la (faible) population résidente.

Suite aux observations paysagères et aux informations fournies puis partagées par les différents acteurs locaux rencontrés, la vallée a révélé ses potentiels et faiblesses en termes d'attractivité économique et de richesse productive. Afin de mieux révéler les caractéristiques partagées de la vallée de la Levrière, il est vite apparu important pour les acteurs d'intégrer ses dynamiques économiques et ses activités agricoles à une échelle plus large, prenant en compte la vallée de l'Epte, constitutive d'une même entité paysagère et la ville de Gisors qui structure le territoire. Cette échelle de réflexion semblait plus pertinente dans l'hypothèse d'une volonté de différenciation et de rayonnement des produits locaux à l'échelle départementale, régionale, voire même nationale. La vallée de la Levrière se situe dans une situation d'entre deux, de lisière, et manque en elle-même d'identité propre puisqu'elle se situe à la frontière entre le Pays de Bray et le Vexin Normand.

Le territoire des vallées de la Levrière et de l'Epte représente un bassin de production agricole important et diversifié (élevage de porcs ou de bovins (lait et viande), la pisciculture, les cultures (betteraves, lin, colza, luzerne, orge, blé...), les vergers, l'exploitation forestière (dont la finalité nous était, encore à ce stade, inconnue), l'héliciculture (Figure 1). Mais ces produits peinent à être valorisés localement malgré la présence d'industries agro-alimentaires : une boulangerie-fromagerie artisanale, une sucrerie et une usine de déshydratation. L'un des freins à un potentiel développement local semble la forte ouverture vers l'extérieur : forte tendance à l'exportation (demande essentiellement extérieure et notamment, importants bassins de consommation à Rouen et Paris) et services de transformation situés à l'extérieur de la vallée, comme c'est le cas pour le service d'abattage.

L'esquisse d'un circuit local est néanmoins à relever : quelques productions sont en vente directe mais restent très fortement concurrencées par les supermarchés de Gisors. Le marché de Gisors est aussi touché par le phénomène d'attraction de ces établissements et observe une nette baisse de fréquentation et une faible représentation des produits locaux à la vente. Finalement, ces initiatives, encore timides et isolées, sont le fait d'un potentiel manque de perspective de rentabilité à l'échelle de la vallée. Ceci, en raison de la faible densité démographique, d'un pouvoir d'achat contrasté, mais aussi du manque de mise en réseau et de structuration (en termes de commercialisation) des différents producteurs et éleveurs locaux qui ne parviennent pas à concurrencer l'offre extérieure des grands distributeurs et à créer une identité particulière à leurs produits.

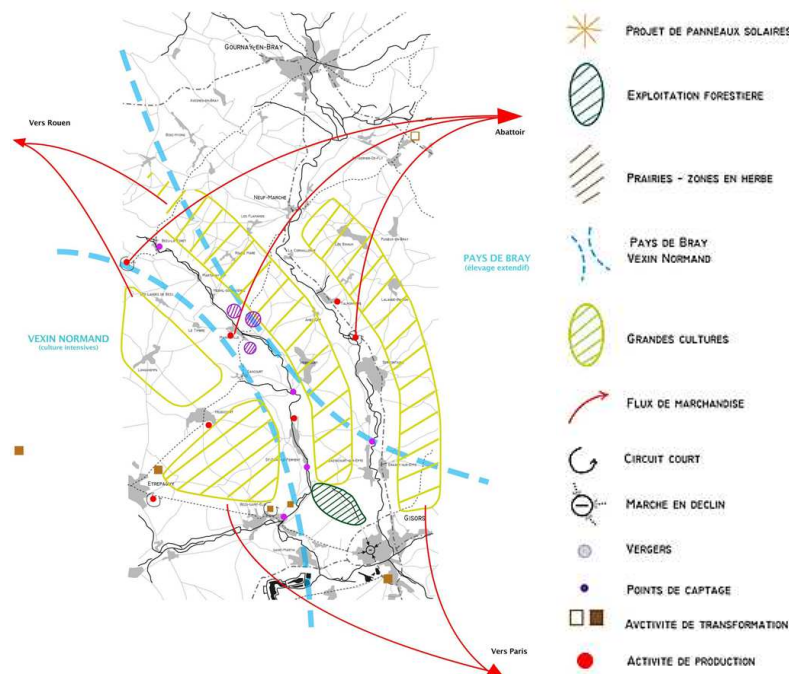


FIGURE 1. Filière locale : un bassin d'activités majoritairement exportées

3.3. La prospective et les idées d'action

Répondant à la question “Quels circuits pour la Vallée en 2040 ?”, trois tables ont réfléchi, en parallèle, à des scénarii et des pistes d'action pour la vallée. Nous rendons compte brièvement des propositions de l'une d'entre elles.

Les acteurs de la table 2 se sont focalisés sur une “Vallée Buzy/Bises/Bizz Ness”. Trois thèmes majeurs et incontournables sont ainsi représentés par ce titre : l'économie (Buzy), le social et la convivialité (Bises) et l'écologie (Bizz, rappelant le son des abeilles et ainsi, de manière indirecte la préservation de la biodiversité). Il s'agirait de faire de la vallée, un territoire convivial, productif et durable.

Les acteurs ont discuté du périmètre dans lequel devaient s'inscrire ces différentes réflexions. Ils se sont finalement accordés sur deux échelles imbriquées : l'une centrée autour de la vallée de la Levrière et l'autre plus large et englobant la vallée voisine et les villes alentours. L'idée étant de parvenir à créer une dynamique locale sur la vallée, grâce à une concertation, voire une unification communale autour des différents thèmes abordés, tout en s'inscrivant dans une zone d'influence plus large permettant de mutualiser les actions, les réglementations et les projets.

Pour les acteurs, la vallée se doit d'être créatrice de richesse et donc être productive. Mais ces richesses, en termes de ressources, de productions, doivent pouvoir parvenir à être valorisées localement. Pour cela, certains acteurs ont mentionné le besoin d'une meilleure communication autour des produits locaux. L'idée d'un label, d'un site internet permettant d'améliorer la visibilité de ces derniers a ainsi été évoquée. Pour autant, le maintien des échanges extérieurs (exportations notamment) est apparu essentiel lors de la discussion.

Une autre thématique abordée a été l'écologie, la préservation de la Vallée et de son authenticité. Plusieurs idées ont pu se dégager au fur et à mesure du jeu et des discussions : la mise en place de jardins partagés (bien que leur utilité ait été débattu, certains préférant parler de "maraîchage collectif"), la création de lisières (zones tampons autour des bâtis urbains), la plantation de vergers le long de la Levrière, le développement des mobilités douces (traçage de pistes cyclables, développement des transports en commun) etc... L'idée de "sensibilisation" et "d'éducation" des locaux a aussi été évoqué par les acteurs, dans le sens où cela permettrait de soutenir cette volonté de préservation et la démarche écologique de la vallée.

La dernière thématique abordée, particulièrement essentielle pour certains acteurs, est le lien social et l'accès aux services. En effet, le manque de services au sein de la vallée est très rapidement apparu comme une problématique incontournable sur laquelle travailler (l'idée de marché fixe et mobile a été discutée). De plus, les joueurs ont pu faire part d'une volonté de renforcer le lien social à travers l'aménagement de "place publique", pour en faire des lieux de rencontre et de convivialité. Une autre réflexion, plutôt transversale, a pu être abordée : la volonté d'une piste à circulations douces et écologiques (vélos, trottinettes électriques, transport en communs électriques) reliant les différents villages et longeant le ruisseau de la Levrière. En outre, il a été abordé le souhait (voir l'obligation) d'être une vallée "connectée" grâce à la fibre optique.

3.4. La restitution

La restitution a permis de faire la synthèse des 3 scénarii et de dégager cinq points-clés :

- Que ce soit par la création d'un "parc naturel des deux vallées", d'une "communauté de communes" ou d'un regroupement "Villes ouvertes à la campagne" (titre du scénario 3), on constate dans l'ensemble des scénarios la volonté de création d'une entité englobante. Celle-ci ayant pour objectif de fédérer et unifier les communes, en particulier des deux vallées, afin de mener un développement commun de ce territoire.

- Un autre élément récurrent est la volonté de créer un axe structurant qui suit le lit de la Levrière, par exemple par la création d'une piste cyclable. En parallèle de cet axe principal, l'ensemble des scénarios aborde le besoin de créations d'axes d'échanges internes sous de multiples formes (jonctions entre vallées de la Levrière et de l'Epte, circuits vélos, bornes pour transports électriques, transports en communs).

- Outre le besoin de se déplacer dans la vallée, il y a la nécessité de créer des points centraux, des centres névralgiques qui offrent des services (marché local, magasin de producteurs, commerces, maisons médicales, complexe sportif) et permettent de créer du lien social (lieu de rencontre, place publique). Ces services (médecins, marché) ont besoin d'être "mobiles" entre les communes de la vallée.

- Il a été convenu par l'ensemble des acteurs que ces marchés et commerces doivent s'ancrer dans une dynamique de filières locales afin de dynamiser l'économie de la vallée.

- Enfin, bien qu'unifiée, cette vallée ne se veut pas fermée sur elle-même. Il a été exposé la volonté de renforcer les connexions avec les territoires extérieurs afin d'inscrire la vallée au sein d'une entité plus grande.

Le jeu de territoire a permis de faire émerger un répertoire d'actions opérationnelles. Bien entendu, ce « répertoire des possibles » comprend des actions plus ou moins difficiles à mettre en œuvre. Il apparaît que la mise en place de certaines actions nécessite un temps assez long et un engagement politique fort. Néanmoins, beaucoup d'entre elles peuvent être exécutées sur le court terme et produire des résultats positifs et visibles au niveau local.

4. Discussion et conclusion

4.1. L'itinéraire méthodologique comme architecture d'un projet de territoire

Cette approche empirique, qui part du terrain et des objectifs pédagogiques de formation des étudiants, pour répondre aux objectifs de l'action et réinterroger les concepts et méthodes de la recherche, donne un fil directeur pour penser la genericité de l'itinéraire méthodologique et un cadre d'analyse pour les transitions territoriales. Ainsi, tout au long de l'itinéraire méthodologique (Lardon et al. 2002), il pourrait y avoir des traitements géomatiques complémentaires pour consolider les analyses, extraire les connaissances et faciliter la participation.

L'analyse réalisée par les étudiants s'est principalement appuyée sur des méthodes qualitatives, de recueil des ressentis et des points de vue d'acteurs, mais aussi d'imprégnation de terrain des étudiants, mettant en avant les dimensions sensibles (Olmedo, 2015). Une fois n'est pas coutume, on peut inverser le regard en partant des données chaudes pour les étayer par des données froides, la démarche habituelle en géomatique étant plutôt d'élaborer des traitements quantitatifs voire des modèles, éventuellement illustrés par des verbatim d'acteurs. C'est ainsi que l'on pourrait faire appel aux méthodologies suivantes : Traitements cartographiques SIG pour les évolutions, combinaisons, ... ; Analyse d'images pour donner à voir la structure physique du territoire, la morphologie de l'habitat, l'évolution de la végétation, le contraste parcellaire ; Analyse spatiale pour reconnaître des motifs dans le paysage ; Analyse de réseau pour traiter les informations sur les relations entre acteurs.

Les connaissances produites dans ce type de démarches pourraient bénéficier des méthodologies spécifiques déjà existantes, pour systématiser l'analyse et la restituer

aux acteurs et aux chercheurs. Ainsi, ce sont les interfaces qui sont en jeu plus que les méthodes en elles-mêmes, les infrastructures de données plus que les données en elles-mêmes (Feyt et Noucher, 2014). On peut penser à l'analyse des discours et des cartes qui peut être appliquée autant aux entretiens individuels d'acteurs, qu'aux productions collectives, mais aussi aux impressions des étudiants, aux lectures de paysage ou à la reconnaissance d'objets spatiaux. Pour cela, on pourrait utiliser les ontologies, les méthodes d'extraction textuelles. Un travail plus spécifique de visualisation pourrait être nécessaire.

Pour aider au raisonnement des acteurs et accompagner les changements dans les territoires, il importe de tenir à la fois un horizon, pour orienter l'action, mais aussi de tenir le cap de manière souple, pour s'adapter aux nouvelles données. Cela nécessite un constant aller-retour entre la vision du futur et la réalité du présent. On pourrait ainsi simuler les scénarios d'évolution tout en élargissant le champ des possibles, en utilisant différentes approches : la construction chorématique, telle que menée sur la prospective eau (Casanova Enault et Chatel, 2017 a et b) ; la simulation multi-agents (Bousquet et al., 2002), les jeux de reconstruction spatiale (Ramadier et Bronner, 2006).

4.2. Une nouvelle façon de penser les démarches de développement territorial

En s'appuyant sur la conception d'un itinéraire méthodologique de diagnostic de territoire adapté au questionnement de terrain, nous pouvons concevoir un outil d'aide à la décision pour les acteurs. Ce faisant, nous élaborons aussi des éléments de questionnement pour les chercheurs disciplinaires et interdisciplinaires (Borderon et al., 2015), afin qu'ils intègrent progressivement différents systèmes de connaissances et différentes compétences (Stock and Burton, 2011). Ce sont aussi des éléments d'apprentissage pour les étudiants (Lardon et Mainguenaud, 2006).

Une telle approche pourrait fédérer les nombreuses méthodologies, trop souvent déployées indépendamment les unes des autres. Ainsi, des tas de données qualitatives ne sont pas traitées quand de nombreux traitements quantitatifs n'atteignent pas leurs interlocuteurs. Il y a une profusion d'acteurs qui refont les mêmes choses et les mêmes erreurs, alors qu'il existe un corpus riche d'expériences réalisées, voire modélisées, mais dont ils n'ont pas connaissance. Dans leurs domaines respectifs, les chercheurs restent aussi trop souvent sur les méthodes qu'ils connaissent et ont du mal à les mettre à l'épreuve d'autres.

Le renouvellement des outils de la géomatique gagnerait à s'appuyer sur de nouvelles conceptions en sciences sociales. Ainsi, l'hybridation des mondes (Gwiazdzinski, 2016), la coexistence des modèles (Hervieu & Purseigle, 2015), les forums hybrides (Callon et al., 2001) ... relèvent d'une science en société qui met les outils numériques au service des humains et qui ouvrent l'espace en partage (Bonny et al., 2016).

Remerciements

Les étudiants DEBATS, tous les acteurs participants, l'AVL, le CAUE.

Bibliographie :

- Amblard L., Berthomé K., Houdart M., Lardon S. (2018). L'action collective dans les territoires : questions structurantes et fronts de recherche. *Revue Géographie Economie Société* 2018, Vol. 20, n°2, p. 227-246.
- Benoît M., Deffontaines J. P., Lardon S. (2006). *Acteurs et territoires locaux – vers une géoagronomie de l'aménagement*. Quae Editions, Paris.
- Bonny Y., Bautès N., Gouëset V. (dir.), 2016. L'espace en partage. Approche interdisciplinaire de la dimension spatiale des rapports sociaux. *Presses Universitaires de Rennes*, Rennes.
- Borderon M., Buchs A., Leblan V., Vecchione E. (2015). Réflexivité et registres d'interdisciplinarité. Une boussole pour la recherche entre natures et sociétés. *Natures Sciences Sociétés*, vol. 23, n. 4, p. 399-407.
- Bousquet F., Barreteau O., d'Aquino P., Etienne M., Boissau S., Aubert S., Le Page C., Babin D., and Castella J.-C. (2002). Multi-agent systems and role games: an approach for ecosystem co-management. *Complexity and ecosystem management: the theory and practice of multi-agent approaches*. Edward Elgar, Cheltenham, p. 248-285.
- Brook, I. (1998). Goethean science as a way to read landscape, *Landscape Research*, vol. 23, n° 1, pp. 51-69. <https://doi.org/10.1080/01426399808706525>
- Caron P., Valette E., Wassenaar T., Coppens G., Papazian V. (2017). *Des territoires vivants pour transformer le monde*. Editions QUAE.
- Casanova Enault L., Chatel C. (2017a). Une grille chorématique des dynamiques spatiales pour expliquer l'évolution des territoires, *Mappemonde* n° 119, <http://mappemonde.mgm.fr/119as3/>
- Casanova Enault L., Chatel C. (2017b). La modélisation graphique de phénomènes émergents pour répondre aux besoins de la prospective, *Mappemonde*, n°119, <http://mappemonde.mgm.fr/119as4/>
- Callon M., Lascoumes P., Barthe Y. (2001) *Agir dans un monde incertain. Essai sur la démocratie technique*. Le Seuil, Paris.
- Debolini M., Marraccini E., Rizzo D., Galli M., Bonari, E. (2013). Mapping local spatial knowledge in the assessment of agricultural systems: A case study on the provision of agricultural services. *Applied Geography*, vol. 42, p. 23-33.
- Feyt G. et Noucher M. (2014), La gouvernance informationnelle, outil et enjeu stratégiques des recompositions territoriale : vers l'émergence de nouveaux référentiels géographiques ?, Actes du Colloque CIST « Fronts et frontières des sciences du territoire », Paris, 27-28 mars 2014.
- Gwiazdzinski L. (2016). L'hybridation des mondes. Territoires et organisations à l'épreuve de l'hybridation. *L'innovation autrement*. Elya Editions.

- Hervieu B., Purseigle F. (2015). The sociology of agricultural worlds: from a sociology of change to a sociology of coexistence. *Review of Agricultural and Environmental Studies*, 96(1).
- Lardon S. (2013). Le “jeu de territoire”, un outil de coordination des acteurs locaux. *Revue FaçSADe*, vol. 38, p. 1-4.
- Lardon S., Mainguenaud M. (2006). Parcours dans un itinéraire méthodologique de diagnostic de territoire : exploration par un navigateur. *Revue Internationale de Géomatique*, vol. 16, n. 2, p. 147-161.
- Lardon S., Marraccini E., Filippini R., Gennai Schott S., Rizzo D., Johany F. (2016). Prospective participative pour la zone urbaine de Pise. L’eau et l’alimentation comme enjeux de développement territorial. *Cahiers de Géographie du Québec*, vol. 20, n°170, p. 265-286.
- Lardon S., Noucher M. (2016). Prospective territoriale participative et usages de l’information géographique. *Cahiers de Géographie du Québec*, Vol. 60, n° 170, p. 209-360.
- Lardon S., Piveteau V. (2005). Méthodologie de diagnostic pour le projet de territoire : une approche par les modèles spatiaux. *Géocarrefour*, vol. 80, n° 2, p. 75-90.
- Lardon S., Therond O. (2018). Retours sur l’école-chercheurs FormaSciences « Prospective, participation et modélisation spatiale pour la gestion des ressources dans les territoires ». *Cahiers de Géographie du Québec*, vol. 62, n° 175, p. 193-200.
- Olmedo E. (2015). *Cartographie sensible. Tracer une géographie du vécu par la recherche-crédation*, Thèse en géographie, Université Paris I Panthéon Sorbonne.
- Ramadier T., Bronner A-C. (2006). Knowledge of the environmental and spatial cognition: JRS as a technique for improving comparisons between social groups. *Environment and Planning B: Planning and Design*, 33, p. 285-299.
- Stock P., Burton R.J.F. (2011). Defining Terms for Integrated (Multi-Inter-Trans-Disciplinary). *Sustainability*, 3, 1090–1113.

Panorama d'un *geoweb* des flux et réseaux

**Bahoken F.^{1,4}, Le Campion G.², Jégou L.^{3, 4}, Maisonobe M.⁴,
Côme E.⁵**

1. AME-SPLOTT, IFSTTAR / Université Paris-Est
Cité Descartes, 14-20 Boulevard Newton, F-77420 Champs-sur-Marne, France
francoise.bahoken@ifsttar.fr
2. UMR 5319 Passages, Équipe ARD / CNRS
Maison des Suds, 12 Espl. des Antilles, F-33600 Pessac, France
Gregoire.LECAMPION@cnrs.fr
3. UMR 5193 LISST / Université de Toulouse 2 / CNRS
Maison de la Recherche, 5, allées Antonio-Machado, F-31058 Toulouse Cedex 9,
France. jegou@univ-tlse2.fr
4. UMR 8504 Géographie-Cités / Équipe PARIS / CNRS
13, rue du Four, F-75006 Paris, France
marion.MAISONOBE@cnrs.fr
5. COSYS-GRETTIA, IFSTTAR / Université Paris-Est
Cité Descartes, 14-20 Boulevard Newton, F-77420 Champs-sur-Marne
etienne.come@ifsttar.fr

RÉSUMÉ. L'analyse de la dynamique des aires urbaines ou des métropoles, la délimitation de leurs aires fonctionnelles, la comparaison spatio-temporelle de leurs motifs est souvent freinée par l'insuffisance de données relationnelles ouvertes et l'absence jusque récemment de dispositifs d'analyse et de géo-visualisation dédiés. Au-delà des questions d'ouverture des données (géo)numériques, nous proposons un panorama d'un *geoweb* spécifique aux flux et réseaux localisés. L'éclairage ainsi apporté sur les pratiques cartographiques actuelles révèle trois grandes familles d'applications web ainsi que les besoins d'une communauté, restreinte mais dynamique, d'analyser librement ses propres jeux de données.

ABSTRACT. The analysis of the dynamics of urban areas or metropolises, the delimitation of their functional areas, the spatio-temporal comparison of their patterns is often limited by the lack of open relational data and the absence until recently of dedicated analysis and geo-visualization frameworks. Beyond the questions of opening (geo)numeric data, we propose a panorama of a *geoweb* specific to flows and networks. The insights provided on current mapping practices reveal three main families of web applications as well as the needs of a small but dynamic community to freely analyze its own data sets.

MOTS-CLÉS : *geoweb, flux, réseaux, application web, cartographie thématique.*

KEYWORDS: *geoweb, flows, spatial networks, web applications, thematic cartography.*

1. Introduction

L'analyse de la dynamique des aires urbaines ou des métropoles, la délimitation de leurs aires fonctionnelles, la comparaison spatio-temporelle de leurs motifs est souvent freinée par deux catégories de problèmes inhérents aux données et aux outils. L'insuffisance de données ouvertes décrivant interactions et interrelations territoriales, pourtant cruciales pour l'aménagement du territoire et la planification, freine en effet les possibilités d'analyses empiriques. L'absence de dispositifs d'analyse et de géo-visualisation dédiés contribue, de plus, à éluder du paysage cartographique actuel nombre d'images présentant pourtant un intérêt visuel et analytique fort (Bahoken, 2016). Au-delà des questions liées à la mise en ligne des données relationnelles qui relèvent de problèmes spécifiques, nous souhaitons attirer l'attention sur les possibilités actuelles de leur analyse et représentation cartographique en ligne, avant de proposer une nouvelle application qui leurs sera consacrées.

Notre fondons notre hypothèse principale sur l'engouement suscité depuis les années 2005 par le *geoweb*, une notion définie par T. Joliveau (2010, 2011) comme le processus de création de cartes dans le contexte du web 2.0, fondé sur la participation, l'expérience utilisateur et la compétence informatique. Nous considérons que cette informatisation générale de la société (cf. Goodchild, 2007) est susceptible de profiter à l'analyse cartographique des flux et des réseaux localisés. L'ouverture des possibilités de stockage et de traitement de données, pouvant être massives, facilitant l'appropriation des formats matriciels pourrait étendre, par là même, leurs possibilités d'analyse et de représentation géo(carto)graphique.

Le support *web* a en effet entraîné un renouvellement général de la fabrique cartographique qui a permis d'introduire de nouvelles possibilités de représentation des changements (Cunty et al. 2017). Les possibilités d'animation autorisent une visualisation moins naïve des mouvements, celles d'interactions et de cartographies dynamiques admettant une exploration active des informations facilitant leur appropriation. Ces nouvelles fabriques qui conduisent à la production de nouvelles images cartographiques numériques (Joliveau et al., 2013) présentent un risque élevé d'uniformisation (Noucher, 2017) dans ce contexte du *Tous cartographes !* Elles accompagnent toutefois un renouvellement des usages de la carte en même temps qu'elles redéfinissent la figure du cartographe (Zanin, 2013), à moins d'en introduire une nouvelle : l'utilisateur-concepteur.

Le fait est que les matrices décrivant interrelations et interactions sont déjà largement mobilisées dans l'analyse des réseaux d'acteurs, localisés ou non, en particulier dans leur représentation sous la forme de graphe, à la faveur d'une appétence sans limite pour des réseaux sociaux, qu'ils soient appréhendés comme média social, comme objet d'étude ou comme cadre théorique et conceptuel d'analyse de systèmes relationnels. Cependant, force est de constater que les productions réalisées ne sont, hélas, toujours pas convaincantes (Baron et al., 2012) dans l'ensemble, si ce n'est utiles. La raison tient au contexte pollué par une surenchère technique et visuelle qui, si elles font le lit de la *dataviz* et d'un *data journalism* très

en vogue, ne tiennent pas compte des méthodes éprouvées, de leur réception par différents publics, et des possibilités de leur usage dans un cadre analytique.

À l'heure du *Tous cartographes !*, il semblerait qu'une meilleure connexion entre développeurs de dispositifs du *geoweb*, thématicien·ne·s et géomaticien·ne·s - cartographes positionné·e·s à l'interface de ces deux communautés, pourrait contribuer à remédier à cette situation. Le développement d'une algorithmie spécifique en réponse à des questionnements spécifiques, combinant informations statistiques, géographiques et cartographiques, incluant des aspects de sémiologie pour améliorer leur lisibilité, pourrait conduire à renouveler la production visuelle de ces dispositifs.

Dans cet objectif, nous avons engagé un projet interdisciplinaire : *gflowiz* (*geographic flow visualisation*)¹. Nous en présentons ici les premiers résultats, avec, dans un premier temps, un panorama du *geoweb* des flux et réseaux origine-destination qui révèle, dans un second temps, trois grandes familles d'applications, ouvrant la voie à une enquête à propos des *desiderata* pour une nouvelle application, qui feront l'objet d'une troisième section.

2. Un corpus unique d'applications de web cartographie

Dans le cadre du projet *gflowiz*, nous avons commencé par réaliser un état de l'existant : c'est-à-dire par recenser un ensemble assez varié et aussi large que possible d'applications relevant du *geoweb*. Le corpus ainsi constitué recense près de 80 cartes et applications web dédiées aux flux et réseaux localisés. Notre première observation à l'issue de ce travail de recension invalide une hypothèse de départ fondée sur un faible intérêt apparent pour ce type de document, considérant que la production cartographique correspondante disponible en ligne, comme dans sa version statique, relèverait davantage d'une seule analyse des localisations – et non des interactions ou interrelations – quand bien même les données seraient disponibles ; que la représentation graphique privilégiée correspondante serait celle de la carte choroplèthe, moins celle de la carte ou du graphe de flux, de la matrice ordonnée. Sachant que l'une des raisons régulièrement évoquées pour justifier le caractère secondaire de la carte de flux étant souvent d'ordre technologique (en raison du coût d'accès aux outils existants pour les non initié·e·s) sans compter les problèmes de méthodes, l'hypothèse que le développement d'un *geoweb* spécifique puisse faciliter les choses apparaît plus que séduisante.

2.1. Constitution du corpus et possibilités d'exploration

Une large curation de sites web a été réalisée en réponse à la question suivante : Quelles sont les possibilités de géo visualisation de flux, trafics et mouvements observés dans un espace géographique en ligne ?

Deux phases de collecte sont remarquables. La première d'entre elles, réalisée en 2009 relève de la veille générale. Si les applications et mises en ligne de ces cartes font

¹ <https://tinyurl.com/gflowiz>

l'objet d'une recension en ligne², la curation effectuée ne saurait être considérée comme régulière. Il faudra en effet attendre 2015 pour cela, avec une sauvegarde systématique sur le réseau social *SeenThis*³. Les 80 applications collectées de 2017 à janvier 2019 ont fait l'objet d'un examen minutieux à l'aide d'une première grille de lecture permettant leur classement manuel en six grands thèmes ainsi définis : A- Contexte et sujet ; B- Type de représentation principale ; C- Qualité logicielle ; D- Public visé ; E- Technologies mobilisées et F- Métadonnées. L'exploration de la liste d'applications obtenue a été rendue possible grâce au développement d'une première version d'un tableau de bord interactif réalisé sous R/Shiny, initialement dans le cadre du stage de fin d'études de A. Nguyen (2018), enrichie et désormais maintenue (Le Campion, Nguyen, 2019).

Cette première grille de lecture étant apparue trop détaillée pour permettre son exploitation simple, des regroupements ont par la suite été réalisés, donnant lieu à une seconde version du tableau de bord actuellement en ligne⁴. La fabrique cartographique spécifique au *geoweb* peut ainsi être explorée à l'aide d'un double système de filtrage interactif (voir Fig. 1).

The screenshot shows the 'Dashboard GFlowiz' interface. On the left, there are several filter menus: 'Domaine', 'Contexte de développement', 'Type de représentation principal', 'Logique de représentation', 'Dimension spatiale', 'Public visé', and 'Date de mise en ligne'. The main content area displays a list of applications with columns for 'Id', 'Acronyme/Titre', 'Auteur', 'Contexte_de_developpement', 'Domaine', 'Type_de_deplacements', and 'Thématique'. A table below the list shows the first four entries:

Id	Acronyme/Titre	Auteur
1	Visualizing massive data streams : a transport use case	Alexandre Mas
2	Automated layout of origin-destination flow maps: U.S. county-to-county migration 2009-2013	Daniel M. Stepi
3	Canvas Flowmapping Layer with JSAPI	Sarah Bellum
4	When Napoleon ventured east	TASS

FIGURE 1. Tableau de bord des applications recensées

Une première possibilité est accessible sur la partie latérale gauche, elle permet de sélectionner la ou des variables à visualiser. La seconde possibilité de filtrage proposée permet une exploration résultant du croisement entre une sélection de variables (domaine, contexte de développement, type de représentation principal...) et quelques grands thèmes de la grille.

2.2. Description du corpus d'applications

La fabrique cartographique spécifique au *geoweb* est examinée d'un point de vue thématique, sémiologique ou technologique, au regard des possibilités d'actions sur

² <https://www.scoop.it/t/analyse-des-matrices-de-flux>

³ <https://seenthis.net/people/fbahoken>

⁴ <http://37.187.79.5:3838/gflowiz/DashB2/>

les différentes composantes de l'image, sachant que différents problèmes généraux liés à la visualisation d'informations scientifiques ont été présentés par Keim et al. (2008).

2.2.1. Un intérêt indéniable pour le support numérique

Contre toute attente, l'hypothèse que le développement d'un *geoweb* spécifique aux flux et réseaux localisés repositionne les cartes correspondantes dans le paysage géonumérique actuel est largement validée. Le développement de l'informatisation apparaît profiter à l'analyse (carto)graphique de flux, réseaux et mouvements de toutes sortes depuis les années 2010 (voir Fig. 2).

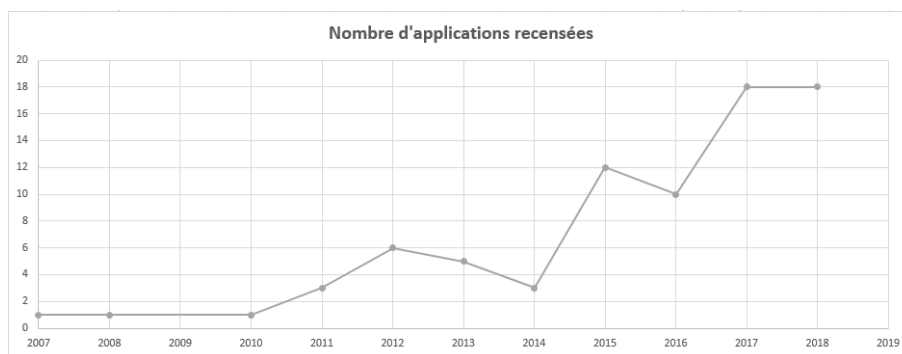


FIGURE 2. Date de mise en ligne des applications recensées (décembre 2018)

La forte croissance apparente du nombre d'applications recensées est à nuancer, dans la mesure où il n'est pas possible de réaliser un bilan exhaustif de ces applications, ni d'estimer leur part par rapport à un ensemble de sites de cartographie 2.0. La figure 2 décrit donc surtout l'évolution des mises en lignes d'applications web repérées, proposant des géovisualisations de données de flux et réseaux, sous différentes formes.

Les dispositifs géo numériques sont assez variés, mais trois types d'applications se dégagent. Celles qui sont de type « thématiques », dédiées à un jeu de données particulier, sont les plus nombreuses, avec 50 éléments recensés (sur 80). Ces applications, qui tiennent essentiellement lieu de vitrine, sont cependant souvent fermées ; elles s'opposent aux dispositifs moins fréquents qui permettent d'ajouter des jeux de données externes pour réaliser ses propres cartes : 8 dispositifs de type « logiciels web ») ou simplement les mettre en scène : 7 dispositifs de type « *storymaps* ». Cette diversité témoigne de la levée ces dernières années d'une bonne partie des verrous techniques qui avaient pu être observés empiriquement lors de la première phase de recension, en lien avec un intérêt collectif pour le support cartographique que renforce le choix de la thématique.

2.2.2. Des applications monothématiques

Si la Figure 1 montre une nette augmentation de la fréquence de mise en ligne de cartes de flux et réseaux à partir de 2015, elle serait liée à un effet du contexte de la « crise migratoire » des politiques européennes qui suscita la mise en ligne d'applications de cartographie de phénomènes migratoires, en réponse aux attentes de la société civile sur un sujet éminent politique. L'étalement visible sur les années suivantes serait quant à lui lié au décalage entre la production de l'information, sa mise à disposition et son exploitation cartographique.

Les applications mobilisant des flux de type démographique, concernant des populations humaines en général, sont les plus représentées (avec 20/80), elles sont suivies par celles qui décrivent des flux économiques (10) et de transport (10).

Ces variations thématiques influent sur le type de représentation principale (carte de flux ou réseaux origine-destination, ou de flux affectés sur réseaux) et d'implantation spatiale. Les applications décrivant des liens valués, en implantation ponctuelle (bi localisée) sont ainsi plus fréquentes que celles qui mobilisent une implantation (poly)linéaire. Cela étant, les dispositifs dédiés au transport sont les plus riches en termes cartographiques. Les modes concernés sont majoritairement terrestres (ferroviaires et routiers, pour moitié d'entre eux), dans une moindre mesure maritimes ou aériens. Les applications de transport décrivent pour moitié des flots de véhicules en précisant le mode, collectif (bus, métro, navire) ou individuel (voiture, vélos en libre-service). Un petit quart des applications de transport concerne des flux de personnes (passagers, migrants, sportifs, etc.).

2.3. Une ouverture récente vers des possibilités d'analyses thématiques

L'éventail des possibilités actuelles d'exploration de ces données est symptomatique de l'engouement soutenu d'une communauté de « néocartographes ». Si l'effort a pendant longtemps été limité à la seule visualisation directe des informations, on observe la mise en ligne récente d'applications spécifiquement dédiées à une analyse thématique configurable par l'utilisateur, qui ouvre la voie vers une première approche de statistique descriptive.

La première d'entre elles apparaît fondée sur des objectifs de présentation en ligne sous une forme cartographique numérique de gros corpus de données, c'est le cas de *Resource Trade* (<https://resourcetrade.earth/>) par exemple. Les applications concernées, qui présentent des niveaux de sophistication variés, n'autorisent toutefois pas de nouvelles analyses *via* l'importation de données externes.

Une seconde approche va plus loin en permettant une géo-exploration visuelle de données de mobilités dédiées, selon différentes modalités graphiques comme *Coscimo* (<http://www.coscimo.net/>, 2013).

Enfin, certaines applications autorisent le chargement de données externes, tel *Flowmap.blue* (<https://flowmap.blue/>, 2018), et la co-visualisation comme *The Vistorian* (<https://networkcube.github.io/vistorian/> (2018) ou encore *Netscity* (<https://www.irit.fr/netscity/>, 2019) ; d'autres présentent en plus des fonctionnalités

avancées de rendu cartographique de flux, tel le module dédié aux flux de mobilités de *France découverte Geoclip Air* (<https://france-decouverte.geoclip.fr/>, 2019).

Une troisième approche semble davantage consacrée à des objectifs de cartographie en ligne, en réponse aux lacunes des éditeurs actuels. Elles sont révélatrices d'un besoin de renouvellement des pratiques de cartographie tant chez les cartographes thématiques traditionnels, dans le cas de *Magrit* (<http://magrit.cnrs.fr>, 2017), que chez les néocartographes spécialisés dans le développement web, avec l'exemple de *Kepler* (<https://kepler.gl/>, 2018). Il est intéressant de noter que les solutions proposées dans les deux cas sont réalisées en lien avec une pratique montante du développement de fonctions voire d'outils libres, reproductibles (Giraut, Lambert, 2017), spécifiquement dédiés à la manipulation de données statistiques à des fins de cartographie thématique.

3. Typologie des applications web dédiées aux flux et réseaux

Vu l'importance du corpus d'applications décrites par une cinquantaine de champs, il nous a semblé intéressant de réaliser une typologie permettant de dégager les principales familles. Une classification ascendante hiérarchique (CAH) précédée par une analyse des correspondances multiples (ACM) a été réalisée dans cet objectif. La méthode étant très sensible à la fois aux données manquantes et aux modalités « rares », la liste des champs a été scindée en deux sous-corpus, l'un relevant des aspects technologiques et l'autre des aspects descriptifs : leur liste est présentée dans l'Annexe 1.

La réalisation d'une première ACM a révélé deux individus très particuliers, au sens où ils contribuent fortement (35,2%) à la construction du plan factoriel (voir Fig. 3).

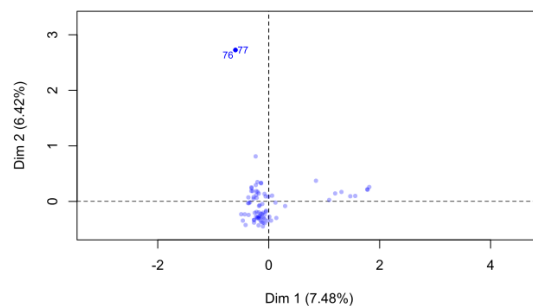


FIGURE 3. Contribution des applications de *Anagraph* à la construction du plan factoriel

Il s'agit du système d'applications développé par *Anagraph* pour décrire 24h du trafic du Grand Montréal, à travers deux dispositifs distincts : le premier porte sur des flux : *MTL Trajets*, identifiant 76 dans la figure 3, (<https://mtl-trajet.anagraph.io/flow>)

et le second sur des trajets : *MTL trajets - trip explorer*, identifiant 77, <https://mtl-trajet.anagraph.io/trips>).

Ces deux dispositifs *MTL Trajets* font figure d'exceptions en raison de leurs très riches caractéristiques. Les modalités qui participent de leur richesse sont les suivantes, présentées par ordre d'importance. Elles présentent 1) fond géo numérique de type *OpenStreetMap* sur lequel est projeté 2) un déplacement de type *flux matériel* relevant de 3) tous *modes de transport*. 4) La logique de représentation sous-jacente est celle d'un *mouvement* qui mobilise dans sa sémiologie 5) des *variables rétinienne dynamiques* sous la forme d'animation pour illustrer 6) un *transport de personnes*. Le dispositif est de type 7) *application web* (non thématique) déployé à 8) une échelle *locale* sur 9) *une seule couche* cartographique, 10) en 2018.

Pour analyser au mieux l'ensemble de notre corpus, nous avons choisis d'écarter ces deux individus singuliers avant de relancer l'analyse.

3.1. Entre réseaux d'acteurs localisés et réseaux planaires

La seconde ACM apporte des résultats intéressants qui discriminent deux grands types d'applications fondamentalement différentes sur le plan théorique : la première relève de l'analyse de réseaux d'acteurs au sens de la *Social network analysis* (SNA) et la seconde de l'analyse de réseaux planaires et non planaires.

Une analyse préliminaire de ce sous-corpus d'applications et de variables permet d'affirmer que la première dimension isole des applications dont la thématique relève plutôt du *réseau d'acteurs*, dans le domaine *Science*, avec des *représentations graphiques mixtes* pour décrire un lien de type *relation scientifique entre lieux*.

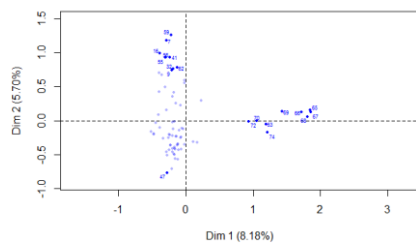


FIGURE 4. Applications de type SNA

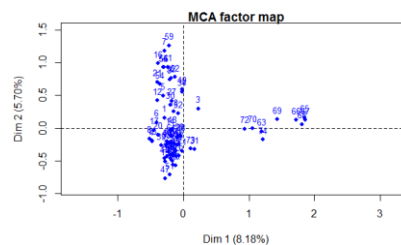


FIGURE 5. Applications de type réseaux planaires et non planaires

La deuxième dimension discrimine nos individus sur des applications dont le mode de représentation principale est soit la *carte de réseau*, soit celle de flux, réalisées en *implantations linéaires* et *poly linéaires* ; la logique de représentation étant le *lien*, valué ou non, avec des *types de déplacement* qui opposent *matériel* et *immatériel*.

3.2. Trois grandes familles d'applications

La meilleure partition obtenue, c'est-à-dire celle qui suggère des groupes où les individus qui les composent sont plus homogènes entre eux qu'avec les individus extérieurs, est formée de trois classes qui correspondent aux sous-groupes déjà identifiés lors de l'ACM (voir Fig. 6).

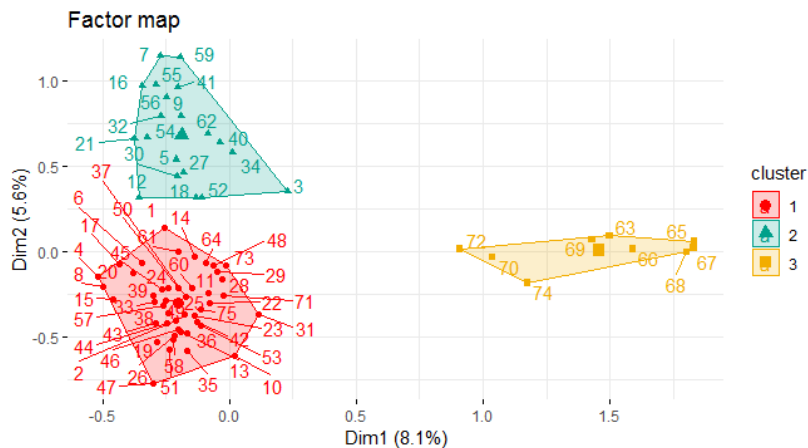


FIGURE 5. Plan des clusters d'applications

Comme l'indiquent F. Husson et al. (2010), les individus remarquables caractérisant les différents clusters peuvent être classés à l'aide des parangons, qui correspondent aux individus « moyens » de chaque classe, situés le plus au centre du nuage de points ; et à l'aide des individus « extrêmes », les plus distants des autres classes qui représentent une sorte « d'idéal-type » au sens d'accentuation des caractéristiques distinctives.

Plus encore que ces individus particuliers, l'analyse des modalités caractéristiques de chacune des classes apporte une idée précise de la construction de nos clusters, les variables contribuant à la formation de chacun des trois clusters sont présentées dans l'Annexe 2.

Le premier cluster se compose des applications dont la représentation principale est la *carte de flux*, qui représente principalement des *déplacements immatériels*, et où l'on retrouve *plusieurs représentations secondaires* et *plusieurs modalités graphiques* ; sont également présents des indicateurs statistiques qui sont représentés par la variable rétinienne *Taille* (symboles proportionnels). Le parangon de cette classe est l'application : *Global trade : Africa. The world's import from and export to Africa visualized* (<https://lab.interactivethings.com/global-trade-africa/>).

En revanche, le deuxième cluster est composé plutôt par des applications dont la représentation principale est la *carte de réseau*, réalisée avec une logique de représentation qui est le *lien en implantation polylinéaire* - alors que dans le cluster précédent nous étions sur une logique de flux et une implantation ponctuelle (bi

localisée). On ne trouve pas d'indicateurs associés à ces représentations de type réseau planaire - le type de déplacement est *matériel*. Les applications du cluster n°2 relèvent surtout du domaine du *transport* ou de la thématique de l'accessibilité aux réseaux. Son parangon est l'application *Travel Time Tube Map* (http://www.tom-carden.co.uk/p5/tube_map_travel_times/applet/tube_map_travel_times.js.pde)

Le troisième cluster est un sous-groupe particulier. Les applications qui le composent sont produites exclusivement dans un contexte de recherche ; elles représentent plutôt des réseaux d'acteurs structurés par des relations scientifiques ; le domaine associé est donc la Science. Ce cluster est significativement associé à des représentations mixtes (graphe/carte/diagramme) sans sémiologie particulière : les variables rétinienne *Couleur* et *Taille* ne sont pas mobilisées. Pour ce dernier cluster le parangon est l'application *Palladio : Mapping the Republic of Sciences* (<http://hdlab.stanford.edu/palladio/>).

4. Desiderata pour une nouvelle application web

Les souhaits pour une nouvelle application ont été recueillis dans le cadre d'une *Enquête sur les besoins pour une application web de géo-visualisation de flux et de réseaux*, menée en ligne au second semestre de 2018. Elle a permis de collecter un peu plus de 200 réponses auprès d'utilisateurs potentiels de l'application, pour apprécier les besoins et les fonctionnalités utiles. Elle nous permet de préparer la conception d'une nouvelle application, en particulier de prendre en compte la forte demande d'analyse.

Les répondants sont en grande majorité des personnes qui utilisent déjà des outils d'analyse et de représentation, dans le domaine des transports, de l'enseignement ou de l'environnement, mais qui expriment le besoin de disposer d'un outil plus complet et plus adapté, notamment pour créer des représentations interactives et multiples, concurrentes, de leurs propres données. La cartographie de flux, couplée à d'autres représentations (cartographie thématique, graphes et diagrammes) revient dans la majorité des demandes, dans le cadre d'une application qui serait accessible parce qu'en ligne, gratuite, ergonomique et *open-source*.

La demande de représentations est également importante pour des applications qui soient esthétiquement agréables, basées sur des fonds personnalisables, partageables et diffusables. Ces besoins ne semblent pas comblés ou même pris en compte par les applications actuelles, de bureau ou en ligne, ce qui justifie notre projet.

5. Conclusion

La compilation de près de 80 *web-applications* apporte un éclairage sur les pratiques cartographiques actuelles, en termes de possibilités et de développement.

Peu d'applications semblent complètement inscrites dans le paradigme de la « cartographie de visualisation » tel que défini par A. Mac Eachren en 2004. Celles qui ont été recensées ici relèvent encore peu de la (géo)visualisation proprement dite

associée à un traitement amont des informatiques statistiques et géographiques. Et pourtant, l'heure semble être au développement de représentations techniquement plus abouties qui laissent entrevoir des possibilités intéressantes sur le support web 2.0. C'est pourquoi le projet *gflowiz* se propose de contribuer à remédier à cette situation, en proposant une nouvelle application accessible en ligne.

Les possibilités de co-visualisation des données selon plusieurs modalités graphiques à l'instar de *Coscimo* (Jégou et Maisonobe, 2013) demeure timides, alors qu'elles sont dans bien des cas complémentaires pour explorer des données relationnelles. En particulier, le graphe de flux donne à voir la morphologie d'un système de relations entre des acteurs pouvant être localisés dans un espace, là où la carte décrit leur géographie et le diagramme, souvent, leur temporalité.

Le calcul d'indicateurs statistiques complémentaires, par exemple, devant enrichir qualitativement les représentations demeure toutefois encore largement absent des applications. Ces fonctionnalités sont souvent demandées par les personnes enquêtées, dans le cadre d'outils qui permettraient l'analyse de jeux de données personnels, leur représentation dans des conditions paramétrables, avec une reproductibilité garantie.

Le défi à résoudre dans le cadre du développement d'une nouvelle application web dédiée aux flux et réseaux consisterait donc à considérer ces lacunes pour proposer, idéalement, un ensemble d'outils d'analyse et de représentation de matrices d'interactions géographiques qui sont par essence complexes, parce qu'elles sont denses (concernent de nombreux acteurs) et riches, en étant catégorielles et temporelles.

Remerciements :

Les résultats présentés ici ont été obtenus dans le cadre de gflowiz, une collaboration interdisciplinaire financée par l'Institut français des sciences et technologies des transports, de l'aménagement et des réseaux (IFSTTAR), départements Aménagements, mobilités, Environnement (AME) et Composants et systèmes (COSYS), avec la participation du projet Fédérateur Mobilités et transitions numériques. Elle a été réalisée en collaboration avec l'UMR Lisst-Cieue, l'UMR Passages et l'UMR Géographie-cités du Centre national de la recherche scientifique (CNRS).

Bibliographie

- Bahoken F. (2016), *Contribution à la cartographie d'une matrice de flux*, Mémoire de Thèse en Géographie - Sciences des territoires, Université Paris Diderot (Paris 7).
- Baron M., Eckert D., Jegou, J. (2012), Peut-on démêler l'écheveau mondial des collaborations scientifiques ? *Revue Mappemonde*, Rubrique Internet, n°102.
- (Le) Campion G., Nguyen A. (2019) *gFlowiz DB : tableau de bord d'exploration interactive d'applications web de visualisation de flux et mouvements*, Université Paris-Est, AME-SPLOTT / COSYS-GRETTIA, IFSTTAR et UMR Passages.

- Cunty C., Mathian H. et Groupe Cartomouv (2017), Les pratiques de cartographie animée pour représenter le changement, *Revue M@ppemonde*, n°120.
- Giraud, T., Lambert, N. (2017), Reproducible cartography, *Advances in Cartography and GISsciences, Proceedings of the International Cartographic Conference, ICACI'2017*, Springer, p.173-183.
- Goodchild M., (2007), Citizens as sensors: Web 2.0 and the volunteering of geographic information », *Revue GeoFocus*, vol 7, p. 1-10.
- Husson F., Lê S., Pagès J. (2010). *Exploratory Multivariate Analysis by Example Using R*. Chapman & Hall.
- Joliveau T. (2010), *Le GéoWeb pour les Nuls*, <https://mondegeonumerique.wordpress.com/2010/06/24/le-geoweb-pour-les-nuls/>
- Joliveau T. (2011), Le géoweb, un nouveau défi pour les bases de données géographiques, *Revue L'Espace géographique*, vol. 40, p. 154-163.
- Joliveau T., Noucher M., Roche S. (2013), La cartographie 2.0, vers une approche critique d'un nouveau régime cartographique, *Revue L'Information géographique*, Armand Colin, vol. 77, p. 29-46
- Keim D., Andrienko G., Fekete J.-D., Görg C., Kohlhammer J., Melançon G., (2008), *Visual Analytics: Definition, Process, and Challenges*, In: Kerren A. & al. (Eds.): *Information Visualization*, Springer-Verlag Berlin Heidelberg, LNCS 4950, pp, 154–175,
- Mac Eachren A., (2005), *How Maps Work, Representation, Visualization, and Design*, New-York, The Guildford Press,
- Noucher M. (2017). *Les petites cartes du web. Approches critiques des nouvelles fabriques cartographiques*, Éditions de la rue d'Ulm – Presses de l'École normale supérieure, Paris.
- Zanin C. (2013), *Cartographie et analyse territoriale*, Mémoire d'habilitation à diriger des recherches en Géographie, Université Paris Diderot (Paris 7).

Annexe(s)

Annexe 1. Liste des variables du sous-corpus d'application analysés

Les variables décrivant le sous-groupe d'applications qui a fait l'objet de l'ACM et de la CAH présentées dans le texte sont les suivantes : le titre, le nombre de couches cartographiques (de 1 à 231), le nombre de représentations graphiques (de 1 à 7), le dispositif géo-numérique (java, web (collecte), web thématique, web interactif, storymap, prototype, logiciel web, package R, ...), le type de représentation principale (carte de flux, carte de réseau, graphe, diagramme à coordonnées parallèles), la logique de représentation (lieux, liens, flux, réseaux), l'implantation spatiale (ponctuelle, linéaire, polylinéaire, surfacique), le mode de représentation (vecteur et/ou raster, mixte), l'indicateur statistique représenté, la présence d'une variable rétinienne dynamique ou non, de la variable visuelle *Taille* et/ou de la teinte (*Couleur*) incluant l'opacité et l'opposition colorée et/ou de la *Forme* et/ou du *Cignotement* et/ou de l'*Animation* et/ou du *Mouvement*, le nombre de représentations secondaires, la présence d'une légende, le langage logiciels (Javascript, MapBox,

PHP, Python, ...) le nombre de dimensions spatiales (1D, 2D, 3D, 3D globe), le fonds géonumérique (OSM, Mapbox, Thunder Forest, BingMaps, Google Maps, ...), le public visé, la date de réalisation, le contexte de développement (recherche, public, privé), le domaine (économie, démographie, environnement, littérature, politique, transports, télécommunication, santé, ...), l'échelle géographique, la thématique (mobilités, migrations, sport, armement, ...), le type de déplacement, la population (statistique) concernée et le mode de transport.

Annexe 2. Formation des clusters d'applications

TABLE 1. Formation du cluster 1

Variable	Cla/Mod	Mod/Cla	Global	p.value	v.test
Implantation_lineaire_OUI	97,14	73,91	46,66	3,92e-10	6,25
Type_representation_principale_Carte de flux	85,41	89,13	64	1,25e-08	5,69
Indicateur_OUI	87,80	78,26	54,66	2,14e-07	5,18
Logique_representation_Flux	83,33	76,08	56	1,22e-05	4,37
Type_depl_immateriel	90	58,69	40	2,24e-05	4,23
VV_taille_OUI	85,29	63,04	45,33	0,0001	3,88
Nb_representations_secondaires_2	100	34,78	21,33	0,0001	3,85
Domaine_Demographie	94,73	39,13	25,33	0,0003	3,60
Nb_representations_graphiques_3	100	26,08	16	0,001	3,17
VVD_OUI	89,47	36,95	25,33	0,003	2,96

Cla/Mod : pourcentage d'individus possédant la modalité qui appartient à la classe,

Mod/Cla : pourcentage des individus de la classe qui prennent la modalité,

Global : pourcentage global des individus (i.e. de toute la population) qui prennent la modalité,

p-value : significativité de la différence entre les proportions Mod/Cla et Global,

v.test: une mesure d'association entre variables, mesure l'association de la modalité avec la classe,

TABLE 2. Formation du cluster 2

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Implantation_polylineaire_OUI	100	80	21,33	5,66e-13	7,20
Type_representation_principale_Carte de reseau	100	80	21,33	5,66e-13	7,20
Logique_representation_Liens	100	55	14,66	3,42e-08	5,51
Implantation_lineaire_NON	47,5	95	53,33	6,06e-06	4,52
Indicateur_NON	50	85	45,33	3,56e-05	4,13
Logique_representation_Autre	100	30	8	0,0001	3,72
Domaine_Transport	54,54	60	29,33	0,0009	3,31
Mode_de_transport_Ferroviaire	85,71	30	9,33	0,001	3,25

Geoweb des flux et réseaux

Type_depl_materiel	64,28	45	18,66	0,001	3,21
Thematique_Accessibilite au reseau	100	20	5,33	0,003	2,87

Cla/Mod : pourcentage d'individus possédant la modalité qui appartient à la classe,

Mod/Cla : pourcentage des individus de la classe qui prennent la modalité,

Global : pourcentage global des individus (i.e, de toute la population) qui prennent la modalité,

p-value : significativité de la différence entre les proportions Mod/Cla et Global,

v,test: une mesure d'association entre variables, mesure l'association de la modalité avec la classe,

TABLE 3. Formation du cluster 3

	Cla/Mod	Mod/Cla	Global	p, value	v, test
Type_representation_principale_Mixte	100	66,66	8	4,17e-07	5,06
Thematique_Reseau d'acteurs	100	55,55	6,66	7,30e-06	4,48
Population_Relations scientifiques	100	44,44	5,33	0,0001	3,88
Type_depl_NC	100	44,44	5,33	0,0001	3,88
Contexte_de_developpement=Recherche	50	66,66	16	0,0003	3,60
VV_couleur_NON	32	88,88	33,33	0,0004	3,50
Echelle=Echelle_NC	80	44,44	6,66	0,0004	3,48
Domaine=Domaine_Science	80	44,44	6,66	0,0004	3,48
Domaine=Domaine,NA	100	33,33	4	0,001	3,22
implantation_lineaire_NON	22,5	100	53,33	0,002	3,06
VV_taille_NON	21,95	100	54,66	0,002	2,99

Cla/Mod : pourcentage d'individus possédant la modalité qui appartient à la classe,

Mod/Cla : pourcentage des individus de la classe qui prennent la modalité,

Global : pourcentage global des individus (i.e, de toute la population) qui prennent la modalité,

p-value : significativité de la différence entre les proportions Mod/Cla et Global,

v,test: une mesure d'association entre variables, mesure l'association de la modalité avec la classe,

Estimer des flux de navetteurs avec un modèle gravitaire

Application géomatique en région Provence Alpes Côte d'Azur

Safaâ Amarouch¹, Jean-Baptiste Durand¹,
Didier Josselin², Louissette Garcin¹, Nathalie Brachet²,
Frédéric Audard³, Samuel Carpentier-Postel³

1. Master Géomatique et Conduite de Projets Territoriaux, Coursus Master
Ingénierie "Modélisation Géographique des Territoires"

jean-baptiste.durand@alumni.univ-avignon.fr

2. UMR 7300 ESPACE, Université d'Avignon

didier.josselin@univ-avignon.fr

3. UMR 7300 ESPACE, Aix Marseille Université

frederic.audard@univ-amu.fr

RÉSUMÉ. Dans ce poster, nous présentons le modèle gravitaire sous différents modèles en mettant en évidence les effets de distance et les occasions interposées. Le modèle généralisé de Newton est utilisé et donne des estimations relativement corrects des flux intercommunaux en région PACA.

ABSTRACT. In this poster, we present the gravity model in geography, with different modelling including distance effect and intervening opportunities. The generalized Newton model is used to estimate population flows between municipalities in the French Provence Alpes Côte d'Azur Region. It provides rather good results.

MOTS-CLÉS : modèle gravitaire, occasions interposées, flux de populations intercommunaux, modèle de Newton généralisé

KEYWORDS: gravity model, intervening opportunities, population flows, Newton generalized modelling

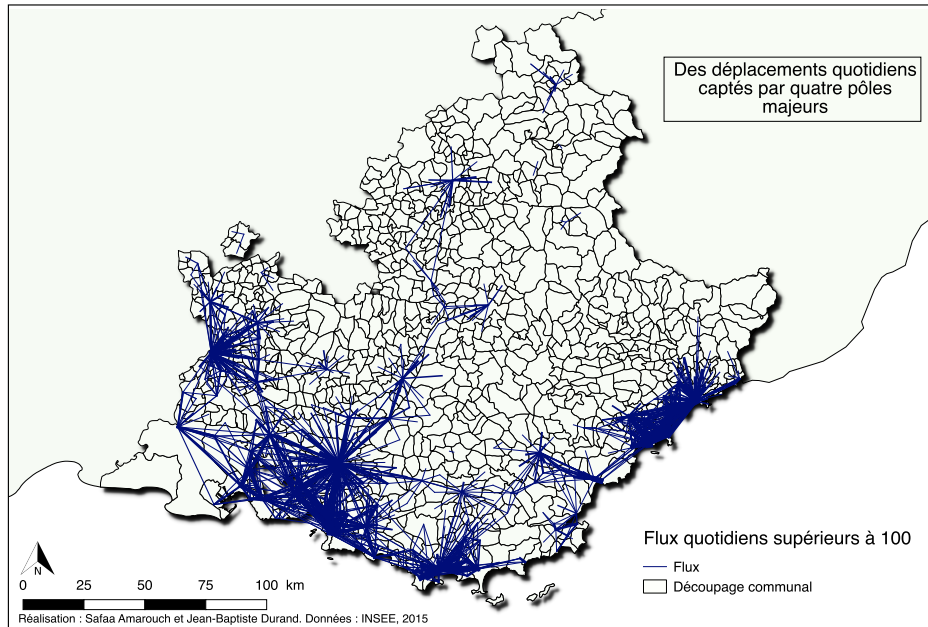


FIGURE 1. Exemples de flux de navetteurs en région PACA

1. Contexte et objectifs

Le modèle gravitaire découle de la propriété très générale selon laquelle deux masses s'attirent ou entretiennent des échanges privilégiés d'autant plus fortement qu'elles sont importantes et proches. Cette propriété se décline dans l'espace, qu'il soit intersidéral ou géographique. Évidemment, les formes de cette interaction et de cette attractivité sont largement différentes selon le contexte considéré. Cela peut en effet se traduire par une force d'attraction, comme par des flux entre les masses. Plus particulièrement, dans l'espace géographique apparaissent plusieurs éléments : la friction, c'est à dire la difficulté à se déplacer, qui est différenciée à cause de l'anisotropie spatiale, mais également la non symétrie des échanges entre deux masses équivalentes (des villes ou des stations de service de transport, par exemple) et l'effet de la topologie d'un ensemble de masses en interactions concurrentielles, constituant un réseau particulièrement complexe à appréhender. La carte de la Figure 1 présente des flux observés journaliers de plus de 100 personnes en région Provence Alpes Côte d'Azur.

Ce poster présente les principales approches dérivées du modèle gravitaire, les discute en mettant en évidence leur spécificité. Ce travail est réalisé dans le cadre d'un projet de recherche action avec la région Provence Alpes Côtes

Estimer des flux de navetteurs avec un modèle gravitaire

d'Azur qui participe à la prospective territoriale régionale en termes de mobilité à l'horizon 2020 et 2040, dans le cadre du Schéma Régional d'Aménagement, de Développement Durable et d'Égalité des territoires. À ce titre, nous avons été notamment sollicités pour analyser et modéliser les flux régionaux. C'est là qu'intervient le modèle gravitaire, dont l'usage peut, dans une certaine mesure, donner une idée des grands mouvements de personnes sur un semis de générateurs de flux et de pôles urbains.

2. Les modèles gravitaires

Auteur	Objectif	Formule	Masse	Distance ($d_{ij} > 0$)	Commentaire
Tobler (1970)	Estimer flux pendulaires	$F_{ij} = \frac{1}{d_{ij}^\alpha}$	Produit de masses ; Asymétrique si exposants ; effet positif	Effet inverse	Calcul log-linéaire ; problème division par 0
Reilly (1931)	Estimer bassin d'attraction	$B_{ij} = \frac{d_{ij}}{1 + \sqrt{\frac{M_i}{M_j}}}$	Rapport de masse j/i ; effet inverse	Au numérateur ; effet positif car calcul de bassin/frontière	Pas de calcul de flux, plutôt bassin
Newton (1684)	Estimation de flux pendulaires	$F_{ij} = \frac{M_i^\alpha M_j^\beta}{d_{ij}^\gamma}$	Produit de masses ; asymétrique si exposants différents ; effet positif	Effet inverse	Calcul des élasticités avec modèle loglinéaire
Wilson (1971)	Estimation de flux migratoires	$F_{ij} = aM_i \frac{M_j f(C_{ij})}{\sum_{k=1}^n M_k f(C_{ik})}$	Produit de masses normalisé ; effet positif	Effet indirect (à travers fonction de coût) ; effet négatif	Introduction indirecte des opportunités interposées
Stouffer (1940)	Estimation de flux par intégration d'occasions interposées dans un rayon r	$F_{ij} = -a.log(f(\Delta)) + c$	à travers $f(s)$ (s = nombre cumulé d'occasions interposées)	Effet négatif	Pas d'explicitation de la distance

FIGURE 2. *Quelques modèles gravitaires - Tableau comparatif*

Dans la Figure 2 nous présentons et comparons 5 approches complémentaires bien connues du modèle gravitaire :

- la loi de Reilly identifiant les bassins d'attractivité gravitaires (Reilly, 1931) ;
- les occasions interposées mises en évidence par Stouffer (Stouffer, 1940) ;
- la fameuse loi d'attraction universelle de Newton qui est généralisée pour cet usage ;
- la première loi de la géographie de Tobler (Tobler, 1970) ;
- l'approche de Wilson qui fait intervenir une fonction de friction dans l'espace (Wilson, 1971).

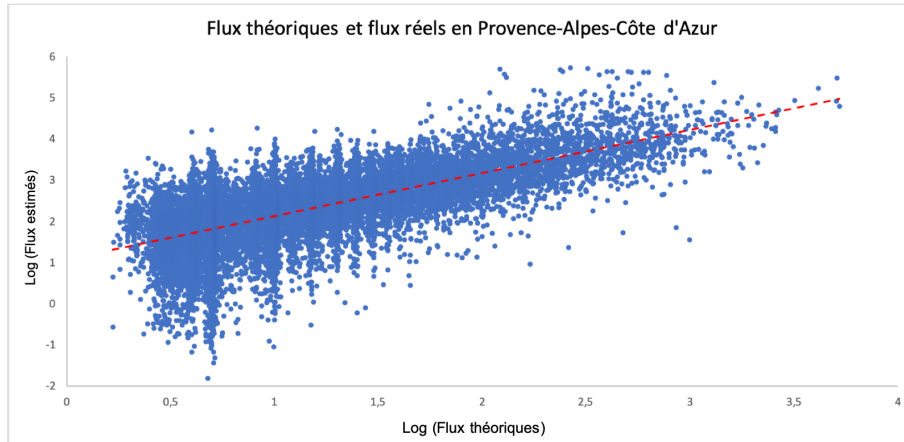


FIGURE 3. *Comparaison entre flux réels et flux théoriques*

Pour Newton, Tobbler et Reilly, la distance est explicite et joue un rôle de frein ou de frontière dans l'attractivité. Pour ces modèles, les entités en relation sont traitées par couples indépendants. Les modèles de Stouffer et Wilson intègrent une forme de friction et prennent en compte l'environnement et les occasions interposées (c'est à dire les entités spatiales disposées entre les deux premières entités en interaction), qui constituent potentiellement une captation de flux théoriques non négligeable, en fonction de la topologie des masses en présence.

3. Premiers résultats

Les traitements ont été réalisés au sein d'un Système d'Information Géographique (QuantumGIS) et d'un logiciel de traitement statistique (R). La figure 4 expose la méthodologie de traitement. Celle-ci a nécessité deux grandes phases. Dans un premier temps, nous constituons une base de données spatiales avec les communes et leurs centroïdes, et nous extrayons les flux observés des fichiers des navettes domicile-travail/étude, ainsi que les distances intercommunales. Dans un second temps, nous calculons les flux théoriques en région Provence Alpes Côte d'Azur à partir des populations communales, en utilisant la formule de Newton généralisée. Ces calculs sont réalisés sous R.

La formule utilisée pour estimer les flux F_{ij} est la forme la plus simple de Newton (cf. Figure 2), où les exposants α et β des masses M_i et M_j sont égaux à 1 et où l'exposant θ de la distance d_{ij} est égal à 2.

$$F_{ij} = \frac{M_i M_j}{d_{ij}^2}$$

Estimer des flux de navetteurs avec un modèle gravitaire

Le nuage de points de la figure 3 présente la qualité de l'ajustement entre les flux estimés et les flux théoriques. Nous obtenons un R^2 de 0,47 ce qui est assez correct et encourageant (car il s'agit d'un R^2). À noter une certaine difficulté rencontrée pour prévoir les petits flux par rapport aux grands flux. Les coefficients en puissance des masses et de la distance pourraient être potentiellement modifiés pour améliorer l'ajustement. D'autres améliorations peuvent être envisagées si l'on prend en compte les occasions interposées, déviant des flux de leur destination théorique optimale.

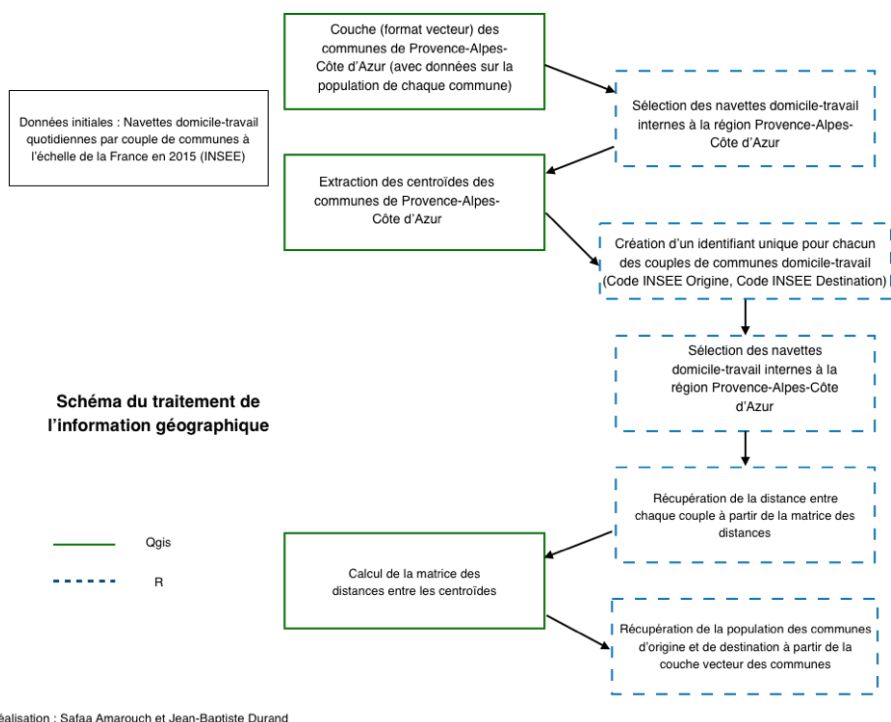


FIGURE 4. *Traitement de l'Information Géographique*

4. Conclusion

Pour de grandes estimations non localisées, comme dans le cadre du projet de la région Provence Alpes Côte d'Azur, les modèles gravitaires constituent une analogie intéressante utilisable pour l'estimation des flux. Les résultats peuvent paraître probants et utilisables pour avoir une idée des grands mouvements et des masses de flux en présence.

Mais des progrès restent à réaliser pour prendre en compte :

SAGEO'2019

- les fortes attractivités aux distances faibles ou nulles (impossibilité de calcul à cause de la division par 0 notamment),
- les occasions interposées déviant ou absorbant des flux de voisinage,
- la topologie et l'organisation spatiale des points de gravité dans l'espace géographique.

L'ensemble des flux gravitaires et leur équilibre, considérés dans un système, est plus que la somme des flux par couple d'entités en interaction.

5. Remerciements

Les auteurs de ce poster remercient vivement la Région Provence Alpes Côte d'Azur pour son partenariat et son soutien financier dans ce projet de recherche-action pédagogique impliquant les étudiant.e.s du Coursus Master Ingénierie "Modélisation Géographique des Territoires" d'Avignon Université.

Bibliographie

- Reilly W. (1931). *The law of retail gravitation*. Knickerbocker Press.
- Stouffer S. A. (1940). Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, vol. 5, n° 6, p. 845–867.
- Tobler W. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, vol. 46, n° 2, p. 234–240.
- Wilson A. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning*, vol. 3, p. 1-31.

MK-MOBIBACK : un dispositif hybride et intégré pour enquêter finement les mobilités quotidiennes des familles

Depeau, S.¹, Bedel, O.², Cherel, P.³, André-Poyaud, I.⁴, Catheline, Y.³, Chardonnel, S.⁴, Gombaud, J.⁵, Jambon, F.⁶, Lepetit, A.¹, Mericskay, B.¹, Quesseveur, E.¹, Leprince F.²

1. UMR ESO CNRS/Université de Rennes
Place du recteur le Moal, 35043 Rennes, France
sandrine.depeau@univ-rennes2.fr, arnaud.lepetit@univ-rennes2.fr,
boris.mericskay@univ-rennes2.fr, erwan.quesseveur@univ-rennes2.fr
2. Alkante
4 rue Alain Colas, 35530 Noyal sur Vilaine
o.bedel@alkante.com
3. RF-TRACK
8 rue de la Rigourdière, 35510 Cesson Sévigné
pierre.cherel@rf-track.com
4. Univ. Grenoble Alpes, CNRS, Science Po Grenoble, PACTE,
38000 Grenoble, France
isabelle.andre-poyaud@univ-grenoble-alpes.fr,
sonia.chardonnel@univ-grenoble-alpes.fr
5. Akhos
34 Rue Frédéric le Guyader, 35200 Rennes
jgombaud@akhos.fr
6. Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG,
38000 Grenoble, France
francis.jambon@imag.fr

RESUME. MK-MOBIBACK est un dispositif hybride et intégré conçu pour enquêter finement et en continu, les mobilités quotidiennes des familles. Cet article présente au préalable les enjeux et le contexte scientifique dans lesquels il s'inscrit, l'histoire du dispositif, résultat d'une longue et riche collaboration entre plusieurs partenaires. Il est ensuite décrit dans ses différentes composantes technologiques, informatiques suivant un cahier des charges précis et adapté. Sont donc présentés 1/ les capteurs spécifiques et leurs fonctionnalités et 2/ une application questionnaire sur tablette supportée par une plateforme de gestion et stockage de

SAGEO'2019

données. Enfin, le dispositif est situé dans le contexte d'application qui a servi à son développement.

ABSTRACT. The article presents MK-MOBIBACK, a hybrid and integrated device allowing to investigate finely and continuously daily mobility of families. Recalling the scientific context and its challenges as well as the scientific and collaborative genealogy through these programmatic phases, the device is described in its different technological and computer components according to precise and adapted specifications. It therefore presents 1/ the specific sensors and their functionalities and 2/ a tablet questionnaire application supported by a data management and storage platform. Finally, the device is presented in its applicative context which served to develop it.

MOTS-CLES : GPS, Capteurs, Appli-questionnaire, Gestion de données, Mobilité quotidienne, Enfants, Famille.

KEYWORDS: GPS, Sensors, Questionnaire-App, Data Management, Daily Mobility, Children, Family.

1. Introduction

1.1. Contexte scientifique et opérationnel : *MOBIKIDS* et ses objectifs

Cet article présente un dispositif méthodologique hybride et intégré de collecte de données de mobilités et activités quotidiennes, développé à la suite d'une série de programmes collaboratifs et révisé dans sa forme finale dans le cadre du programme de recherche MOBIKIDS (ANR-16-CE22-0009). Ce dernier cherche à comprendre le « rôle des cultures éducatives (CEU) dans l'évolution des mobilités quotidiennes et des contextes de vie des enfants » à partir de « la collecte et des analyses de traces géolocalisées et enrichies sémantiquement ». Autrement dit, il s'agit de décrire, d'une part les conditions d'autonomie de déplacements des enfants et d'autre part, de vérifier l'hypothèse de formes de *cultures éducatives urbaines* variant selon les lieux de vie, les situations sociales et modes de vie. Dans cette perspective, et plus spécifiquement dans un contexte de transformations des rapports des familles aux espaces et au temps, la compréhension des conditions d'autonomie de déplacement ne se limite pas au seul niveau de l'individu mais bien dans le système social (l'entourage, la famille), le système temporel (les temps de vie, les horaires des services, des activités) et le système spatial (son environnement, son accessibilité) dans lesquels s'insèrent les mobilités. Ainsi, l'objectif de collecte est de considérer, dans un souci de validité écologique optimale, et de manière concomitante, les temps quotidiens des enfants et des parents (ou adultes de l'entourage) et les espaces de vie. Pour répondre à ces objectifs, les démarches méthodologiques s'appuient sur des protocoles d'enquêtes et de relevés fins de trajectoires, en plein essor depuis une quinzaine d'années, grâce notamment aux apports des technologies numériques et des objets connectés.

Le dispositif MK-MOBIBACK s'appuie, pour sa part, sur un système couplant enregistreur de données GPS et questionnaire sur tablette. Il recouvre certaines particularités fondamentales qui ont guidé les recherches et développements. Notamment l'exigence de validité écologique qui suppose, outre l'autonomie en énergie du dispositif pour la continuité de collecte, la discrétion et l'acceptation des outils de collecte par les personnes enquêtées, la précision spatio-temporelle de la collecte des données et par conséquent, la protection et le respect des données individuelles et la mise en pratique aisée de l'enquête. Enfin, le besoin lié au passage à l'échelle dans la collecte de données, conduisant également à prévoir la gestion d'un grand volume de données, constitue un dernier point de cadrage du développement du dispositif. Ainsi, afin de présenter le dispositif MK-MOBIBACK, nous l'inscrirons dans un premier temps dans le champ des développements méthodologiques actuels associés aux mobilités (notamment enfantines), en retraçant la généalogie des programmes collaboratifs de l'équipe MOBIKIDS qui ont contribué à son développement et à ses améliorations. Ensuite, après avoir décrit le dispositif de collecte dans ces différentes phases (partie 2), seront présentés les différents outils composant le dispositif selon le cahier des charges pour expliciter les diverses fonctionnalités, en partant des différentes étapes de collecte (partie 4).

1.2. Contexte scientifique international des développements méthodologiques dans le champ des mobilités

Depuis une vingtaine d'années, les expérimentations de suivis de mobilité par traqueurs GPS (Wolf & al., 2001) constituent un champ de développement méthodologique prolifique de par les solutions apportées sur le plan opérationnel, social et économique. Ces solutions participent par ailleurs, en France, à la rénovation des enquêtes ménages-déplacements (EMD) via les dispositifs EMC2 du Cerema.

Ciblant plus classiquement des populations adultes, c'est dans le champ des recherches liées aux mobilités enfantines que ces développements posent le plus de défis à la fois éthiques et technologiques. Pour assurer la bonne opérationnalité de la collecte avec les enfants, les dispositifs par capteurs GPS requièrent d'ailleurs des précautions et des méthodes spécifiques (Christensen & al., 2011). Pour autant, et plus encore qu'avec des populations adultes, la collecte de données de mobilité enfantine usant de capteurs GPS permet de dépasser un certain nombre de limites en ce qui concerne d'une part, la qualité des données (la précision, la complétude et la représentativité des données) et d'autre part, la faisabilité ou l'opérationnalité des procédures de collecte (contrat de confiance formalisé avec la famille du fait de la démarche CNIL/RGPD obligatoire, disponibilité des enfants, facilité de compréhension, fatigue et oubli évités, etc.) et d'analyse (Depeau & Quesseveur, 2014). Et si l'un des enjeux importants (pour la qualité des analyses) concerne la taille des échantillons enquêtés, l'usage des capteurs sur un temps de collecte plus long que la période classique définie dans les Enquêtes Ménage Déplacements (1 jour), permet de tolérer un échantillon d'individus plus réduit (ce que suggèrent Stopher & al. 2008 sur la base des travaux de Pas, 1985).

Cependant, ce gain d'échantillonnage conduit à une autre exigence qui a été souvent discutée au départ des travaux utilisant les capteurs GPS, celle de la qualité des informations complétant les données de traces (Stopher & al. 2005). Ce besoin de précisions classiques sur les déplacements et activités (en termes de modes et de motifs) liées aux traces, qui, collectées in vivo sans enrichissement restent muettes et hors contexte donc peu exploitables, conduit depuis quelques années au développement de trois sections importantes relatives au traitement des traces : le nettoyage, le séquençage des traces et l'identification de modes qui sont associés (Stopher & al., 2005, Dalupines & Scott, 2017). Ces trois opérations, par ailleurs centrales dans le dispositif MK-MOBIBACK peuvent, dans certains cas, rendre l'outil de collecte de données captées par GPS auto-suffisant (Schüssler & Axhausen, 2009). Cette limite ne peut être dépassée qu'à l'aide de l'enrichissement des traces qui est opérationnalisée sous des formats très variés (par questionnaire ou carnet de route) en direct (avec des smartphones intégrant un capteur GPS par exemple) ou en différé (via des questionnaires administrés ad hoc).

Dans cette diversité de protocoles hybrides couplant capteurs et questionnaire post-collecte de traces (Shen & Stopher, 2015), l'usage et l'opérationnalité de certains outils est questionnée, notamment ceux développés autour d'applications smartphone. Ces derniers, en permettant d'obtenir en direct des compléments

d'information sur les trajets et activités pourraient sembler a priori plus adaptés. Or, dans notre contexte, celui de populations enfantines, ce dispositif interactif (aux données GPS actives) via le téléphone mobile, n'est pas adapté. En effet, il ne permet pas d'assurer la pleine implication d'enquêtés, du fait de la sollicitation continue qu'il exige et des compétences numériques requises (Zou & Golledge, 2007). Par ailleurs, le système interactif favorisé par l'usage du smartphone rend le dispositif intrusif, et de fait, ne permet pas d'assurer la validité écologique de l'enquête ni de garantir un niveau de confidentialité des données suffisant. En outre, il augmente la pénibilité de la collecte pour les enquêtés et le risque d'abandon.

1.3. Contexte du partenariat et généalogie du dispositif technologique

Dans ce contexte d'évolution rapide des dispositifs de collecte de données de déplacements, la mise en œuvre du protocole de collecte de données de déplacement MK-MOBIBACK (développé autour des capteurs GPS, accéléromètres et associé à un questionnaire de retour sur traces) s'inscrit dans une suite de programmes de recherche collaboratifs : GETM (2005), PUCA (2009), TRACE (2010), MOBITRACE (2011), et SM2D (2014). Ces programmes (Legendre & al., 2005, Depeau & Quesseveur, 2009 ; Christophe & al., 2010 ; Drevon & al., 2014 ; Depeau & Quesseveur, 2014 ; Depeau & al., 2017). Ces derniers, ont permis, outre de renforcer une dynamique collaborative, de définir progressivement une série de verrous à dépasser (technologiques, éthiques, informatiques, d'opérationnalisation de l'enquête) tout en participant à la configuration de la présente méthode intégrée. De là, la première version de la méthode intégrée présentée ici a été développée (dans ses aspects matériels, ergonomiques, technologiques et de couplage informatique) dans le cadre du programme ASTROLLENDRO (2014-2016), avec une enquête test menée auprès d'un échantillon réduit de familles tout venant. Le bilan de cette première version (André-Poyaud & al., 2016) a permis de réviser et d'améliorer certains points (ergonomie, précision des capteurs, automatisation plus étendue, etc.) qui ont servi de base au développement et à son application, pour un passage à l'échelle dans le cadre du programme MOBIKIDS (2017-2021).

2. Présentation générale du dispositif de collecte

Pour rappel, l'enjeu méthodologique était donc de développer un dispositif de collecte de données GPS de type passif (sans transmission des données en temps réel à distance), autonome (du point de vue de la batterie) et couplé à un outil (questionnaire d'enrichissement et compléments des traces sur tablette) qui respecte les principes déontologiques de relations aux enquêtés, les exigences de protection des données personnelles et la mise en pratique aisée de l'enquête envisagée. Cette exigence est d'autant plus fondamentale qu'elle repose sur un gros volume de données à collecter qui requiert, outre un outil de gestion de base de données, des procédés automatisés, notamment pour le séquençage des traces. Enfin, un autre enjeu de taille reposait sur les particularités des populations enquêtées qui ne se

limitent pas aux populations adultes mais tiennent compte de publics spécifiques, comme les enfants.

Le dispositif de collecte d'un point de vue technologique se compose d'un ensemble d'enregistreurs de données, ou « dataloggers » en anglais, intégrés sous la forme de boîtiers {A} comprenant un GPS et un accéléromètre, d'une plateforme centralisée {B} de collecte des données brutes des boîtiers, de traitement (séquençage) et de mise à disposition (via une API) et d'une application Android « questionnaire » pour tablettes {C}. Cette dernière permet à la fois de saisir des informations sur l'enquêté et de recueillir ses commentaires sur l'ensemble des 5 jours de « traces » segmentées (déplacements, arrêts) et de l'interroger sur ses pratiques de mobilité et activités quotidiennes.

La plateforme centralisée s'appuie sur un serveur de fichiers et une base de données spatio-temporelles (PostgreSQL/PostGIS) qui assurent le stockage sécurisé des différentes données collectées et leur mise à disposition pour analyse ultérieure par les différents partenaires du projet. Le schéma suivant (figure 1) illustre le protocole de collecte des traces avec, dans sa partie gauche, les interactions entre les différents éléments du dispositif, et dans sa partie droite, les étapes successives d'une enquête.

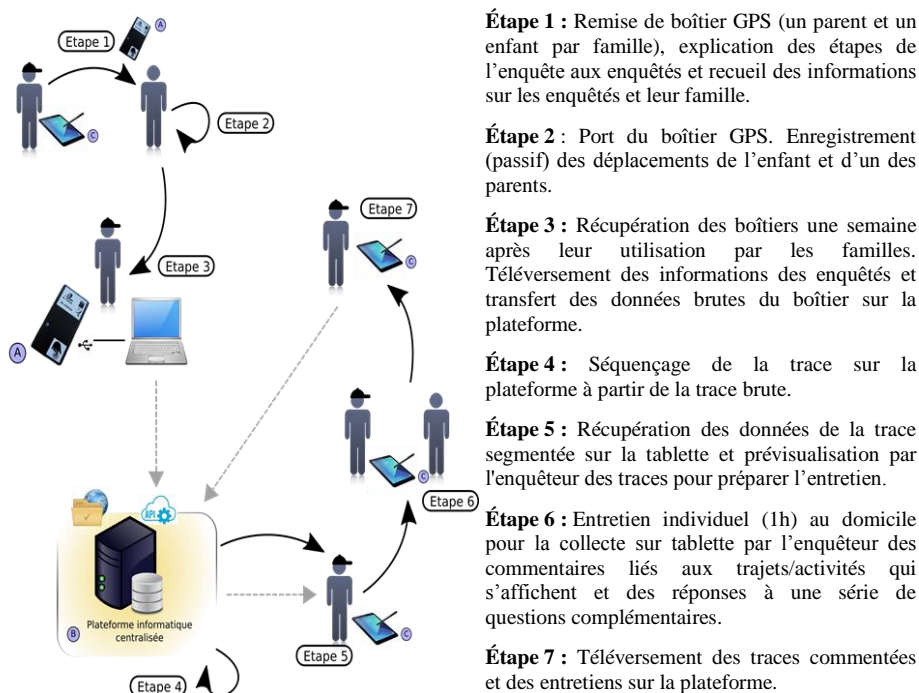


FIGURE 1. Protocole de collecte des traces

3. Développement de boîtiers enregistreurs spécifiques dédiés à des populations spécifiques

3.1. Origine & cahiers des charges

Le boîtier enregistreur de données a été développé par le partenaire RF-TRACK sur la base d'un cahier des charges répondant à des besoins spécifiques adaptés en priorité à des populations d'enfants. Le boîtier d'enregistrement a tenu compte au préalable des différents besoins suivants : **fonctionnels** (autonomie de la batterie d'au moins 5 jours, fréquence de capture de données adaptée) ; **ergonomiques** (taille adaptée à un port en poche, discrétion du boîtier, fixation aisée, LEDs de signalisation-utilisateur, technique de marche-arrêt simple et sécurisée, rechargement électrique simplifié via un port USB) ; **légal** (possibilité d'arrêter volontairement l'enregistrement avec la touche tactile). Ce prototype développé en septembre 2015 et testé dans un premier temps sur un petit échantillon (16 individus d'une population tout-venant et une dizaine de chercheurs/développeurs-experts), a subi ensuite quelques évolutions en janvier 2018 pour permettre d'une part, d'élargir l'étendue des collectes simultanées liée la taille des populations enquêtées et d'autre part, d'améliorer le processus de séquençage.

3.2. Développement et fonctionnalités

Le boîtier (figure 2) permet d'enregistrer sur une mémoire interne les mesures brutes issues des capteurs suivants, à des fréquences prédéfinies : accéléromètre (25 Hz), magnétomètre (5 Hz) et GPS (1 Hz). Chaque mesure est enregistrée de façon continue sur la mémoire interne du boîtier. Ce type de boîtier est qualifié de passif car l'enregistrement est automatique ; il n'y a besoin d'aucune action volontaire de l'individu pour la prise de mesure, et il n'y a aucune transmission à distance en temps réel des mesures. Les données sont encodées dans la mémoire interne sous la forme d'un fichier binaire, au format ad hoc, qui est accessible par USB. Une touche tactile permet de démarrer/arrêter l'enregistrement. C'est sur la base de ces données brutes que sont ensuite appliqués les algorithmes de post-traitement développés par le partenaire Alkante (traitement des données « *a posteriori* »).

La conception des enregistreurs s'est déroulée en plusieurs phases : fabrication d'une carte électronique et élaboration d'un logiciel embarqué associé puis mise en place d'un boîtier, et constitution d'un support de chargement électrique/déchargement électronique des données par port USB. La dernière version permet d'enregistrer un point GPS par seconde (au lieu de toutes les 5 secondes dans sa version initiale). Le composant GPS a été amélioré pour optimiser les performances GPS, notamment le temps de démarrage. Les données de l'accéléromètre servent aujourd'hui à la détection de mouvement pour stopper la fonction GPS (pendant la nuit par exemple) de manière à optimiser l'autonomie et atteindre 5 jours d'utilisation. Les informations du magnétomètre sont collectées,

mais actuellement non encore exploitées. Les données brutes de l'accéléromètre et du magnétomètre pourraient servir de base pour atteindre d'autres objectifs futurs : détection du mode de transport pédestre ou motorisé par exemple, estimation du trajet dans les zones sans couverture GPS, etc. D'autres informations complémentaires sont enregistrées par le boîtier et utilisées pour les phases de post-traitement, notamment pour le séquençage des traces. Ces informations comprennent des métadonnées du composant GPS : nombre de satellites, vitesse instantanée, cap, estimation de précision (« h dop ») et temps de démarrage (« time-to-fix »). Le niveau de la batterie est aussi enregistré régulièrement, ainsi que les moments et les causes (arrêt sur batterie faible ou par l'utilisateur) de démarrage et d'arrêt du boîtier.



FIGURE 2. Boîtier enregistreur (RF-TRACK)

4. Développement d'une application sur tablette adaptée aux exigences d'enquête du programme MOBIKIDS

Le dispositif de collecte et d'enregistrements des données de déplacements (boîtier enregistreur) est complété pour la partie formalisation et enrichissement des traces par une application développée sur tablette pour la collecte de données questionnaires (commentaires des traces des séquences mobilité-immobilité). Le passage des données du boîtier à la tablette s'opère via la plateforme informatique de gestion des données d'enquête (serveur web et base de données, application de séquençage et API de communication). Ce dispositif intégré répond à une revue de besoins spécifiques formulés par les chercheurs d'ESO et leurs partenaires. Il permet ainsi : un **séquençage automatisé** et rapide des traces pour un retour dans un délai réduit aux enquêtés afin de ne pas compromettre leur mémoire et capacité à commenter les traces ; une **fluidité de transmission** en quantité de données et la sécurisation du stockage et du partage de données de nature variées (quanti et quali) ; une **facilitation de la saisie** de données d'enrichissement des traces par un enquêteur en entretien avec la personne enquêtée.

Le prototype a été développé pour ses volets technologique et informatique d'appariement boîtier-tablette et sémantique d'optimisation de formulaires de saisie (visualisation attrayante, synthétisée et séquencée des traces collectées). Il est le résultat d'un long travail de collaboration et d'une série de tests d'enquêtes in-vivo et in-situ au cours du programme ASTROLLENDRO.

L'application a été développée pour tablette sous système d'exploitation Android (version ≥ 6.0), au moyen de l'environnement de développement multiplateforme Ionic (version 1) basé sur la technologie Cordova/PhoneGap. Le code de l'application est écrit en langage JavaScript et s'appuie sur le système AngularJS pour le cœur de l'application et sur la bibliothèque Leaflet pour les interfaces cartographiques.

4.1 Principes de fonctionnement de l'application

L'application tablette est organisée autour du concept d'« enquête » qui permet de caractériser le suivi d'un individu par le dispositif sur la période de 5 jours. L'enquêteur crée donc une nouvelle enquête pour chaque individu suivi, en lui associant notamment un identifiant de boîtier. Plusieurs enquêtes ¹peuvent être éditées sur une même tablette et une même enquête peut être entamée sur une tablette et poursuivie sur une autre. L'application tablette communique avec la plateforme informatique pour transférer une enquête sur la base de données centralisée (téléversement) ou en récupérer une déjà sauvegardée (téléchargement). La saisie des données sur la tablette comprend deux phases successives renseignées à deux moments distincts de l'enquête.

La **phase 1** est renseignée au moment de la remise des boîtiers et donc avant le port de celui-ci par l'enquêté. Elle est composée de rubriques de questions portant sur : l'« individu » (âge, sexe, profession, situation familiale, situation résidentielle), sa « scolarité » ou formation, les équipements de « mobilité », et « communication ». Sont également saisis les points d'intérêts (au moins 5) de la personne enquêtée avec leur localisation géographique. Un point d'intérêt est un lieu que la personne fréquente régulièrement (ex. domicile, lieu de travail, école...). Ils sont utilisés lors de la phase de séquençage pour annoter automatiquement les lieux en début ou en fin de déplacements.

Afin de limiter le temps de recueil d'information et de faciliter la saisie des réponses dans cette phase, différents mécanismes ergonomiques ont été mis en place sur la tablette. L'utilisation de questions fermées ou semi-fermées a été privilégiée pour permettre au maximum une saisie tactile des réponses. Les questions sont organisées en différentes rubriques thématiques visant à favoriser ainsi la succession des écrans au défilement vertical sur la tablette. Une symbolique par pastille rappelle, au sein des différents menus et catégories, la complétude (finalisé/incomplet) des réponses pour la rubrique. Un mécanisme de verrou permet de passer simplement certains formulaires en mode lecture seule pour éviter, lors d'une consultation ultérieure, toutes modifications non désirées par un appui intempestif sur une zone du formulaire. L'interface de saisie des points d'intérêts offre la possibilité de positionner un point sur la carte par pointé direct ou à partir d'une adresse. Un positionnement approximatif peut être précisé en glissant déposant la puce de localisation directement sur la carte.

¹ Le terme « enquête » est entendu ici dans une acception adaptée et réduite signifiant « l'ensemble des données collectées pour un enquêté »

La **phase 2** est complétée après le port et la récupération des boîtiers et le séquençage des traces. L'algorithme utilisé pour le séquençage repose sur trois concepts : 1/ **Arrêt (lieu)** : soit un ensemble de positions regroupées dans un rayon de moins de 50 mètres de distance sur un temps d'au moins 300 secondes ou à moins de 300 secondes d'une mise en veille du boîtier. 2/ **Trajet** (déplacement) : ensemble de positions consécutives entre 2 arrêts. 3/ **Arrêt transitoire** : ensemble de positions d'un trajet, contenues dans un rayon de 50 mètres, durant une période comprise entre 30 et 300 secondes et dont la vitesse est inférieure ou égale à 3 km/h. L'objectif de la phase 2 est double. Elle permet d'une part, de caractériser et d'enrichir sémantiquement les différentes séquences de mobilité/immobilité identifiées par le séquençage. Ces séquences sont organisées sous la forme de lieux, déplacements et arrêt transitoires. Cette détection de « pratiques » est pleinement liée à la thématique de recherche liée à l'enfance et vise principalement à comprendre un ensemble de « pratiques » souvent laissées invisibles dans les enquêtes (Depeau & al., 2017). Cette phase permet d'autre part, de soumettre des questions complémentaires adaptées aux objectifs ou questions de recherche (évaluation de l'autonomie, socialisation, temps libres, habitudes familiales, etc.) et formulées de façon ouverte, fermées et à partir d'échelle d'évaluation. La phase 2 se clôture par un recueil de précisions sur les modalités d'enquête (identité enquêteur, jour, heure, durée et conditions matérielles de passation).

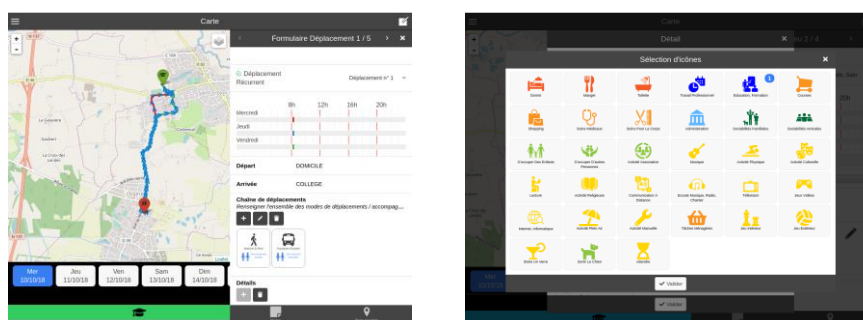


FIGURE 3. Copie d'écran tablette Phase 2, commentaire traces

Pour réduire le temps de saisie, les séquences de mobilité/immobilité de l'enquête ne sont pas présentées de façon chronologique mais sous la forme de trois rubriques en fonction du type de séquence : lieux, déplacements, arrêts-transitoires. Chaque type de séquence est catégorisé selon sa fréquence d'apparition dans les traces (récurrente, secondaire, exceptionnelle), permettant ainsi à l'enquêteur de renseigner en premier et une seule fois les séquences les plus rencontrées (présence au domicile, trajet domicile-travail ou domicile-école...). Chaque séquence est présentée dans une fenêtre réunissant trois vues interactives et synchrones (figure 3) : 1. Une vue cartographique centrale composée d'une carte pour la visualisation de la séquence (un lieu, un déplacement avec ou non le.s arrêt.s transitoire.s) dans la trace ; 2. Une vue temporelle dans la partie basse avec une frise chronologique déroulant l'ensemble des déplacements et activités sur la période enquêtée ; 3. Une

vue sémantique rétractable sur la droite, sous la forme d'un formulaire appelant pour chaque objet (lieu, déplacement, arrêt-transitoire) à : 1/ confirmer la ou les séquences temporelles concernées par la trace à enrichir ; 2/ décrire les lieux (toponymie, fonction, activités associées ; fréquence d'usage) et déplacements (modes d'accompagnement et déplacement associé.s à la chaîne de déplacement, la fréquence de déplacement) ; 3/ décrire les arrêts transitoires (activités associées, confirmation de la temporalité et de la localisation) et le cas échéant, 4/ annoter sous forme de commentaire l'objet à renseigner. Afin de fonctionner sans connectivité internet (wifi ou mobile), l'application embarque un ensemble de fond de plan (carte et photographie aérienne) limité au périmètre de la zone géographique de l'enquête.

5. Conclusion

Le dispositif MK-MOBIBACK, révisé et enrichi dans le cadre du projet MOBIKIDS, couple une technologie passive de collecte des traces de déplacement des individus (boîtier enregistreur GPS) et une application mobile (tablette) d'enrichissement de ces traces par des commentaires post-collecte. Ce dispositif hybride facilite la collecte des données de mobilité de populations spécifiques (adultes, mais aussi enfants), dans leur espace quotidien, et sur une période d'observation conséquente (5 jours). Il a été utilisé dans le cadre d'une enquête réalisée dans le contexte rennais (centre-ville de Rennes et Orgères, une commune dans le sud de la métropole rennaise). Il a pu être éprouvé sur différentes périodes de collecte semi-longitudinales (printemps 2018, automne 2018, hiver 2019) auprès de 89 familles ayant un ou des enfant.s scolarisé.s en CM1/CM2, soit deux échantillons de 91 parents et 91 enfants. L'enquête s'est déroulée par cohorte de 12 à 15 familles simultanément afin de permettre des relevés sur des périodes temporelles identiques.

Cette expérimentation terrain a permis de valider les points forts du dispositif pour réduire le temps de la collecte des données de mobilité. Plus particulièrement, le recueil des commentaires post-collecte est facilité par l'application questionnaire tablette s'appuyant sur des interfaces ergonomiques et majoritairement sur des modalités de réponses fermées. Et plus généralement, la connexion entre les différents éléments du dispositifs (boîtier enregistreur, plateforme centralisée, application questionnaire) a permis de limiter les délais de ruptures de collecte de données, de gagner en confort de collecte et d'interaction avec les enquêtés.

Malgré les apports du dispositif, ce type de protocole de suivi des mobilités quotidiennes nécessite une équipe conséquente d'enquêteurs ainsi qu'un temps important de gestion logistique à la fois du matériel et des enquêtes en cours. Par conséquent, tout élément permettant à la fois de réduire les temps de préparation et de transfert d'information, et d'alléger la charge mentale des enquêteurs et des responsables d'enquête est une source d'amélioration du dispositif. Notamment, la phase de transfert des données brutes du boîtier vers la plateforme pourrait être automatisée et couplée au processus de rechargement électrique du boîtier. De même, la mise en place d'un mécanisme de notification (mail) ou d'un tableau de

bord numérique récapitulatif de l'état d'avancement des différentes enquêtes permettrait également de réduire des étapes de vérifications manuelles parfois coûteuses car réparties sur plusieurs opérateurs.

Remerciements

Les travaux liés au dispositif présenté dans cet article sont en partie financés par les programmes ASTROLLENDRO (Région Bretagne, 14007516) et MOBIKIDS (ANR-16-CE22-0009).

Bibliographie

- Andre Poyaud, I., Bedel, O., Catheline, Y., Chardonnel, S., Depeau, S., Gombaudo, J., Jambon, F., Lepetit, A., Mericskay, B., Quesseveur, E. (2016). Le comportement des piétons au cœur des enjeux méthodologiques liés aux données géolocalisées. Retour critique sur les avancées technologiques et méthodologiques de programmes de recherches. *5ème colloque COPIE 2016 : Des piétons et des villes : connaissances, enjeux, cultures de la marche*, 11-13 juillet 2016, Paris.
- Christensen, P., Mikkelsen, M. R., S. Nielsen, T. A., Harder, H. (2011). Children, Mobility, and Space: Using GPS and Mobile Phone Technologies in Ethnographic Research. *Journal of Mixed Methods Research*, 5 (3), p. 227–246.
- Christophe, S., André-Poyaud, I., Chardonnel, S., Davoine, P.A., Jambon, F., Lbath, A.A., Lutoff, C., Mandran, N., Meillon, B. (2010). Human Behaviour Analysis during Remarkable Flood Episodes: Proposition of Automatic Survey Protocols based on Mobile Technologies, *Proceedings of the Workshop on "Movement Research: are you in the flow?"* AGILE 2010.
- Dalumpines, R. & Scott, D.M. (2017). GIS-based episode reconstruction toolkit (GERT): A transferable, modular, and scalable framework for automated extraction of activity episodes from GPS data. *Travel Behaviour and Society*, vol. 11, n° April 2018, p. 121-130.
- Depeau, S., Chardonnel, S., Andre-Poyaud, I., Lepetit, A., Jambon, J., Quesseveur, E., Gombaudo, J., Allard, T., Choquet, C.A. (2017). Routines and informal situations in children's daily lives. *Travel Behaviour and Society*. vol. 9, n° oct. 2017, p. 70-80.
- Depeau, S. & Quesseveur, E. (2014). A la recherche d'espaces invisibles de la mobilité : usages, apports et limites des techniques GPS dans l'étude des déplacements urbains à l'échelle piétonne. *Netcom*, vol 28, n°1-2, p. 35-54.
- Depeau, S. & Quesseveur, E. (2009). Zone atelier d'expérimentation : de la mobilité encadrée à la mobilité canalisée, dynamique ambulatoire des effets de seuils de la sécurité. In S. Depeau, M. Dumont, L. Grésillon, L. Rougé, S. Stephan, O. Thomas, S. Valorgnes. *Des périphéries urbaines, sous tranquillisants ? Productions et dynamiques des contextes de sûreté dans les territoires de l'Ouest*. Rapport final de recherche, PUCA / Club Ville Aménagement « Territoires urbains et sûreté », octobre 2009, p. 230-260.
- Drevon, G., Jambon, F., Chardonnel, S., Christophe, S., André-Poyaud, I., Davoine, P.A., Lutoff, C. (2014). Évaluation comparée de l'apport de l'assistance GPS aux enquêtes de mobilité. *Netcom*, Vol. 28, n°1/2, 13–34.

- Legendre, A., Boareto, M., Boulanger, D., Cecconi, C., Collavo, N., Delalande, V., Depeau, S., Enaud, C., Hirsch, J., Lavarde, A.-M., Milot, J.-L., Ramadier, T., Savina, Y., Tran, M., & Ulrich, P. (2005). Développement des pratiques urbaines au cours de l'enfance : Apport des nouvelles technologies dans l'analyse spatiale de la mobilité et de l'usage des espaces publics. *Actes du colloque de bilan du programme interdisciplinaire "société de l'information"*. 2005, Lyon: CNRS.
- Pas, E.I. (1986). Multiday samples, parameter estimation and data collection costs for least square regression trip generation models. *Environment and planning, Part A*, n° 18, p. 73-87.
- Schüssler, N., & Axhausen, K. W. (2009). Processing GPS raw data without additional information. *Transportation Research Record*, n° 2105, p. 28-36.
- Shen, & Stopher, (2014). Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews*, Vol. 34, n° 3, p. 316–334
- Stopher, P., K. Kockelman, S.P. Greaves, and E. Clifford (2008). Reducing Burden and Sample Sizes in Multiday Household Travel Surveys, *Transportation Research Record Issue*, n° 2064, 2008, p. 12-18.
- Wolf, J., Guensler, R., Bachman, W. (2001). Elimination of the travel diary: Experiment to derive trip purpose from Global Positioning System travel data. *Transportation Research Record Issue*, n° 1768, 2001, p. 125-134.
- Zhou, J.J. & Golledge, R. (2007). Real-time tracking of activities scheduling/schedule execution with a unified data collection framework. *Transportation Research Part A: policy and practice*, vol. 41, n° 5, p. 444-463.

WIKISemio : un outil de testing collaboratif de cartes : application au risque de piqûre de tiques

Sandro Bimonte¹, Francois Johany², Gabriel Noiret², Isabelle Lebert³, Magalie René-Martellet³, Sylvain Dernet²

1. Université Clermont, TSCF, Irstea
9 Av. Blaise Pascal, 63178, Aubière, France
sandro.bimonte@irstea.fr

2. INRA, UMR Territoires
9 Av. Blaise Pascal, 63178, Aubière, France
francois.johany@inra.fr; sylvain.dernet@inra.fr

3. INRA - VetAgro Sup, UMR EPIA,
63122 Saint Genès Champanelle, France
isabelle.lebert@inra.fr; magalie.renemartellet@vetagro-sup.fr

RESUME. Les outils du Web2.0 (wiki, forum) ont été déjà utilisés avec succès dans les systèmes d'information géographique (SIG) pour l'édition collaborative de contenus géo-référencés. Par contre, aucun travail ne s'est intéressé au testing collaboratif de la sémiologie des cartes. Pour pallier ce manque, nous proposons WIKISemio, un outil web qui intègre un outil WIKI à un SIG pour permettre l'évaluation collaborative des variables sémiologiques de cartes.

ABSTRACT. Web2.0 tools (wiki, forum) have already been used successfully in geographic information systems (GIS) for collaborative editing of geo-referenced contents. On the other hand, no work was performed in collaborative testing of map semiology to date. To compensate for this lack, we propose WIKISemio, a web tool that integrates a WIKI tool with a GIS to allow the collaborative evaluation of the semiological variables of maps.

MOTS-CLES : Sémiologie, systèmes d'information géographique, Wiki, évaluation collaborative

KEYWORDS: Semiology, Geographic Information Systems, Wiki, collaborative evaluation

1. Introduction

Les outils informatiques VGI (Volunteer Geographic Information) permettent la collecte de grandes quantités de données géo-référencées (comme par exemple Openstreet Map, Ushuadi, etc.). Ces outils permettent aux citoyens de contribuer à la définition et à l'ajout d'informations dans des bases de données dans différents domaines : agriculture, environnement, etc. Typiquement la démarche volontaire consiste à considérer les citoyens comme des capteurs de données. Ces données sont ensuite gérées par des organismes ou entreprises qui les retournent vers les citoyens sous un format différent après les avoir analysées et synthétisées. Cette approche « top-down » présente une limitation importante : les indicateurs cartographiés peuvent ne pas correspondre aux attentes des citoyens. S'ils se sentent exclus du processus de prise de décision, ils peuvent se désintéresser du projet voire le critiquer. Cet article présente un outil WIKISemio qui place les citoyens au centre du processus décisionnel pour la définition des variables sémiologiques des cartes afin de favoriser leur participation. WIKISemio est un outil qui permet aux utilisateurs d'échanger entre eux à l'aide d'un WIKI sur les variables sémiologiques de cartes prédéfinies. Il a été développé dans le cadre du projet TELETIQ qui avait pour but d'estimer et comprendre le risque de transmission de maladies liées à l'environnement en utilisant des données de téléphonie mobile et de sciences participatives. Le projet était appliqué aux maladies transmises par les tiques.

2. Motivation

Les sciences participatives ont déjà montré avec succès leur utilité dans différents domaines d'applications (Kullenberg et al., 2016). Des volontaires participent à des travaux récents portant sur des activités de testing d'outils d'analyse visuelle (Borgo et al., 2018). Dans le domaine de l'information géographique, de nombreux travaux s'intéressent à l'édition collaborative de cartes. Pour permettre aux volontaires de discuter et de collaborer entre eux sur le processus de collecte de données et sur les données éditées, certains travaux proposent l'association des outils du Web 2.0 (comme par exemple les forums, le wiki, etc.) aux SIG (Sui, 2008 ; Ciobanu et al., 2007). En effet, les technologies du Web 2.0 ont démontré avec succès leurs capacités à faire collaborer les utilisateurs et à faire émerger en conséquence de nouvelles connaissances. Par contre, dans l'état actuel des connaissances, aucun travail ne s'est intéressé à mettre en place une méthodologie et un outil permettant aux volontaires de collaborer et d'échanger autour des questions relatives à la sémiologie de cartes. En d'autres termes, le testing des outils de geovisualisation dans le domaine des sciences participatives n'a pas encore été abordé.

La solution décrite dans cet article est appliquée à une étude concernant le risque de transmission de maladies par les piqûres de tique (projet TELETIQ). Le risque de piqûre dépend de deux facteurs : la densité de tiques infectées dans l'environnement et de la présence humaine qui par son activité peut s'exposer aux tiques. En effet,

l'activité des tiques dépend des conditions météorologiques et environnementales, et est sensible aux facteurs tels que le climat, l'altitude, l'occupation du sol (forêt, prairie, roche, zones urbaines,...) et la densité de faune sauvage (cerfs, chevreuil, sangliers,...). Ces informations sont combinées et permettent de créer une carte qui estime les conditions favorables à la présence de la tique (Lebert et al., 2017). La carte produite a pour objectif de distinguer les lieux très favorables ou peu favorables à la présence de la tique, afin d'adapter les messages de prévention auprès du grand public. Cette carte doit être lisible et compréhensible par le plus grand nombre de citoyens en fonction de sa perception du risque (Dernat et Johany, 2019). Pour cela, les chercheurs et les cartographes ont proposé plusieurs cartes en utilisant différentes sémiologies pour les soumettre aux volontaires via l'outil WIKISemio. L'outil permettra d'obtenir les commentaires des participants sur la ou les cartes et leurs appréciations quant à la facilité d'utilisation et de compréhension de l'outil.

3. WIKISemio

WIKISemio est un système web cartographique qui intègre un système Wiki. Son architecture est décrite dans la FIGURE 1. La composante SIG est implémentée avec OpenLayers (pour l'affichage), Geoserver (pour le stockage), et le Wiki avec DocuWIKI. Chaque carte et ses variables sémiologiques (couleur, grain, position, taille et forme) sont associées à une page du Wiki via un modèle de données (FIGURE 2) implémenté avec Postgres.

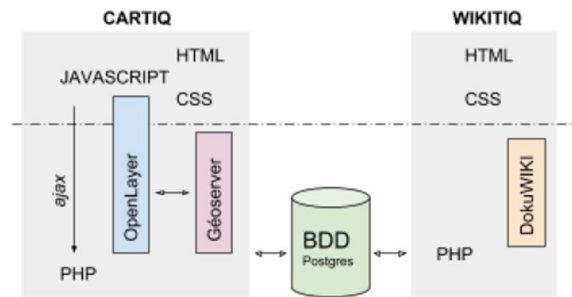


FIGURE 1. Architecture de l'outil WIKISemio

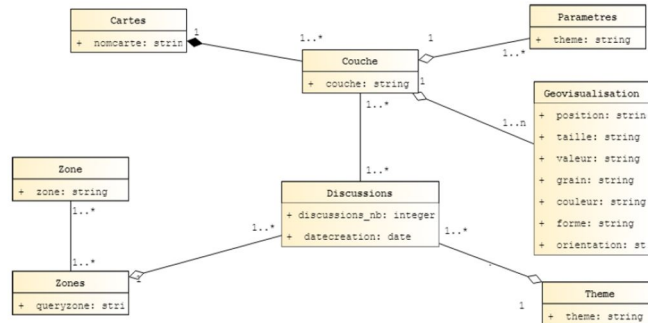


FIGURE 2. Modèle de données mapping SIG-WIKI

La table « Discussions » (FIGURE 2) est la table principale qui contient toutes les discussions. Chaque discussion est identifiée univoquement par la couche géographique (par exemple l'altitude), la zone géographique (par exemple le département de l'Allier) et une géovisualisation (ensemble des variables sémiologiques). Elle pointe donc sur trois tables « Couches », « Zones », et « Theme ». La table « Couche » permet d'avoir des informations sur la carte comme le nom de la couche mais également la carte à laquelle elle est associée (par exemple la carte 'Cartel'), le paramètre de la couche et ses variables sémiologiques.

Un aperçu de l'interface de WIKISemio est montré en FIGURE 3. Dans la partie gauche, les cartes disponibles sont affichées (1) avec des légendes représentant les variables sémiologiques utilisées (2), un menu pour choisir le niveau spatial (3) et un menu pour choisir le thème de discussion (variable sémiologique,...) et le créer ou l'afficher (4). La partie droite correspond à la page wiki qui inclut un forum (5), et donne la possibilité d'évaluer la discussion (6). Il est également possible de personnaliser l'évaluation en ajoutant une liste de questions (7). WIKISemio évalue automatiquement toutes les réponses en fournissant une note globale (la moyenne) à la page et donc l'appréciation de la carte et des variables sémiologiques par les utilisateurs. Les contenus des discussions permettent de comprendre et expliquer la note globale.

WIKISemio, outil de testing collaboratif de cartes

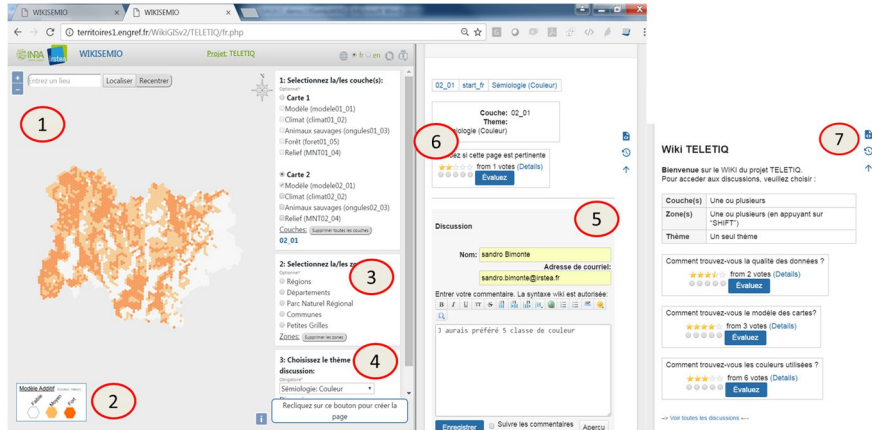


FIGURE 3. Interface de WIKISemio

Dans l'exemple de la FIGURE 3, un utilisateur évalue la « couleur » (variable sémiologique) de la couche représentant le modèle d'estimation des conditions favorables à la présence des tiques, pour la carte Carte2 sur toute la zone géographique modélisée. L'utilisateur peut ensuite attribuer la note « 2 » à cette représentation, note de faible appréciation dont la justification est rédigée dans la fenêtre de discussion : L'utilisateur préfère avoir une connaissance plus fine de la présence de tiques, et aurait préféré une carte présentant 5 classes de couleur.

La possibilité de personnaliser les questions pour l'évaluation de la page Wiki et de la variable visuelle est nécessaire dans un contexte citoyen car la formalisation des questions est fortement liée aux expertises en SIG des volontaires et aux types de cartes qui sont proposées pour évaluation.

WIKISemio comporte un outil d'authentification (FIGURE 4) au système qui permet de sécuriser l'accès à la page d'évaluation (droits d'utilisateur ou d'administrateur pour la visualisation des notes globales et la définition des questions) et donc de garantir la rétractabilité des évaluations.

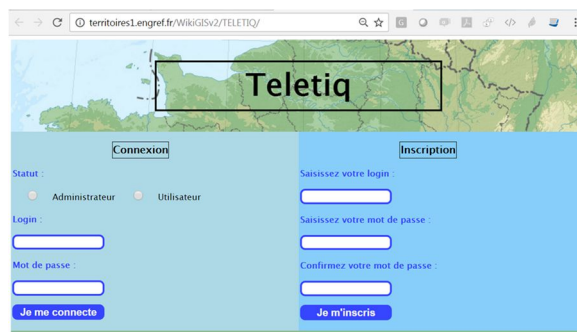


FIGURE 4. Interface d'accès à WIKISemio

4. Conclusion et travaux futurs

Dans cet article nous présentons WIKISemio, un outil pour le testing collaboratif des variables sémiologiques de cartes. WIKISemio associe un système WIKI à un outil web de mapping cartographique. L'efficacité de notre outil sera testé en contexte de sciences participatives dans les prochains mois au cours de « living labs » menés sur la prévention des tiques et maladies à tiques.

Remerciements

Ce travail été réalisé dans le cadre du projet TELETIQ (2017-2018, porteurs du projet G. Vourc'h, I. Lebert, S. Bord, <https://www6.ara.inra.fr/teletiq/>) de l'appel à projet du Programme EMERGENCE I-SITE Clermont. Cette recherche a été financée par l'Agence Nationale de la Recherche du gouvernement français dans le programme « Investissement d'avenir » 16-IDEX-0001 (CAP 20-25). Ce travail est aussi soutenu par le projet ANR VGI4Bio ANR-17-CE04-0012 (vgi4bio.fr).

Bibliographie

- Borgo, R., Micallef, L., Bach, B., McGee, F., Bongshin, L. (2018) Information Visualization Evaluation Using Crowdsourcing. *Comput. Graph. Forum*, 37(3): 573-595.
- Ciobanu, L. D., S. Roche, T. Badar & C. Caron (2007) Du Wiki au WikiSIG. *Geomatica*, 61, 137-145
- Dernat, S., et Johany, F. (2019). Tick Bite Risk as a Socio-Spatial Representation—An Exploratory Study in Massif Central, France, *Land*, 8(3), 46.
- Kullenberg C., Kasperowski D. (2016). What is citizen science? A scientometric meta-analysis, *PLoS ONE*, 11, 1.
- Lebert, I., Cassar, E., Cat, J., Saint-Andrieux, C., Vourc'h, G., Chalvet-Monfray, K., Rene-Martellet, M (2017). *Développement d'un outil de prédiction du risque d'exposition à la tique Ixodes ricinus et aux agents de la Borréliose de Lyme en France*. 14ème rencontre des Microbiologistes du Pôle Clermontois. Clermont-Ferrand, France. 11 avril 2017.

WIKISemio, outil de testing collaboratif de cartes

Sui, D. (2008) The Wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems*, 32, 1-5.

GrAPHiST: un environnement de visualisation pour l'analyse exploratoire de données spatio-temporelles

Jacques Gautier¹, Paule-Annick Davoine², Claire Cuntty³

1. Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France
jacques.gautier@univ-grenoble-alpes.fr

2. Univ. Grenoble Alpes, CNRS, Science Po Grenoble, PACTE, 38000 Grenoble, France
paule-annick.davoine@univ-grenoble-alpes.fr

3. Laboratoire EVS, Université Lyon II, Lyon, France
claire.cuntty@univ-lyon2.fr

RÉSUMÉ. L'analyse exploratoire de données décrivant des phénomènes spatio-temporels est effectuée au travers d'environnements de visualisation, majoritairement créés pour une thématique spécifique. L'augmentation du nombre de données spatio-temporelles disponibles incite au développement d'environnements "génériques", pouvant être utilisés dans différentes thématiques. À travers notre prototype nommé GrAPHiST, nous proposons une approche d'analyse exploratoire de données spatio-temporelles, permettant d'identifier différents types de dynamiques décrivant la structure de phénomènes, relatifs à différentes problématiques.

ABSTRACT. Exploratory analysis of data describing spatio-temporal phenomena is done through visualization environment, mainly created for a specific topic. The increase of available spatio-temporal data incites the development of "generic" environment, able to be used through different topic. With our prototype, called GrAPHiST, we propose an exploratory spatio-temporal data analysis approach, allowing the identification of different types of dynamics describing the structure of phenomena, related to different topics.

MOTS-CLÉS : Analyse exploratoire, Phénomène spatio-temporel, Environnement de visualisation, Dynamique spatio-temporelles

KEYWORDS: Exploratory analysis, Spatio-temporal phenomena, Visualization environment, Spatio-temporal dynamics

Introduction

Les environnements de visualisation, autorisant une analyse exploratoire de données spatio-temporelles, sont majoritairement créés de manière ad hoc dans le cadre d'une thématique particulière (Kaddouri *et al.*, 2014). Or, l'augmentation du nombre de phénomènes décrits par les nouvelles données spatio-temporelles à disposition incite au développement d'approches d'analyse exploratoire "génériques", pouvant être appliquées à différentes problématiques. Afin d'atteindre cet objectif, notre démarche est de passer par l'élaboration d'un environnement de géovisualisation opérationnel, GrAPHiST (pour (environnement de) Géovisualisation pour l'Analyse des PHénomènes Spatio-Temporels).

Événement et phénomènes, dynamiques et patterns

Nous nous focalisons sur quatre types de dynamiques permettant de décrire la structure des phénomènes spatio-temporels: l'existence de pics d'intensité d'un phénomène, localisés dans l'espace et dans le temps (Figures 1.A et 1.B), l'existence de tendances localisées dans la variation temporelle de l'intensité d'un phénomène (Figure 1.C), l'existence d'une évolution spatiale d'un phénomène au cours du temps (Figure 1.D), et l'existence de récurrences cycliques localisées dans certaines zones de l'espace (Figure 1.E).

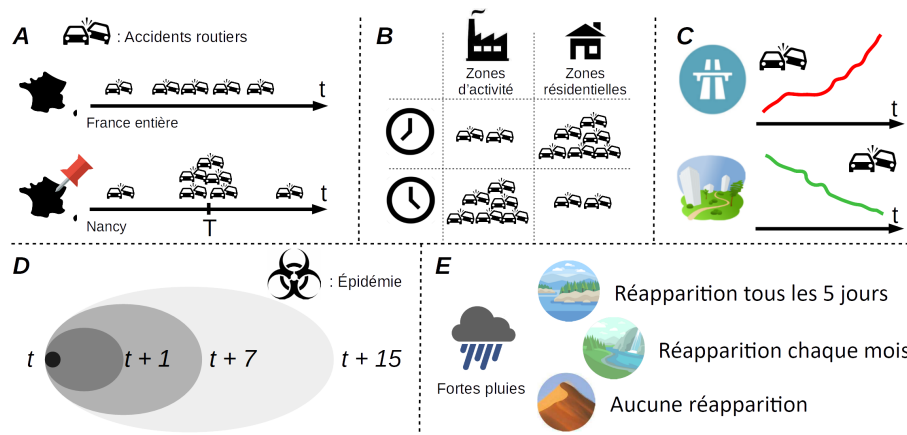


FIGURE 1. *Dynamiques recherche et structures des données observées*

Nous représentons le changement dans l'espace au travers de séries d'événements spatio-temporels, basés sur le Pyramid Model de Mennis (Mennis *et al.*, 2000), lui-même inspiré du Triad Model de Peuquet (Peuquet, 1994). L'événement est considéré comme un objet possédant trois composantes: spatiale, temporelle et thématique. Nous considérons le phénomène spatio-temporel comme un regroupement d'événements traitant de la même problématique, créé selon

les besoins de l'analyste. Les événements que nous représentons sont ponctuels¹, sur le plan spatial et temporel. L'identification des dynamiques structurant un phénomène passe alors par l'identification de patterns dans la distribution spatio-temporelle de séries d'événements.

GrAPHiST

GrAPHiST est un environnement de visualisation créée sous la forme d'une application Web (Figure 2). Sa structure suit celle d'un environnement à multi-fenêtrage. Les événements d'un même phénomène étant proches selon un critère thématique, l'espace graphique disponible est privilégié pour la représentation des dimensions temporelle et spatiale. Chaque fenêtre est occupée par des représentations graphiques permettant d'identifier visuellement des patterns dans les distributions temporelle et spatiale des séries d'événements.

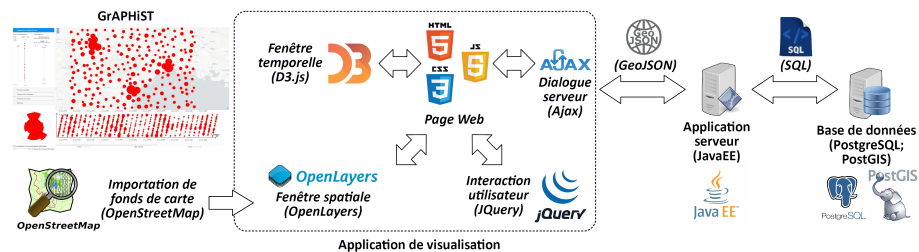


FIGURE 2. GrAPHiST

La fenêtre temporelle est occupée par deux diagrammes. Le premier est un *rose chart*² permettant d'identifier des concentrations temporelles d'événements, selon une échelle temporelle cyclique (Figure 3.A). Le second diagramme est inspiré des *Tiles Map* (Van Wijk, Van Selow, 1999). Le temps y est représenté le long de l'axe des ordonnées selon l'échelle cyclique utilisée dans le *rose chart*, et le long de l'axe des abscisses selon une échelle linéaire. Le diagramme prend la forme d'une courbe en dents de scie, chaque ligne oblique représentant une instance de l'échelle cyclique. Les événements temporellement proches sont agrégés et représentés par des entités graphiques dont la taille est fonction du nombre d'événements agrégés. Ce diagramme permet d'identifier des concentrations temporelles d'événements selon une échelle linéaire (Figure 3.B). En modifiant l'échelle cyclique T représentée de manière progressive, il permet également d'identifier des récurrences cycliques par

1. Les données auxquelles nous avons accès lors des premières implémentations étant majoritairement ponctuelles, nous avons focalisé nos premiers travaux sur ce type de donnée
 2. Lors de futurs travaux, nous souhaitons modifier le *rose chart* afin que l'effectif des classes d'événements soit représenté par l'aire d'une section et non par son rayon.

l'observation d'alignements graphiques, visibles lorsque la durée de l'échelle cyclique représentée approche la période du cycle correspondant (Figure 3.C). Des analyses harmoniques sont effectuées en arrière-plan, afin de suggérer des pistes d'exploration pour la recherche de cycles.

La fenêtre spatiale représente l'ensemble des événements apparaissant durant la période couverte par les diagrammes temporels. Les événements spatialement proches sont agrégés et représentés par des entités graphiques dont la taille est fonction du nombre d'événements agrégés. Cette représentation permet l'identification de concentrations spatiales d'événements (Figure 3.D).

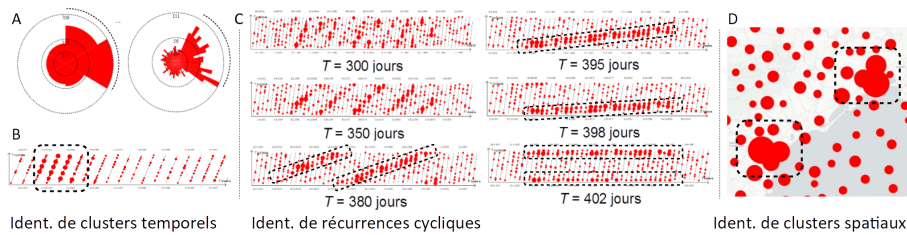


FIGURE 3. *Fonctionnement des représentations graphiques*

L'identification de patterns spatio-temporels s'effectue par la projection de l'une des composantes dimensionnelles des événements dans l'ensemble des représentations de l'interface, au moyen de règles de sémiologie. Ceci peut s'effectuer en projetant le lieu d'apparition, représenté dans la carte, dans les diagrammes temporels (entrée spatiale, Figure 4.A), en projetant le moment d'apparition linéaire, représenté dans le diagramme en dents de scie, dans la carte et le *rose chart* (entrée temporelle linéaire, Figure 4.B), ou en projetant le moment d'apparition cyclique, représenté dans le *rose chart*, dans la carte et le diagramme en dents de scie (entrée temporelle cyclique, Figure 4.C). Les événements sont classés selon différentes périodes, ou zones d'apparition, une couleur ou une valeur étant affectée à chaque classe. Dans la carte et le diagramme en dents de scie, les agrégats sont représentés sous la forme de *pie charts*, dont les sections indiquent la proportion d'événements d'une même classe.

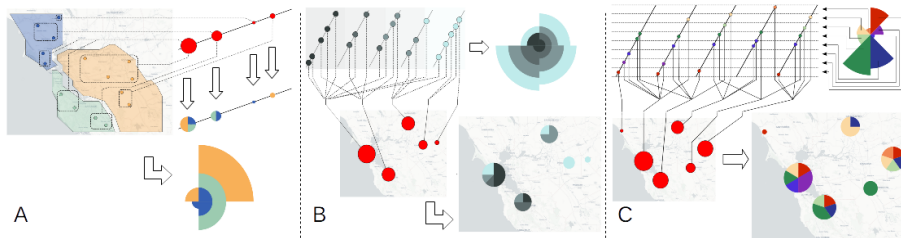


FIGURE 4. *Liens entre les représentations graphiques*

Les événements peuvent être représentés en fonction de leur moment d'apparition - linéaire si l'on souhaite observer, par exemple, les possibles propagations d'un phénomène dans l'espace - cyclique si l'on souhaite observer les comportements cycliques de ce phénomène. Dans la Figure 5 (Military History, 1989), des événements décrivant des batailles de la Guerre de Sécession sont représentés en utilisant une couleur sombre pour les événements les plus anciens, et une couleur claire pour les événements les plus récents. L'observation de gradients de couleur (Figure 5.A)) permet d'identifier un déplacement du lieu d'apparition des événements au cours du temps, correspondant ici à un déplacement des zones de combat au cours du conflit³ (Figure 5.B)).

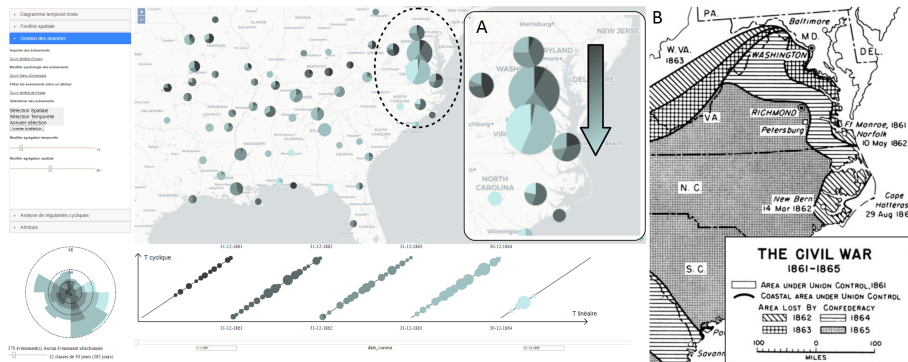


FIGURE 5. An. de batailles de la Guerre de Sécession (Military History, 1989)

L'analyse peut également s'effectuer en deux temps. Dans un premier temps, les événements peuvent être représentés selon leur dimension temporelle afin d'effectuer une analyse sans qu'une organisation spatiale des événements soit suspectée. Sur la Figure 6.A, une classification d'événements d'inondation selon leur mois d'apparition montre une concentration d'événements apparaissant en juin dans la région de Toulouse, tandis que les événements situés au nord de la Gironde apparaissent majoritairement en novembre.

À partir de ces résultats, l'analyste peut émettre une hypothèse sur un lien entre la distribution des événements et un découpage spécifique du territoire. Une seconde analyse peut être effectuée en représentant les événements selon leur dimension spatiale, au moyen de ce découpage, afin d'observer plus précisément la distribution temporelle des événements dans chaque zone. Sur la Figure 6.B, la classification des événements selon le bassin versant où ils appa-

3. Une analyse similaire peut être effectuée sur des données d'épidémiologie, afin d'identifier les axes de propagation d'une épidémie

raissent fait observer, sur le *rose chart*, une forte concentration d'inondations en juin dans bassin de Garonne, et en novembre dans le bassin de la Charente ⁴.

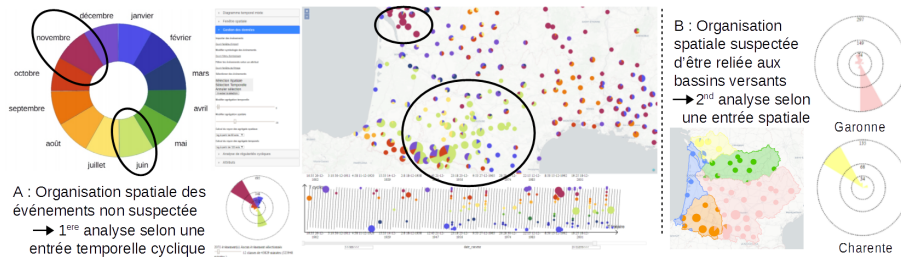


FIGURE 6. Analyse d'événements d'inondation

Conclusion

GrAPHiST ⁵ a fourni un support pour présenter nos propositions de visualisation pour l'analyse exploratoire de données spatio-temporelles (Gautier, 2018), et a été expérimenté sur des données relatives à différentes thématiques (données historiques, criminologiques, d'accidents routiers). Son approche générique nous incite à étendre ce prototype sous la forme d'une librairie JavaScript, afin de faciliter son application à différentes thématiques, voire sa réappropriation. En effet, en tant qu'environnement opérationnel, GrAPHiST représente un socle pour le développement de nouvelles méthodes de visualisation, devant permettre l'analyse d'autres dynamiques spatio-temporelles.

Remerciements

Cette recherche a été soutenue par la Région Auvergne-Rhône-Alpes.

Bibliographie

- Gautier J. (2018). *GrAPHiST: An exploratory analysis approach for the identification of dynamics of spatio-temporal phenomena*. Theses, Université Grenoble Alpes (France). Consulté sur <https://hal.archives-ouvertes.fr/tel-02067484>
- Kaddouri L., Blaise J.-Y., Davoine P.-A., Mathian H., Saint-Marc C. (2014, juin). *État des lieux des représentations dynamiques des temporalités des territoires*. Research Report. UMR 7300 ESPACE ; UMR 3495 MAP - GAMSAU ; Laboratoire d'Informatique de Grenoble. Consulté sur <https://hal.inria.fr/hal-01383450>

4. L'objectif ici n'est pas d'apporter des conclusions sur les phénomènes d'inondation.

5. GrAPHiST est disponible en ligne https://steamer.imag.fr/?page_id=735.

- Mennis J. L., Peuquet D. J., Qian L. (2000). A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science*, vol. 14, n° 6, p. 501–520.
- Military History U. S. A. Office of the Chief of. (1989). *The civil war 1861-1865*. (https://commons.wikimedia.org/wiki/File:Civil_war_1861-1865.jpg)
- 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*, vol. 84, n° 3, p. 441–461.
- Van Wijk J. J., 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*, p. 4–9.

Arabesques, application d'exploration et de géovisualisation de données de flux et de réseaux

Côme E.¹, Bapaume T.¹, Jégou L.², Bahoken F.^{3,4}, Maisonobe M.⁴, Roelandt N.³, Le Campion G.⁵

1. *COSYS-GRETTIA, IFSTTAR / Université Paris-Est
Cité Descartes, 14-20 Boulevard Newton, F-77420 Champs-sur-Marne, France.
etienne.come@ifsttar.fr*
2. *UMR 5193 LISST / Université de Toulouse 2 / CNRS
Maison de la Recherche, 5, allées Antonio-Machado, F-31058 Toulouse Cedex 9,
jegou@univ-tlse2.fr*
3. *AME-SPLOTT, IFSTTAR / Université Paris-Est
Cité Descartes, 14-20 Boulevard Newton, F-77420 Champs-sur-Marne, France
francoise.bahoken@ifsttar.fr*
4. *UMR 8504 Géographie-Cités / Équipe PARIS / CNRS
13, rue du Four, F-75006 Paris, France
marion.MAISONOBE@cnrs.fr*
5. *UMR 5319 Passages, Équipe ARD / CNRS
Maison des Suds, 12 Espl. des Antilles, F-33600 Pessac, France
Gregoire.LECAMPION@cnrs.fr*

RÉSUMÉ. Le projet gflowiz (geographic flow visualisation) est une collaboration interdisciplinaire réalisée dans l'objectif de concevoir et développer Arabesque, une nouvelle application web dédiée à la géo visualisation de données de flux et de réseaux spatiaux. Cette proposition présente différentes possibilités d'actions sur les informations géographiques, statistiques et cartographiques qui concourent à la réalisation de cartographies correspondantes.

ABSTRACT. The gflowiz (geographic flow visualisation) project is an interdisciplinary collaboration carried out with the objective of designing and developing Arabesques, a new web application dedicated to the geo visualization of flow datasets and spatial networks. This proposal presents different possibilities for action on geographical, statistical and cartographic information that contribute to dedicated maps.

MOTS-CLÉS : matrice OD, hyper matrice, graphe multiplexe, geoweb, cartographie.

KEYWORDS: origine-destination matrix, hyper matrix, multiplex graph, geoweb, flow mapping.

1. Introduction

Le projet gflowiz (*geographic flow visualisation*) est une collaboration interdisciplinaire financée par l'Institut français des sciences et technologies des transports, de l'aménagement et des réseaux (IFSTTAR). Réalisée en collaboration avec le Centre national de la recherche scientifique (CNRS), elle a notamment pour objectif de concevoir et développer une nouvelle application web dédiée à la géo visualisation de données de flux et de réseaux spatiaux, *Arabesque*.

Cette application, inscrite dans le cadre de la Loi Lemaire pour une république numérique, relève de l'objectif général visant à accroître la compréhension des déterminants géographiques de la mobilité spatiale de biens et de personnes. Elle souhaite répondre à l'impératif de visualisation des résultats de recherches fondamentales ou appliquées relevant de cadres de développements théoriques et méthodologiques pouvant être considérés comme transversaux, à différents thèmes (population, habitat, environnement, transport, ...), interdisciplinaires par essence (géographie, démographie, environnement, géomatique, sciences pour l'ingénieur, sciences humaines et sociales, ...) en raison des défis sociétaux liés à l'essor de la société numérique qu'ils permettent de relever.

Sur le plan scientifique, cette application ambitionne d'innover dans la manipulation de données numériques décrivant des flux et réseaux localisés. C'est pourquoi elle s'inscrit, d'une part, dans le nouveau paradigme de la « cartographie de visualisation » (Mac Eachren, 2005) consistant à combiner dans un même environnement des dispositifs de géo visualisation et de traitement des informations géographiques et statistiques – à visée descriptive dans un premier temps. En autorisant, d'autre part, la prise en compte de jeux de données relationnelles complexes, pouvant être à la fois volumineux et présenter différentes dimensions des mobilités spatiales (plusieurs catégories thématiques et/ou temporalités). Une attention particulière est en outre apportée au rendu en lui-même, que ce soit pour le dessin ou pour la sémiologie cartographique des figurés. La volonté d'une amélioration qualitative des images produites devrait conduire à développer une approche sensée de la géo visualisation des mobilités et des interactions spatiales.

L'objet de cette démonstration est de présenter le fonctionnement de la toute première version de *Arabesques*, en appui de l'atelier *Le Geoweb des cartes de flux et réseaux* réalisé en marge de la conférence SAGEO'2019 pour le compte de l'Action Prospective *Théoriser, Observer, Analyser, Modéliser le Geoweb* (AP GEOWEB) du GdR MAGIS.

Le développement présenté fait suite à une première phase qui a donné lieu à un état de la question des possibilités actuellement offertes de manipulation de données relationnelles à des fins d'analyse cartographique thématique de flux sur le web (Bahoken et al., 2019). L'application relève ainsi – à ce stade de développement – de la formation du premier des trois *clusters* identifiés par une typologie de l'existant. Il est composé d'applications dont la représentation principale est la *carte de flux*, réalisée pour symboliser principalement des *déplacements immatériels*, c'est-à-dire origine-destination selon une logique de raisonnement cartographique classique,

mobilisant essentiellement la variable rétinienne *Taille* appliquée à des figurés linéaires.

Cette famille d'applications se caractérise également par la présence au sein du même environnement, de *plusieurs représentations secondaires* présentées selon *plusieurs modalités graphiques*. La réalisation d'une enquête en ligne a ensuite permis d'identifier les besoins spécifiques d'une communauté pour ce type d'application, ouverte dans son code source et libre d'utilisation. Elle a enfin abouti à la définition du problème à résoudre : proposer un ensemble d'outils d'analyse et de représentation de matrices d'interactions géographiques ou territoriales complexes, c'est-à-dire denses (concernant de nombreux acteurs) et riches, car catégorielles et temporelles.

Le développement d'*Arabesque* est entièrement réalisé en *Javascript*. Il diffère en ce sens fondamentalement des solutions disponibles dans d'autres environnements, d'une part, dans R sous la forme de packages ; d'autre part, dans les outils de type Système d'Information géographique (SIG) ou dans les applications *standalone* installées localement. Le développement a tiré parti des bibliothèques *Proj4Js*, *D3* et *Openlayers* pour l'essentiel, avec un rendu en *WebGL* permettant l'affichage d'un nombre conséquent d'objets graphiques, de l'ordre de plusieurs milliers.

Cette architecture a permis d'aboutir à un outil web, disponible sans installation, facile à prendre en main et permettant l'importation de ses propres données, *a minima*, un fichier .csv contenant des données origine-destination localisées par défaut à l'aide de codes standards (ISO3, ISO2, COG, etc.) ou de géolocalisations fournies par l'utilisateur.

Arabesque permet l'exploration, le filtrage, l'agrégation, l'affichage et la représentation cartographique de données de flux ou de réseaux, à différentes échelles, dans la perspective d'une analyse géo localisée thématique. Son interface est actuellement composée de trois parties (Fig. 1) : à gauche, une zone de gestion des couches géographiques à afficher ; au centre, un espace d'exploration cartographique ; à droite, une zone de sélection et de filtrage des données de flux.

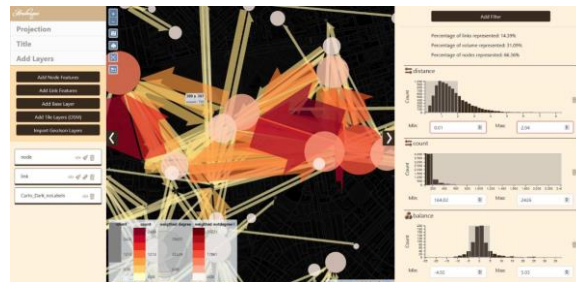


FIGURE 1. Interface de l'application *Arabesque*

La démonstration¹ proposée va ainsi focaliser l'attention sur les principales possibilités d'action proposées sur les composantes géographiques, statistiques et cartographiques offertes par cette toute nouvelle application.

2. L'information géographique

La première composante relève de l'information géographique qui sert de support (de fond de carte) ou d'habillage de la cartographie.

2.1. Jeux d'échelles et résolution

L'éventail des possibilités de représentations graphiques offertes aux usagers de cette nouvelle application est indépendant de l'échelle géographique des données à analyser. En effet, les possibilités de dessin qui sont proposées sont les mêmes quel que soit le niveau géographique.

Afin d'éviter d'éventuelles divergences entre les tailles des plus petits et des plus gros figurés, nous avons partiellement suivi une suggestion générale de Waldo Tobler d'ajuster le dessin des figurés linéaires à la résolution spatiale. Notre solution a pour objectif de garantir une perception locale du plus petit élément et globale de l'ensemble des motifs à peu près équilibrée, y compris en cas de zoom / dé zoom.

Cet ajustement des tracés est rendu possible grâce à un ratio défini empiriquement, fixé par défaut à 0,02% de la plus grande dimension du rectangle d'encombrement maximal du jeu de données de flux - de l'information statistique et non géographique (voir *infra*).

2.2. Jeux de données géographiques

L'application autorise l'import de ses propres jeux de données, aux formats .JSON et geoJSON, leur reprojection dans différents systèmes et leur personnalisation pour la représentation (voir §4.).

Des fonds géonumériques sont proposés à titre complémentaire, à des fins d'habillage ; il s'agit à ce stade de couches issues de *OpenStreetMap* et de *NaturalEarthData*.

En cas d'absence d'informations géographiques associées à un jeu de données statistique, il est possible de géocoder un fichier .CSV sous réserve que ce dernier comporte *a minima* un identifiant d'unités spatiales valide. À défaut de cela, les nœuds correspondants seront supprimés lors d'une vérification de l'intégrité du fichier.

Plusieurs possibilités de géocodage sont proposées pour différents niveaux territoriaux généraux, tels que les codes ISO et UNEP pour les pays, régions ou sous-régions du monde ; les identifiants du découpage européen NUTS 1,2,3 ; les codes

¹ La version de développement est accessible à l'adresse : <http://vlsstats.ifsttar.fr/gflowiz/>

postaux ou COG pour les différents découpages français ; les identifiants de territoires spéciaux, comme le Grand Paris, le *greater London* notamment.

3. L'information statistique

La seconde composante consiste à interagir avec une information statistique selon différentes modalités : exploration, calcul d'indicateurs, filtres interactifs numériques et/ou graphiques en soutien à l'exploration visuelle.

3.1. Jeux de données statistiques

Ces possibilités sont offertes pour toutes les dimensions d'une ou de plusieurs matrices empilées formant un même jeu de données, que l'information soit valuée ou non, disponible pour son cœur (des liens) et/ou en ses marges (des nœuds).

Les matrices peuvent en effet être simples ou complexes :

– Matrice simple (ordinaire) : il n'existe qu'une seule relation, éventuellement bilatérale ou orientée, pour un même couple de lieux ;

– Matrice complexe, est une hyper matrice (ou matrice empilée) équivalente à un graphe multiplexe qui comporte plusieurs liens bilatéraux ou non, pour un même couple de nœuds.

Quel que soit leur type, les matrices doivent être fournies dans un format long présentant *a minima* trois colonnes (identifiants des lieux d'origine, de destination et valeur (mesure ou modalité) du lien).

3.2. Calcul d'indicateurs basiques

Outre les paramètres classiques résumant les distributions statistiques (min, max, ...), le calcul de quelques indicateurs de base est réalisé par défaut lors du chargement des données ; il s'applique aux liens et aux nœuds. Fondés sur des approches mixtes, de réseaux et gravitaires, ces calculs proposent par exemple au nombre de degrés (valué ou non), à des mesures de centralité, au calcul d'un volume ou d'une balance (soldes) des échanges, à la mesure de leurs espacements. En présence d'hyper matrices ou de graphe multiplexe, des opérations supplémentaires sont réalisées (voir § 3.3).

3.3. Possibilités de filtrage

3.3.1. Filtrage numérique

Le filtrage numérique s'applique aux données quantitatives de stock (absolues) ou de taux (continues ou pseudo-continue). Il est réalisé soit visuellement à l'aide d'une fenêtre de sélection sur un histogramme interactif à fenêtre coulissante (*slider*), soit de manière numérique (en indiquant un seuil). La part d'information sélectionnée, les valeurs minimales et maximales sont indiquées automatiquement sur la figure. Pour

les liens, la possibilité d'un filtrage selon la distance géodésique parcourue est en outre proposée.

3.3.2. Filtrage temporel

Les possibilités de filtrage des matrices temporelles ne sont actuellement disponibles que pour les liens. Elles portent sur le choix d'une date ou d'une période, sous une forme numérique ou graphique (*slider*). Trois formats de date sont disponibles (chaîne, horaire, horodatage).

3.3.3. Filtrage catégoriel

Le filtrage catégoriel s'applique aux données qualitatives nominales disponibles dans les jeux de données de nœuds et/ou de liens. Trois possibilités sont proposées graphiquement et numériquement :

- Sélection de variables multiples : l'utilisateur choisit une ou plusieurs des variables qualitatives disponibles (par exemple, les préfectures et sous-préfectures s'il dispose d'une variable spécifiant le profil administratif des villes françaises) ou le type de liens (par exemple, import ou export pour des données d'échanges commerciaux) ;

- Sélection d'une variable unique : l'utilisateur choisit une seule variable (pour les nœuds : par exemple, les flux au départ et à destination de Clermont-Ferrand pour des données de mobilité ; les flux de type import pour des données commerciales) ;

- Sélection par suppression d'une variable unique : l'utilisateur choisit une ou plusieurs variables portant sur les nœuds et les liens à supprimer. Si la suppression sélection porte sur des nœuds, tous les liens correspondants (par exemple, la région Ile-de-France pour une analyse à l'échelle nationale) sont supprimés.

3.4. Possibilités d'agrégation

Des fonctions d'agrégation attributaires, relatives à la gestion des hyper matrices ou graphes multiplexes, sont proposées pour réduire le nombre de figurés à représenter - ne pas représenter, par exemple, les n liens qui correspondent aux n dates disponibles.

Par défaut, c'est la fonction somme qui est appliquée au chargement des données, en l'absence de précisions. L'utilisateur peut toutefois choisir d'appliquer une fonction moyenne, minimum, maximum ou médiane calculée sur l'ensemble des matrices ou des graphes fournis. Il est également possible de choisir une seule date ou bien d'agréger les données, selon une fonction donnée, sur une période ou pour des catégories.

Cette fonction est importante, car elle définit par défaut l'intensité des teintes et l'opacité des figurés correspondants (voir *infra*).

4. L'information cartographique

La troisième composante de *Arabesques* est dédiée au rendu cartographique, par une action sur la symbolisation (ponctuelle ou linéaire) des données ou sur les fonds servant de support ou d'habillage.

Le dessin des différentes couches et de leurs objets est finement paramétrable, afin de tenir compte de l'éventuelle complexité de l'information, en termes de densité, nécessitant une gestion particulière des superpositions et des dispositions des liens et nœuds (voir Fig. 2). Les liens les plus gros sont placés au premier plan par défaut à l'inverse des cercles les plus larges

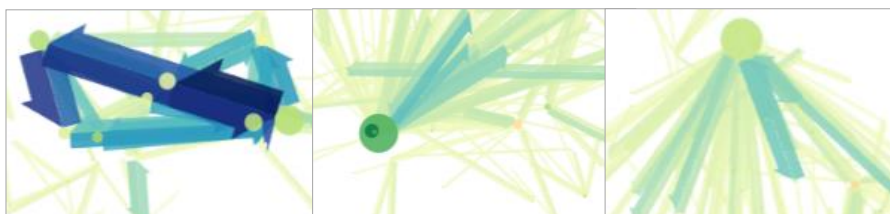


FIGURE 2. *Dispositions et gestion des superpositions*

La représentation cartographique relève notamment de la richesse (type de caractères, catégories, temporalités) des informations symbolisées par les différents objets géographiques représentés : des liens, des nœuds. Il est possible de contextualiser leur représentation par des informations relatives à leur structure *via* la réalisation de cartes choroplèthes (par exemple de statistiques socio-économiques, environnementales ou autres, si les données correspondantes sont fournies), ou bien l'affichage d'informations complémentaires d'habillage disponibles par défaut. Les différents éléments de paramétrage sont ensuite repris dans une légende placée sur la vue principale.

Les éléments classiques de la conception cartographique thématique (échelle graphique, légende), de même que la possibilité d'ajouts graphiques complémentaires (titre, etc.) sont disponibles par défaut.

4.1. *Symbologie des nœuds*

Le dessin des nœuds est enrichi grâce à une fenêtre interactive permettant de paramétrer leur apparence : leur *Taille*, leur *Forme* (cercle uniquement), leur *Teinte* et leur *opacité* ; ainsi que les labels correspondants (voir Fig. 3).

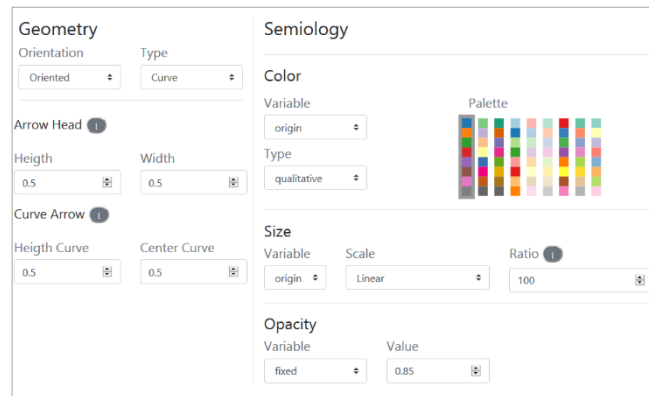


FIGURE 3. Fenêtre de paramétrage des nœuds

4.1. Symbologie des liens

Le dessin des liens est plus riche que celui des sommets. Notre application propose donc une possibilité d'action sur leur géométrie qui enrichit leur sémiologie, par l'ajout de la *Forme* et l'*Orientation*, qui définit leur style (voir Fig. 4).

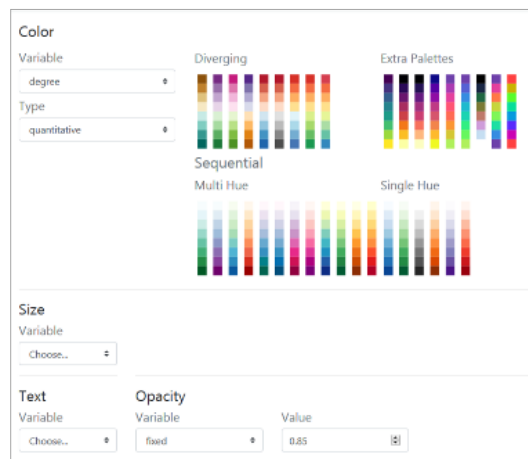


FIGURE 4. Fenêtre de paramétrage des liens

Il est en effet possible d'agir sur toutes les dimensions géométriques du lien :

- la *Taille* permet le paramétrage de l'épaisseur du lien en fonction d'une valeur ;
- la *Valeur*, associée à la *Couleur*, permet le paramétrage de la teinte du lien, éventuellement de son intensité en fonction de variables respectivement catégorielles ou numériques (de taux) ;
- l'*Opacité* des liens, qui est fondée sur l'intensité de la teinte calibrée pour les matrices complexes est fixée par défaut à 0,85, mais peut être modifiée ;
- l'*Orientation*, lorsqu'elle est présente, permet le paramétrage des dimensions de la tête de la demi-flèche (voir Fig. 2) ;
- la *Forme* qui autorise le dessin de flèches (voir Fig. 5) prenant l'apparence d'une
 - d'une droite ;
 - d'une courbe, avec paramétrage de sa hauteur (point maximal atteint) et de sa pente à l'aide d'un point d'inflexion ;
 - d'un triangle droit ;
 - d'un triangle courbe, sous la forme de goutte d'eau.



FIGURE 5. Exemple de variation de formes de flèches

- le ratio, déjà évoqué, dont la valeur par défaut peut être modifiée afin d'optimiser le dessin en fonction de la résolution spatiale.

Conclusion

Le projet gflowiz (*geographic flow visualisation*) est un projet innovant visant à créer *Arabesques*, une application web dédiée à la géo visualisation de données de flux et de réseaux spatiaux. Il tient compte des informations collectées grâce à un inventaire

des applications existantes dans le domaine et d'une enquête réalisée auprès de professionnels pour collecter leurs besoins en termes d'application web.

Une attention particulière a été apportée aux composantes géographiques, statistiques et cartographiques de l'application, afin d'obtenir un ensemble homogène et cohérent avec la littérature actuelle.

Arabesques offre de nombreuses possibilités tant dans les données en entrées qu'elle permet de charger et manipuler (matrices simples ou complexes), que dans les possibilités de filtrage, d'agrégation, de représentation et de symbologie.

Remerciements :

L'équipe de gflowiz souhaite remercier les porteur.e.s de l'Action Prospective Théoriser, Observer, Analyser, Modéliser le Geoweb du Groupement de recherche 2340 du CNRS : Méthodes et Applications pour la Géomatique et l'Information spatiale (GDR MAGIS), pour leur accueil chaleureux, le support et le soutien pour l'organisation de la session de la conférence SAGEO'2019, dédiée à la première présentation de cette application.

Référence

Bahoken F., Le Champion G., Jegou L., Maisonobe M., Côme E. (2019), Panorama d'un geoweb des flux et réseaux, in : *Actes du colloque Spatial Analysis and GEomatics (SAGEO 2019)*, Clermont-Ferrand, 13-15 novembre 2019, 14 p.

Mac Eachren A., (2005), *How Maps Work, Representation, Visualization, and Design*, New-York, The Guildford Press,

Polluscope - Vers un observatoire participatif de l'exposition individuelle à la pollution de l'air et de ses effets sanitaires

Mariam Brahem¹, Mohamed Chachoua²,
Hafsa El Hafyani¹, Zoubida Kedad¹, Ahmad Ktaish^{1,4},
Souheir Mehanna¹, Cyril Ray³, Yehia Taher¹,
Rémy Thibaud³, Laurent Yeh¹, Karine Zeitouni¹

1. Laboratoire DAVID, Université de Versailles Saint-Quentin-en-Yvelines -
Université Paris-Saclay

prenom.nom@uvsq.fr

2. LASTIG, Ecole des Ingénieurs de la Ville de Paris - University Paris Est

chachoua@eivp-paris.fr

3. Institut de Recherche de l'Ecole Navale

prenom.nom@ecole-navale.fr

4. LSCE, Laboratoire des Sciences du Climat et de l'Environnement

CNRS-CEA-UVSQ,IPSL

RÉSUMÉ. La qualité de l'air est un déterminant important de la santé des personnes. Sa dégradation continue, notamment en environnement urbain et industrialisé devient une préoccupation centrale pour nos sociétés. Les capteurs/stations fixes existants permettent une évaluation et une prédiction à moyenne-large échelle de cette pollution mais qui sont cependant affectées par des distributions hétérogènes de ces capteurs sur le territoire. Les micro-capteurs connectés émergents offrent l'opportunité de mesurer l'exposition réelle des individus à la pollution atmosphérique en tout lieu tout au long de leurs activités journalières. L'objectif du projet ANR Polluscope¹ est d'évaluer l'exposition individuelle à la pollution de l'air et de ses effets sur la santé par l'exploitation combinée de ces capteurs fixes et mobiles. Le projet aborde ainsi des verrous multiples, notamment en termes de métrologie, de protocole de collecte, d'intégration

1. Site web du projet Polluscope : <http://polluscope.uvsq.fr>

aux données de mobilité, de traitement et d'analyse de données imparfaites, de confidentialité des données personnelles, etc. Cet article présente brièvement les objectifs scientifiques, la méthodologie de mise en œuvre et les principes de conception du système d'information conçu pour le traitement des données de mobilités et de pollution.

ABSTRACT. Air quality is a key factor of human health. Its continuous degradation, especially in urban and industrialized environments, is becoming a major concern for our societies. Existing fixed sensors/stations allow a medium- to large-scale assessment and prediction of this pollution, but are limited by heterogeneous and sparse distributions of these sensors across the urban spaces. Emerging connected and embedded micro-sensors provide an opportunity to measure the actual exposure of individuals to air pollution anywhere, anytime, anywhere, during their daily activities. The objective of the ANR project Polluscope is to assess personal exposure to air pollution and its health effects through the combined use of these fixed and mobile sensors. The project thus addresses multiple research issues, particularly in terms of metrology, data collection protocol, integration with mobility data, processing and analysis of imperfect data and privacy, etc. This paper aims to presents the scientific objectives, methodology and design principles of the information system proposed for the processing of mobility and pollution data.

MOTS-CLÉS : Information géographique volontaire, micro-capteurs environnementaux, analyse de données spatio-temporelles, plateforme cloud, villes intelligentes et durables

KEYWORDS: Volunteered Geographic Information, environmental micro-sensors, spatio-temporal data analysis, cloud-based platform, smart and sustainable cities

1. Contexte et problématique

Que ce soit en milieu intérieur ou dans un environnement extérieur, la qualité de l'air que nous respirons est l'une des principales préoccupations de la population, notamment en zones urbaines denses en raison de l'importance des risques sanitaires de la pollution atmosphérique. Cependant, l'exposition individuelle réelle à cette pollution est encore mal connue. La connaissance de l'exposition individuelle est primordiale pour une meilleure compréhension des conséquences sanitaires des polluants, mais il n'existe quasiment pas de données de ce type. En effet, actuellement la mesure de la qualité de l'air, se fait via des stations fixes, qui ne fournissent que des valeurs moyennes par zone, et ne prennent pas en considération les milieux intérieurs (e.g. habitations, centres commerciaux). Il en résulte que cette méthode ne suffit pas pour quantifier l'exposition réelle à l'échelle individuelle, qui change selon les modalités de déplacement et selon les activités propres à chacun, le moment et le lieu, à l'extérieur comme à l'intérieur.

Afin de répondre à ce problème, des recherches se sont orientées vers le développement de micro-capteurs mobiles, parfois communicants. Ainsi, l'avènement du concept d'information participative d'une part, et l'émergence de cette nouvelle technologie de micro-capteurs environnementaux portables et connectés d'autre part, rendent envisageable la mesure de l'exposition individuelle, d'une manière continue, en tout lieu et à tout moment. Cette technologie a suscité un intérêt croissant de la communauté scientifique (sciences environnementales, santé, géosciences et informatique notamment), engendrant un nombre croissant de travaux. C'est dans cette optique que s'inscrit le projet ANR Polluscope. Il repose sur l'exploitation des nouveaux capteurs personnels afin d'observer l'exposition individuelle et l'impact sanitaire (Annesi-Maesano *et al.*, 2019). Plus exactement, Polluscope vise à concevoir et développer une plateforme de collecte, d'intégration, de gestion et d'analyse de données issues de micro-capteurs environnementaux individuels et mobiles. Cette plateforme contribue à améliorer la précision des mesures spatio-temporelles de la pollution atmosphérique et la compréhension de ses effets sanitaires. L'intégration avec de données contextuelles externes, essentiellement géo-spatiales, permet d'enrichir ces données et potentiellement de généraliser les connaissances extraites.

2. Campagne de mesure

Le déploiement, l'évaluation et l'usage de la plateforme est réalisé "sur le terrain" à l'aide de campagnes de mesures. Une campagne de mesure consiste à équiper des volontaires avec des capteurs portatifs de divers polluants (PM2.5, PM10, NO2, Noir de carbone), de température et d'humidité, de capteurs de santé (spiromètre et oxymètre) et d'une tablette connectée, incluant un GPS. Une cohorte d'environ cent soixante volontaires est recrutée et suivie sur deux

saisons, répartis sur deux années. La moitié servira pour l'étude épidémiologique incluant des patients asthmatiques ou BPCO (bronchopneumopathie chronique obstructive) et des témoins sains, l'autre moitié comprend des volontaires habitant l'agglomération de Versailles Grand Parc. Les données collectées sont donc exploitées à des fins de validation de deux scénarios : l'un est une étude épidémiologique, tandis que le second correspond à un usage de type sciences participatives.

3. Évaluation des capteurs

La première phase du projet a permis une étude minutieuse et une évaluation des capteurs, avant de pouvoir en choisir trois pour la campagne (Languille *et al.*, 2019). Le bilan de cette phase a révélé la grande diversité dans les formes, la qualité et les prix. La plupart des capteurs qui ont été évalués par rapport à un instrument de référence s'est révélée non fiable, au point de ne pas être exploitables. Certains très chers sont de très bonne qualité. D'autres plus abordables permettent de mesurer une tendance et donnent des mesures reproductibles. Le critère d'ergonomie est passé au second plan pour privilégier la fiabilité des mesures. Le protocole d'évaluation mis au point, les métriques et les outils ont été rassemblés dans un article actuellement en soumission. Vu la spécificité et la diversité des équipements, le protocole de la campagne a demandé plusieurs ajustements, comme l'ajout de tablettes pour pallier aux problèmes de connectivité et de positionnement GPS.

4. Traitement et analyse des données

La plateforme Polluscope permet d'agréger les données hétérogènes en entrée et d'adapter le mode d'extraction et de chargement selon les possibilités de chaque capteur, leur format spécifique et leur fréquence d'échantillonnage et de transmission. De plus, avec le temps, la collecte continue de données génère une volumétrie importante de données, amenant à une problématique *Big Data*.

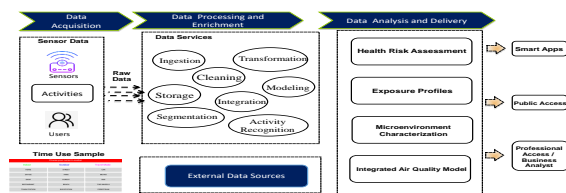


FIGURE 1. Architecture globale de la plateforme Polluscope

Dans ce contexte, le *Cloud Computing* constitue un choix important en tant que plateforme *Big Data* générique. Les technologies *Cloud* offrent l'avantage de créer des services d'infrastructure évolutifs pouvant être configurés à la demande et dynamiquement en fonction de la charge de travail requise et du volume de données en question. Nous avons opté pour *Docker*, une technologie *Cloud* permettant la virtualisation de tous les composants de la plateforme Polluscope, notamment les serveurs d'ingestion, traitement, analyse et visualisation de données, sous formes des images *Docker* composant à la fois des services de types *SaaS* et des plateformes de type *PaaS*, prêtes à déployer sur une infrastructure *Cloud* (Haque *et al.*, 2017). La figure 1 montre les principaux composants structurels de la plateforme Polluscope, regroupés autour des étapes de la chaîne de traitement décrites ci-dessous :

- **Ingestion de données:** Cela permettra la transformation, la normalisation, la distribution et chargement de données dans la plateforme de traitement et d'analyse de données. L'ingestion est effectuée via des APIs. Dans une future version, l'usage d'une plateforme distribuée pour la collecte de données telle que Kafka, est envisagé.

- **Traitement des données:** Plusieurs outils de traitement de données, tel que SPARK, R, etc., sont utilisés pour effectuer des tâches spécifiques liées au flux de traitement des données, telles que le nettoyage de données, leur enrichissement, etc. Les données collectées sont notamment enrichies par l'utilisation de sources ouvertes telles qu'OpenStreetMap et des données de pollution issues de la modélisation et fournies par Airparif.

- **Exportation de données:** Ceci permet la transformation, normalisation, et l'intégration des données de sortie vers une ou plusieurs plateformes d'analyse et de visualisation de données.

- **Analyse et visualisation:** Il s'agit d'explorer et d'analyser les données collectées dans toutes leurs dimensions (spatiale, temporelle, de mesures quantitatives ou sémantiques, etc.) et à différents niveaux de granularité. Grafana² est utilisé pour la visualisation interactive des niveaux de pollution rencontrés tout au long des activités et des trajectoires des participants. La figure 2 présente un exemple d'une trajectoire avec un lien dynamique entre l'emplacement du participant (le point bleu sur la carte) et la concentration de polluants selon la droite glissante rouge sur les courbes de mesures.

5. Conclusion

La qualité de l'air est un sujet d'actualité qui est au cœur préoccupations de la population en raison de l'importance des risques sanitaires de la pollution atmosphérique. Cette préoccupation est notamment accrue en zones urbaines

2. plateforme open source de visualisation et d'analyse des données: <https://grafana.com>

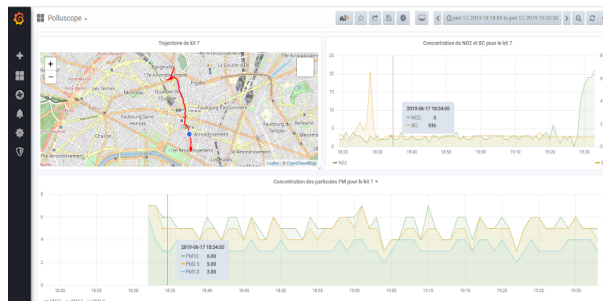


FIGURE 2. Analyse visuelle d'une trajectoire et des mesures associées

denses où les épisodes d'alerte sont récurrents. Cette démonstration présente la première version de la plateforme Polluscope, fruit d'une collaboration pluridisciplinaire impliquant les sciences environnementales, les géo-sciences, et l'informatique. Ce premier socle de l'architecture est amené à s'étoffer avec de nouvelles méthodes et fonctionnalités afin de transformer les données brutes en connaissances exploitables. Évalués dans un cadre expérimental réel impliquant entre 160 et 200 volontaires, ces travaux ambitionnent de contribuer à réduire les risques de certaines maladies et à conduire les personnes concernées à changer leurs comportements.

6. Remerciements

Ce travail a été réalisé dans le cadre du projet Polluscope, ANR-15-CE22-0018, soutenu par l'Agence Nationale de Recherche, dans le cadre de l'appel à projet générique CE22 – Plan d'action 2016, Défi 6 – Mobilité et systèmes urbains durables. Les auteurs tiennent à remercier l'ensemble des membres du projet pour leur collaboration.

Bibliographie

- Annesi-Maesano I. *et al.* (2019). Vers un observatoire participatif de l'exposition individuelle à la pollution de l'air et de ses effets sanitaires. *Colloque National Capteurs et Sciences Participatives (CASPA), Paris, 2019.*
- Haque R., Taher Y., Zeitouni K. (2017). Towards aircare: A highly scalable data management system for mobile sensor driven exposure analysis to air pollution. *Sensor Driven Exposure Analysis to Air Pollution. IEEE ComSoc Newsletter.*
- Languille B., Gros V., Bonnaire N., Pommier C., Honoré C., Debert C. *et al.* (2019). A methodology for the characterization of portable sensors for air quality measure with the goal of deployment in citizen science. *To appear in STOTEN Journal, Science of the Total Environment (Elsevier).*

Analyse en ligne des données de biodiversité en milieu agricole

Elodie Edoh-Alove¹, Aurélien Besnard², Catherine Brenon⁶, Sandro Bimonte³, Benoit Fontaine⁴, Ali Hassan⁵, Yassine Hifdi³, Roseline Preud'homme⁴, Iandry Rakotoniaina⁴, Amir Sakka³

1. GEOSYSTEMS France,
6 rue Jean Pierre Timbaud, 78180 Montigny-Le-Bretonneux, France
edoh-alove@geosystems.fr
 2. LPO Aquitaine, 433 Chemin de Leysotte, 33140 Villenave-d'Ornon, France
aurelien.besnard@lpo.fr
 3. Université Clermont, TSCF, Irstea, 9 Av. Blaise Pascal, 63178, Aubière, France
{[sandro.bimonte](mailto:sandro.bimonte@irstea.fr), [amir.sakka](mailto:amir.sakka@irstea.fr)}@irstea.fr
 4. UMS Patrinat (AFB CNRS MNHN), UMR 7204, Muséum national d'Histoire naturelle, 43 rue Buffon, 75005 Paris, France
{[benoit.fontaine](mailto:benoit.fontaine@mnhn.fr), [rose-line.preudhomme](mailto:rose-line.preudhomme@mnhn.fr), [iandry.rakotoniaina](mailto:iandry.rakotoniaina@mnhn.fr)}@mnhn.fr
 5. Équipe R&D, Umanis,
7-9 rue Paul Vaillant Couturier, 92300, Levallois-Perret, France
ahassan@umanis.com
 6. Chambre de l'Agriculture de l'Allier,
60 Cours Jean Jaurès, 03000 Moulins, France
cbrenon@allier.chambagri.fr
-

RESUME. La conservation de la biodiversité et sa relation avec les pratiques agricoles représentent actuellement un défi majeur, car elles touchent à des enjeux environnementaux, sociaux et économiques. Les systèmes VGI ne fournissent pas d'outils graphiques d'analyses complexes. Dans cet article nous présentons donc l'implémentation d'un entrepôt de données spatiales concernant la biodiversité en milieu agricole.

ABSTRACT. The conservation of biodiversity and its relation to agricultural practices currently represent a major challenge, since they affect environmental, social and economic issues. VGI systems do not offer effective analysis tools. Therefore, in this paper we present the implementation of a spatial data warehouse concerning biodiversity data in the agriculture context.

1. Introduction

La conservation de la biodiversité et sa relation avec les pratiques agricoles représentent actuellement un défi majeur, car elles touchent des enjeux environnementaux, sociaux et économiques (Bommarco et al., 2013). Les données d'observation faunistiques et floristiques à grande échelle spatiale ou temporelle englobant un large éventail de situations permettent des analyses statistiquement pertinentes. Cela implique que des centaines ou des milliers d'observateurs soient mobilisés, à un coût qui serait prohibitif s'ils devaient être payés (Regnier et al., 2015). Le travail présenté dans cet article s'inscrit dans le cadre du projet ANR VGI4Bio. VGI4Bio a pour but de définir un ensemble d'outils statistiques et de modèles de comportement d'observateurs pour extraire et visualiser des données précises et pertinentes à partir de la masse de données opportunistes (données Volunteer Geographic Information – VGI (Sui et al., 2013)) afin de produire des indicateurs significatifs de la biodiversité dans le contexte agricole.

De plus, comme les systèmes VGI ne fournissent pas d'outils d'analyses complexes, dans VGI4Bio nous utiliserons l'OLAP spatial (SOLAP) pour analyser ces bio-indicateurs agricoles (Sautot, et al., 2015). L'intégration des données spatiales dans les entrepôts de données (ED) et les systèmes OLAP conduisent au concept de Spatial OLAP (SOLAP) (Andrienko, 1999). Un système SOLAP a été défini comme « *Une plate-forme visuelle spécialement conçue pour supporter l'analyse et l'exploration spatio-temporelles rapides et faciles des données multidimensionnelles à l'aide d'affichages cartographiques aussi bien qu'à l'aide de tableaux et diagrammes statistiques* » (Andrienko, 1999).

Dans ce travail, nous présentons un nouveau système SOLAP développé dans le cadre du projet VGI4Bio. Cet outil implémente de nouvelles techniques de géo-visualisation conçues explicitement pour prendre en compte la particularité des requêtes SOLAP : multi-échelles spatiales (hiérarchies spatiales) et plusieurs indicateurs numériques. Nous montrons comment notre outil SOLAP permet d'explorer et d'analyser les données de biodiversité en contexte agricole de façon simple et intuitive à travers deux cas d'études issus du projet VGI4Bio: cas des données de l'Observatoire Agricole de la Biodiversité (OAB) et cas des données Faune-Aquitaine de la Ligue de Protection des Oiseaux (LPO).

1. OAB et Faune-Aquitaine

Les sciences participatives ont déjà montré avec succès leur intérêt dans différents domaines d'applications.

Pour étudier les impacts de l'agriculture sur la biodiversité, avec des jeux de données à grande échelle spatio-temporelle, le ministère de l'agriculture a mis en

3. Exemples d'analyse

Nous illustrons nos propositions de géo-visualisation à travers deux exemples de requête basés sur les données du protocole pollinisateur de l'OAB.

Classiquement une requête OLAP permet de visualiser dans une table pivot des mesures de données de même dimension mais à des niveaux de granularité différents (Ex: jour->mois, département->région etc.). Ce type de visualisation est impossible dans le SIG car les niveaux spatiaux sont géométriquement emboîtés.

Par exemple, concernant une requête : Abondance d'abeilles dans les départements et régions, la géo-visualisation proposée (cf. Figure 2) est une carte choroplèthe pour les deux niveaux. Le niveau spatial le plus fin (départements) est représenté par des polygones pleins et le niveau le plus élevé (régions) est représenté par une zone tampon

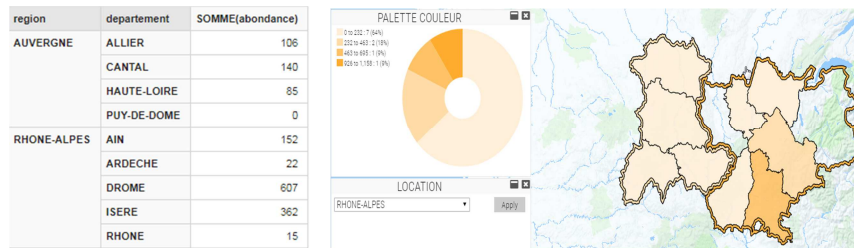


Figure 2. Géo-visualisation pour 2 niveaux spatiaux.

spatiaux.

departement	type_traitement	SOMME(abondance)	SOMME(nbParcellesVisitees)
YVELINES	PAS DE TRAITEMENT	155	6
	TOUT TRAITEMENT	114	9

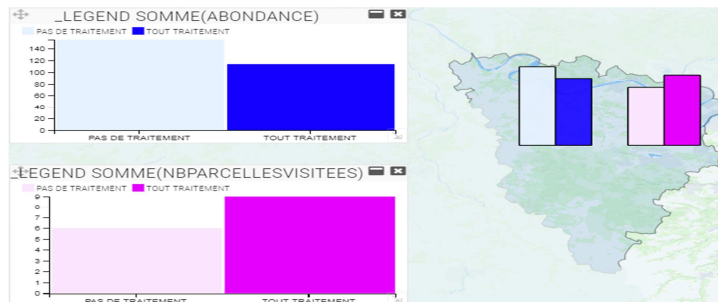


Figure 3. Géo-visualisation pour n mesures.

Contrairement aux SIG et outils SOLAP existants, en ce qui concerne les mesures, nous avons mis en place une politique de distribution de barres sur les cartes, qui tient compte du nombre de mesures, pour regrouper tous les indicateurs

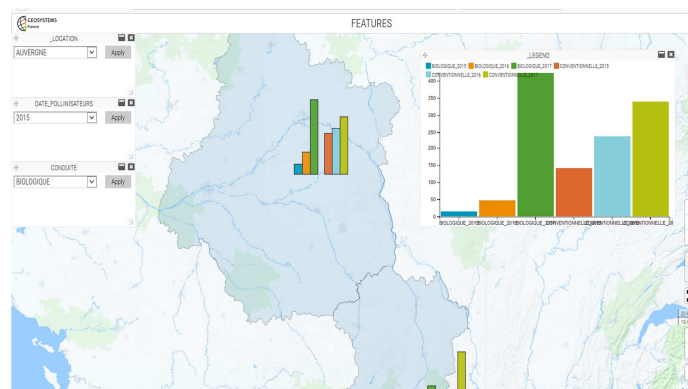
Entreposage et analyse en ligne des données de biodiversité

de la même mesure. Un exemple est présenté dans la Figure 3 concernant la requête : Abondance d'abeilles et nombre de parcelles visitées en fonction du type de traitement.

Enfin, contrairement aux SIG et outils SOLAP existants, nous proposons aussi une géovisualisation qui correspond à l'ordre des colonne de table pivot, car cette ordre défini des requêtes différentes, meme si en utilisant les mêmes données. Considérons la table pivot de Figure 4a. Elle permet l'analyse de la distribution temporelle de la somme de l'abondance pour chaque type de conduite. Par contre, En inversant la dimension temporelle et la dimension conduite, le décideur pourrait visualiser l'impact de chaque conduite sur l'abondance par année. Nous avons donc choisi de représenter ce groupement défini par la dimension plus à droite de la table pivot, aussi sur la carte en créant des groupes de barres. Dans cette exemple, la géovisualisation proposée est montrée en Figure 4b, où deux groupes, associés à "biologique" et "conventionnelle", contiennent 3 barres associées aux années 2015, 2016 et 2017.

region	conduite	date_y	SOMME(abondance)
AUVERGNE	BIOLOGIQUE	2016	5
		2017	220
	CONVENTIONNELLE	2015	96
		2017	544
CENTRE	BIOLOGIQUE	2015	24
		2016	135
		2017	626
	CONVENTIONNELLE	2015	316
2016		362	
2017	469		
	ILE-DE-FRANCE	BIOLOGIQUE	2015
2016			3
CONVENTIONNELLE		2015	13
		2016	114
		2017	6

(a)



(b)

Figure 4. Géovisualisation pour n dimensions.

Nous pouvons aussi remarquer que des widgets pour la sélection des membres de dimensions utilisées dans la requête OLAP sont automatiquement affichés aussi dans la carte. Le couplage des cartes interactives et ces outils de sélection permet de supporter pleinement le paradigme par excellence d'analyse visuelle "overview, zoom filter and details on demand" (Andrienko et al., 1999).

4. Conclusion et travaux futurs

Motivé par l'importance des outils de géo-visualisation dans l'analyse SOLAP, nous avons présenté dans cette contribution une implémentation d'un entrepôt de données spatiales de biodiversité. Nos travaux en cours sont la finalisation des applications SOLAP pour les autres protocoles de l'OAB et l'implémentation d'autres fonctionnalités de géo-visualisation dans l'outil SOLAP.

Remerciements. Ce travail est soutenu par le projet ANR VGI4Bio (ANR-17-CE04-0012), l'association Symbiose Allier et le projet ANR programme "Investissements d'Avenir" (16-IDEX-0001 CAP 20-25).

Bibliographie

- Andrienko G., Andrienko N. (1999). Interactive maps for visual data exploration. *International Journal of Geographical Information Science*, vol. 13, n° 4, p. 355-374.
- Bimonte S. (2014). A generic geovisualization model for spatial OLAP and its implementation in a standards-based architecture. *Ingénierie des Systèmes d'Information*, vol. 19, n° 5, p. 97-118.
- Bimonte S., Hassan A. (2016). OLAP2Map: Un outil OLAP Spatial. *Atelier Démo SAGEO 2016*.
- Bommarco R., Kleijn D., Potts SG. (2013). Ecological intensification: harnessing ecosystem services for food security. *Trends in ecology and evolution*. vol. 28, p. 230-238.
- Regnier, C., Achaz, G., Lambert, A., Cowie, R.H., Bouchet, P., Fontaine, B. (2015). Mass extinction in poorly known taxa. *Natl. Acad. Sci. U. S. A.*, vol. 112, p. 7761-7766.
- Sautot L., Faivre B., Journaux L., Molin P. (2015). The hierarchical agglomerative clustering with Gower index: A methodology for automatic design of OLAP cube in ecological data processing context. *Ecological Informatics*, vol. 26 n° 2, p. 217-230.
- Sui, D.Z., Elwood, S., Goodchild, M. (2013). Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in *Theory and Practice*. Springer.

Géomorphologie à Tetiaroa : Prémices d'un observatoire de données scientifiques et environnementales.

Benoît Stoll¹ et Manaarii Longine

1. Laboratoire GEPASUD, Université de la Polynésie française
BP6570, 98702, Faa'a – Tahiti, Polynésie française
Benoit.stoll@upf.pf

RESUME. En 1966, l'atoll de Tetiaroa a été acheté par Marlon Brando. Il a souhaité qu'il devienne un modèle écologique et un site dédié à la conservation, l'éducation et la recherche. La Tetiaroa Society accueille des équipes de chercheurs locaux et internationaux, elle a exprimé un besoin en termes de collecte, stockage, partage et représentation des données scientifiques produites sur l'atoll.

Le traitement géomorphologique des données raster historiques remontant à 1955 constitue les prémices d'une base de données géographiques multi-thématique qui fera à terme de Tetiaroa un atoll sous surveillance environnementale et scientifique.

Cet article décrira la cartographie de l'occupation des sols et du trait de côte des motu de Tetiaroa, une analyse diachronique permettra de vérifier le comportement géodynamique de ceux-ci et un suivi de l'érosion sera proposé.

ABSTRACT. The atoll of Tetiaroa has been acquired in 1966 by the actor Marlon Brando. He wished Tetiaroa to become an ecological model and a conservation, education and research site. The Tetiaroa Society hosts teams of local and international researchers, it expressed a strong need in terms of collecting, stocking, sharing and representing the scientific data produced on the atoll.

The geomorphologic processing of the historical raster database from 1955 up to 2014 constitute the firstfruits of a multi-thematic geographical database which will make of Tetiaroa an atoll under environmental and scientific surveillance.

This article describes the land cover and coast line mapping of the Tetiaroa's motu, a diachronic analysis checks their geodynamic behavior, and an erosion and scarps monitoring is proposed.

MOTS-CLES : Géomorphologie, Atoll de Tetiaroa, Observatoire, Surveillance environnementale, Système d'information géographique.

KEYWORDS: Géomorphology, Tetiaroa atoll, Observatory, Environmental monitoring, Geographical information system.

1. Introduction

L'île de Tetiaroa est le seul atoll des îles du vent dans l'archipel de la société en Polynésie française, il est situé à 50km au nord de l'île de Tahiti par 17°00' Sud et 149°34'Ouest (Figure 1). Cet atoll fermé possède 12 « *motu* » ou îlots coralliens, de surface cumulée de 5,5km² (Table 2).



FIGURE 1. Atoll de Tetiaroa – îles du vent (gauche) – Motu de Tetiaroa (droite)

La Tetiaroa Society, organisation à but non lucratif est chargée de promouvoir la préservation de l'environnement, la recherche scientifique et l'éducation. L'éco-station sert de lieu d'accueil pour les équipes de recherche internationales.

Avec la multiplication des programmes de recherches sur l'atoll, la Tetiaroa Society a exprimé le besoin de centraliser et partager les données issues de ces recherches dans une base de données pérenne : celle-ci permettra de mettre à disposition des fonds de cartes, des couches vectorielles spécifiques et de stocker les données des différents chercheurs et associations dans un système commun afin d'optimiser le traitement et la valorisation des données scientifiques.

Le système d'information géographique (SIG) étant un outil particulièrement adapté à cette problématique, un SIG multi-thématique a commencé à être mis en place depuis 2014. Pour assurer la compatibilité des différents types de données, il faut garantir une précision spatiale des données suffisante. Des méthodes de positionnement par GPS différentiel (DGPS) ont été étudiées concrètement sur Tetiaroa. Les données raster disponibles, remontant jusqu'en 1955, ont été géo-référencées lorsque nécessaire (Table 1). Leur traitement a permis de commencer à alimenter l'observatoire de données de Tetiaroa.

Nous développerons dans cet article le descriptif de la base de données raster historique ainsi que le protocole GPS utilisé. Nous poursuivrons par la cartographie de l'occupation des sols, l'analyse diachronique de la géomorphologie des *motu* puis le suivi du trait de côte et de l'érosion. Nous finirons par des conclusions et perspectives.

2. Base de données raster et géolocalisation GPS

La base de données comprend trois types de données raster : des prises de vue aériennes (PVA), des images satellites ainsi que de la donnée lidar. (Table 1)

TABLE 1. Base de données raster de Tetiaroa

Date	Type	Capteur	Précision	Droits
1955	PVA	Photo N&B	1m20	Service de l'urbanisme - PF
1981	PVA	Photo N&B	1m	Service de l'urbanisme - PF
2001	PVA	Photo couleur	50cm	Service de l'urbanisme - PF
2004	PVA	Photo couleur	80cm	Service de l'urbanisme - PF
2010	Image satellite	Worldview2	50cm	Service de l'urbanisme - PF
2012	Image satellite	Worldview2	50cm	Laboratoire GEPASUD
2014	Image satellite	Pléiades	70cm	ETH Zurich - IDEA
2017	Image héliportée	Lidar	-	EsTH Zurich - IDEA

Les ortho-photos de 1955 à 2004 ne nécessitent pas de corrections géométriques, mais elles ont du être géo-référencées par un recalage sur l'image Pléiades 2014 en utilisant comme point d'amer des blocs coralliens n'ayant ni grossi, ni été déplacés depuis 1955, et situés à une faible profondeur.

En 2017 le projet Moorea IDEA (Davies *et al.*, 2016) a financé une campagne de cartographie lidar de l'atoll de Tetiaroa (Gruen *et al.*, 2017). De façon classique, la donnée lidar permet de produire un modèle numérique de surface (MNS) et un modèle numérique de terrain (MNT). Sur Tetiaroa, un modèle bathymétrique du lagon (jusqu'à -20m) a aussi été réalisé.

Loin de tout réseau de bases GPS permanentes, la géolocalisation est effectuée à l'aide d'un mobile Trimble Geo7x (antenne Zéphyr2 L1/L2) pour les mesures de terrain, et la post-correction utilise les données d'une base Trimble GeoXH (antenne Zéphyr L1/L2). La précision obtenue est de 2cm en milieu ouvert et de l'ordre de 50cm sous la végétation (Stoll and Longine, 2019).

3. Géomorphologie

3.1. Occupation des sols

La bande infrarouge de l'image Pléiades 2014 a mis en évidence le corail vivant sous l'écume des vagues, la couronne récifale a ainsi pu être vectorisée par photo-interprétation.

(Cozannet *et al.*, 2013) a relevé le trait de côte en 1955, 1981 et 2001. Nous étendons cette étude aux images de 1955 à 2014 (Table 1). L'occupation des sols (Figure 2) a été vectorisée par photo-interprétation. Les classes considérées sont : l'aragonite, les débris coralliens, le sable corallien, la lentille d'eau douce, la végétation, et la végétation de marécages (Stoddart, 1980).

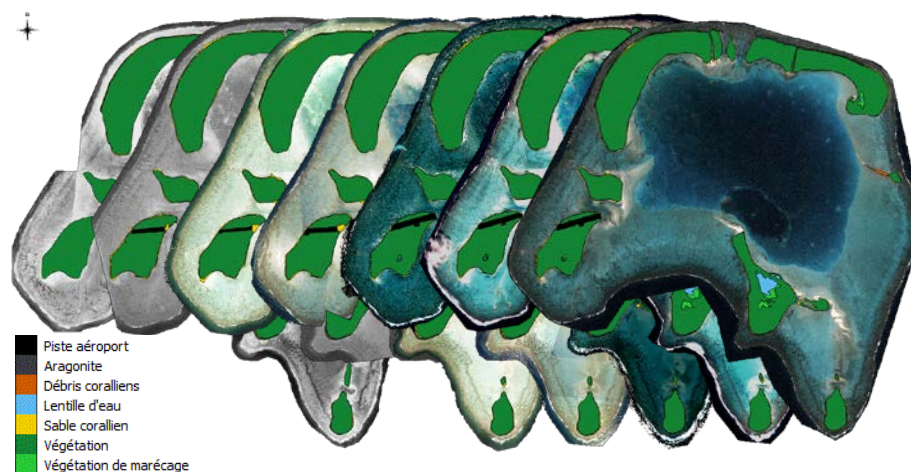


FIGURE 2. Occupation des sols de 1955 à 2014(de gauche à droite)

3.2. Analyse diachronique

A partir de ces cartographies, l'évolution des *motu* depuis 1955 a été étudiée en termes de surface et de trait de côte. Les surfaces en 1907 de la Table 2 sont issues d'un document notarial.

TABLE 2. Surface des *motu* de Tetiaroa (en Hectares)

	1907	1955	1981	2001	2004	2010	2012	2014
<i>Aie</i>	2,9	2,06	2,07	2,27	2,31	2,31	2,3	2,75
<i>Auroa</i>	5,2	4,85	4,8	4,65	4,67	4,65	4,67	4,62
<i>Hiraanae</i>	38,45	36,66	36,64	36,62	36,32	36,57	36,41	36,52
<i>Honuea</i>	29,1	28,01	30,7	31,35	31,16	30,8	30,66	30,59
<i>Motu one</i>	<u>3,1</u>	<u>1,65</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>	<u>0</u>
<i>Onetahi</i>	76,9	75,62	78,6	78,57	78,57	77,93	77,48	78,54
<i>Oroatera</i>	83,2	87,17	90,3	89,64	89,64	90,24	89,99	90,31
<i>Reiono</i>	26,3	23,83	Abs.	23,84	23,84	23,71	23,65	23,99
<i>Rimatu'u</i>	90,7	88,72	95,77	94,5	94,5	92,86	93,1	93,62
<i>Tahuna iti</i>	<u>5,1</u>	<u>5,01</u>	<u>9,25</u>	<u>8,17</u>	<u>8,17</u>	<u>8,37</u>	<u>7,21</u>	<u>7,11</u>
<i>Tahu. rahi</i>	<u>5,15</u>	<u>3,54</u>	<u>Abs.</u>	<u>1,26</u>	<u>1,26</u>	<u>0,97</u>	<u>0,97</u>	<u>0,88</u>
<i>Tauini</i>	7,4	8,22	8,18	8,05	8,05	8,19	8,22	8,17
<i>Tiaraunu</i>	211,5	171,36	175,14	174,55	174,55	174,44	174,35	174,95
Total	580	538,34	534,09	547	555,69	553,34	551,63	553,5

La surface des *motu* sur socle corallien (en italique Table 2) n'évolue pas et à l'exception de certains bancs de sable, ne voient pas leur trait de côte évoluer. A *contrario*, d'autres *motu*, sablonneux (soulignés et grisés Table 2), voient une forte évolution (Figure 3). Ceux-ci se déplacent au grès des aléas climatiques (forte houle, cyclones, etc.) par un phénomène d'érosion/accrétion (Cozannet *et al.*, 2013).

Géomorphologie à Tetiaroa : prémices d'un observatoire de données scientifiques et environnementales

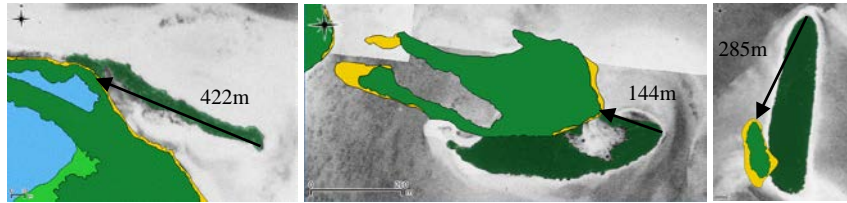


FIGURE 3. Evolutions de Motu one (gauche), Tahuna iti (centre), Tahuna rahi (droite) entre 1955 (transparent) et 2014 (plein)

3.3. Suivi du trait de côte et de l'érosion

Les changements géomorphologiques conséquents de certains *motu* imposent un suivi régulier. Les traits de côte de 1955 à 2014 sont extraits des occupations des sols précédentes. Un suivi du trait de côte par relevés GPS est entamé en 2016 et 2017 en notant le type de substrat parmi : *Aragonite – Construction – Débris coralliens – Sable corallien – Escarpements et Végétation* (Berney *et al.*, 2016).

Depuis quelques années, une érosion importante sur certaines plages a été observé (Jeanson *et al.*, 2014). Celle-ci touche l'habitat et le lieu de ponte de nombreuses espèces sensibles (oiseaux, tortues, ...). Si nous ne pouvons en vérifier l'historique, il est possible d'en cartographier les escarpements. Un calcul de pentes sur le modèle numérique de terrain (MNT) lidar de 2017 a permis de localiser 9443,5 mètres linéaires d'escarpements dont il convient de surveiller l'évolution.

4. Conclusions et perspectives

L'observatoire de données de Tetiaroa a été initié par l'intégration des données raster historiques de 1955 à 2017 dont le traitement a permis de produire des données géomorphologiques (occupation des sols, trait de côte). Plusieurs programmes scientifiques ont déjà commencé à alimenter la base de données : citons ici l'élimination des moustiques (Bossin *et al.*, 2016; Stoll and Bossin, 2019), l'observation des tortues marines (Petit *et al.*, 2016; Stoll *et al.*, 2019) et l'inventaire des sites archéologiques (Molle and Hermann, 2017; Molle *et al.*, 2019).

Concernant l'aspect méthodologique, un serveur hébergeant la base de données doit être mis en place et une politique de stockage des données scientifiques des chercheurs accueillis à la Tetiaroa Society doit être formalisée. A court terme il est envisagé de mettre au point un protocole de modélisation 3D des plages par drone ; celui-ci permettra le suivi des nids de tortues (Stoll *et al.*, 2019) ainsi que la surveillance des escarpements dus à l'érosion des plages.

Remerciements :

Merci au SAU de la Polynésie française, au projet IDEA pour les données raster et lidar. Armin Gruhn, Tao Guo, Serkan Ural, Frank Murphy pour leur aide.

Bibliographie

- Berney, J., Eland, I., Krantz, T., Moore, S., Smith, D., Strout, N., Wilgers, L., 2016. The Tetiaroa Atoll Research Project: Documenting Changes in Island Morphology and Biodiversity through GIS-Based Field Research. Ocean Solutions, Earth Solutions.
- Bossin, H., Petit, H., Marie, J., Stoll, B., Cheong Sang, M., 2016. Suppression of an isolated population of the mosquito vector *Aedes polynesiensis* on the atoll of Tetiaroa, French Polynesia, by sustained release of *Wolbachia*-incompatible male mosquitoes. ICE - XXV International Congress of Entomology, Orlando, Florida, USA.
- Cozannet, G.L., Garcin, M., Petitjean, L., Cazenave, A., Becker, M., Meyssignac, B., Walker, P., Devilliers, C., Brun, O.L., Lecacheux, S., Baills, A., Bulteau, T., Yates, M., Wöppelmann, G., 2013. Exploring the relation between sea level rise and shoreline erosion using sea level reconstructions: An example in French Polynesia. 12th International Coastal Symposium. Journal of Coastal Research, Plymouth, England, pp. 2137-2142.
- Davies, N., Field, D., Gavaghan, D., Holbrook, S.J., Planes, S., Troyer, M., Bonsall, M., Claudet, J., Roderick, G., Schmitt, R.J., Amaral Zettler, L., Berteaux, V., Bossin, H., Cabasse, C., Collin, A., Deck, J., Dell, T., Dunne, J., Gates, R., Harfoot, M., Hench, J.L., Hopuare, M., Kirch, P., Kotoulas, G., Kosenkov, A., Leichter, J.J., Lenihan, H., Magoulas, A., Martinez, N., Meyer, C., Stoll, B., Swalla, B., Tartakovsky, D.M., Teavai Murphy, H., Turyshv, S., Valdinos, F., Williams, R., Wood, S., Consortium, I., 2016. Simulating social-ecological systems: the Island Digital Ecosystem Avatars (IDEA) consortium Gigascience - Biomed Central. The Open Publisher.
- Gruen, A., Guo, T., Ural, S., Troyer, M., Kocaman, S., 2017. DSM/DTM-related investigations of the Moorea Avatar project. Asian Conference on Remote Sensing, New Delhi, India.
- Jeanson, M., Anthony, E., Etienne, S., Dolique, F., 2014. Morphodynamic characterization of beaches on a Pacific atoll island: Tetiaroa, French Polynesia. 13th International Coastal Symposium. Journal of Coastal Research Durban, South Africa, pp. 176-181.
- Molle, G., Hermann, A., 2017. Archaeological Research on Teti'aroa (Society islands, French Polynesia) Final Report - Phase 1.
- Molle, G., Hermann, A., Lagarde, L., Stoll, B., 2019. The long-term history of Teti'aroa (Society Islands, French Polynesia): New archaeological and ethnohistorical investigations. Journal of Pacific Archaeology
- Petit, M., Gaspar, C., Leport, G., Esposito, C., Stabile, V., 2016. Saisons de ponte 2014-2015 et 2015-2016 de la tortue verte (*Chelonia mydas*) sur l'atoll de Tetiaroa. Te Mana O Te Moana.
- Stoddart, D., 1980. Land cover map of Tetiaroa's motu. Gump station - Moorea.
- Stoll, B., Bossin, H., 2019. Observatoire de données de Tetiaroa : Contribution à la modélisation de l'environnement des moustiques., SAGEO, Clermont Ferrand.
- Stoll, B., Hart, M., Gaspar, C., 2019. Contribution de la géomatique à l'observation des nids de tortues marines de Tetiaroa., SAGEO, Clermont Ferrand.
- Stoll, B., Longine, M., 2019. Géomorphologie à Tetiaroa : Prémices d'un observatoire de données scientifiques et environnementales. SAGEO, Clermont Ferrand.

Observatoire de données de Tetiaroa : Contribution à la modélisation de l'environnement des moustiques.

Benoît Stoll¹ et Hervé Bossin²

1. Laboratoire GEPASUD, Université de la Polynésie française

BP6570, 98702, Faa'a – Tahiti, Polynésie française

Benoit.stoll@upf.pf

2. Institut Louis Malardé

Papeete – Tahiti, Polynésie française

hbossin@ilm.pf

RESUME. En 2014 l'institut Louis Malardé de Papeete entame une opération innovante de lutte anti-vectorielle sur le motu Onetahi de l'atoll de Tetiaroa. Afin de suivre la dynamique de la population de moustiques *Aedes polynesiensis*, un réseau de pièges à moustiques est installé et géolocalisé précisément. Dans une optique de modélisation, l'environnement des moustiques a été entièrement cartographié en considérant les différents types d'occupation des sols, sable et végétation, ainsi que les éléments anthropiques comme les bâtiments divers de l'hôtel « The Brando », les habitations du personnel, les bâtiments techniques, les réseaux techniques, les routes, Chaque couche vectorielle se voit attribuer dans sa table d'attribut, un triplet d'indices GOR (« Gite », « Obstacle », « Repos ») représentant la probabilité des moustiques d'effectuer l'un de leurs 3 comportements de base. Cette cartographie exhaustive alimente l'observatoire de données scientifiques et environnementales de Tetiaroa.

ABSTRACT. In 2014, the Institut Louis Malardé from Papeete, launched an innovative vector control experiment on motu Onetahi, Tetiaroa atoll. In order to monitor the *Aedes polynesiensis* mosquito population, a network of mosquito traps has been deployed and precisely geo-located. To modelize the experiment, the mosquito environment has been completely mapped considering the various land use such as sand, vegetation, and the various anthropic elements such as "The Brando" hotel buildings, staff accommodations, technical facilities, roads, etc. A BOR index triplet ("Breeding", "Obstacle", "Resting") which represents the probabilities of one of the three mosquito's basic behavior is specified for each vector layer in its attribute table. This exhaustive map feeds the Tetiaroa scientific and environmental data observatory.

MOTS-CLES : Lutte anti-vectorielle, cartographie, environnement des moustiques, modélisation, observatoire de données scientifiques et environnementales, Atoll de Tetiaroa .

KEYWORDS: Vector fight, mapping, mosquito environment, modeling, scientific and environmental data observatory, Tetiaroa atoll.

1. Introduction

L'île de Tetiaroa est un atoll des îles du vent dans l'archipel de la société en Polynésie française. Cet atoll fermé, situé à 50km au nord de l'île de Tahiti par 17°00' Sud et 149°34'Ouest, comporte 12 « *motu* » (îlots coralliens) (Figure 1).

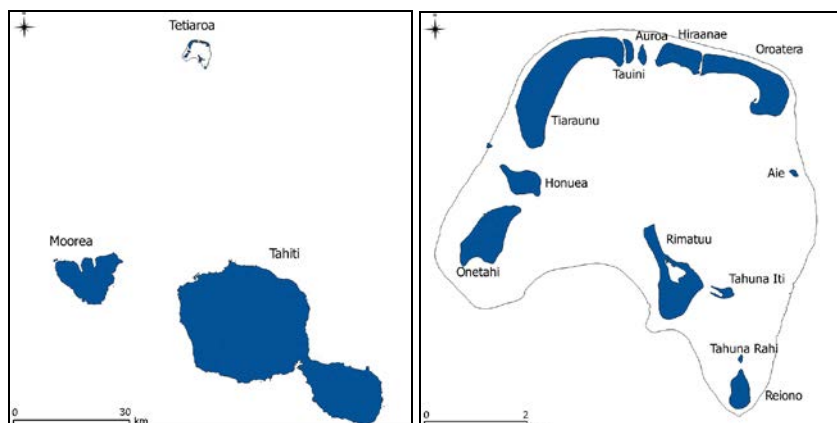


FIGURE 1. Atoll de Tetiaroa – Archipel de la société – Polynésie française

La Tetiaroa Society, créée en partenariat avec l'hôtel de grand luxe « The Brando », est une organisation à but non lucratif chargée de mettre en avant la préservation de l'environnement, la recherche scientifique et l'éducation sur l'atoll de Tetiaroa (Tetiaroa_Society, 2016). Le projet Moorea IDEA (Davies *et al.*, 2016; Gruen *et al.*, 2017) a pour but de modéliser l'environnement des îles de Moorea et Tetiaroa.

Depuis 2014, l'institut Louis Malardé conduit un programme innovant de lutte anti-vectorielle sur le *motu* Onetahi de l'atoll de Tetiaroa (Limb Kemp *et al.*, 2013; Bossin *et al.*, 2016a, b). Le but étant l'élimination du moustique *Aedes polynesiensis*, espèce vecteur de maladies vectorielles. L'institut Louis Malardé souhaite modéliser le comportement des moustiques afin d'optimiser les paramètres de l'opération pour améliorer la stratégie des futures campagnes de lutte anti-vectorielle.

Cet article décrit la méthodologie de cartographie de l'environnement des moustiques qui servira de base à cette modélisation. Les couches vectorielles produites sont intégrées dans le SIG multi-thématique de la Tetiaroa Society, qui est le prémice d'un nouvel observatoire de données scientifiques et environnemental (Stoll and Longine, 2019).

Nous décrivons dans un premier temps le protocole du programme de lutte anti-vectorielle, puis l'aspect méthodologique : images satellites, mesure GPS et création des différentes couches vectorielles, suivi de conclusions et perspectives.

2. Programme de lutte anti vectorielle

Le procédé consiste à produire au laboratoire et à lâcher dans les zones infestées, de façon régulière et en nombre suffisant, des moustiques mâles porteurs de la bactérie *Wolbachia* (technique dite de l'insecte incompatible ou TII). Les moustiques mâles ne piquent pas, ils ne sont donc pas source de nuisance et ne transmettent pas de pathologies infectieuses. Une femelle moustique inséminée par l'un de ces mâles est stérile à vie, l'incompatibilité par *Wolbachia* induisant l'arrêt du développement embryonnaire dans les œufs qu'elle pond. Incapable de se renouveler, la population de moustiques ainsi traitée s'effondre en l'espace de quelques mois.

Le *motu* Onetahi a été choisi pour deux raisons principales : d'une part, il est de taille modeste (78Ha), ce qui réduit le nombre de moustiques mâles stérilisants à produire, ainsi que l'effort de cartographie. D'autre part, il est suffisamment éloigné des autres *motu* pour que les moustiques soient isolés des autres populations de moustiques, ce qui limite le risque de réinfestation et facilite l'évaluation de l'efficacité du procédé TII sur le terrain.

Un suivi précis de la dynamique de la population de moustiques avant, pendant et après l'opération est réalisé. Pour cela, un réseau de pièges (pièges BG-Sentinel et pièges GAT) a été positionné sur le *motu* Onetahi ainsi que sur les motus Honuea, Tiarauu, Auroa, Rimatuu et Reiono qui font office de *motu* témoins.

3. Méthodologie

3.1. Données images et protocole GPS

Le projet Moorea IDEA a fourni une image Pléiades tri-stéréoscopique de Tetiaroa (R, V, B, IR à 70cm) de 2014 et le modèle numérique de surface (MNS) afférent. Cette image a servi de fond de carte ainsi que de base pour la carte de la végétation.

Concernant le positionnement, un GPS Trimble Geo7x est utilisé pour les relevés terrain, la post-correction étant effectuée en utilisant une base GPS Trimble GeoXH avec antenne Zéphyr L1/L2 (Stoll and Longine, 2019). La précision obtenue est de 2cm avec un ciel ouvert et de l'ordre de 50cm sous la végétation.

3.2. Couches vectorielles et cartographie

La cartographie est réalisée à partir d'outils SIG gratuits (Qgis). Ces outils permettent d'associer des données Raster (images satellites) et des données vectorielles. L'ensemble des éléments cartographiques produits dans le cadre de cette étude sont au format vectoriel afin de pouvoir associer des informations complémentaires dans la table d'attribut. La cartographie finale sera la superposition des différentes couches vectorielles décrites ci-après.

Il s'agit de cartographier précisément l'environnement des moustiques, c'est-à-dire tous les éléments paysagers et architecturaux qui interagissent avec ces derniers. Citons les éléments naturels comme les plages, la végétation, mais aussi les éléments anthropiques comme la piste d'aéroport, les bâtiments et routes, les chambres techniques, ainsi que l'emplacement de chaque piège à moustiques de l'opération.

3.2.1. *Eléments naturels*

Les plages ont été extraites de l'occupation des sols (Stoll and Longine, 2019). La végétation sur le *motu* Onetahi, est de deux types : naturalisée au centre du *motu*, et fortement anthropisée (plantée et/ou entretenue) partout ailleurs.

La cartographie de la végétation utilise un jeu de classes « orienté moustique » basé non pas sur les espèces végétales mais sur le type de végétation. En effet le moustique *Aedes polynesiensis* affectionne les forêts fermées, ombragées, humides avec des sous-bois arbustifs, il reste loin des zones ouvertes, ensoleillées et venteuses. Cette cartographie a été produite par photo-interprétation de l'image Pléiades 2014 appuyée sur la hauteur de la végétation tirée du MNS. Ce travail a été appuyé par une campagne intensive de relevés terrains au GPS (1600 points).

3.2.2. *Eléments anthropiques*

Cartographier les éléments anthropiques a nécessité de numériser l'ensemble des bâtiments (hôtel, zone technique, zones de vie, containers de stockage, ...) mais aussi les routes et pistes ainsi que les chambres techniques pouvant abriter des gîtes.

Les plans d'architecte de l'hôtel ont été vectorisés et recalés. Les 289 bâtiments ainsi que les 156 chambres techniques ont été positionnés au GPS, ou par recalage sur l'image Pléiades 2014. De même, les 7989 mètres linéaires de routes et pistes ont été relevés au GPS et les zones de défrichage (4 mètres autour des routes) ont été extrapolées par un tampon fixe. Enfin, l'emplacement précis des 49 pièges BG-Sentinels et des 78 pièges GAT du réseau a été relevé.

3.2.3. *Indices comportementaux G.O.R.*

Le moustique a trois types d'interactions comportementales avec son environnement : il se déplace, il se (re)pose, et il exploite des gîtes larvaires. Dans un but de modélisation, un triplet d'indices « *G.O.R.* » pour « Gîte », « Obstacle » et « Repos » est ajouté dans la table attributaire de chaque couche vectorielle afin de quantifier la probabilité respective de ces trois comportements de base.

La valeur probabiliste de ces indices peut être déterminée par la connaissance *a priori* des experts sur le comportement des moustiques, ou dans certains cas, être estimée par une expérience de marquage/lâcher/recapture permettant de quantifier le degré de déplacement d'un moustique au travers de certains types de végétation. Par anticipation, et comme les deux espèces de moustiques vecteurs présentes sur le

Observatoire de données de Tetiaroa : Contribution à la modélisation de l'environnement des moustiques

motu Onetahi ont des comportements différents, nous avons implémenté deux triplets d'indice *G.O.R.* : *Gap*, *Oap* et *Rap* pour l'*Aedes polynesiensis* et *Gaa*, *Oaa* et *Raa* pour l'*Aedes aegypti* (Figure 2). Les bâtiments étant constitués de différentes parties avec des valeurs *G.O.R.* distinctes, la vectorisation a pris en compte ce fait.

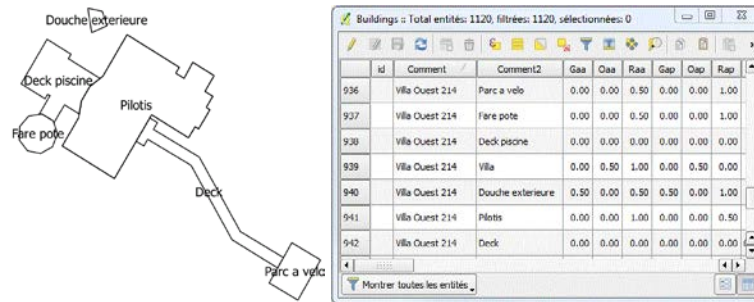


FIGURE 2. Indices *G.O.R.* d'une villa de l'hôtel

La Figure 3 rassemble l'ensemble des couches vectorielles précitées et constitue la cartographie de l'environnement des moustiques sur le *motu* Onetahi.

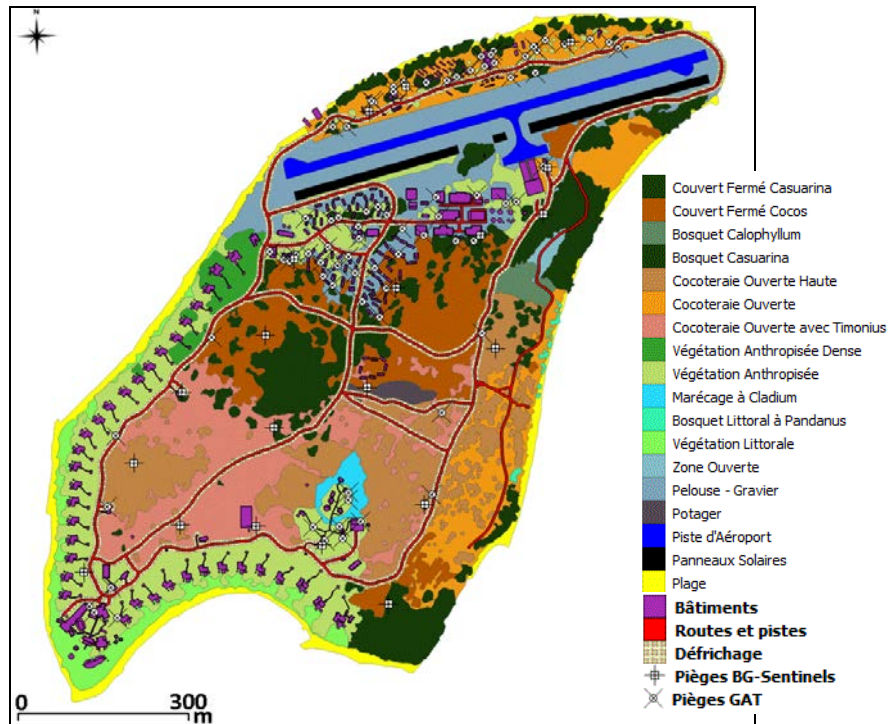


FIGURE 3. Cartographie de l'environnement des moustiques sur Onetahi

4. Conclusions et perspectives

Nous avons effectué la cartographie de l'environnement des moustiques sur le *motu* Onetahi de Tetiaroa où se déroule l'opération pilote d'élimination des moustiques de l'institut Louis Malardé. Pour ce faire, nous avons cartographié l'ensemble des éléments naturels mais aussi des éléments anthropiques présents sur le *motu*. Un triplet d'indices *G.O.R.* a été proposé, dans la table d'attribut des couches vectorielles, pour quantifier les probabilités d'occurrence des trois types d'interactions entre les moustiques et leur environnement.

Cette carte va servir de base pour modéliser le comportement des moustiques et ainsi permettre l'optimisation des différents paramètres de l'opération. Elle est aussi intégrée à l'observatoire de données scientifiques et environnementales de la Tetiaroa Society.

Remerciements :

Merci au projet IDEA pour les données Pléiades2014, à l'institut Louis Malardé, à Xenia Jost pour sa contribution et à la Tetiaroa Society pour son aide logistique.

Bibliographie :

Bossin, H., Petit, H., Marie, J., Stoll, B., Cheong Sang, M., 2016a. Releases of *Wolbachia* mosquito males to sterilize and suppress a population of *Aedes Polynesiensis* on Tetiaroa Atoll, French Polynesia. International Congress for Tropical Medicine and Malaria, Brisbane - Australia.

Bossin, H., Petit, H., Marie, J., Stoll, B., Cheong Sang, M., 2016b. Suppression of an isolated population of the mosquito vector *Aedes polynesiensis* on the atoll of Tetiaroa, French Polynesia, by sustained release of *Wolbachia*-incompatible male mosquitoes. ICE - XXV International Congress of Entomology, Orlando, Florida, USA.

Davies, N., Field, D., Gavaghan, D., Holbrook, S.J., Planes, S., Troyer, M., Bonsall, M., Claudet, J., Roderick, G., Schmitt, R.J., Amaral Zettler, L., Berteaux, V., Bossin, H., Cabasse, C., Collin, A., Deck, J., Dell, T., Dunne, J., Gates, R., Harfoot, M., Hench, J.L., Hopuare, M., Kirch, P., Kotoulas, G., Kosenkov, A., Leichter, J.J., Lenihan, H., Magoulas, A., Martinez, N., Meyer, C., Stoll, B., Swalla, B., Tartakovsky, D.M., Teavai Murphy, H., Turyshchev, S., Valdivinos, F., Williams, R., Wood, S., Consortium, I., 2016. Simulating social-ecological systems: the Island Digital Ecosystem Avatars (IDEA) consortium. Gigascience - Biomed Central. The Open Publisher.

Gruen, A., Guo, T., Ural, S., Troyer, M., Kocaman, S., 2017. DSM/DTM-related investigations of the Moorea Avatar project. Asian Conference on Remote Sensing, New Delhi, India.

Limb Kemp, H., Cheong Sang, M.A., Tuterarii, P., Bossin, H., 2013. SUPPORTIVE OUTCOME OF A WOLBACHIA-BASED RANGING TRIAL AGAINST *AEDES POLYNESESIENSIS* ON THE ATOLL OF TETIAROA, FRENCH POLYNEsia. *Pathogens and Global Health* 107, 457-457.

Stoll, B., Longine, M., 2019. Géomorphologie à Tetiaroa : Prémices d'un observatoire de données scientifiques et environnementales. SAGEO, Clermont Ferrand.

Tetiaroa_Society, 2016. TETIAROA CONSERVATION AND SUSTAINABLE USE PLAN (CASUP).

Un système d'information géographique décisionnel pour la surveillance épidémiologique de la tuberculose en Afrique Subsaharienne : Cas du Gabon

Raymond Ondzigue Mbenga^{1,2,3}, Thomas Devogele¹, Sydney Maghendji^{2,3}, Veronika Peralta¹, Edgard Brice Ngoungou^{2,3}

1. Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT), Université de Tours, 3 Place Jean-Jaures 41000 Blois

2. Département d'Epidémiologie, Biostatistique, Informatique Médicale (DEBIM), Santé Publique, Médecine du Travail et Légale (DSPMTML), BP 4009 Faculté de Médecine, Université de Sciences de la Santé (USS), Libreville-Owendo, Gabon

3. Unité de Recherche en Epidémiologie des Maladies Chroniques, Santé et Environnement (UREMCSE), BP 11587 Faculté de Médecine, Université de Sciences de la Santé (USS), Libreville-Owendo, Gabon

RÉSUMÉ. Dans cette étude, un état de l'art de l'apport des systèmes d'information géographique (SIG) dans la surveillance épidémiologique de la tuberculose a été réalisé. Une analyse comparative aux niveaux SOLAP, technique et métier de ces SIG a été faite. Les résultats de cette analyse ont permis de proposer une architecture SOLAP pour client mobile et un modèle multidimensionnel d'entrepôt de données pour la mise en place d'un SOLAP tuberculose au Gabon.

MOTS-CLÉS : SIG décisionnel, Surveillance épidémiologique, Tuberculose, Gabon

ABSTRACT. In this study, a state-of-the-art study on the contribution of geographic information systems (GIS) to tuberculosis epidemiological surveillance was carried out. A comparative analysis at the SOLAP, technical and business levels of these GIS was carried out. The results of this analysis made it possible to propose a SOLAP architecture for mobile clients and a multidimensional data warehouse model for the implementation of a tuberculosis SOLAP in Gabon.

KEYWORDS: Decisional SIG, epidemiologic Monitoring, Tuberculosis, Gabon.

1. Introduction

Un système d'information de surveillance épidémiologique a pour objectifs d'aider à la prise de décision pour le contrôle des maladies, d'organiser les circuits d'information nécessaires (utilisateurs-sources) et d'améliorer la pertinence et la fiabilité des informations (Olascoaga, 1982). Il facilite la communication en temps réel ou quasi temps réel des faits épidémiologiques importants aux décideurs qui les analysent pour l'élaboration des recommandations appropriées pour l'application des mesures de prophylaxie¹ par les unités de décision. Selon (Green, 2012), l'application de systèmes d'information géographique (SIG) à la santé publique est très prometteuse. Car, un SIG permet d'automatiser les processus de surveillance épidémiologique des maladies. Aussi, grâce à ces puissants outils d'analyses spatiales, il permet d'améliorer la compréhension des causes de problème de santé complexes et d'élaborer des stratégies de lutte efficaces. (Colby et Johnson, 2002) ont affirmé que le couplage d'un SIG à un outil d'analyse multidimensionnelle OLAP (Online Analytical Processing) ou SOLAP constitue un puissant outil d'aide à la décision en santé publique. Un SIG-OLAP ou SOLAP aide à obtenir en temps réel ou quasi temps réel des données y compris leurs historiques pour faire des calculs prévisionnels et des analyses des tendances permettant d'établir des bonnes stratégies de prévention, de suivi et de planification des interventions pour réduire l'incidence d'une maladie. Un SOLAP est un moyen efficace de visualisation et d'analyse des données épidémiologiques de la tuberculose pour révéler les zones à fort risque de contamination, d'identifier les personnes à risque et suivre sa propagation dans l'espace et son évolution dans le temps.

Cet article est structuré de la façon suivante : La section 2 décrit la problématique. La section 3 présente une analyse de l'état de l'art des SIG appliqués à la surveillance épidémiologique de la tuberculose avant de proposer, en section 4, une architecture SOLAP et un modèle multidimensionnel d'entrepôt de données pour la surveillance de la tuberculose (TB).

2. Problématique

La tuberculose est l'une des premières causes de mortalité et de morbidité au monde. L'organisation mondiale de la santé (OMS) a estimé qu'il y a eu 10,4 millions de nouveaux cas et 1,4 million de décès en 2017. Au Gabon, cette maladie reste un sérieux problème de santé publique. Car, il y a eu une forte augmentation du nombre de cas pulmonaire entre 2016 et 2017, respectivement 126 et 512 cas. La situation de la tuberculose multirésistante reste tout aussi très préoccupante. En effet, sur les 121 cas dépistés entre 2016 et 2018, seule 45 ont pu bénéficier d'un traitement. Cette recrudescence de la tuberculose est en grande partie liée à l'absence d'un système d'information de surveillance efficient et efficace. La grande

¹ . La prophylaxie désigne le processus actif ou passif ayant pour but de prévenir l'apparition, la propagation ou l'aggravation d'une maladie.

masse de données produite par les centres de soins est dispersée dans des outils au format papier (registres des tuberculeux, laboratoire, carte d'identité du patient tuberculeux, fiche suivi du tuberculeux etc.). Ce qui rend très fastidieux les analyses des données pour le calcul des indicateurs permettant d'aider les autorités sanitaires à la prise de décision pour des actions de prévention efficaces. Pour permettre aux autorités sanitaires de réduire l'incidence la tuberculose, nous proposons d'élaborer un système d'information géographique décisionnel. Sa mise en place permettra aux autorités sanitaires de disposer d'une « plate-forme visuelle supportant l'exploration et l'analyse spatio-temporelle des données, à la fois facile et rapide, selon une approche multidimensionnelle, à plusieurs niveaux d'agrégation, via un affichage cartographique, tabulaire ou en diagramme statistique » (Bédard et al, 1997).

La présente étude constitue le premier SIG décisionnel dans le domaine la surveillance épidémiologique de la tuberculose au Gabon.

3. Etat de l'art

Nous avons fait une revue de la littérature sur les systèmes d'information géographique (SIG) appliquée à la surveillance épidémiologique de la tuberculose.

Une analyse comparative des SIG qui ont retenu notre attention a été réalisée selon trois grands groupes avec un ou des critère(s) : (1) SOLAP (critères : analyse des données selon différents grains, affichage des données sous formes de tableaux croisés, graphiques et cartes) ; (2) Technique (critère : utilisation d'un client SOLAP mobile) ; (3) Métier (critères : détection automatique des grappes de tuberculose toutes formes, géolocalisation des cas de tuberculose multirésistante, suivi de la propagation de la tuberculose, prédiction du risque tuberculose, prise en compte d'indicateurs de stocks d'antituberculeux, suivi dans l'espace et dans le temps de l'évolution de la résistance aux antituberculeux). Le critère « *utilisation d'un client SOLAP mobile* » renvoi à l'utilisation des technologies Web et sans fil pour accéder, diffuser, traiter et analyser des données géolocalisées sur la tuberculose pour l'aide à la décision en tout lieu et dans toutes les situations via des appareils mobiles.

La Table 1 résume l'analyse comparative ces différents SIG. Elle montre que, au niveau SOLAP, la majorité des systèmes d'information étudiés ne font pas d'analyse selon différents grains (66,67%), et bien qu'ils affichent les données sous forme de graphiques (66,67%) et de cartes (83,33%), seulement la moitié (50%) affiche les données sous forme de tableaux croisés. Pour ce qui est du niveau technique, aucun système n'utilise de client SOLAP mobile. Au niveau métier, un très grand nombre de systèmes proposés dans la littérature ne font pas détection automatique des grappes de tuberculose toutes formes (83,33%), de géolocalisation de la tuberculose multirésistante (100%), de suivi de la propagation de la tuberculose (100%), de prédiction de risque tuberculose (83,33%), de prise en compte d'indicateurs de stocks d'antituberculeux (100%) et de suivi dans l'espace et dans le temps de l'évolution de la résistance aux antituberculeux (100%).

TABLE 1. *Analyse comparative des SIG tuberculose (TB) de la littérature*

Groupes	Critères	(Moonan et al, 2004)	(Lewetal, 2006)	(Bananabouetal, 2013)	(Zemri et al, 2015)	(Guthrie et al, 2016)	(Smith et Hayward, 2016)
SOLAP	Analyse diff. grains	Non	Non	Oui	Oui	Non	Non
	Aff. tableaux croisées	Non	Oui	Non	Non	Oui	Oui
	Aff. graphiques	Non	Non	Oui	Oui	Oui	Oui
	Aff. cartes	Oui	Non	Oui	Oui	Oui	Oui
Technique	SOLAP mobile	Non	Non	Non	Non	Non	Non
Métier	Détection Grappes tuberculose toutes formes	Non	Non	Non	Oui	Non	Non
	Géolocalisation TB Multirésistante	Non	Non	Non	Non	Non	Non
	Suivi propagation TB	Non	Non	Non	Non	Non	Non
	Prédiction risque TB	Non	Non	Non	Oui	Non	Oui
	Prise en compte indicateurs stocks anti tuberculeux	Non	Non	Non	Non	Non	Non
	Suivi spatio-temporel de la résistance aux antituberculeux	Non	Non	Non	Non	Non	Non

4. Contribution

4.1. Architecture Système d'information géographique décisionnel Tuberculose

L'architecture du système proposé est illustrée dans la Figure 1. Elle est composée de quatre niveaux : L'ensemble des sources de données tuberculose,

l'entrepôt de données spatiales multidimensionnelles, le serveur SOLAP et le client SOLAP.

L'entrepôt de données spatiales multidimensionnelles (EDSM) sera réalisé en utilisant un système de gestion de base de données spatial. C'est à ce niveau que seront stockées les données alphanumériques et spatiales.

Le serveur SOLAP est responsable de la mise en œuvre des opérateurs SOLAP pour calculer et naviguer au sein des cubes de données spatiales. Ces cubes de données sont représentés par un schéma SOLAP qui définit une correspondance entre les concepts de cube données spatiales (dimensions, hiérarchies, mesures, etc.) et le schéma relationnel (tables et colonnes utilisées pour représenter les faits et les dimensions) (Bimonte et al, 1997).

Le client SOLAP est composé d'un client OLAP responsable de la visualisation tabulaire des données et d'un client SIG responsable de la visualisation cartographique (Bimonte et al, 1997).

Dans cette architecture, le module ETL (Extract-Transform-Load) assure l'extraction des données brutes structurées et non structurées des différentes sources de données (cliniques, laboratoire, stock d'antituberculeux, etc.) et leur transformation avant leur chargement dans l'entrepôt de données.

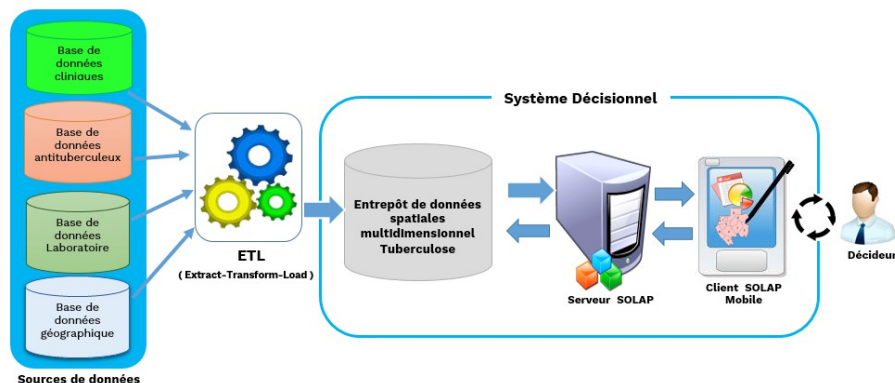


FIGURE 1. Proposition d'architecture SOLAP tuberculose

4.2. Modèle d'entrepôt de données multidimensionnelles tuberculose

Nous avons utilisé le Dimensional Fact Model (DFM) pour la modélisation de l'entrepôt de données multidimensionnel tuberculose. Ce modèle repose sur un formalisme de type entité-relation et permet de représenter des faits, mesures, dimension et hiérarchies.

La modélisation des faits de l'entrepôt de données tuberculose est proposés dans la Figure 2. Les « faits » pris en compte pour la construction de l'entrepôt sont : la mobilité du patient, le traitement de la tuberculose, les examens et les ruptures de stocks d'antituberculeux.

4.2.1. Fait « traitement tuberculose »

Le fait « traitement » a pour dimensions le patient (agrégé selon sexe, âge et profession), le résultat d'examen (TB positif ou négatif), le type de tuberculose. A ces dimensions s'ajoutent les dimensions temporelle (semestre et année) et spatiale (ville, département sanitaire et région sanitaire). L'évaluation de ce fait permettra de suivre les issues de traitement des patients sous traitement antituberculeux.

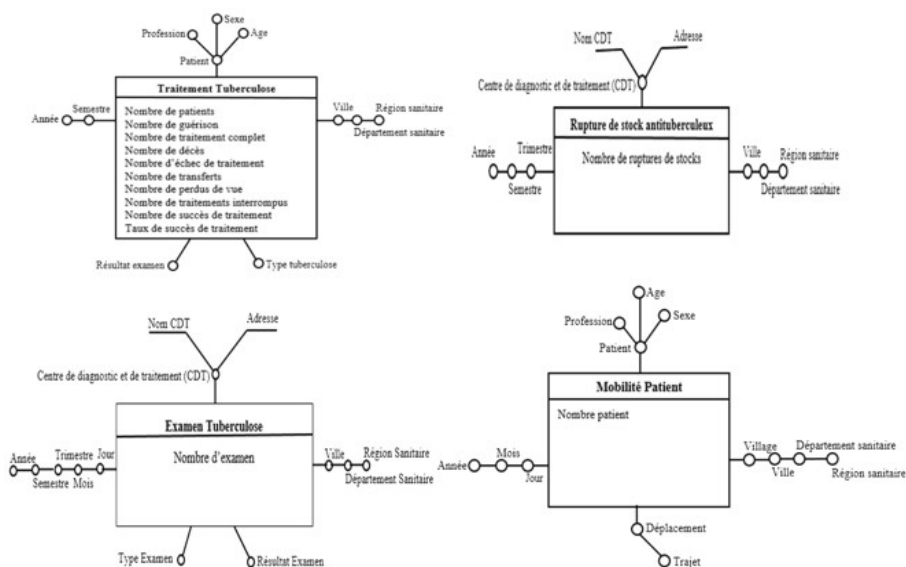


FIGURE 2. Schéma multidimensionnel

4.2.2. Fait « Rupture de stocks d'antituberculeux »

Le fait « stocks d'antituberculeux » a pour dimension le centre de diagnostic et de traitement (CDT). A cette dimension s'ajoutent les dimensions temporelle (trimestre, semestre et année) et spatiale (ville, département sanitaire et région sanitaire). L'analyse de ce fait va permettre suivre les ruptures de stocks d'antituberculeux dans l'espace et dans le temps.

4.2.3. Fait « Examen tuberculose »

Le fait « Examen tuberculose » a pour dimensions le centre de diagnostic et de traitement (CDT), le type d'examen (culture ou microscopique) et résultat d'examen. A ces dimensions s'ajoutent les dimensions temporelle (jour, mois, trimestre, semestre et année) et spatiale (ville, département sanitaire et région sanitaire). L'analyse de ce fait permettra de quantifier les examens réalisés par CDT dans le temps et dans l'espace.

4.2.4. Fait « mobilité patient souffrant de tuberculose »

Le fait « mobilité du patient souffrant de tuberculose » a pour dimensions le patient et le déplacement des patients souffrant de tuberculose (agrégé par trajet). A ces dimensions s'ajoutent les dimensions temporelle (jour, mois et année) et spatiale (village, ville, département sanitaire et région sanitaire). L'analyse de ce fait permettra de suivre les mouvements spatio-temporels des patients souffrant de tuberculose.

5. Conclusion et perspectives

Dans cet article, une analyse de l'état de l'art des SIG appliqués à la surveillance épidémiologique des maladies et de la tuberculose en particulier a été réalisée. Elle a permis de montrer l'inexistence d'un SIG décisionnel intégré pour client SOLAP-mobile de surveillance épidémiologique de la tuberculose. Les résultats cette analyse nous ont permis de proposer une architecture SIG décisionnel client mobile et un modèle préliminaire d'entrepôt de données multidimensionnelles intégré.

Comme perspective, nous allons mener des enquêtes et entretiens semi-directifs auprès des décideurs de la surveillance de la tuberculose afin de raffiner le modèle, intégrer des nouvelles sources de données et spécifier des indicateurs métiers. Dans un second temps, nous passerons à la phase d'implémentation du système.

Ce travail de recherche se réalise dans le cadre du projet « renforcement du système national d'information sanitaire » du Gabon. Il est financé par la Banque Mondiale.

Références

- Benabbou A, Bouamrane K, Hamdadou D. (2013). *Mise en place d'un système d'information spatio-temporel pour la surveillance épidémiologique.*
- Bouba FA. (2016). *Système d'information décisionnel sur les interactions environnement-santé : cas de la Fièvre de la Vallée du Rift au Ferlo (Sénégal).*
- Bédard Y et al., (1997). Spatial olap. 2^e Forum annuel sur R-D, Géomatique VI : Un monde accessible. 13-14 Novembre, 1997, Montréal.

SAGEO'2019

- Moona PK et al., (2004). *Using GIS Technology to identify areas of tuberculosis transmission and incidence.*
- Olascoaga RC, (1982). *Système d'information et de surveillance épidémiologique des maladies vésiculeuses dans les Amériques.*
- Bédard Y et al., (2005). Enrichissement du OLAP pour l'analyse géographique : exemple de réalisations et différentes possibilités technologiques.
- Colby MM et Johnson YJ, (2002). *Potential uses for geographic information system-based planning and decision support technology in intensive food animal production.*
- Green Chris. (2012). *Systèmes d'information géographique et santé publique : avantages et difficultés.*
- Guthrie JL et al., (2016). *Technology and tuberculosis control: the OUT-TB Web experience.*
- Lew WJ et al., (2006). *An Internet-based surveillance system for tuberculosis in Korea.*
- Golfarelli et al., (1998). *Conceptual Design of Data Warehouses from E/R Schemes.*
- Moona PK et al., (2004). *Using GIS Technology to identify areas of tuberculosis transmission and incidence.*
- Parmanto B et al., (2008). *Spatial and multidimensional visualization of Indonesia's village health statistics.*
- Smith CM et Hayward, (2016). *DotMapper: an open source tool for creating interactive disease point map.*
- Younsi FZ, (2016). *Mise en place d'un Système d'Information Décisionnel pour le Suivi et la Prévention des Epidémies.*
- Zemri FA et al., (2015). *Vers un système d'Aide à la Décision Multicritères et Spatiotemporel pour la Surveillance Epidémiologique.*
- Bimonté S et al., (2016). Une architecture orienté services pour l'OLAP Spatial. *In EDA 2016*, Vol. RNTI-B-12, pp.17-24.

Web mapping comme outil support dans la dynamique participative au développement territorial

Line Carelle MONKAM-DJAMBOU¹, Georges KOUAMOU², Chrispin PETTANG¹

1. Laboratoire d'aménagement urbain, Ecole Polytechnique
B.P. 8390 Yaoundé Cameroun

2. Département de Génie Informatique, Ecole Polytechnique
B.P. 8390 Yaoundé Cameroun

RESUME. La construction des plateformes de gestion de développement territorial en Afrique fait face la plupart du temps au manque de moyens matériels et humains de collecte des données pertinentes. Et pour cause les organismes chargés de recueil de données sur le terrain ne sont pas suffisamment pourvus de ressources humaines de qualité et en quantité. Cette insuffisance est accentuée par l'absence d'infrastructures de communication qui facilitent une meilleure accessibilité afin de rendre l'information fiable pour une bonne prise de décision. Grâce au développement technologique, des approches participatives de recueil de données sont proposées à l'instar des instituts nationaux de la statistique qui ont opté pour l'auto-dénombrement comme méthode de recensement.

Cette étude s'inspire de cette démarche pour proposer la construction d'un outil d'aide à la gestion participative du développement territorial au Cameroun. Il s'agit d'une approche collaborative de mise en œuvre d'un SIG (Système d'Informations Géographiques) web interactif faisant intervenir les populations dans la collecte des données, destiné à créer une dynamique participative au développement et à la gestion territoriale.

ABSTRACT. The construction of Territorial Development Management Platforms in Africa mostly faces the lack of material and human resources for collecting relevant data. And because data collection agencies in the field are not sufficiently provided with human resources of quality and quantity. This insufficiency is accentuated by the absence of communication infrastructures that facilitate a better accessibility in order to make the information reliable for a good decision-making. Thanks to technological development, participatory approaches to data collection have been proposed, such as the National Statistical Institutes, which have opted for self-enumeration as a census method.

This study takes advantage on this approach to propose the construction of a tool for participatory management of territorial development in Cameroon. It is a collaborative project for the development of an interactive web-based GIS involving the populations in the collection of data, intended to create a participative dynamic in territorial development and management.

MOTS-CLES : SIG web, développement territorial, collecte participative des données.

KEYWORDS: web GIS, territorial development, interactive, data collection.

1. Introduction

La décentralisation qui impose aux collectivités territoriales de maîtriser leurs espaces pour une meilleure planification des actions est aujourd'hui un facteur clé du développement territorial. L'un des outils à maîtriser pour l'atteinte du développement local est l'information géographique qui s'intègre progressivement dans les systèmes d'information gouvernementaux. Cette information souvent inaccessible dans les PED (Pays En Développement), lorsqu'elle est traitée, stockée et restituée sous forme de carte, constitue un outil efficace pour accompagner les politiques de gestion et d'aménagement des territoires.

Il est donc important de mettre en œuvre des méthodes d'accès et de gestion rapides et fiables des données. On constate que très peu des organismes chargés de recueil de données sur le terrain sont pourvus d'un personnel de qualité et en quantité suffisante. Cette insuffisance est accentuée par l'absence d'infrastructures de communication permettant une bonne accessibilité afin de rendre l'information géographique fiable à même de faciliter une bonne prise de décision. Néanmoins, le développement des plateformes telles que OpenStreetMap et Google Maps ont démocratisé l'accès aux données cartographiques.

En se basant sur les principes d'auto-dénombrement en cours d'exploration dans les instituts de statistique et de crowdsourcing, nous proposons dans cet article une approche de mise en œuvre des web mapping basée sur la participation de la population dans la collecte des données. L'accès aux données collectées par les populations sur place pourra se faire au moyen des API (Application Programming Interface) de Google map. La gestion, le stockage et la restitution des données géographiques nécessitent l'élaboration des SIG en ligne au moyen des plateformes de web mapping. L'étape développée dans le cadre de ce travail est la mise en place d'un outil s'inscrivant dans la dynamique participative du développement territorial.

Dans la suite de cet article, la section 2 présente les exigences de la décentralisation et son impact sur le développement territorial. La section 3 l'incontournable nécessité du recours à l'information géographique. La section 4 les difficultés d'accès à cette information. La section 5 la proposition d'une solution basée sur les atouts technologiques et enfin la section 6 l'importance de cet outil.

2. Les exigences de la décentralisation et du développement territorial

Généralement définie comme le transfert d'autorité et de responsabilités de l'Etat vers des personnes morales de droit public distinctes de lui, la décentralisation rend effective la participation des citoyens à la gestion des affaires locales. En effet, elle impose un partage de responsabilités entre l'Etat et les collectivités territoriales. Les populations doivent rechercher avec beaucoup d'engagement et d'implication, le bien être de toute la communauté. L'amélioration de leurs conditions de vie, la meilleure gestion du patrimoine et du territoire requiert leurs accompagnements dans toutes les démarches de développement. La population doit donc occuper une place importante dans la quête du développement local.

L'apport de la décentralisation dans le développement est de permettre aux communautés locales de contribuer à la conception et à la mise en œuvre des projets et outils afin de réduire leurs souffrances. Le but étant toujours d'offrir aux citoyens des services de qualité.

La réussite d'une démarche de développement territoriale basée sur la décentralisation repose sur la dynamique du partenariat et l'écoute mutuelle entre les autorités étatiques et la population. Les interventions d'appui au développement local se fondent essentiellement sur la gestion concertée d'un territoire en valorisant les projets et initiatives des habitants, les plans de développement des collectivités territoriales, la contractualisation, la participation, toutes choses qui concourent à l'amélioration de la gouvernance locale.

La révélation des ressources locales, leur valorisation et leur spécification par la coordination des acteurs sont les facteurs clés de l'émergence des territoires et du développement local. [Kahina Moussaoui et Khelloudja Arabi, 2014]. Donc le développement territorial basé sur la maîtrise des espaces aménageables, la gestion du patrimoine et l'analyse spatiale du territoire dans l'optique d'une meilleure planification des actions nécessite l'accès aux informations géographiques fiables.

3. L'incontournable nécessité du recours à l'information géographique

En effet, l'information géographique représente un enjeu majeur pour la gestion et l'aménagement des territoires. Elle constitue pour l'Etat et les autorités publiques un moyen essentiel pour l'amélioration, la conception, la mise en œuvre, le suivi et l'évaluation des politiques.

L'appropriation d'un territoire ainsi que des enjeux qu'il soulève nécessite une connaissance exhaustive de ce dernier. L'accès à l'information géographique paraît alors comme une étape sine qua non dans le processus de développement. La maîtrise de l'information d'un territoire apparaît comme une composante essentielle des politiques de développement. [Macodou GUEYE, 2017]. Mener à bien le développement des collectivités locales nécessite la disponibilité régulière des informations fiables et de qualité devant éclairer les prises de décision des dirigeants.

Une population, désireuse d'accéder au développement de son territoire doit donc absolument maîtriser son espace en prenant connaissance des problèmes qui s'y pose. La satisfaction de cette exigence passe par l'accès à l'information géographique.

3.1. Les difficultés d'accès à l'information géographique fiable et de qualité dans les PED

Les PED voient leurs développements freinés car l'accès aux données géographiques y sont difficiles et complexes du fait du faible ratio des routes praticables, du manque de moyens et de personnels fiables.

3.1.1. Problèmes de couverture routière

Les PED font face à une insuffisance des voies de circulation praticables. Le Cameroun par exemple, comptait en 2016 près de 51 350 km de routes principales dont seulement 6 114 km bitumée [MINTP 2015] pour une superficie de 475 650 km² soit environ 13 m de route bitumée pour une surface d'un kilomètre carré. Cette proportion indique clairement que de multiples zones du territoire camerounais sont enclavées.

De plus les routes non bitumées sont presque impraticables en saison de pluie et celles bitumées sont dégradées. La praticabilité des routes existantes est donc réduite ce qui rend difficile l'accès à de multiples zones du territoire et par conséquent à l'information territoriale.

3.1.2. L'insuffisance des ressources

Une autre des faiblesses des PED est l'insuffisance, voire l'absence des structures uniquement chargées de la collecte des données. Au Cameroun par exemple, les instituts chargés de collecte, lorsqu'ils existent, sont rattachés à des ministères et leurs actions sont donc parcellaires et cloisonnées car liées aux projets de ce ministère et non à la collecte générale des données.

L'information est également rendue difficile d'accès au Cameroun car ce pays connaît une exploitation irrationnelle des ressources humaines. Ce qui conduit à un manque de méthodes efficaces de recueil des informations adéquates pouvant être utiles pour des instituts tels que l'Institut Statistique. Loin de là, les enquêtes et les descentes sur le site prennent trop de temps. La couverture du territoire national peut se faire en plusieurs mois ce qui pourrait rendre les premières informations recueillies obsolètes.

Il est donc urgent pour le Cameroun et les autres PED en quête de l'atteinte des ODD de penser à une nouvelle méthode de collecte des données géographiques et de gestion de ces dernières.

4. Proposition d'une solution basée sur les atouts technologiques

4.1. Etat de l'art de la géomatique et du web mapping

Le mot géomatique est issu de la contraction des termes géographie et informatique. La géomatique est donc étroitement liée à l'information géographique. Elle consiste en au moins trois activités distinctes : collecte, traitement et diffusion des données géographiques. Elle a été au centre des travaux de plusieurs chercheurs depuis son apparition en 1960. En effet, aujourd'hui on retrouve plusieurs outils permettant de faire de la géomatique, à l'instar de : ArcGIS d'ESRI (Environmental Systems Research Institute), MapINfo importé par Pitney Bowes... Avec l'avènement du web, on retrouve désormais plus de cartographie en ligne que de cartographie de bureau. Plusieurs outils de web mapping ont été élaborés à l'instar de : My Maps de Google Map, Arcgis online du groupe ESRI, des outils Openstreetmaps et des Openlayers. La cartographie participative ayant été au centre de multiples réflexions, des outils tels que Wikimapia ont été mis sur pieds.

4.2. Présentation de la solution proposée basée sur les atouts du territoire

L'un des piliers de la technologie est le développement du mobile. Le Cameroun en 2013 déjà possédait 15 189 233 abonnés aux opérateurs téléphoniques, [ART 2013] ; pour une population de 21.66 millions soit un ratio de 70 % de la population abonnée à un opérateur. De plus, le marché du mobile connaît une extension considérable ces dernières années, les utilisateurs se servent plus du smartphone ou de la tablette que d'un ordinateur. Ce développement du mobile s'accompagne également du développement des applications pour de multiples usages.

L'autre atout de la technologie est le développement de l'Internet qui influe remarquablement tous les aspects de l'informatique en partant de la structuration des programmes jusqu'au mode d'accès aux applications. L'un des outils touchés par ce développement est le SIG. On quitte des SIG desktop vers des web mapping. La diffusion de données cartographiques par le biais de site web rend dynamique l'accès à l'information puisqu'elle est construite par un traitement déclenché à la demande de l'utilisateur. Son architecture est représentée sur la figure 1 ci-dessous :

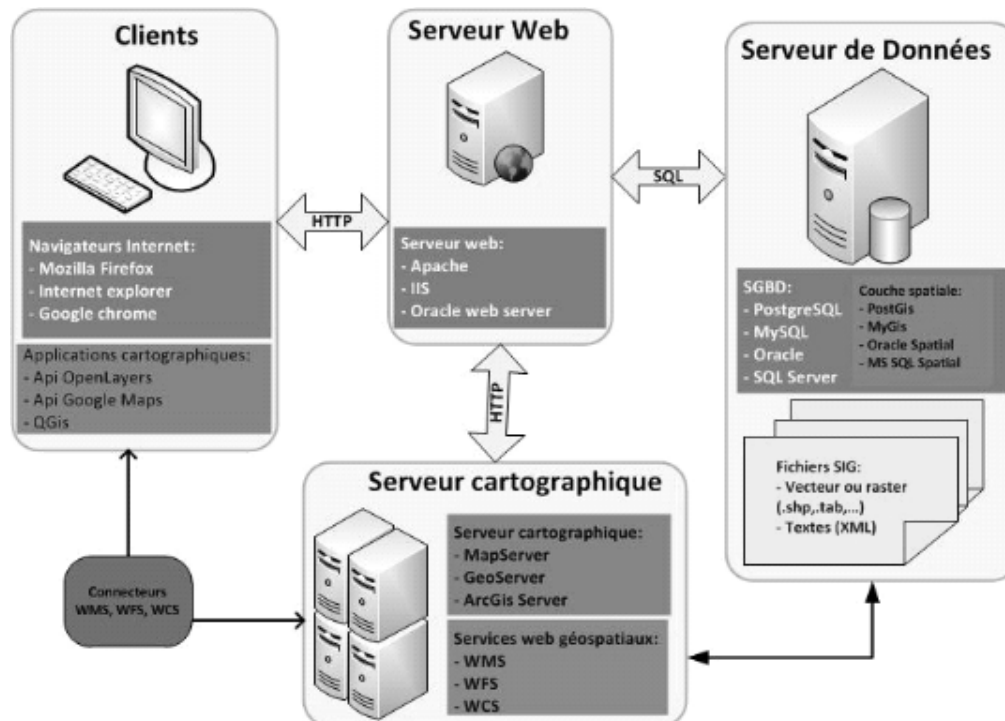


Figure 1 : Architecture d'un web mapping [Frédéric MILLOGO, 2012]

Le web mapping, moyen innovant de créer et de partager des cartes, offre un accès multiple et simultané à une base cartographique avec la possibilité de fédérer plusieurs sources de données. Son ubiquité permet l'hétérogénéité des terminaux et un affichage indépendant de l'appareil utilisé. De plus il est convivial.

Toutefois on constate que dans certaines contrées, la collecte des données par des équipes sur le site s'avère être très coûteuse. L'extraordinaire potentiel de la web-cartographie fascine, mais fait souvent oublier la préexistence, sous des formes plus traditionnelles, tant de l'interaction que de la participation. [Gilles Palsky, 2010].

Pour pallier à ce problème de collecte des données, tout d'abord les statisticiens élaborent des questionnaires sur la base des informations nécessaires aux autorités locales. Ces questionnaires sont ensuite codifiés sous forme de formulaires par les informaticiens. Les données recueillies au moyen des formulaires sont regroupées dans des feuilles Excel, puis vérifiées avant d'être importées dans la base de données cartographique. Elles sont ensuite restituées et donc mises à la disposition des populations dans Google MyMaps qui sert de socle pour le prototype actuel.

My maps, service de web mapping permet de créer des cartes personnalisées en ligne que l'on pourra insérer par la suite sur des sites web en fonction de l'utilisation de ces dernières

Le web mapping conçu à partir de My maps est en effet un SIG web à plusieurs couches parmi lesquelles : la délimitation de la commune de Bangangté importée à partir d'un .kml ; les services (hôpitaux, établissements scolaires, églises) et les sites touristiques obtenues à partir de la base de données géographiques obtenue après le remplissage des formulaires par les populations autochtones, puis représenté sur la carte de base en mode satellite. Pour sa conception, une phase de sensibilisation de la population de la commune de Bangangté a été effectuée. Suite à cela, des formulaires ont été conçus et mis à la disposition des populations au moyen d'un site web réalisé à cet effet. Après la vérification de ces données, elles ont été importées sur la carte réalisée sur MyMaps. Le processus de collecte est décrit sur la figure 2.

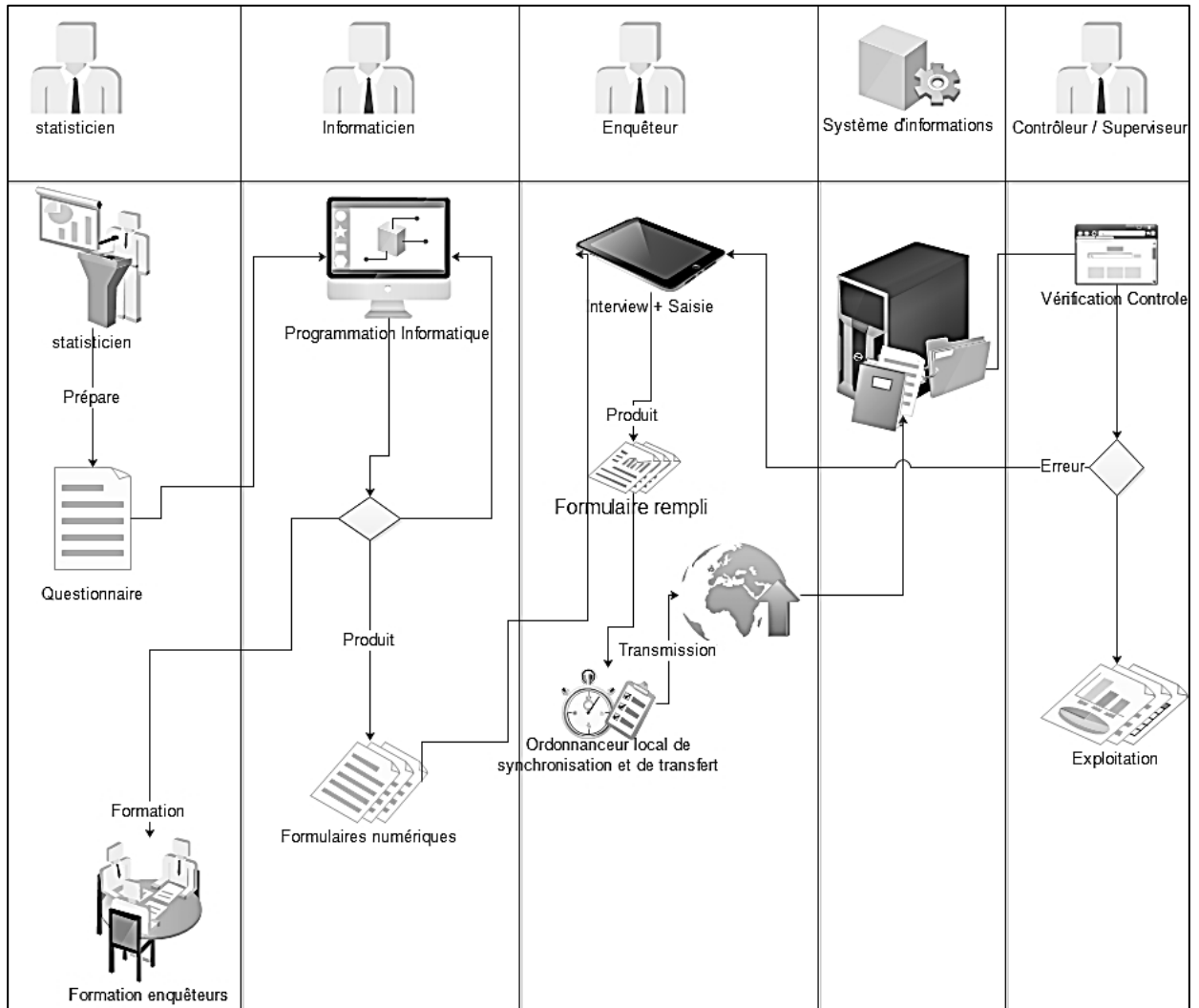


Figure 2. Processus de collecte

L'interface de cette cartographie web est présentée sur la figure 4 suivante

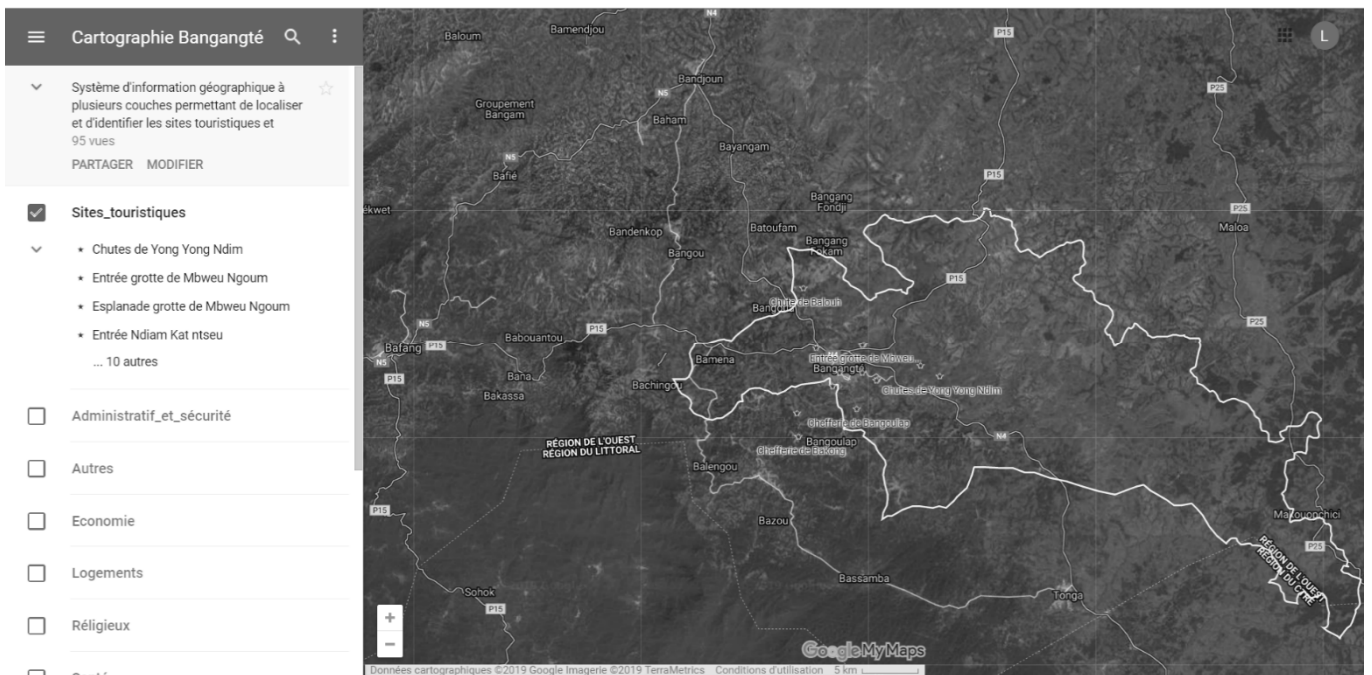


Figure 3 : Interface prototype

Le choix de My Maps comme serveur de cartographie en ligne repose sur le fait qu'il offre une facilité de gestion des couches et permet l'ajout et l'accès aisé aux photographies descriptives du lieu. Un moyen rapide de développement territorial et de mise en valeur des patrimoines locaux (sites touristiques, ...). Un clic sur un lieu enregistré permet d'obtenir des photos descriptives du lieu, ses coordonnées géographiques et sa description. On peut dérouler la liste des images descriptives, elles s'ouvriront dans un nouvel onglet.

Loin de là, My Maps et Google Maps étant des outils d'un seul producteur, il est possible de ramener la carte créée via My maps sur l'application google maps et permettre ainsi aux utilisateurs d'adapter la carte à leurs expériences personnelles et de profiter simultanément des fonctionnalités de google maps et celle de la carte.

5. Importance de cet outil dans le développement participatif du territoire

L'appropriation d'un tel outil et son intégration dans la gouvernance locale passe par une phase de sensibilisation des responsables locaux et des populations. Créer un outil tel que décrit dans ce document permettrait l'implication des populations dans la collecte des données et donc dans la conception des SIG en ligne. Le web mapping conçu permet de transcrire les connaissances locales des populations obtenues au moyen des formulaires sur des cartes virtuelles, utilisées comme instruments interactifs d'apprentissage géospatial et d'aide à la prise de décision. Le web mapping réalisé présente l'avantage qu'il donne les positions des sites touristiques et autres services à toute personne curieuse ne se trouvant pas forcément dans la localité. L'analyse spatiale, l'aménagement du territoire, socles du développement territorial seront donc plus aisés.

Mettre en place un web mapping interactif faisant intervenir les populations dans la collecte des données c'est mettre en place l'un des socles du développement participatif du territoire.

6. Conclusion et perspectives

Ce travail bâti autour de la décentralisation et le développement territorial, l'importance de l'information géographique, la complexité d'accès à cette dernière dans les PED avait pour objectif la mise en place d'un outil informatique basé sur une collecte participative des données géographiques pour rendre plus simple la constitution de la database et accroître la vitesse de développement du territoire. Ce qui nous a le plus préoccupé dans l'échantillon réalisé pour la commune de Bangangté est la mise en place et la gestion de la feuille d'enregistrement des informations recueillies à partir de l'API. Nous nous intéressons pour la suite à l'extension de cet outil au territoire national voire les autres PED. Ceci se fera au moyen des méthodes de sensibilisation des peuples du territoire et à la mise en place d'un portail devant permettre l'importation automatique de la base de données géographiques quotidiennement actualisée et vérifiée.

Bibliographie

- ART, Recueil des données statistiques période de 2012-2013. 80 pages.
- Boris Mericskay, Roche Stéphane. Cartographie et SIG à l'ère du Web 2.0 : Vers une nouvelle génération de SIG participatifs. *Conférence internationale de Géomatique et Analyse Spatiale*, Novembre 2010, France. PP 228-242.
- Frédéric MILLOGO, 2012. *Mise en place d'une application web mapping de géolocalisation des points d'intérêt de la ville de Ouagadougou*, Mémoire de Master, Université de Douala et de l'AUF.
- Giacomo Rambaldi : *Cartographes aux pieds nus et SIG participatifs* SIG participatifs. N° 27, septembre 2005. 01 page
- Gilles Palsky : *Cartes participatives, Cartes collaboratives : La cartographie comme maïeutique*. CFC N°305. Septembre 2010. 11 pages
- Henri Pomon (2015) : *SIG : La dimension géographique du système d'information*. 2e édition DUNOD Paris, 2015.
- Kahina Moussaoui et Khelloudja Arabi (2014) : *Le développement des territoires dans un contexte de restructuration de l'Etat*. CIREC-Canada.
- Karima Boudedja : Les acteurs et le développement local : outils et représentations. Cas des territoires ruraux au Maghreb. Géographie. Université Paul Valéry – Montpellier III, 2013. Français. 390 pages
- Macodou GUEYE : « Maîtrise de l'information territoriale : un impératif pour optimiser l'efficacité et la performance des collectivités locales. » Octobre 2017. 01 page.
- MINTP, Annuaire statistique du secteur des infrastructures. Edition 2015 PP.17. 215 pages.
- Thomas PIERGIOVANNI 2012 : « Système d'Information Géographique et développement local : la réalisation du diagnostic territorial du premier arrondissement de la ville de Moundou au Tchad » Août 2012. 29 pages

Un modèle statique de risque d'incendie de forêt induit et subi.

Eric Maillé¹, Cécille Duveillier¹

1. RECOVER/EMR, Irstea Aix en Provence
CS40061, F-13182 Aix en Provence, France
eric.maill@irstea.fr

RESUME. De nombreux risques naturels ou technologiques sont issus de la propagation d'un danger depuis une zone source, vers une zone d'enjeux menacés. La zone source induit un risque sur les zones d'enjeux qui le subissent. Le risque d'incendie de forêt a la particularité de présenter tant des zones sources, les zones d'ignition potentielle, que d'enjeux (zones péri-urbaines, zones forestières elles-mêmes), de géométrie surfaciques. Par ailleurs, celles-ci sont parfois confondues. Nous proposons un modèle spatial d'évaluation des risques d'incendie de forêt induit et subi. Bien que statique, celui-ci repose sur les principes de la propagation du feu dans les zones combustibles, utilisés notamment dans les modèles dynamiques de comportement du feu. A l'état de prototype pour la recherche, le modèle se fonde sur la représentation de la propagation dans l'espace géographique des différents potentiels qui constituent les composantes cumulatives du risque, en particulier le potentiel d'occurrence mais aussi la vulnérabilité. Le modèle ne préjuge toutefois pas des modes d'évaluation de ces grandeurs. En ce sens, il s'agit d'un cadre nécessitant des développements expérimentaux pour atteindre un niveau d'opérationnalité permettant son transfert auprès des gestionnaires du risque d'incendie.

ABSTRACT. Many natural or technological risks stem from the propagation of the danger from a source toward some threatened stakes. The source induces a risk on the stakes, which are subjected to that risk. The particularity of the forest fire risk is that both sources and stakes have a surface geometry. Moreover, these are sometimes mingled. We propose a spatial model designed to assess both induced and subjected risks. Although static, the model is based on the principles of fire propagation used in dynamic fire behaviors models. The model is a research prototype, based on the representation of spatial spread of cumulative components of the risk, especially the potential of occurrence potential and also the vulnerability. However, the model does not state on the modalities to assess these components. In that way, it is a framework that requires further experimental developments in order to reach an operational level that will allow transferring it to decision makers in charge of forest fire risk management.

MOTS-CLES : Risques naturels, incendie de forêt, risque induit, risque subi, aléa, vulnérabilité.
KEYWORDS: Natural risk, forest fire, induced risk, subjected risk, hazard, vulnerability

1. Introduction

La zone Européenne à risque d'incendie de forêt progresse tant en latitude qu'en altitude sous l'effet du changement climatique (Chatry 2010, Barbero & *al.* 2015), et des changements de l'occupation des sols (Maillé & *al.* 2014). Des événements de grande magnitude sont à attendre durant les prochaines années (Tédim 2019). La gestion du risque d'incendie de forêt fondée sur des outils d'évaluation pour l'aide à la décision est appelée à devenir un enjeu de plus en plus prégnant de la gestion des territoires locaux (Finney 2005).

Nous proposons de formaliser l'évaluation des risques d'incendie *induit* et *subi* au moyen d'un modèle formel implémenté. Nous précisons le cadre conceptuel utilisé dans la section 2 de cet article, puis, dans la section 3, nous explicitons les principes du modèle. Dans la section 4 nous proposons une première application du modèle au feu de Rognac de 2016.

2. Le cadre conceptuel de modélisation

Le cadre conceptuel de modélisation utilisé est fondé sur une définition du risque à deux composantes uniquement : l'aléa et la vulnérabilité (Hardy 2005, Haight *et al.*, 2004) (figure 1).

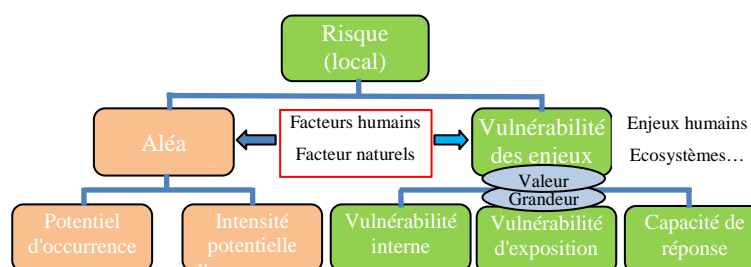


FIGURE 1. Le schéma conceptuel de risque utilisé

2.1. La vulnérabilité

La vulnérabilité est un attribut des classes d'« enjeux ». Elle représente le potentiel de dommages sur l'enjeu en cas de réalisation d'un aléa (Blanchi & *al.* 2011). La vulnérabilité d'un enjeu a trois composantes : interne, d'exposition, et proactive (Pugnet & Maillé 2013). Un coefficient de « grandeur » représente l'importance d'une instance d'enjeu par rapport aux autres instances de la même classe (par exemple, la fréquentation, en nombre de visites par jour, d'un ERP). Un coefficient de *valeur* doit aussi être associé à chacune des classes d'enjeux : il représente la valeur (économique, patrimoniale, paysagère, éthique, écosystémique, etc.) assignée à la classe d'enjeux par la société.

On propose, à des fins conceptuelles mais sans objectif d'évaluation, la formulation suivante pour le calcul de vulnérabilité V d'une instance d'enjeu i de classe ci :

$$V_i = kValeur_{ci} * kGrandeur_i * [VInterne_i * Exposition_i * (1 - CapacitéReponse_i)]$$

La vulnérabilité d'une portion d'espace (un pixel, ou un territoire communal, par exemple) est la somme des vulnérabilités des enjeux (ou des fractions d'enjeux, pour les enjeux linéaires ou surfaciques) co-localisés sur cet espace.

2.2. L'aléa

L'aléa a deux composantes : l'intensité potentielle locale du feu et le potentiel d'occurrence (Hardy 2005). Le calcul de l'intensité potentielle locale d'un feu utilise l'équation Byram (Byram 1959, Alexander 1982). Celle-ci prend en entrée le taux de propagation (rate of spread – ROS), dont les moyennes locales peuvent être statiquement évaluées en fonction du combustible et des conditions météorologiques locales instantanées. Pour cela on utilise des conditions météorologiques de « références » et les cartes de végétation.

Le potentiel d'occurrence locale est déterminé par la probabilité d'éclosion d'un incendie à cette localisation (Catry et *al.*, 2009) et la probabilité qu'un incendie s'y propage depuis une localisation voisine. C'est donc cette grandeur propagée qui justifie la distinction entre le risque induit et le risque subi. La propagation du potentiel d'occurrence est simulée en utilisant des équations inspirées de celles de la dynamique de propagation du feu, mais sans la dimension temporelle (modélisation statique).

De même que pour la vulnérabilité, l'aléa est nul si la probabilité d'occurrence ou l'intensité potentielle locale est nulle. Enfin, le risque est défini comme le produit des estimateurs normalisés (entre leurs minima et maxima respectifs sur la zone d'étude) de l'aléa et de la vulnérabilité : il est donc nul si la vulnérabilité ou l'aléa est nul.

2.3. Quantification

Les grandeurs ci-dessus (intensité potentielle locale, potentiel d'occurrence, vulnérabilité, etc.) sont estimées par des "modèles métiers". On choisit de normaliser toutes les grandeurs sur l'espace de simulation étudié. On appelle « potentiels » les valeurs normalisées entre 0 et 1, sur la zone d'étude, des sorties des modèles métiers spécialisés. Le modèle obtenu n'est donc pas générique, mais valable uniquement sur la zone où il est spécifié. Il ne permet que de hiérarchiser les localisations de l'espace par rapport aux niveaux de risques (induits et subis), sans permettre d'en calculer une valeur absolue. La normalisation introduit également des biais liés à la non-linéarité des "modèles métiers".

3. Propagation du potentiel d'occurrence

Afin de calculer les risques induit et subi locaux pour chaque localisation du territoire étudié, un potentiel d'occurrence est « propagé » au travers d'une grille raster.

3.1. Propagation locale

Le potentiel d'occurrence dû à la propagation du feu venant des localisations voisines est calculée au moyen d'un modèle de propagation locale du feu très simple, dont les principaux facteurs sont, comme pour un modèle dynamique de propagation du feu, la charge combustible de propagation locale, la pente locale et le vent local, étant donné un vent général sur la zone simulée dans des conditions de référence (Pugnet *et al.*, 2013).

3.2. Propagation à longue distance des potentiels d'occurrence

On procède au calcul intégral des potentiels d'occurrence pour toutes les localisations de la zone étudiée, en fonction du potentiel d'éclosion de l'ensemble des localisations et du potentiel de propagation cumulée jusqu'aux localisations cibles. Donc pour chaque localisation le potentiel d'éclosion, puis le potentiel de propagation cumulé jusqu'à l'ensemble des autres localisations est calculé.

Le parcours de la grille suit une géométrie centrifuge isotrope. S'agissant d'un modèle statique pour simuler la propagation d'un potentiel d'occurrence et non d'un feu, la géométrie de propagation implémentée est uniquement calculatoire, sans impact notable sur la forme de la distribution spatiale du potentiel obtenue. On choisit donc une géométrie de propagation déterminée par la forme du voisinage de la grille raster, de forme carrée.

3.2. Cumul des grandeurs composant le risque lors de la propagation

Le potentiel d'occurrence obtenu pour chaque localisation permet le calcul des différentes composantes du risque, dont l'aléa (en multipliant le potentiel d'occurrence par l'intensité potentielle locale normalisée) et la vulnérabilité locale. Ces grandeurs sont sommées de façon intégrale sur toute la grille, puis affectées à la localisation source (pour le risque induit) ou cible (pour le risque subi) selon des modalités un peu différentes dans les deux cas.

Par exemple le risque induit somme les vulnérabilités de toutes les localisations susceptibles d'être affectées par l'aléa, alors que le risque subi ne dépend que de la vulnérabilité locale. De même pour l'aléa induit qui somme les intensités potentielles maximales des localisations potentiellement atteintes, alors que l'aléa

subi ne dépend que de l'intensité potentielle maximale et du potentiel d'occurrence locaux.

4. Une application au feu de Rognac de 2016

Le feu périurbain de Rognac du 10 août 2016 a brûlé 2655ha dans la région de Marseille (Bouches-du-Rhône). Il a affecté 130 bâtiments et a détruit 36 maisons. Le modèle développé a été appliqué sur la zone du feu de Rognac, à des fins d'illustration de son fonctionnement. Le poster illustre quelques exemples de sorties produites, ainsi que les cartes finales des risques induit et subi.

Les zones à fort risque induit sont principalement les zones à potentiel de propagation fort en amont des zones à enjeux, dans la direction du vent dominant (ici le vent du nord-ouest – Mistral). Les zones à risque subi fort sont plutôt les zones à enjeux elles-mêmes, même si celles-ci ne présentent pas un aléa subi très important. Les résultats produits par le prototype non finement calibré, confirment surtout la nécessité d'appréhender les deux types de risques, induits et subis, leur cartes respectives présentant des patterns spatiaux très différents.

5. Conclusion

Le prototype de modèle proposé démontre la possibilité de formalisation de concepts d'experts très utilisés par les décisionnaires opérationnels tels que le sont les risques subi et induit. Sur les zones à fort risque induit, les gestionnaires devront prioritairement mener des actions visant à limiter le potentiel d'éclosion. Sur les zones à fort risque subi, ils devront insister sur la limitation de la vulnérabilité, la protection, les moyens de lutte.

Toutefois, en l'état actuel de développement du prototype, plusieurs limites à son utilisation opérationnelle existent :

- Le prototype ne permet jamais de produire des indices absolus, mais seulement de représenter leur distribution spatiale relative sur le territoire. Les cartes obtenues sur des territoires différents ne sont pas comparables en valeur absolue.
- Le modèle doit être validé par des analyses statistiques sur un échantillon de feux suffisant. Un seul (ou un très petit nombre) de points d'éclosion par feu peut être testé pour évaluer son risque induit, et seul un petit nombre de points de dommage (sur des enjeux anthropiques) peuvent être testés pour évaluer leur risque subi. La validation complète du modèle nécessite donc un grand nombre de données sur les feux réels pour l'instant encore non disponibles.

Au-delà de son utilisation opérationnelle, le modèle peut aussi avoir une utilisation de médiation entre les différents gestionnaires intervenant dans la gestion du risque d'incendie de forêt. Il doit notamment aider à une meilleure coordination d'acteurs plus impliqués dans la prévention du risque induit, en particulier les aménageurs des territoires, avec ceux plus impliqués dans la protection contre le risque subi, en particulier les services de lutte.

Bibliographie

- Alexander M.E. (1982). Calculating and interpreting forest fire intensities. *Canadian Journal of Botany*, vol. 60, pp.349-357.
- Barbero R., Abatzoglou J. T., Kolden C. A., Hegewisch K. C., Larkin N. K. and Podschwit H., (2015). Multi-scalar influence of weather and climate on very large-fires in the Eastern United States. *Int. J. Climatol*, 35: 2180–2186. doi:10.1002/joc.4090
- Blanchi R., Leonard J., Leicester R., Lipkin F., Boulaire F. et McNamara, C., (2011). Assessing vulnerability at the urban interface. *The 5th International Wildland Fire Conference*, Sun City, South Africa.
- Byram G.M. (1959). Combustion of forest fuels. *Forest fire control and use*, Davis, K.P. (Ed.) New York, McGraw-Hill.
- Catry F. X., Rego F. C., Bação F. L., Moreira F., (2009). Modeling and mapping wildfire ignition risk in Portugal. *International Journal of Wildland Fire* 18(8) 921-931 <https://doi.org/10.1071/WF07123>
- Chatry C. Le Gallou J.Y., Le Quentrec M., Lafitte J.J., Laurens D., Creuchet et Grelu J., (2010). Rapport de la mission interministérielle changement climatique et extension des zones sensibles aux feux de forêts.- 189p
- Finney M. A. (2005). The challenge of quantitative risk analysis for wildland fire. *Forest Ecology and Management* 211, pp. 97–108.
- Ganteaume A., Camia A., Jappiot M., San Miguel-Ayanz J., Long-Fournel M. (2013). A review of the main driving factors of forest fire ignition over Europe. *Environmental Management* 51, 2013.
- Haight R. G., Cleland D. T., Hammer R. B., Radeloff V. C., Rupp T. S. (2004). Assessing fire risk in wildland-urban interface. *J. For.* 102(7):41-48
- Hardy C. C. (2005). Wildland fire hazard and risk: problems, definitions, and context. *Forest Ecology and Management* 211, pp. 73–82.
- Maillé E., Bellet A., Long-Founel M., Pugnet L., Morge D., (2014). Dynamique d'interface rural-urbain et du risque d'incendie de forêt. Simulations sur la Basse Provence. *Conférence SAGEO2014, 24-27 novembre 2014, Grenoble*. DOI:10.3199/JESA.45.1-n © Lavoisier 2012
- Pugnet, L. et Maillé, E. (2013). Assessment of wildland–urban interface vulnerability to fire using a multi-criteria analysis. *International conference on forest fire risk modelling and mapping - Vulnerability to forest fire at wildland-urban interfaces*, 30th of September, 1st and 2nd of October 2013, Aix en Provence, France.
- Pugnet, L., Chong, D.M., Duff, T.J. et Tolhurst, K.G. (2013). Wildland–urban interface (WUI) fire modelling using PHOENIX Rapidfire: A case study in Cavaillon, France. *20th International Congress on Modelling and Simulation*, Adelaide, Australia, 1-6 December 2013. www.mssanz.org.au/modsim2013
- Tedim F., (2019). Extreme wildfires events: an escalating worldwide problem. *6th International Conference on Fire Behavior and Fuels, 29th of April-3rd of May 2019, Marseille, France*.

Une technique de spécification abstraite pour la publication de géodonnées liées

Claudine Métral¹, Gilles Falquet¹

1. Centre universitaire d'informatique, Université de Genève
7 route de Drize, CH-1227 Carouge, Suisse
prenom.nom@unige.ch

RÉSUMÉ. La publication de géodonnées liées pour le Web sémantique peut impliquer des opérations relativement complexes pour transformer les données sources vers leur version publiée et pour relier celles-ci à d'autres ensembles de données. Dans cet article, nous proposons un formalisme basé sur des règles de transformations de graphes RDF pour spécifier abstraitement la génération de géodonnées RDF liées. Ce formalisme a été testé sur un cas d'utilisation, dans le domaine des cyclones tropicaux, pour montrer qu'il est suffisant pour exprimer des opérations de publication et de liaison complexes.

ABSTRACT. The publication of linked geodata for the Semantic Web may involve relatively complex operations to map source data to their published version and to link these ones to other datasets. In this paper we propose a formalism, based on RDF graph transformation rules, to abstractly specify the generation of RDF linked geodata. The formalism has been tested on a use case in the domain of tropical cyclones to show that it is sufficient to express complex publication and linking operations.

MOTS-CLÉS : Géodonnées liées, Web sémantique, RDF, SPARQL, Métadonnées de publication

KEYWORDS: Linked geodata, Semantic Web, RDF, SPARQL, Publication metadata

1. Introduction

Le Web sémantique est en train de devenir un moyen pratique pour rendre les données, et en particulier les données géoréférencées, disponibles pour un large éventail d'applications. La publication de données sur le Web sémantique consiste (1) à transformer les données source en un graphe RDF, c'est-à-dire un ensemble de triplets (*sujet, prédicat, objet*), et (2) s'assurer que ce graphe est relié à d'autres graphes de manière à réutiliser les données existantes. Plusieurs méthodologies ou éléments méthodologiques ont été développés pour aider à la réalisation de ce processus de publication. Mais la plupart des méthodologies existantes ne proposent

pas de moyen de documenter formellement le processus de publication avec pour conséquence la perte de l'information de provenance. De plus, le processus est difficile à répéter.

Nous proposons un modèle de métadonnées pour représenter (1) les transformations qui ont produit les données publiées à partir des sources disponibles, (2) comment les entités de différentes sources ont été appariées et reliées. Dans cet article, nous présentons le modèle développé et un cas d'utilisation lié aux cyclones tropicaux, sur lequel le modèle a été testé.

1.1 Contexte et travaux connexes

Des méthodologies ont été proposées pour la publication de données ou de géodonnées sur le Web sémantique. Par exemple, Vilches-Blázquez et al (2014) proposent un processus de publication qui comprend les étapes suivantes :

1. la spécification des besoins (identification et analyse des sources de données géospatiales, création des IRI) ;
2. la modélisation ontologique (choix ou création d'une ontologie qui définit le vocabulaire RDF à utiliser pour la publication des données) ;
3. la génération du RDF ;
4. la génération de liens vers d'autres ensembles de données ;
5. la publication et l'exploitation des données et des métadonnées.

La génération du RDF peut se faire avec des outils permettant de définir les transformations, exprimées selon une structure et un vocabulaire cible choisis par l'utilisateur. Celui-ci peut aussi utiliser le vocabulaire défini dans des ontologies. Les outils de transformation dépendent du type des données source. On peut citer RML (RDF Mapping Language) pour des données CSV, XML ou JSON (Dimou et Vander Sande, 2014), ou R2RML (RDB to RDF mapping language) pour des données issues de bases de données relationnelles (Das *et al.*, 2012).

Il est crucial de noter que dans de nombreux cas, la publication d'un ensemble de données pour le Web sémantique ne se réduit pas à une traduction directe d'un modèle de données à RDF mais nécessite des transformations plus sophistiquées. En effet, les données publiées doivent être adaptées aux usages envisagés (d'autres usages, non prévus, peuvent apparaître par la suite). Ces transformations peuvent inclure des changements d'unité ou d'échelle, des simplifications, des agrégations, le pré-calcul de certaines valeurs (pour éviter leur re-calcul par les utilisateurs), etc.

Des formalismes de transformation de modèles ou de graphes ont été publiés dans le domaine du génie logiciel, en particulier pour le *MDD-Model Driven Development* (Czarnecki et Helsen, 2016). Des études théoriques formelles ont également été réalisées sur les formalismes de type grammaire de graphes (Ehrig *et al.*, 2006). Dans le cadre du Web sémantique, plusieurs auteurs (Huang et Harrie, 2019), (Scheider *et al.*, 2019) (Métral et Falquet, 2014) ont remarqué que le langage de requêtes SPARQL sur les graphes RDF, avec quelques extensions, fournit une alternative, en général plus simple et lisible, que les formalismes précédents.

1.2 Cas d'utilisation

Il existe actuellement deux sources de données qui constituent des ressources majeures pour les cyclones : (1) la plateforme PREVIEW¹, qui vise à fournir des données spatiales sur les risques mondiaux liés à divers aléas naturels, dont les cyclones tropicaux, (Giuliani & Peduzzi 2011), (2) la base de données EM-DAT² qui donne accès aux données portant sur plus de 22.000 catastrophes majeures dans le monde, de 1900 à nos jours. Bien que cette base de données donne accès à de nombreuses statistiques sur les pertes humaines et financières, les événements et les chiffres d'EM-DAT ne sont pas géoréférencés avec précision car les seules informations spatiales disponibles sont les limites administratives. Ces deux sources de données sont complémentaires et pourraient bénéficier d'une interconnexion.

L'ensemble de données PREVIEW que nous avons utilisé comprend l'évolution de l'étendue spatiale des cyclones tropicaux entre 1970 et 2014, modélisés sous forme de buffers. Chaque événement distinct est associé à des propriétés générales (nom, date, ...) et à un ou plusieurs buffers décrivant les différentes étapes de son évolution. Chaque buffer est lui-même associé à la surface d'un pays touché ou à la surface d'une mer, ainsi qu'à une catégorie Saffir-Simpson. L'objectif est, entre autres, de publier des données sur les cyclones sous la forme de cartes, comme l'illustre la figure 1.



Figure 1 : Carte de l'ouragan Katrina en 2005

2. Spécification du processus de publication

L'approche que nous présentons ici concerne le point 3 du processus de Vilches-Blázquez et al., c'est-à-dire la génération du graphe RDF qui sera publié. Dans le modèle que nous proposons la génération du graphe publié à partir d'un ensemble de

¹ PREVIEW Global Risk Data Platform <http://preview.grid.unep.ch>

² Emergency Events Database : <http://www.emdat.be>

sources de données $\{S_1, \dots, S_n\}$ est le résultat de la composition de quatre transformations :

$$\{S_1, \dots, S_n\} \xrightarrow{rdf} \{G_1, \dots, G_n\} \xrightarrow{\cup} G_{src} \xrightarrow{\text{génération}} G_{pub0} \xrightarrow{\text{liaisons et équivalences}} G_{pub}$$

La première étape (rdf) a pour but de transformer les sources pour produire des graphes RDF à partir des données sources (exprimées selon différents modèles). Elle peut être spécifiée dans l'un des formalismes décrit en 1.1. On prend ensuite l'union de ces graphes pour produire le graphe source G_{src} .

Le but de la génération est de produire un graphe RDF qui utilise le vocabulaire de publication pour les noms de classes et de propriétés. De plus cette transformation doit créer les IRI des nouveaux nœuds (donc générer des IRI) et calculer les valeurs des nouveaux nœuds littéraux à partir de valeurs du graphe source (par exemple avec des opérations spatiales)

Cette transformation est spécifiée par un ensemble de règles de transformation de la forme [**source:** m_s , **group:** g , **IRI:** i , **val:** v , **pub:** m_p] où

- m_s est un motif de graphe tel que défini dans le langage SPARQL.
- g est un ensemble de variables de regroupement.
- i est un ensemble d'expressions de la forme $y_i = IRI(x_{i1}, \dots, x_{ik})$ où les x_i sont des variables du motif m_s .
- v est un ensemble d'expressions de la forme $z_j = E(x_{j1}, \dots, x_{jl})$ où E est une expression construite à l'aide de fonctions arithmétiques et d'opérateurs spatiaux.
- m_p est un ensemble de motifs de triplets contenant uniquement des variables apparaissant dans m_s , i ou v .

La sémantique d'une telle règle peut être décrite par sa transformation en SPARQL. Étant donné un graphe source G_{src} l'interprétation d'une règle est l'ensemble des triplet produits par l'évaluation de l'expression

```
construct  $m_p$  where {  $m_s$ 
    bind( $IRI_1(x_{i1}, \dots, x_{ik})$  as  $y_1$ ) ...
    bind( $E(x_{j1}, \dots, x_{jl})$  as  $z_j$ ) ... }
group by  $g$ 
```

sur le graphe G_{src} . Les IRI_k sont des fonctions injectives qui génèrent des IRI à partir de nœuds (IRI ou littéraux). Elles servent à construire les nœuds du graphe de publication qui n'existent pas dans le graphe source. Les expressions $E(x_{j1}, \dots, x_{jl})$ servent à spécifier le calcul des objets géométriques, représentés par des littéraux au format WKT ou autre.

Les transformations d'équivalence et liaison ont pour but d'identifier les nœuds du graphe qui correspondent à une même entité géographique et de lier les nœuds provenant de différentes sources entre eux ou avec des d'autres données publiées sur le Web. La correspondance et la liaison peuvent être traitées de manière semblable

puisque'il s'agit dans les deux cas de créer une relation (d'équivalence ou autre) entre deux objets. La spécification est formée de règles qui sont des paires [**condition:** $c(x,y)$, **relation:** $x r y$]. Une telle règle spécifie qu'un triplet $x r y$ doit être généré si la condition c est vraie pour x et y . La condition est une conjonction d'un motif sur le graphe de publication et d'expressions booléennes sur x , y et les variables du motif.

3. Application au cas d'utilisation

Nous présentons ci-dessous deux exemples d'utilisation du formalisme sur le cas d'utilisation.

L'un des objectifs était de publier pour chaque cyclone un buffer par degré d'intensité. Ce buffer peut être obtenu en fusionnant tous les buffers (un par pays) pour une intensité donnée trouvés dans la source PREVIEW. Dans le formalisme proposé cette transformation peut être exprimée comme suit :

```

source: e a pv:Event. e pv:hasBuffer pb. pb pv:SSCategory cat.
           pb pv:hasGeometry/asWKT pg.
           // pour chaque événement e on sélectionne ses buffers, avec pour chacun sa
           // catégorie cat et sa géométrie pg en WKT
group: e, cat
IRI:    c = IRI(e); b = IRI(e, cat) ; g = IRI(e, cat);
           // création des IRI des nouveaux nœuds à partir des IRI de nœuds source
val:    w = toWKT(Simplify(Merge(pg))) // fusion et simplification des polygones
pub:    c a sweet:TropicalCyclone. c tc:hasBuffer b.
           b sweet:SSScale cat. b geo:hasGeometry g. g geo:asWKT w.
           // association de chaque cyclone c à un buffer avec sa géométrie et sa catégorie
    
```

Un autre objectif étant de publier les informations sur les cyclones issues de PREVIEW enrichies des données d'EMDAT, il était nécessaire de faire correspondre les entités cycloniques des deux sources. Dans ce cas, le processus de correspondance était basé sur le nom, l'année, le mois de début et les propriétés du pays affecté.

```

condition:
           {tc a sweet:TropicalCyclone ; tc:name tcName.
           h a yago:Hurricane111467018; dpb:name hName};
           Similar(tcName, hName)
relation: tc owl:sameAs h
    
```

5. Conclusion et perspectives

Nous avons présenté un formalisme qui contribue à définir des descriptions abstraites des processus de publications de données liées à partir de source de géodonnées. Par conséquent, les utilisateurs de ces données peuvent avoir une compréhension de haut niveau de la façon dont les données publiées ont été obtenues, et donc de leur sémantique, sans avoir besoin de se référer à la documentation spécifique aux outils techniques utilisés.

Ces spécifications peuvent facilement être publiées sous forme de métadonnées RDF qui accompagnent les données liées publiées. Les règles de correspondance et de liaison sont liées à une ontologie des opérations spatiales. Par conséquent, certaines propriétés des données publiées (en termes de provenance, de précision ou de perte d'information) peuvent être déduites par raisonnement sur les opérations impliquées dans les règles. L'expérience que nous avons menée sur un cas d'utilisation a montré que cette technique de spécification est suffisamment expressive pour décrire de manière compacte un processus de publication relativement complexe.

La proximité entre ce formalisme et le langage (geo)SPARQL fait qu'il est aisé de passer de la spécification à une implémentation opérationnelle, pour autant qu'on écrive les fonction de génération d'IRI et de correspondance nécessaires.

Bibliographie

- Czarnecki K., Helsen S. (2006). Feature-based survey of model transformation approaches. *IBM Systems Journal*, 45(3), 621–645.
- Das S., Sundara S., Cyganiak R. (Eds) (2012). R2RML: RDB to RDF Mapping Language - W3C Recommendation 27 September 2012, <https://www.w3.org/TR/r2rml/>.
- Dimou A., Vander Sande M. (Eds) (2014). RDF Mapping Language (RML) - Unofficial Draft 17 September 2014, <http://rml.io/spec.html>.
- Ehrig H., Ehrig K., Prange U., Taentzer G. (2006). *Fundamentals of Algebraic Graph Transformation*. Springer.
- Giuliani G., Peduzzi P. (2011). The PREVIEW Global Risk Data Platform : a geoportal to serve and share global data on risk to natural hazards. *Natural Hazards and Earth System Science*, 11 (1):53-66.
- Hofer B., Mäs S., Brauner J., Bernard L. (2017). Towards a knowledge base to support geoprocessing workflow development. *International Journal of Geographical Information Science*, 31(4), 694–716.
- Huang W., Harrie L. (2019). Towards knowledge-based geovisualisation using Semantic Web technologies: a knowledge representation approach coupling ontologies and rules. *International Journal of Digital Earth*, 1–22.
- Knublauch, H. (2013). *SPIN - Syntaxe SPARQL*. <http://spinrdf.org/sp.html>.
- Knublauch H., Kontokostas D. (Eds) (2017). *Shapes Constraint Language (SHACL) – W3C Recommendation 20 July 2017*, <https://www.w3.org/TR/shacl/>.
- Métral C., Falquet G. (2014) Prototyping Information Visualization in 3D City Models: a Model-based Approach. 3DGeoInfo 2014 Conference, Dubai, 12-13 November 2014
- Scheider S., Ballatore A., Lemmens R. (2019). Finding and sharing GIS methods based on the questions they answer. *International Journal of Digital Earth*, 12(5), 594–613.
- Vilches-Blázquez L.M., Villazón-Terrazas B., Corcho O., Gómez-Pérez A. (2014). Integrating geographical information in the Linked Digital Earth. *International Journal of Digital Earth*, 7 (7):554-575.

Approche ontologique pour la modélisation de carte contextuelle

Bilal Koteich¹, Eric Saux¹, Arnaud Grancher²

1. Ecole navale, Institut de Recherche de l'Ecole Navale (IRENav), BCRM
Brest, Lanvéoc-Poulmic, CC 600, 29240 BREST Cedex 9 (France)

bilal.koteich@ecole-navale.fr,eric.saux@ecole-navale.fr

2. Géomaticien, Freelance

agrancher@gmail.com

RÉSUMÉ. Cet article présente tout d'abord un état de l'art des approches permettant la modélisation du contexte dans un système d'information, d'un point de vue général. Nous proposons ensuite une approche fondée sur une ontologie spatio-temporelle afin d'automatiser l'étape de la sélection des données dans un contexte plus spécifique de confection de cartes personnalisées. L'approche proposée tient compte du profil de l'utilisateur et de ses préférences, ainsi que des événements afférents à un environnement. Nous souhaitons ensuite, par raisonnement logique, pouvoir en déduire une sélection de catégories et d'objets qui contribuera à la confection d'une carte personnalisée.

ABSTRACT. This paper first presents a state of the art of context modeling approaches in an information system from a general point of view. We then propose an approach based on a spatial-temporal ontology to automate the data selection step in the process of making personalized maps. The proposed approach of context modeling takes into consideration the user's profile and preferences, as well as events related to an environment. We then aim, by logical reasoning, to be able to deduce a selection of categories and objects that will contribute to the creation of a personalized map.

MOTS-CLÉS : Ontologie, Modélisation du contexte, Cartographie, Raisonnement spatio-temporel

KEYWORDS: Ontology, Context Modeling, Mapping, Spatio-Temporal Reasoning

1. Introduction

Depuis l'émergence des technologies sans fil et des appareils mobiles, une nouvelle classe d'applications appelées services sensibles au contexte a suscité

SAGEO'2019 - Clermont Ferrand, 13-15 Novembre 2019

un vif intérêt dans le domaine de l'informatique ubiquitaire. L'utilisation de tels services dans le domaine de la cartographie numérique nous paraît cruciale. Une information peut-être utile dans un certain contexte d'utilisation et inutile dans un autre, d'où la problématique de la sélection adéquate des informations. Ces problématiques sont au cœur du processus de fabrication de la carte à la demande.

Afin de réduire les coûts liés à la production de cartes personnalisées, une solution consiste à automatiser certains processus entrant dans la confection de ces cartes. La conception d'un dispositif de carte à la demande est un domaine de recherche pluridisciplinaire dont le but est d'élaborer des mécanismes qui soient capables, sans assistance humaine, de récolter un ensemble d'exigences utilisateur et de les interpréter pour confectionner une carte personnalisée.

Nous nous intéressons à proposer un modèle de représentation des données dont l'objectif principal sera la sélection initiale des informations cartographiques nécessaires à l'automatisation de la confection d'une carte.

Cet article présente un modèle de représentation du contexte fondé sur les ontologies. Après avoir introduit les objectifs de cette recherche, la deuxième partie présente la notion de contexte selon différents points de vue, les applications capables d'adapter leur comportement selon le contexte, ainsi que les différentes approches pour le modéliser. La troisième partie introduit ensuite notre modèle conceptuel fondé sur une ontologie. Un scénario d'exécution est alors présenté. La dernière partie conclut ce travail tout en proposant des évolutions futures.

2. Modélisation du contexte

2.1. Définition des applications sensibles au contexte

La sensibilité au contexte est la capacité d'un système à tirer profit, interpréter et utiliser des informations issues du contexte et adapter sa réponse en fonction du contexte d'utilisation. Cette définition correspond à peu près à la première définition de la sensibilité au contexte proposée par Schilit (Schilit, Theimer, 1994). Selon Dey (Dey, 2001), un contexte se définit comme « toute information en mesure d'être utilisée pour caractériser la situation d'une entité. Une entité est une personne, un lieu ou un objet concerné dans l'interaction entre un utilisateur et une application ».

La notion du contexte en cartographie numérique a été reconnue par Nivala (Nivala, Sarjakoski, 2003) et Sarjakoski (Sarjakoski, Nivala, 2005) dans le cadre de leurs travaux sur les cartes numériques mobiles. Les chercheuses se sont d'abord appuyées sur les définitions de contexte proposées par Schilit (Schilit, Theimer, 1994) et Dey (Dey, 2001). Puis, en s'appuyant sur les travaux de Chen (Chen, Kotz, 2000), elles ont proposé une classification du contexte, comprenant cinq catégories de haut niveau, adaptée aux cartes mo-

biles pour caractériser un contexte cartographique : informatique (système de navigation), utilisateur (tâche à effectuer, profil, préférences, etc.), physique (environnement, lieu, orientation), temps (heure de la journée, saison, etc.), historique (historique de la navigation).

2.2. Représentation du contexte

On distingue six grandes approches pour modéliser le contexte : les modèles clés-valeurs, les modèles fondés sur des langages de balises (XML), les modèles graphiques (UML), les modèles orientés objet, les modèles logiques et les modèles reposant sur les ontologies. Afin d'évaluer les différentes approches de modélisation du contexte, Strang (Strang, Linnhoff-Popien, 2004) a évalué ces six approches par rapport à plusieurs exigences des services omniprésents dans les systèmes informatiques. Les auteurs concluent que le modèle à base d'ontologie présente l'approche la plus expressive ainsi qu'une approche très prometteuse pour la modélisation du contexte dans le domaine de l'informatique. Les ontologies introduisent non seulement une représentation de connaissances partageables et réutilisables, mais permettent également l'ajout de nouvelles connaissances sur le domaine. Selon Wang (Wang *et al.*, 2004), nous pouvons résumer les avantages des ontologies pour la modélisation du contexte par : le partage des connaissances, l'inférence logique et la réutilisation des connaissances.

3. Modélisation contextuelle : la carte à la demande

3.1. Approche de modélisation ontologique

Le modèle proposé est fondé sur quatre classes principales : la classe « User », la classe « Activity », la classe « Environment », et la classe « Event ». La classe « User » décrit les caractéristiques et les préférences de l'utilisateur final de la carte, c'est-à-dire la personne qui en fait la demande. Cette classe est une sous-classe de la classe « Person » de l'ontologie FOAF¹. Pour chaque instance de la classe « User », nous sommes intéressés par ses *capacités* et ses *incapacités*, sa *profession*, la *communauté* à laquelle il appartient, son *positionnement*, l'ensemble de ses *intérêts*, son *historique de navigation* et son *activité*.

La classe « Activity » a pour objectif de modéliser la tâche ou la mission de l'utilisateur. Chaque instance de cette classe est caractérisée par un *temps* de déroulement qui est une instance de la classe « TemporalEntity » de l'ontologie OWL-Time², une *zone géographique* qui est une instance de la classe « SpatialThing » de l'ontologie GeoSPARQL³, le *niveau d'expertise* de l'utilisateur et un *environnement* dans lequel se déroule l'activité.

1. <http://xmlns.com/foaf/spec/>

2. <https://www.w3.org/TR/owl-time/>

3. <https://www.opengeospatial.org/standards/geosparql>

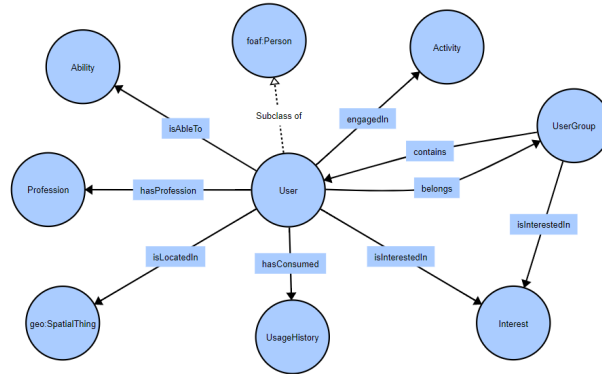


FIGURE 1: Ontologie de l'utilisateur.

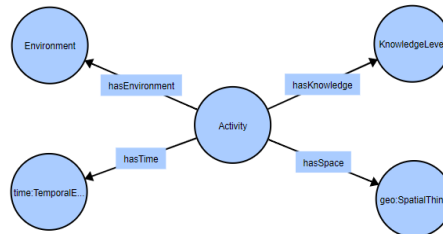


FIGURE 2: Ontologie de l'activité.

Un « Environment » se compose d'un *environnement physique*, pour représenter les conditions environnementales dites naturelles, et d'un *environnement personnel* pour décrire le *support de visualisation*, la *taille de l'écran*, etc. Il est associé à une *zone géographique*, et un *événement* auquel il peut être exposé.

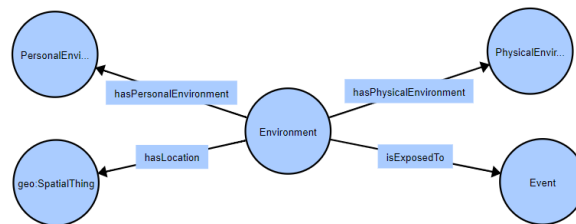


FIGURE 3: Ontologie de l'environnement.

Un « Event » peut impliquer un changement de *situation* qui, à son tour, peut conduire à un nouveau *contexte*, par exemple un nouveau contexte de navigation dans le cas de la confection d'une carte marine personnalisée. Une situation peut concerner une ou plusieurs *entités*. L'événement, la situation et l'entité peuvent avoir des *états* qui peuvent être de nature *spatiale*, *temporelle* ou *sémantique*.

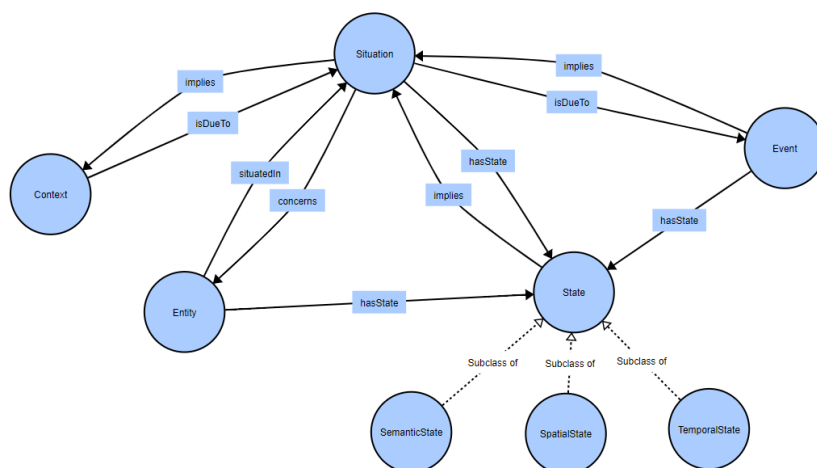


FIGURE 4: Ontologie de l'événement.

3.2. Exemple de scénario d'exécution

Dans cette section, nous présentons un scénario d'exécution afin d'illustrer l'approche de modélisation ontologique ci-avant. Un des cas d'utilisation est la confection d'une carte marine contextuelle. Ainsi, après avoir recueilli les préférences de l'utilisateur, les conditions environnementales dans lesquelles il évolue, ainsi que les informations relatives à son activité de navigation maritime côtière (plaisance, pêche, transport, militaire, etc.), l'objectif est alors de sélectionner, par raisonnement logique, les objets cartographiques entrant dans la confection d'une carte marine adaptée à son contexte d'utilisation. En se limitant à l'ontologie des événements, la *nuit* est considérée comme une instance d'un « Event ». Cet événement implique une situation de *faible visibilité* pour la navigation, ce qui aboutit à un contexte de *navigation nocturne*. Par inférence, l'instance *navigation nocturne* nécessite la sélection d'entités ayant un feu (ou lumière dans la terminologie classique) comme les instances de la catégorie des bouées cardinales. Ces dernières, selon leur type (Nord, Sud, Est, Ouest) ont un état temporel qui est relatif au rythme de leur feu (feu à occultations régulières, feu à deux éclats groupés, etc.), un état spatial relatif à leur positionnement géographique ainsi qu'un état sémantique qui, selon leur type, indique la zone de sécurité à emprunter pour la navigation (c.-à-d. une cardinale Nord indique que la navigation doit se faire au Nord de cette bouée, le danger se situant au Sud).

4. Conclusion

Dans cet article, nous avons présenté différents modèles de représentation du contexte permettant à un système d'information de répondre efficacement à un besoin utilisateur. Parmi ceux-ci et selon les critères de comparaison proposés par Malik (Malik, Jain, 2017), les modèles à base d'ontologie semblent appropriés à répondre à la problématique de la sélection des informations perti-

nentes dans le processus de confection de cartes. Notre modélisation est fondée sur la classification du contexte comprenant les cinq catégories de haut niveau proposées par Sarjakoski (Sarjakoski, Nivala, 2005), ainsi qu'une représentation des événements afférents à un environnement. L'approche présentée dans cet article est en cours de développement et s'appuie sur des ontologies de haut niveau telles que GeoSPARQL, OWL-Time et FOAF pour ne citer qu'elles. L'étape ultérieure consiste à enrichir les ontologies et à formaliser les concepts complexes afin de raisonner sur les axiomes de la base de connaissances. Un raisonnement axiomatique sera probablement insuffisant pour satisfaire notre besoin d'offrir un service contextuel propre à un utilisateur. C'est la raison pour laquelle, nous étudierons en complément, et alternativement au raisonnement axiomatique (pour satisfaire le principe de décidabilité) un raisonnement à base de règles.

Bibliographie

- Chen G., Kotz D. (2000). *A Survey of Context-Aware Mobile Computing Research*. Rapport technique. Hanover, NH, USA.
- Dey A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing Journal*, vol. 1, n° 5, p. 4-7.
- Malik S., Jain S. (2017). Ontology based context aware model. In *Proceedings of the international conference on computational intelligence in data science (ICCIDS)*, p. 1-6.
- Nivala A., Sarjakoski L. T. (2003). Need for context-aware topographic maps in mobile devices. In *Proceedings of ScanGIS'2003 - the 9th scandinavian research conference on geographical information science, 4-6 june 2003, Espoo, Finland*, p. 15-29.
- Sarjakoski L. T., Nivala A.-M. (2005). Adaptation to context — a way to improve the usability of mobile maps. In L. Meng, T. Reichenbacher, A. Zipf (Eds.), *Map-based mobile services: Theories, methods and implementations*, p. 107-123. Springer Berlin Heidelberg.
- Schilit B., Theimer M. (1994). Disseminating active map informations to mobile hosts. *Institute of Electrical and Electronics Engineers (IEEE) Networks.*, vol. 5, p. 2232-2242.
- Strang T., Linnhoff-Popien C. (2004). A context modeling survey. In *Workshop on advanced context modelling, reasoning and management, UbiComp 2004 - the sixth international conference on ubiquitous computing, Nottingham, England*.
- Wang X. H., Zhang D. Q., Gu T., Pung H. K. (2004). Ontology based context modeling and reasoning using OWL. In *Proceedings of the second IEEE annual conference on pervasive computing and communications workshops*, p. 18-22.

Règles d'association pour l'analyse de l'évolution d'un territoire

Asma GHARBI¹, Cyril DE RUNZ², Herman AKDAG¹, Sami FAIZ³

1. LIASD, Université de Paris 8, Saint-Denis, France

Herman{asma.haj-ali-gharbi}@univ-paris8.fr

2. LIFAT Université de Tours, IUT de Blois, Blois, France

cyril.derunz@univ-tours.fr

3. ISAMM, La Manouba, Tunisie

sami.faiz@isa2m.rnu.tn

RÉSUMÉ. Dans cet article, nous partons de l'hypothèse que les dynamiques spatiales et les usages des objets géographiques peuvent, en partie, être expliqués voire anticipés par leurs historiques de changements de fonctions et de co-localisations. Ainsi, nous proposons une démarche qui, se basant sur la recherche de règles d'association tout en considérant l'asymétrie des données, permet de générer des modèles explicatifs voire prédictifs de l'évolution d'un territoire. Appliquées sur des données issues de Corine Land Cover, les méthodes de fouille proposées ont fait leurs preuves dans la gestion du déséquilibre des données et les règles générées ont mis en évidence des dynamiques urbaines réalistes.

ABSTRACT. This work starts from the hypothesis that spatial dynamics and the change of geographical objects functions could be, partly, explained or anticipated by the history of their functions and co-locations changes. Hence, a process, based on association rules-based approaches and considering data asymmetry, for the extraction of an explicative/predictive model of territorial evolution is proposed. Applied on data extracted from Corine Land Cover, the proposed mining methods proved their worth in the management of data imbalance and the generated rules highlight realistic urban dynamics.

MOTS-CLÉS : Règles d'association, MSApriori, règles d'évolution, dynamique spatiale

KEYWORDS: Association rules, MSApriori, Evolution rules, spatial dynamics

1. Introduction

Un territoire évolue, se bâtit et se transforme au fil du temps. Ainsi, nos sociétés ont besoin de modèles permettant d'expliquer, voire d'anticiper, quelles seront les évolutions futures de nos territoires (urbains, rurales, etc.). Un large panel de travaux, se basant sur la fouille de données spatio-temporelles pour traiter la question des changements d'occupation/usage du sol (e.g. (Qiang, Lam, 2015)). La plupart de ces travaux visent essentiellement à identifier et à caractériser ces changements ainsi qu'à découvrir les relations qu'ils ont avec les différentes variables naturelles et anthropogéniques. Bien qu'efficace, cette démarche spécifique au type de couverture suivi (e.g. végétations, bâti), se concentre sur les caractéristiques internes des zones étudiées et néglige les relations spatiales et temporelles des données (i.e., une telle propriété est spatialement voisine à une telle propriété, une telle propriété apparait avant une telle propriété sur une même localisation). Notre approche tente de répondre à ces problématiques en explorant les dépendances entre l'historique de l'évolution de l'objet spatio-temporel (relations temporelles de succession de fonctions), l'historique de ses co-localisations (relations spatiales de voisinage) et sa future fonction (future occupation/usage du sol). Pour cela, nous cherchons à extraire des règles d'association spatiotemporelles pour construire un possible modèle explicatif/descriptif de l'évolution d'un territoire. L'approche que nous proposons est constituée d'une première phase de suivi de l'évolution des objets pour la création de la base d'apprentissage, puis d'une seconde phase de recherche de règles d'association sur la base d'apprentissage créée. La base d'apprentissage générée étant dans les faits asymétrique, nous exploitons une méthode adaptée à cette problématique – MSAPriori (Liu *et al.*, 1999) – pour laquelle nous proposons d'utiliser pour valeurs de seuil des estimateurs statistiques de position.

2. Suivi de l'évolution et construction de la base d'apprentissage

L'objectif de cette première phase consiste à partir d'une série temporelle de cartes vectorielles décrivant un territoire à des dates données, pour définir un modèle de données permettant le suivi de l'évolution de celui-ci.

Dans ce contexte, nous considérons :

- L'espace comme étant un support immuable où se situent les objets et les relations exprimant leurs configurations spatiales.
- Le temps comme étant linéaire sous forme d'une succession d'évènements de changement de fonction et/ou de forme ou de topologie.
- L'objet géographique comme le produit de trois composantes à savoir, une dimension spatiale (S) (l'étendu spatial de l'objet), une dimension temporelle (T) (i.e., la date d'observation de l'objet), et une dimension sémantique (F) (la fonction de l'objet) (Perret *et al.*, 2015).

Tenant compte de ces éléments nous, considérons que l'évolution d'un objet est caractérisée par la modification de la valeur d'au moins une de ses composantes ou attributs. Ce changement donne, ainsi, lieu à la génération d'une nouvelle version de l'objet, portant une nouvelle identité. Ainsi, son évolution est décrite par ces différentes versions. Afin de garder le lien entre elles et de pouvoir en suivre l'évolution, comme proposé, dans (Kauppinen, Hyvönen, 2007), nous plaçons le traceur de l'identité sur l'un des attributs de l'objet : l'empreinte spatiale. Ainsi, l'objet est identifié par sa zone spatiale allouée et ses successeurs peuvent être repérés à l'aide des requêtes spatiales de recouvrement et de voisinage pour reconstruire sa trajectoire de vie (cf. figure1.a).

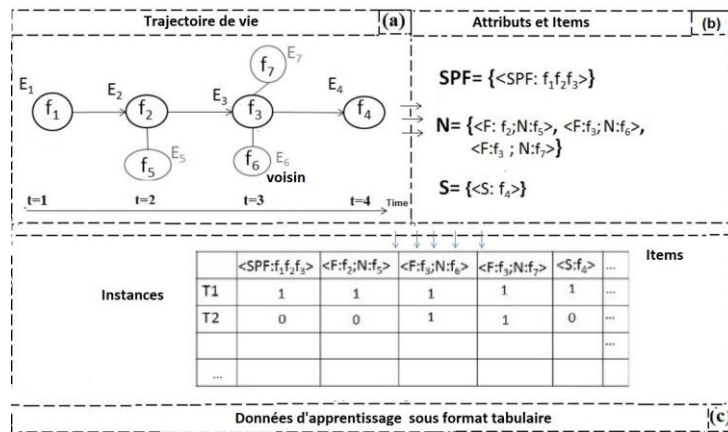


Figure 1. Exemple illustrant la génération de données d'apprentissage.

Dans le contexte du présent travail, nous stipulons que l'évolution d'un territoire peut être en partie explicitée voire anticipée par l'historique des fonctions de ses objets et de leurs configurations spatiales. À cet égard, notre trajectoire de vie (cf. figure1.a) peut être décrite sous forme d'une transaction composée de trois types d'information (cf. figure1.b). Le premier type (SPF) correspond à la succession de ses fonctions et dans notre exemple a pour valeur $\langle f_1, f_2, f_3 \rangle$. Le deuxième type (N) donne les valeurs des voisins au cours du temps. Par exemple, la relation de voisinage entre l'objet E_2 de fonction f_2 et l'objet E_5 de fonction f_5 , sera inscrite dans l'ensemble des voisinages comme $\langle F : f_2; N : f_5 \rangle$. Enfin, le dernier type (S) correspond à la fonction de l'objet successeur, dans l'exemple $\langle S : F_4 \rangle$. Nos différents attributs sont transcrits sous forme de table binaire indiquant leurs présence/absence dans la transaction (cf. figure1.c). Chaque attribut présent dans une transaction est appelé item. Faisant le choix que chaque transaction ait un seul attribut S et un seul SPF, une trajectoire de vie pourra donc être transcrite par une ou plusieurs transactions dans la base d'apprentissage suite au possible fusion/scission d'un

objet au cours du temps. Il convient aussi de noter que les voisinages sont identifiés selon la relation topologique «se toucher».

3. Extraction des règles d'association pour la modélisation de l'évolution territoriale

L'une des approches de modélisation, évoquée dans la littérature, notamment dans (Silva, Wu, 2012), est la modélisation à base de règles. L'un des challenges ici, c'est la définition des meilleures combinaisons de règles de transitions lorsqu'il y a beaucoup de variables à gérer. Dans ce contexte, plusieurs méthodes basées sur les algorithmes d'apprentissage ont été proposées dont l'analyse des associations qui permet de découvrir les relations entre différentes variables stockées dans une base de données (Solanki, Patel, 2015). Ces méthodes, notamment Apriori et MSApriori (Liu *et al.*, 1999), fonctionnent en deux étapes. Une première étape cherchant à regrouper les valeurs présentes (appelées items) dans des ensembles d'items, appelés itemsets. Seuls sont regroupés les items dont la fréquence d'apparition est supérieur à un seuil (minSup) ou à plusieurs seuils (nommés *SMI*), dans le cas de MSApriori, pour gérer des possibles déséquilibres dans les données. Ces items sont donc considérés comme fréquents, et les itemsets associés aussi. Sur ces itemsets fréquents, une approche de génération de règles d'association est effectuée afin de récupérer les possibles règles de co-occurrences.

Une première application sur nos données d'Apriori a révélé une asymétrie de données. En effet, seules des règles composées exclusivement par des items de type N sont générées (cf. tableau 1). Ceci s'explique par le fait que les items de voisinage sont très fréquents par rapport aux autres types d'items (S, SPF). Si on définit un minsup assez faible les items de type S et SPF apparaîtraient dans les candidats fréquents, toutefois, ceci mènera à un problème d'explosion combinatoire, i.e., un très grand nombre de candidats. Au travers ses multiples valeurs de support minimum, MSApriori rend possible d'extraire des règles comportant des éléments plus rares que si on avait un seuil global unique tout en réduisant la complexité combinatoire si un unique seuil bas serait défini. Cependant, dans cette méthode d'affectation de *SMI* l'expert doit définir la valeur des seuils en fonction d'une évaluation subjective de sa satisfaction à l'égard des modèles générés. Dans cet article, nous cherchons à minimiser la supervision du processus d'affectation des seuils par l'utilisation d'estimateurs de position dans un échantillon.

Méthodes à base de quartiles (QuartilesBased) : Partant de l'hypothèse qu'ils peuvent servir de frontières entre les valeurs "basses" et "élevées", nous nous intéressons aux estimateurs robustes de position que sont la médiane (Q2), les premier et troisième quartiles (Q1 et Q3). En utilisant ces valeurs, la population est divisée en quatre groupes de tailles équivalentes correspondant à quatre intervalles semi-ouverts. Pour chaque groupe, sa médiane est affectée comme le *SMI* de tous les items qui lui appartiennent. Nous appelons cette première

variante *QuartilesBased*.

Méthode à base de clustering (ClusterBased) : Afin d'affiner le partitionnement, nous pouvons nous servir d'algorithmes de clustering permettant de considérer les fréquences des items et les comparer mutuellement afin de regrouper les items par similarité de leurs fréquences. Ne souhaitant pas présumer du nombre de groupes de fréquences, nous avons opté, dans la variante *ClusterBased*, pour l'algorithme Espérance-Maximisation (EM) ne nécessitant pas ce paramétrage.

Vers la prise en compte de la sémantique des items

Pensant que la sémantique (S, SPF, N) des items doit pouvoir jouer un rôle dans la démarche, nous proposons une variante de *QuartilesBased* appelée *QuartilesBasedSem* ainsi qu'une variante de *ClusterBased* appelée *ClusterBasedSem*.

QuartilesBasedSem : Dans cette variante, nous partitionnons dans un premier temps les items selon leur sémantique afin d'obtenir 3 groupes (groupe d'item SPF, groupe d'item N, groupes d'item S). Pour chaque groupe, la valeur du quartile adéquat (Q1, Q2, Q3) est calculée puis attribuée en tant que *SMI* de tous les items de ce groupe. Pour les attributs les moins fréquents (S et SPF), la médiane (Q2) est adoptée comme *SMI*. Pour les attributs les plus fréquents (N), le quartile supérieur (Q3) est adopté.

ClusterBasedSem : Dans cette variante, le clustering se fait sur des données ayant pour attributs la fréquence des items et leur sémantique. La médiane de chaque groupe d'items lui sert de *SMI*.

4. Résultats

Nous avons appliqué nos propositions sur des données réels extraites de Corine Land Cover pour Paris. Quatre cartes vectorielles datant de 1990, 2000, 2006 et 2012 sont fournies. Les trois premières sont utilisées pour la génération des modèles et la quatrième pour l'évaluation de ceux-ci. La base de données d'apprentissage générée à partir de ces cartes vectorielles, compte 3913 instances, 190 items (97 N et 93 S et SPF). Dans le tableau 1, la proportion d'items SPF, N sur le nombre total d'items extraits montre que toutes les méthodes proposées ont réussi à faire apparaître les items S et SPF. Nous constatons que les méthodes à base de quartiles surpassent celles à base de clustering, et que le fait de prendre en compte la sémantique des éléments dans l'affectation des minsups améliore ces ratios (ex. 0,55 pour *QuartilesBased* contre 0,75 pour *QuartilesBasedSem*). Cependant, nous n'avons pu extraire de règles dans un temps raisonnable pour *QuartilesBasedSem*. Nous distinguons deux types de règles : les règles de stabilité (pas de changement de fonction entre les instants t et $t-1$); et les règles de transition où la fonction d'un objet change. Nous pourrions constater dans le tableau 1 que les trois méthodes pour lesquelles nous avons pu extraire des règles fournissent le même nombre de règles de transition mais pas le même nombre de règles de stabilité.

Table 1. Performances in terms of generation of S and SPF items

	$ ItemsSetSPF $ $ Items $	Règles Pertinentes générées	Règles de Stabilité	Règles de transition
US : Apriori (minsup=40%)	0	0	0	0
US : MSApriori ($\beta=0.5$)	0.086	0	0	0
MS:QuartilesBased	0.550	24816	2469	127
MS:QuartilesBasedSem	0.753	-	-	-
MS:ClusterBased	0.462	6264	6137	127
MS:ClusterBasedSem	0.5	9363	9236	127

5. Conclusion

Dans cet article, nous avons extrait des relations spatio-temporelles de voisinage et de succession d'usages des objets géographiques pour produire un modèle prospectif de l'évolution d'un territoire. Pour cela, à partir des différents snapshots, notre approche identifie des trajectoires de vies des objets, transcrivent celles-ci dans une base d'apprentissage binaire. Sur cette base d'apprentissage et afin de gérer l'asymétrie des données, nous avons appliqué MSApriori avec différentes stratégies d'affectation des minSup basées sur des estimateurs de position. Cela nous a permis d'extraire des règles d'évolution et de stabilité.

Bibliographie

- Kauppinen T., Hyvönen E. (2007). Modeling and reasoning about changes in ontology time series. In R. Sharman, R. Kishore, R. Ramesh (Eds.), *Ontologies: A handbook of principles, concepts and applications in information systems*, p. 319–338. Boston, MA, Springer US.
- Liu B., Hsu W., Ma Y. (1999). Mining association rules with multiple minimum supports. In *Proceedings of the fifth acm sigkdd international conference on knowledge discovery and data mining*, p. 337–341.
- Perret J., De Runz C., Rodier X., Varet-Vitu A., Dumenieu B., Saligny L. *et al.* (2015). Études des dynamiques de l'occupation du sol : questionnement, simplification et limites. *Revue Internationale de Géomatique*, vol. 25, n° 3, p. 301–330.
- Qiang Y., Lam N. S. N. (2015). Modeling land use and land cover changes in a vulnerable coastal region using artificial neural networks and cellular automata. *Environmental Monitoring and Assessment*, vol. 187, n° 3, p. 57.
- Silva E., Wu N. (2012). Surveying models in urban land studies. *CPL bibliography*, vol. 27, n° 2, p. 139-152.
- Solanki S. K., Patel J. T. (2015). A survey on association rule mining. In *2015 fifth international conference on advanced computing & communication technologies*, p. 212–216.