Fouille de trajectoires, des données aux connaissances

Laurent Etienne

DOCTORAT SCIENCES BRETAGNE DE LA MER LOIRE ET DU LITTORAL

Université de Bretagne Occidentale

HABILITATION À DIRIGER DES RECHERCHES

L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE

ÉCOLE DOCTORALE N° 598
Sciences de la Mer et du Littoral
Spécialité : *géomatique*

Par

Laurent ETIENNE

Fouille de trajectoires, des données aux connaissances

HdR présentée et soutenue à
Télé-Amphitéatre du Pôle Numérique du Bouguen, 6 Rue du Bouguen, 29200 Brest
avec retransmission en visio-conférence, le 11/07/2023 à 10h00
Unité de recherche : LabISEN, YNCREA OUEST, 20 rue Cuirassé Bretagne, Brest

Rapporteurs avant soutenance :

Christophe CLARAMUNT	Professeur, Ecole Navale, IRENav, Lanvéoc
Aldo NAPOLI	Directeur de recherche, MINES ParisTech, CRC, Valbonne
Ana-Maria OLTEANU-RAIMOND	Directrice de recherche, Université Gustave Eiffel, IGN-ENSG, LASTIG-MEIG, Saint-Mandé

Composition du Jury :

Examinateurs :	Sonia CHARDONNEL	Directrice de recherche, HdR, UMR PACTE, Grenoble
	Christophe CLARAMUNT	Professeur, HdR, Ecole Navale, IRENav, Lanvéoc
	Iwan LE BERRE	Maître de conférences, HdR, UBO/IUEM, LETG, Plouzané
	Aldo NAPOLI	Directeur de recherche, HdR, MINES Paris - PSL, CRC, Sophia Antipolis
	Ana-Maria OLTEANU-RAIMOND	Directrice de recherche, HdR, IGN-ENSG, LASTIG, Saint-Mandé
Invités :	Ayman AL FALOU	Professeur de l'enseignement supérieur, HdR, LabISEN, Brest
	Thomas DEVOGELE	Professeur, Université de Tours, LIFAT, Blois

Table of Contents

INTRODUCTION

Grâce à l'essor des objets connectés et au développement des technologies de géolocalisation, nous disposons aujourd'hui de masses de données décrivant, de manière plus ou moins détaillée, nos activités quotidiennes.

La communauté scientifique s'intéressant à l'étude des déplacements est une communauté riche et multidisciplinaire étudiant ce thème de recherche sous de nombreux prismes. Au cours de mon parcours académique, j'ai eu la chance de collaborer avec de nombreux chercheurs issus de sciences très variées telles que la sociologie, l'écologie, la géographie, l'urbanisme, les mathématiques et les statistiques, l'informatique et la géomatique.

Ces collaborations ont été menées lors de projets de recherche interdisciplinaires tels que MOBIKIDS, SMARTLOIRE présentés au chapitre 4 ainsi que dans le cadre de groupements de recherche tels que le GDR CNRS MAGIS ou le réseau européen COST MOVE.

Les données de mobilité, collectées tout au long de ces projets variés, peuvent provenir d'une multitude de sources différentes, souvent hétérogènes et parfois contradictoires, incomplètes ou erronées. Dans certains cas, la combinaison et la fusion de ces sources de données permettent d'enrichir la représentation des mouvements réalisés en y ajoutant des descriptions sémantiques porteuses de sens en fonction de la thématique étudiée. Cependant, la multiplicité des attributs décrivant ces déplacements et leur hétérogénéité rend leur analyse particulièrement complexe. Une étude sur les apports des objets connectés pour la caractérisation et la segmentation automatique de trajectoire de véhicules de secours est présentée dans le chapitre 3.

Lorsque ces données de mouvement sont sauvegardées sur le long terme, il est alors possible d'exploiter ces corpus de données pour en extraire des connaissances concernant nos habitudes et d'y détecter des anomalies. Pour extraire ces connaissances, il est nécessaire de proposer de nouvelles mesures permettant de comparer la mobilité selon différents axes (spatial/temporel/sémantique). Le chapitre 5 de ce mémoire porte sur différentes mesures de similarité de trajectoires intégrant ces différents aspects.

À l'aide de ces nouvelles mesures de similarité, il est alors possible de constituer des ensembles de trajectoires (clusters) et chercher à en déduire des comportements types. Les

techniques de fouille de données peuvent être employées pour caractériser ces ensembles de trajectoires. Dans ce mémoire, le chapitre 6 présente les travaux réalisés sur la génération de patrons spatio-temporels de trajectoires via une extension du concept de statistiques descriptives des boites à moustaches (boxplot).

Enfin le contexte dans lequel les personnes évoluent impacte leurs choix de mobilité. Ici encore, le développement des objets connectés, des réseaux de capteurs in situ et les outils de télédétection permettent de disposer de masses de données de plus en plus fines et détaillées décrivant l'environnement à des échelles spatiales et temporelles toujours plus précises. Le programme européen d'observation de la Terre Copernicus met à disposition de vastes quantités de données mondiales, couvrant les espaces terrestres, aériens et marins, accessibles gratuitement et librement. Couplées aux données de mobilité, ces données environnementales viennent enrichir et décrire le contexte des trajectoires réalisées. Dans le contexte maritime, le contexte environnemental est particulièrement intéressant pour définir la notion de risque lié à la navigation. En effet, la mer comporte de nombreux dangers (risque d'échouage sur des hauts fonds, mauvaise visibilité, tempêtes, courants marins, glaces de mer...) que les marins cherchent à éviter en optimisant leurs trajectoires. Le chapitre 7 s'intéresse à la notion de risque et à son impact sur la mobilité en termes de planification et d'optimisation.

Chapitre 1

RAPPORT D'ACTIVITÉ

1.1 État civil

Identité	
Prénom	Laurent
Nom	ETIENNE

Coordonnées professionnelles	
ISEN YNCREA OUEST	
20 Rue Cuirassé Bretagne	
29200 BREST	
FRANCE	
Téléphone	+33 (0)2 98 03 84 00
Email	laurent.etienne@yncrea.fr

1.2 Parcours professionnel

Depuis 2019	**ISEN Yncrea Ouest, LabISEN équipe KLaIM**	Brest
Enseignant Chercheur en informatique		
• Responsable de l'équipe de recherche KLaIM du LabISEN		
• Responsable du domaine professionnel NEDD (niveau M1/M2)		
(Numérique Environnement et Développement Durable)		
2014-2019	**Polytech Tours, LIFAT/BDTLN**	Tours
Maître de conférences en informatique		
• Responsable adjoint de l'équipe de recherche BDTLN du LIFAT		
2013-2014	**ENSAM/École Navale, groupe MoTim**	Lanvéoc / Brest
Attaché temporaire d'enseignement et de recherche		
2012-2013	**Dalhousie, MARIN Lab, projet PASSAGES**	Halifax, Canada
Enseignant-chercheur post-doctoral		
2008-2012	**École Navale, groupe SIG**	Lanvéoc
Assistant d'enseignement et de recherche		
2007-2008	**École Navale, groupe SIG, projet LOCOSS**	Lanvéoc
Ingénieur de recherche		
2004-2007	**SARL CEFTI , SARL 3A**	Bretagne
Gérant créateur d'entreprise et Administrateur systèmes		
2004-2005	**CNAM**	Rennes
Enseignant vacataire en informatique		

1.3 Formation universitaire et qualifications

2011	**Qualification aux fonctions de Maître de Conférences**	
	Section	27 (informatique)
2008-2011	**Doctorat en géomatique**	
Thèse	Université de Brest	
Titre	Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime.	
Mention	Très Honorable	
Distinction	Prix de thèse du GDR MAGIS 2012	
Directeur		Dr. Alain Bouju (L3I, La Rochelle)
Co-directeur		Pr. Thomas Devogele (IRENav, Lanvéoc)
Jury	Président	Pr. Jacques Tisseau (Lab-STICC, Brest)
	Rapporteurs	Pr. Hervé Martin, (LIG, Grenoble)
		Pr. Karine Zeitouni (PRiSM, Versailles)
	Examinateur	Pr. Christophe Claramunt (IRENav, Lanvéoc)
		Pr. Pascal Poncelet (LIRMM, Montpellier)
2002-2003	**Diplôme d'études approfondies en informatique**	
DEA	Université de Rennes 1	
	Option	Imagerie numérique et intelligence artificielle
Titre	Étude de l'interaction des sens de la vision et du toucher en réalité virtuelle	
Laboratoire d'accueil	Institut de Recherche en Informatique et Systèmes Aléatoires	
Directeur de stage	Bruno Arnaldi (IRISA)	
Tuteur	Anatole Lecuyer (IRISA)	
2001-2002	**Maîtrise en informatique**	
Maîtrise	Université de Rennes 1	
	Option	Calculs distribués et compression de données
2000-2001	**Licence en informatique**	
Licence	Université de Rennes 1	
1998-2000	**DUT Génie Électrique et Informatique Industrielle**	
DUT	Université de Rennes 1	

1.4 Encadrement d'activité de recherche

1.4.1 Responsabilités administratives

Responsabilité de groupes de recherche

- **Co-porteur de l'action de recherche AR6 Mobilité et impacts socio-environnementaux** du projet 2022-2026 du GDR CNRS MAGIS (Méthodes et Applications en Géomatique et Information Spatiale) depuis 2022.
- **Responsable de l'équipe de recherche Knowledge Learning and Information Modelling (KLaIM)** du LabISEN (Équipe de 7 enseignants-chercheurs) depuis 2019.
- **Responsable adjoint du groupe de recherche Base de Données et Traitement du Langage Naturel (BDTLN)** du Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT) de 2017 à 2019.
- **Membre du conseil scientifique du Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT)** représentant de l'équipe Base de Données et Traitement du Langage Naturel de 2015 à 2018.

Responsabilités pédagogiques

- **Responsable du Domaine Professionnel NEDD** (M1/M2) Numérique Environnement et Développement Durable (NEDD) à l'ISEN Brest de 2021 à aujourd'hui.
- **Directeur des études** du Parcours des Écoles d'Ingénieurs Polytechnique de l'Université de Tours de 2015 à 2019.
- **Membre du Conseil de Perfectionnement** du département Aménagement et Environnement de l'École Polytechnique de l'Université de Tours de 2015 à 2016.
- **Correspondant ERASMUS** du département Aménagement et Environnement de l'École Polytechnique de l'Université de Tours de 2015 à 2016 (Gestion des échanges avec 9 universités partenaires).
- **Correspondant ESRI** gestionnaire de la licence de site éducation pour l'Université de Tours (2014-2019).
- **Correspondant IGN** du département Aménagement et Environnement de Polytech Tours (2014-2019).

1.4.2 Encadrement de thèses de doctorat

- **Thèse 1 : Frédérick BISONE**
  Thèse soutenue le 31 mai 2021
  Sujet : Extraction de trajectoires sémantiques à partir de données multi capteurs : application à des véhicules de secours
  Financement : CIFRE (Société GRUAU)
  Organisme d'accueil : Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT)
  Directeur de thèse : Pr. Thomas DEVOGELE
  Ratio de co-encadrement : 60%.
- **Thèse 2 : Clément MOREAU**
  Thèse soutenue le 25 novembre 2021
  Sujet : Fouille de séquences de mobilité sémantique. Sur l'élaboration de mesures pour la comparaison, l'analyse et la découverte de comportements
  Financement : ANR MOBIKIDS / PRIR Smart Loire
  Organisme d'accueil : Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFAT)
  Directeur de thèse : Pr. Thomas DEVOGELE
  Ratio de co-encadrement : 30%.
- **Thèse 3 : Mark STODDARD**
  Soutenance prévue en 2023
  Sujet : Measuring Arctic Maritime Remoteness using Risk-Based Ship Transit Times in Ice
  Financement : Canadien (DRDC Canada)
  Organisme d'accueil : Dalhousie University, MARIN Lab, Canada
  Directeur de thèse : Pr. Ronald PELOT
  Ratio de co-encadrement : 30%.

1.4.3 Membre de comités de suivi de thèse

- Clément IPHAR : 2015
- Clay PALMERA : 2017

- Yasmine BOUNAAS : 2020
- Maryam MASLEK ELAYAM : 2020, 2021, 2022

1.4.4 Encadrement de stage de recherche

- Gabriel Ahtune (Mai à Septembre 2010) : Implémentation d'outils de qualification de trajectoires de navires.
- Ajith Winston Bryn (Janvier à Mai 2013) : Design of an interactive website for Carbon Capture Multi-Criteria Decision Making.
- François-Xavier Guillaud (Octobre 2014 à Avril 2015) M2 : La mobilité révélée par GPS : MobiRev, outil de traitement automatique des traces GPS.
- Maxence Esnault 2016 L3 : Optimisation algorithmique de la distance de Fréchet.
- Florian Lardy 2017 L3 : Optimisation et parallélisation de la distance de Fréchet.
- Frédérick Bisone (Février à Août 2017) M2 : Analyse sémantique de trajectoires de touristes.
- Imad Bader, Nader Yantani (Septembre à Novembre 2017) M2 : Fouille de données de positions de navires.
- Xiaomeng Wang, Yanwan Yao (Octobre 2017 à Avril 2018) M2 : Analyse comparative du déplacement de pigeons voyageurs.
- Bruno Dunas 2018 L3 : Suivi des parcours de visite de touristes à l'aide de smartphones. Clément Moreau (Février à Août 2018) M2 : Fouille de Trajectoires Sémantiques.
- Valentin Rall, Yann Cariou (Octobre à Mars 2020) M2 : Évaluation de la qualité de l'air à l'aide d'une flotte de capteurs mobiles.
- Hugo Neveu (Juin à Septembre 2021) M1 : Reconstruction de modèle de terrain 3D à l'aide de capteur Intel RealSense.

1.4.5 Projets scientifiques et industriels collaboratifs

Partenariat Hubert Curien (PHC) du ministère des Affaires étrangères

- Polonium, Gdansk University of Technology (2010-2011)
- Ulysses, National University of Ireland, Maynooth (2011-2012)

Projets de Recherche d'Intérêt Régional (PRIR)

- LOCOSS (2006-2009)
- MARMOUTIER II (2016-2019) 410 k€
- SMART LOIRE (2017-2020) 412 k€

Projet Fondation de France

- DACTARI (2013-2015)

Projets ANR

- MOBIKIDS (2017-2020) 510 k€
- Dé-AIS (2014-2017) 351 k€

Projets internationaux

- Projet bilatéral Allemagne/Canada : PASSAGES (2012-2015) 2 M$
- Projet européen ITEA2 (EUREKA cluster on information technology) : RECONSURVE (2011-2012)

1.4.6 Membre de comité d'organisation de conférences

- Conférence internationale "Conference on Spatial Information Theory" (COSIT), l'Aber Wrac'h, Septembre 2009.
- Workshop international "Information Fusion and Geographical Information : Towards the Digital Ocean" (IF&GIS'11), Brest, Mai 2011.
- Conférence Internationale GeoSpatial Semantics (GeoS'11), Brest, Mai 2011.
- Session Chair, Marine Big Data Workshop MBDW, 21st IEEE International Conference on Mobile Data Management (MDM 2020).

1.4.7 Membre de comité de programme de conférences

- Spatial Analysis and GEOmatics (SAGEO 2018, 2019, 2021)

- International Conference on Future Networks and Distributed Systems (ICFNDS 2017, 2018, 2019, 2020, 2021)
- Marine Big Data Workshop MBDW, IEEE International Conference on Mobile Data Management (MDM 2020, 2022)

1.4.8 Relecteur scientifique pour les revues et conférences

- Géo-Regards
- Netcom. Réseaux, communication et territoires
- Revue Internationale de Géomatique (RIG)
- ISPRS International Journal of Geo-Information (IJGI)
- Symmetry
- Sustainability
- Knowledge and Information Systems (KAIS)
- Pacific-Asia Conference on Knowledge Discovery and Data Mining (2017)
- Conférence EGC Atelier Gestion et Analyse des données Spatiales et Temporelles (GAST 2018)
- Conférence ACM SIGAPP SAC GIA (2019, 2020)

1.4.9 Membre de réseau de recherche

Participant et animateur de l'action de recherche "Mobilité et impacts socio-environnementaux" du GDR CNRS MAGIS 2022-2027. Participant à l'action prospective "Mobilités et Trajectoires" du GDR CNRS MAGIS 2017-2021. Membre du groupe de travail "recherches polaires et sub polaires" du CNRS depuis 2019. Participant au programme européen COST IC0903 (European Cooperation in Science and Technology) MOVE (Moving object and Knowledge Discovery) de 2009 à 20013. Membre du comité d'expert du projet ANR PORTIC de 2019 à 2022. Participant du GIS ITS Bretagne de 2011 à 2013. Membre de l'AFIGEO.

Table 1.1 – Nombre de publications par catégories

Publications	
Ouvrages, chapitres	8
Revues	13
Conférences internationales	18
Conférences nationales	8
Rapports de recherche	4
Brevets, dépôts APP	2
Total	**53**

1.5 Publications scientifiques

Cette section présente une liste complète de mes publications. Cette liste inclut 51 publications toutes natures confondues. La stratégie de publication a consisté principalement à publier vers les communautés Géomatique et Intelligent Transportation Systems (ITS) au sens large avec une volonté d'adresser les aspects mobilité, fouille de données, visualisation et Big Data. Cette liste est mise à jour régulièrement sur ma page HAL [1]

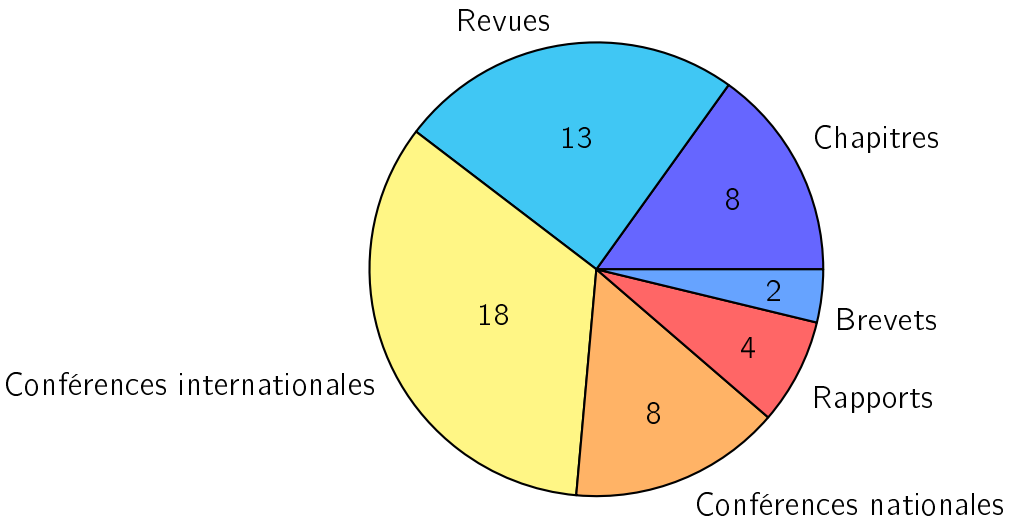


Figure 1.1 – Diagramme de répartition des publications par catégories.

1. https://cv.archives-ouvertes.fr/letienne

1.5.1 Chapitres d'ouvrages

[**Ch1**] **Etienne**, Laurent, Ray, Cyril, Camossi, Elena et Iphar, Clément. **fév. 2021**, « Maritime Data Processing in Relational Databases », in : *Guide to Maritime Informatics*, Springer International Publishing, p. 73-118, doi : 10.1007/978-3-030-61852-0_3, url : https://hal.archives-ouvertes.fr/hal-03137050 (cf. p. 26).

[**Ch2**] Faury, Olivier, **Etienne**, Laurent, Fedi, Laurent, Rigot-Müller, Patrick, Cheaitou, Ali et Stephenson, Scott. **sept. 2019**, « L'impact de la gestion du risque sur l'attractivité du passage du nord-est », in : *Baltic Arctic – Strategic Perspective*, Les collections Océanides, Editions EMS, p. 169-190, url : https://hal.archives-ouvertes.fr/hal-02885901.

[**Ch3**] Stoddard, Mark, **Etienne**, Laurent, Pelot, Ronald, Fournier, Melanie et Beveridge, Leah. **août 2018**, « From Sensing to Sense-Making : Assessing and Visualizing Ship Operational Limitations in the Canadian Arctic Using Open-Access Ice Data », in : *Sustainable Shipping in a Changing Arctic*, url : https://hal.archives-ouvertes.fr/hal-01897714.

[**Ch4**] **Etienne**, Laurent, Fournier, Melanie, Beveridge, Leah, Stoddard, Mark et Pelot, Ronald. **2017c**, « Vessel navigation constraints in Canadian Arctic waters », in : *Advances in Shipping Data Analysis and Modeling Tracking and Mapping Maritime Flows in the Age of Big Data*, url : https://hal.archives-ouvertes.fr/hal-01627377.

[**Ch5**] **Etienne**, Laurent, Alincourt, Erwan et Devogele, Thomas. **2015a**, « Maritime network monitoring : from position sensors to shipping patterns », in : *Maritime Networks : spatial structures and time dynamics*, url : https://hal.archives-ouvertes.fr/hal-01627357.

[**Ch6**] **Etienne**, Laurent, Hjelmfelt, Allen, Pelot, Ronald et Fournier, Melanie. **2014d**, « Global maritime situational awareness », in : *Global maritime security : new horizons*, url : https://hal.archives-ouvertes.fr/hal-01740722.

[**Ch7**] Devogele, Thomas, **Etienne**, Laurent et Ray, Cyril. **août 2013**, « Maritime monitoring », in : *Mobility Data : Modelling, Management, and Understanding*, Cambridge University Press, p. 224-243, url : https://hal.archives-ouvertes.fr/hal-01170999.

[**Ch8**] **Etienne**, Laurent, Devogele, Thomas et Alain, Bouju. **2012b**, « Spatio-temporal Trajectory Analysis of Mobile Objects Following the same Itinerary », in : *Advances in Geo-Spatial Information Science, Chapter Modeling Space and Time*, sous la dir. de Shi, Goodchild, Lees et Leung, p. 47-58, isbn : 978-0-415-62093-2, url : https://hal.archives-ouvertes.fr/hal-01740733.

1.5.2 Articles de revues

[**Re1**] Cheaitou, Ali, Faury, Olivier, **Etienne**, Laurent, Fedi, Laurent, Rigot-Müller, Patrick et Stephenson, Scott. **nov. 2022**, « Impact of CO2 emission taxation and fuel types on Arctic shipping attractiveness », in : *Transportation Research Part D : Transport and Environment* 112, p. 103491, doi : 10.1016/j.trd.2022.103491, url : https://hal.archives-ouvertes.fr/hal-03912714.

[**Re2**] Rigot-Müller, Patrick, Cheaitou, Ali, **Etienne**, Laurent, Faury, Olivier et Fedi, Laurent. **jan. 2022**, « The role of polarseaworthiness in shipping planning for infrastructure projects in the Arctic : The case of Yamal LNG plant », in : *Transportation Research Part A : Policy and Practice* 155, p. 330-353, doi : 10.1016/j.tra.2021.11.009, url : https://hal.archives-ouvertes.fr/hal-03549995.

[**Re3**] Duroudier, Sylvestre, Chardonnel, Sonia, Mericskay, Boris, Andre-Poyaud, Isabelle, Bedel, Olivier, Depeau, Sandrine, Devogele, Thomas, **Etienne**, Laurent, Lepetit, Arnaud, Moreau, Clément, Pelletier, Nicolas, Ployon, Estelle et Tabaka, Kamila. **2020a**, « Diagnostic qualité et apurement des données de mobilité quotidienne issues de l'enquête mixte et longitudinale Mobi'Kids », in : *Revue Internationale de Géomatique* 30.*1-2*, p. 127-148, doi : 10.3166/rig.2020.00105, url : https://hal.archives-ouvertes.fr/hal-03254291.

[**Re4**] Fedi, Laurent, Faury, Olivier et **Etienne**, Laurent. **avr. 2020**, « Mapping and analysis of maritime accidents in the Russian Arctic through the lens of the Polar Code and POLARIS system », in : *Marine Policy* 118, p. 103984, doi : 10.1016/j.marpol.2020.103984, url : https://hal.archives-ouvertes.fr/hal-02885952.

[**Re5**] Faury, Olivier, Fedi, Laurent, **Etienne**, Laurent, Rigot-Müller, Patrick, Stephenson, Scott et Cheaitou, Ali. **2019f**, « Polaris, Quelle influence sur la sécurité de la na-

vigation en Arctique ? », in : *Journal de la Marine Marchande 5093*, p. 10-11, url : https://hal.archives-ouvertes.fr/hal-02110303.

[**Re6**] Moreau, Clément, Devogele, Thomas et **Etienne**, Laurent. **jan. 2019**, « Calcul de similarité sémantique entre trajectoires », in : *Revue Internationale de Géomatique* 29.*1*, p. 107-127, doi : 10.3166/rig.2019.00077, url : https://hal.archives-ouvertes.fr/hal-02885967.

[**Re7**] Bisone, Frédérick, **Etienne**, Laurent et Devogele, Thomas. **déc. 2018**, « Modélisation et extraction de la sémantique des trajectoires à partir de données multi-capteurs », in : *Revue Internationale de Géomatique* 28.*4*, p. 461-483, doi : 10.3166/rig.2018.00065, url : https://hal.archives-ouvertes.fr/hal-02110261.

[**Re8**] Fedi, Laurent, **Etienne**, Laurent, Faury, Olivier, Rigot-Müller, Patrick, Stephenson, Scott et Cheaitou, Ali. **2018b**, « Arctic navigation : stakes, benefits and limits of the polaris system », in : *Journal of Ocean Technology* 13.*4*, url : https://hal.archives-ouvertes.fr/hal-02110281.

[**Re9**] **Etienne**, Laurent, Devogele, Thomas, Buckin, Maike et Mcardle, Gavin. **2016c**, « Trajectory Box Plot : a new pattern to summarize movements », in : *International Journal of Geographical Information Science*, Analysis of Movement Data 30.*5*, p. 835-853, doi : 10.1080/13658816.2015.1081205, url : https://hal.archives-ouvertes.fr/hal-01215945.

[**Re10**] **Etienne**, Laurent, Devogele, Thomas et Mcardle, Gavin. **2014b**, « Oriented spatial box plot, a new pattern for points clusters », in : *International Journal of Business Intelligence and Data Mining* 9.*3*, http://www.inderscience.com/info/inarticle.php?artid=68367, doi : 10.1504/IJBIDM.2014.068367, url : https://hal.archives-ouvertes.fr/hal-01172635.

[**Re11**] **Etienne**, Laurent. **2013b**, « Motifs spatio-temporels de trajectoires, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime. », in : *Cartes & géomatique*, url : https://hal.archives-ouvertes.fr/hal-01627361.

[**Re12**] Devogele, Thomas et **Etienne**, Laurent. **2012a**, « Mesures de similarité de trajectoires basées sur l'utilisation de patrons spatio-temporels », in : *Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes*

d'Information 17.*1*, p. 11-34, url : https://hal.archives-ouvertes.fr/hal-01171533.

[**Re13**] **Etienne**, Laurent, Devogele, Thomas et Bouju, Alain. **2009**, « Analyse de similarité de trajectoires d'objets mobiles suivant le même itinéraire : Application aux trajectoires de navires », in : *Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information* 14.*5/2009*, pp.85-106, url : https://hal.archives-ouvertes.fr/hal-00389481.

1.5.3 Conférences internationales

[**CI1**] Moreau, Clément, Devogele, Thomas, De Runz, Cyril, Peralta, Veronika, MOREAU, Evelyne et **Etienne**, Laurent. **juill. 2021**, « A Fuzzy Generalisation of the Hamming Distance for Temporal Sequences », in : *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Luxembourg, France : IEEE, p. 1-8, doi : 10.1109/FUZZ45933.2021.9494445, url : https://hal.archives-ouvertes.fr/hal-03319236.

[**CI2**] Faury, Olivier, Cheaitou, Ali, **Etienne**, Laurent, Fedi, Laurent et Rigot-Muller, Patrick. **juin 2020**, « The loading capacity of convoy for the transit of container along the Northeast Passage », in : *28th Annual Conference of the International Association of Maritime Economists*, Hong Kong, China, url : https://hal.archives-ouvertes.fr/hal-03355136.

[**CI3**] Fedi, Laurent, Faury, Olivier, Cheaitou, Ali, **Etienne**, Laurent et Rigot-Muller, Patrick. **juin 2020**, « Application and analysis of the IMO taxonomy on casualty investigation over 20 years of marine events in the Russian and Norwegian Arctic », in : *28th Annual Conference of the International Association of Maritime Economists*, Hong Kong, China, url : https://hal.archives-ouvertes.fr/hal-03355149.

[**CI4**] Moreau, Clément, Devogele, Thomas, Peralta, Veronika et **Etienne**, Laurent. **mars 2020**, *A Contextual Edit Distance for Semantic Trajectories*, ACM Symposium On Applied Computing, Poster, doi : 10.1145/3341105.3374125, url : https://hal.archives-ouvertes.fr/hal-02382303.

[**CI5**] Bisone, Frédérick, Devogele, Thomas et **Etienne**, Laurent. **nov. 2019**, « From Raw Sensor Data to Semantic Trajectories », in : *5th ACM SIGSPATIAL In-*

ternational Workshop on the Use of GIS in Emergency Management (EM-GIS 2019), Chicago, United States, doi : 10.1145/3356998.3365777, url : https://hal.archives-ouvertes.fr/hal-02380444.

[**CI6**] Faury, Olivier, Cheaitou, Ali, **Etienne**, Laurent, Fedi, Laurent, Rigot-Muller, Patrick et Stephenson, Scott. **juin 2019**, « How attractive is the Northern Sea Route for container shipping ? An economic model », in : *27th Annual Conference of the International Association of Maritime Economists*, Athens, Greece, url : https://hal.archives-ouvertes.fr/hal-03354313.

[**CI7**] Fedi, Laurent, Faury, Olivier, **Etienne**, Laurent et de FERRIERE le VAYER, Amaury. **juin 2019**, « Mapping and analysis of maritime claims in the Russian Arctic based on POLARIS System », in : *27th Annual Conference of the International Association of Maritime Economists*, Athens, Greece, url : https://hal.archives-ouvertes.fr/hal-03354308.

[**CI8**] Rigot-Müller, Patrick, **Etienne**, Laurent, Faury, Olivier et Stephenson, Scott. **juin 2018**, « Ship routing and scheduling for the assembly of a LNG plant in the arctic : a decision support system », in : *25th EUROMA Conference*, Budapest, Hungary, url : https://hal.archives-ouvertes.fr/hal-03355195.

[**CI9**] Devogele, Thomas, **Etienne**, Laurent, Esnault, Maxence et Lardy, Florian. **nov. 2017**, « Optimized Discrete Fréchet Distance between trajectories », in : *6th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, Redondo Beach, United States, 9 pages, doi : 10.1145/3150919.3150924, url : https://hal.archives-ouvertes.fr/hal-01631192.

[**CI10**] Boulakbech, Marwa, Messai, Nizar, Sam, Yacine, Devogele, Thomas et **Etienne**, Laurent. **juin 2016**, « SmartLoire : A Web Mashup Based Tool for Personalized Touristic Plans Construction », in : *2016 IEEE 25th International Conference on Enabling Technologies : Infrastructure for Collaborative Enterprises (WETICE)*, Paris, France : IEEE, p. 259-260, doi : 10.1109/WETICE.2016.66, url : https://hal.archives-ouvertes.fr/hal-01512953.

[**CI11**] Stoddard, M A, **Etienne**, Laurent, Fournier, Melanie, Pelot, R et Beveridge, L. **2015b**, « Making sense of Arctic maritime traffic using the Polar Operational Limits Assessment Risk Indexing System (POLARIS) », in : *9th Symposium of the International Society for Digital Earth (ISDE)*, t. 34, IOP Conf. Series : Earth and Environmental Science 1, Halifax, Canada, p. 012034, doi : 10.1088/1755-

1315/34/1/012034, url : https://hal.archives-ouvertes.fr/hal-01512696.

[**CI12**] Stoddard, Mark, **Etienne**, Laurent et Pelot, Ronald. **2015c**, « From Sensing to Sense-Making : Assessing and visualizing ship operational limitations in the Canadian Arctic using open-access ice data », in : *Arctic Council International Conference on Safe and Sustainable Shipping in a Changing Arctic Environment (ShipArc 2015)*, Malmö, Sweden, url : https://hal.archives-ouvertes.fr/hal-01627369.

[**CI13**] **Etienne**, Laurent, Devogele, Thomas et Mcardle, Gavin. **juin 2014**, « State of the Art in Patterns for Point Cluster Analysis », in : *Computational Science and Its Applications–ICCSA*, Guimaraes, Portugal, doi : 10.1007/978-3-319-09144-0_18, url : https://hal.archives-ouvertes.fr/hal-01118544.

[**CI14**] **Etienne**, Laurent et Pelot, Ronald. **2013c**, « Simulation of maritime paths taking into account ice conditions in the Arctic », in : *11th International Symposium for GIS and Computer Cartography for Coastal Zone Management (CoastGIS)*, Victoria, Canada, url : https://hal.archives-ouvertes.fr/hal-01740751.

[**CI15**] Ray, Cyril, Grancher, Arnaud, Thibaud, Rémy et **Etienne**, Laurent. **2013e**, « Spatio-Temporal Rule-based Analysis of Maritime Traffic », in : *Third Conference on Ocean & Coastal Observation : Sensors and Observing Systems, Numerical Models and Information (OCOSS)*, Nice, France, url : https://hal.archives-ouvertes.fr/hal-01627352.

[**CI16**] **Etienne**, Laurent, Ray, Cyril et Mcardle, Gavin. **2011b**, « Spatio-temporal visualisation of outliers », in : *international workshop on Maritime Anomaly Detection (MAD)*, t. 1, Tilburg, Netherlands, p. 1-2, url : https://hal.archives-ouvertes.fr/hal-01740748.

[**CI17**] **Etienne**, Laurent, Devogele, Thomas et Bouju, Alain. **mai 2010**, « Spatio-temporal trajectory analysis of mobile objects following the same itinerary », in : *Joint International Conference on Theory, Data Handling and Modelling in GeoSpatial Information Science*, Hong Kong, Hong Kong SAR China, pp 86-91, url : https://hal.archives-ouvertes.fr/hal-00495484.

[**CI18**] Lécuyer, Anatole, Burkhardt, Jean-Marie et **Etienne**, Laurent. **2004**, « Feeling bumps and holes without a haptic interface : the perception of pseudo-haptic textures », in : *SIGCHI conference on Human factors in computing systems*,

New York, United States, url : https://hal.archives-ouvertes.fr/hal-01627372.

1.5.4 Conférences nationales

[**CN1**] Duroudier, Sylvestre, Chardonnel, Sonia, Mericskay, Boris, André-Poyaud, Isabelle I., Bedel, Olivier, Depeau, Sandrine, Devogele, Thomas, **Etienne**, Laurent, Lepetit, Arnaud, Moreau, Clément, Pelletier, Nicolas, Ployon, Estelle et Tabaka, Kamila. **nov. 2019**, « Données hétérogènes de mobilités quotidiennes : protocole de diagnostic qualité et d'apurement à partir de la base MOBI'KIDS », in : *Spatial Analysis and GEOmatics*, SAGEO, Clermont-Ferrand, France, url : https://halshs.archives-ouvertes.fr/halshs-02327295.

[**CN2**] Moreau, Clément, Devogele, Thomas et **Etienne**, Laurent. **nov. 2018**, « Extraction de motifs de trajectoires sémantiques similaires », in : *Spatial Analysis and Geomatics*, Montpellier, France, url : https://hal.archives-ouvertes.fr/hal-02110019.

[**CN3**] Bisone, Frédérick, **Etienne**, Laurent et Devogele, Thomas. **nov. 2017**, « Extraction automatique de la sémantique des trajectoires », in : *Spatial Analysis and GEOmatics 2017*, INSA de rouen, Rouen, France, url : https://hal.archives-ouvertes.fr/hal-01643365.

[**CN4**] Devogele, Thomas, Esnault, Maxence et **Etienne**, Laurent. **nov. 2016**, « Distance discrète de Fréchet optimisée », in : *Spatial Analysis and Geomatics (SAGEO)*, Nice, France, url : https://hal.archives-ouvertes.fr/hal-02110055.

[**CN5**] **Etienne**, Laurent et Devogele, Thomas. **jan. 2014**, « Trajectoires médianes », in : *14 ème conférence Extraction et Gestion des Connaissances, Ateliers fouille de données spatiales et temporelles & construction, enrichissement et exploitation de ressources géographiques pour l'analyse de données*, Rennes, France, url : https://hal.archives-ouvertes.fr/hal-02110086.

[**CN6**] Devogele, Thomas et **Etienne**, Laurent. **nov. 2010**, « Mesures de similarité de trajectoires suivant le même itinéraire », in : *Conférence internationale de Géomatique et Analyse Spatiale (SAGEO) 2010*, Toulouse, France, p. 155-169, url : https://hal.archives-ouvertes.fr/hal-03357887.

[**CN7**] **Etienne**, Laurent, Devogele, Thomas et Bouju, Alain. **jan. 2010**, « Analyse temps réel du comportement d'objets mobiles évoluant dans un espace ouvert », in : *5ème atelier Représentation et raisonnement sur le temps et l'espace (RTE)*, Caen, France, url : `https://hal.archives-ouvertes.fr/hal-03357906`.

[**CN8**] **Etienne**, Laurent, Devogele, Thomas et Bouju, Alain. **déc. 2008**, « Outil d'aide aux décideurs concernant le suivi de navires : suivi de trajectoires relatives entre navires et détection de trajectoires inhabituelles », in : *7ème journées scientifiques et techniques du CETMEF*, Paris, France, url : `https://hal.archives-ouvertes.fr/hal-03357871`.

1.5.5 Rapports

[**Ra1**] **Etienne**, Laurent. **2019c**, *Schéma Départemental d'Analyse et de Couverture des Risques (2019-2023)*, Rapport opérationnel, Service Départemental d'Incendie et de Secours d'Indre-et-Loire, 150 p.

[**Ra2**] **Etienne**, Laurent, Pelot, Ronald et Cecilia, Engler. **2013d**, *Analysis of Marine Traffic along Canada's Coasts : Phase 2 - Part 2 : A spatio-temporal simulation model for forecasting marine traffic in the Canadian Arctic in 2020*, Rapport de recherche, Defense R&D Canada, Centre for Operational Research & Analysis, 182 p.

[**Ra3**] **Etienne**, Laurent. **déc. 2011**, « Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime. », Thèse de doctorat, Université de Bretagne occidentale - Brest, 209 p., url : `https://tel.archives-ouvertes.fr/tel-00667953`.

[**Ra4**] **Etienne**, Laurent. **2003**, *Étude de l'interaction des sens de la vision et du toucher en réalité virtuelle*, Thèse de master, Université de Rennes 1, 96 p.

1.5.6 Brevets et dépôts

[**Br1**] Lécuyer, Anatole, **Etienne**, Laurent et Arnaldi, Bruno. **2011c**, « Logiciel PseudoHaptik »,
IDDN.FR.001.260014.000.S.P.2011.000.10000 (France).

[**Br2**] Lécuyer, Anatole, **Etienne**, Laurent et Arnaldi, Bruno. **avr. 2005**, « Modulation of Cursor Position in Video Data for a Computer Screen », WO 2005/031556 A1 (France), url : `https://hal.archives-ouvertes.fr/hal-03361547`.

1.6 Bilan de l'activité d'enseignement

Mon activité d'enseignement en informatique et en géomatique a été principalement réalisée en écoles d'ingénieurs (CNAM, École Navale, Lycée Naval, Dalhousie University - Canada, Polytech Tours, ISEN Brest). Les enseignements réalisés sont diversifiés et incluent des cours magistraux, des travaux dirigés, des travaux pratiques et des encadrements de projets/stages pour un volume horaire total d'environ 1900h. J'ai eu l'occasion d'enseigner à des publics variés de niveaux différents dont la répartition est présentée dans le Tableau 1. Cette multitude de niveaux et de publics m'a permis d'adapter mes cours et ma pédagogie à différents degrés de complexité allant de l'initiation à la formation avancée et au perfectionnement. De plus, j'ai encadré et évalué de nombreux projets d'élèves ingénieurs en informatique et en géomatique ainsi que des stagiaires de niveau master. Au cours de ma carrière, j'ai également eu la responsabilité éditoriale de plusieurs unités d'enseignements et cours en informatique et géomatique dont certains ont été conçus et enseignés en anglais. J'ai également publié un chapitre d'ouvrage pédagogique traitant de la manipulation de données maritimes au sein des bases de données relationnelles [**Ch1**].

Table 1.2 – Répartition de l'activité d'enseignement

Établissement	Situation	Période	Public concerné	Volume
ISEN Brest	Permanent	Depuis 2019	Élèves ingénieurs (L2-M2)	250h/an
Polytech Tours	Permanent	2014 - 2019	Élèves ingénieurs (L1-M2)	192h/an
Dalhousie University	Vacataire	2012 - 2013	Élèves ingénieurs (L3/M1)	100h
Lycée Naval Brest	Permanent	2013 - 2014	Classes préparatoires	90h
École Navale	Permanent	2008 - 2014	Élèves ingénieurs (L3/M1)	150h/an
CNAM Rennes	Vacataire	2004 - 2005	Formation continue (L2)	66h

Chapitre 2

REPRÉSENTATION DES TRAJECTOIRES

L'étude des déplacements peut être réalisée à différentes échelles spatiale, temporelle et sémantique [**67**]. La notion même de trajectoire nécessite donc d'être définie en fonction de ces échelles et plus particulièrement du contexte de l'étude.

En effet, en fonction des modèles d'abstraction utilisés, différentes algèbres pourront être mobilisées pour raisonner sur les différentes composantes des trajectoires (temps, espace, sémantique).

2.1 La représentation temporelle des séquences d'activités

Une première forme d'abstraction de la représentation des déplacements peut être réalisée en se focalisant sur l'aspect temporel.

La représentation la plus simple consiste à définir la relation d'ordre temporel entre différents évènements constituant une séquence comme proposé par Lamport dans [**118**]. Cet ordre temporel est basé sur la notion de *précédence*. Un évènement a précède un évènement b si a s'est produit à un moment antérieur à b. Cette relation transitive est alors notée $a \rightarrow b$.

Une séquence de mobilité $\langle e_1, ..., e_n \rangle$ est constituée des n évènements $e_{i \in \{1,...,n\}}$ si $\forall i, j \in [\![1, n]\!], i < j \Rightarrow e_i \rightarrow e_j$.

Dans cette représentation, la notion de durée n'est pas clairement définie. Il est possible d'étendre la séquence de mobilité en y intégrant la notion d'intervalle disposant d'un début et d'une fin. Ces débuts et fins font référence à un temps absolu ou un temps relatif.

L'ajout de la notion d'intervalle temporel complexifie les relations entre les éléments constitutifs d'une séquence de mobilité. L'algèbre de Allen[**117**] étend la notion de *précédence* en définissant de nouvelles relations entre intervalles temporels présentées dans la

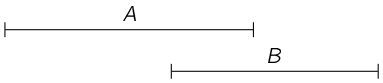
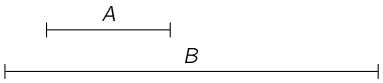
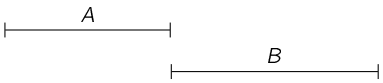
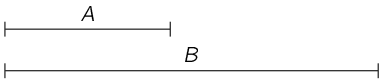
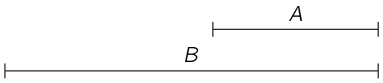
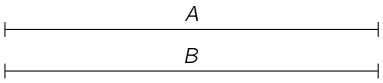
Nom			Définition	Image
se déroule avant	$b(A,B)$	=	$a^+ < b^-$	
chevauche	$o(A,B)$	=	$a^- < b^- \wedge b^- < a^+ \wedge a^+ < b^+$	
se déroule pendant	$d(A,B)$	=	$b^- < a^- \wedge a^+ < b^+$	
rencontre	$m(A,B)$	=	$a^+ = b^-$	
débute	$s(A,B)$	=	$a^- = b^- \wedge a^+ < b^+$	
termine	$f(A,B)$	=	$a^+ = b^+ \wedge b^- < a^-$	
est égal à	$a(A,B)$	=	$a^- = b^- \wedge a^+ = b^+$	
se déroule après	$bi(A,B)$	=	$b(B,A)$	
est chevauché par	$oi(A,B)$	=	$o(B,A)$	
contient	$di(A,B)$	=	$d(B,A)$	
est rencontré par	$mi(A,B)$	=	$m(B,A)$	
commence par	$si(A,B)$	=	$s(B,A)$	
se termine par	$fi(A,B)$	=	$f(B,A)$	

Table 2.1 – Relations entre intervalles temporels de l'algèbre d'Allen. $A = [a^-, a^+]$ et $B = [b^-, b^+]$

table 2.1 extraite de [**4**]. Dans cette représentation, un élément A dispose d'un début noté a^+ et d'une fin a^-, d'un intervalle temporel $[a^-, a^+]$ et d'une *durée* δ.

Ces relations temporelles entre éléments constitutifs d'une séquence de mobilité sont particulièrement intéressantes. Elles permettent d'affiner la notion de *précédence* et d'assurer la continuité et l'unicité temporelle dans les séquences via l'usage de la relation rencontre (meet). Ainsi, tout évènement d'une séquence débute immédiatement après le précédent. La fin d'un évènement correspondant toujours au début d'un autre. Cependant, dans certains cas particuliers, il est parfois possible de réaliser plusieurs activités en même temps. Enfin cette algèbre temporelle permet également de comparer les intervalles de séquences d'activités différentes et de définir des relations entre ces activités [**90**].

Ces intervalles peuvent être qualifiés avec des attributs sémantiques décrivant plus ou moins finement l'activité réalisée comme par exemple, les notions de STOP et de MOVE

qui permettent de représenter la mobilité sous la forme d'une séquence d'activités comme exposé dans le chapitre 3. Des descriptions sémantiques plus fines s'appuyant sur des ensembles d'attributs sont présentées au chapitre 4.

2.2 Le modèle Time-Geography

Dans les années 1970, le modèle Time-Geography de Hägerstrand [**125**] s'est intéressé à la représentation des déplacements et activités des individus ainsi que les relations à leurs environnements. Différentes notions de contraintes exercées par l'environnement et/ou l'individu sont introduites. Ces contraintes ayant alors un impact sur la mobilité et les activités journalières réalisées :

- des contraintes de capacité limitent les activités de l'individu en fonction de ses capacités physiques (manger, dormir, moyen de déplacement),
- des contraintes d'interaction qui nécessitent la présence conjointe de plusieurs individus dans le même espace et au même temps afin de mener une activité particulière (par exemple le travail, conduire les enfants à l'école),
- des contraintes d'autorité qui restreignent l'accès à certains espaces spatio-temporels (horaires d'ouverture d'un magasin par exemple).

L'axe Z d'un espace tridimensionnel peut être utilisé pour représenter le temps. Cet espace est nommé cube spatio-temporel (space-time cube) [**31**]. Le parcours spatio-temporel (space-time path) décrit un ensemble d'activités réparties dans le temps et l'espace et réalisables dans une durée limitée. Les stations d'activités sont représentées sur la figure 2.1(a) par des tubes décrivant leur emplacement dans l'espace et leur disponibilité dans le temps. Le concept du prisme spatio-temporel (space-time prism) permet de représenter la zone spatio-temporelle dans laquelle un individu est susceptible de se déplacer entre deux localisations connues nommées ancres spatio-temporelles (prism anchor) et en fonction de ses capacités de déplacement (mode de transport) et d'un budget temps alloué à une activité statique (figure 2.1(b)). Ce concept est nécessaire lorsque l'on ne dispose pas du parcours détaillé réalisé par l'individu entre ces deux localisations. Dans le chapitre 6, ces concepts ont été étendus pour représenter un ensemble de trajectoires évoluant dans un cube spatio-temporel.

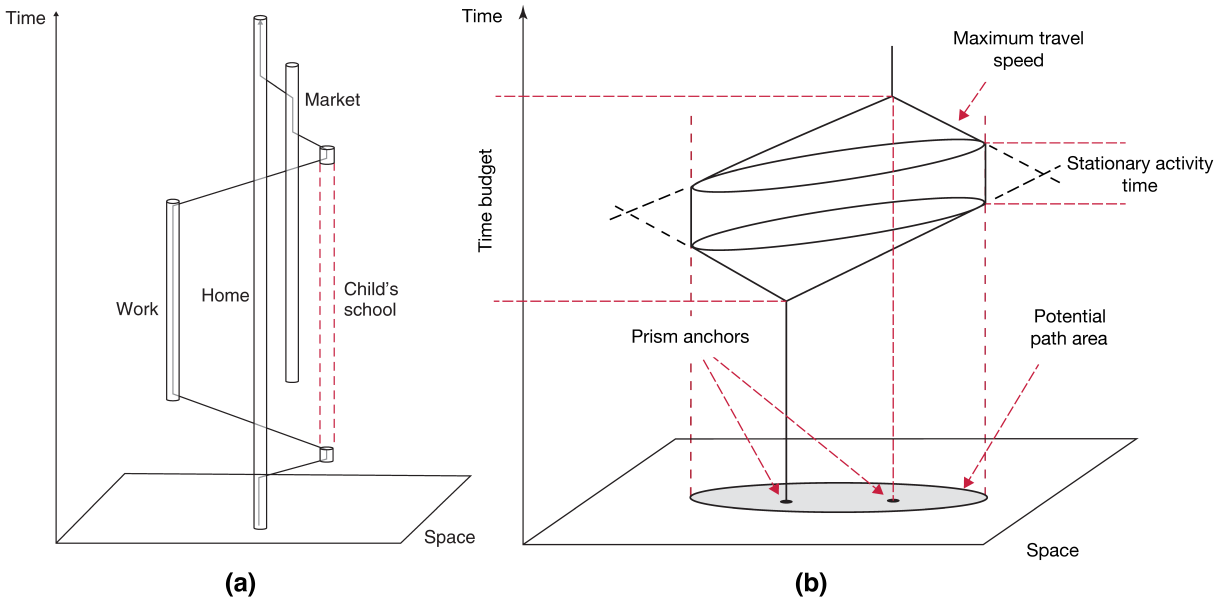


Figure 2.1 – Concepts fondamentaux de la Time-Geography : (a) le parcours spatio-temporel et (b) le prisme spatio-temporel [**31**]

Cependant, ce modèle vise principalement à représenter les contraintes physiques permettant de décrire la mobilité des individus dans leur milieu. Il ne permet pas d'en déduire des comportements de mobilité.

2.3 Trace spatiale

Avec l'essor des objets connectés et des technologies de géolocalisation Global Navigation Satellite System (GNSS), les positions des individus sont désormais accessibles de manière plus détaillée (spatialement et temporellement). Ces positions peuvent être vues comme des données "brutes" provenant d'un capteur Global Positionning System (GPS).

Les objets mobiles évoluent dans un certain espace (2D, 3D, aérien, sous-marin, réseau routier...). La localisation géographique de l'objet au sein de cet espace peut être représentée à l'aide de coordonnées c formulées en fonction d'un repère spécifique. Ces coordonnées peuvent être exprimées au format géographique World Geodesic System 1984 (WGS84) sous la forme d'une latitude et d'une longitude $c = (\varphi, \lambda)$, dans le repère cartésien $c = (x, y, (z))$ ou dans n'importe quel autre Système de Coordonnées de Référence (SCR) permettant de localiser la position de l'objet dans l'espace. Lorsqu'un objet se déplace, ses coordonnées évoluent au cours du temps. La position p de l'objet combine

l'estampille temporelle t et les coordonnées c de l'objet à cet instant. On pourra alors simplifier la notation d'une position sous la forme d'un tuple $p = (x, y, z, t)$.

La suite temporellement ordonnée de positions discrètes des individus permet de reconstruire leur trace spatiale en connectant toutes les positions par des segments de lignes [**69**]. Il est également possible de déduire de ces données de positions certains paramètres caractérisant la mobilité de l'individu tels que sa vitesse instantanée, son accélération, sa direction, mais également des paramètres agrégés sur des intervalles de temps (vitesse moyenne, angle de rotation, sinuosité) [**85**, **73**, **65**].

Dans certains cas, l'échantillonnage régulier des systèmes de positionnement par satellite n'est pas forcement le plus adapté pour représenter la trace spatiale parcourue par un objet. En effet, lorsqu'un objet reste longtemps au même endroit, il peut être intéressant de simplifier la représentation spatiale en supprimant les doublons. De même lorsqu'un objet se déplace à vitesse constante sur une ligne droite, il n'est nécessaire de conserver que le premier et le dernier élément de la ligne droite. Le filtre de Douglas et Peucker [**124**] et son extension temporelle [**71**, **75**] sont particulièrement efficaces pour limiter le nombre de positions nécessaire à la représentation d'une trajectoire tout en conservant une marge d'erreur inférieure à un paramètre ϵ fixé.

Les travaux portant sur l'étude et la comparaison de ces traces spatiales sont présentés plus en détail dans le chapitre 5.

2.4 Intégration de la sémantique

Pour mieux comprendre les comportements en lien avec la mobilité, il est nécessaire de s'intéresser à l'aspect sémantique de celle-ci.

Les travaux de Peuquet [**106**] se sont intéressés à cet aspect sémantique en introduisant une triade conceptuelle basée sur une représentation de l'objet (What), de l'emplacement (Where) et du temps (When). Ces travaux ont été étendus par Claramunt, Parent et Thériault dans [**95**] pour y intégrer le comment (How) et le pourquoi (Why) un individu agit.

Avec l'essor des objets connectés et des technologies de géolocalisations, les positions précises des individus sont désormais collectées et sauvegardées au sein de bases de données d'objets mobiles ouvrant la voie à l'analyse et à la fouille de données.

Ces processus de fouille de données visent à extraire des comportements et des motifs

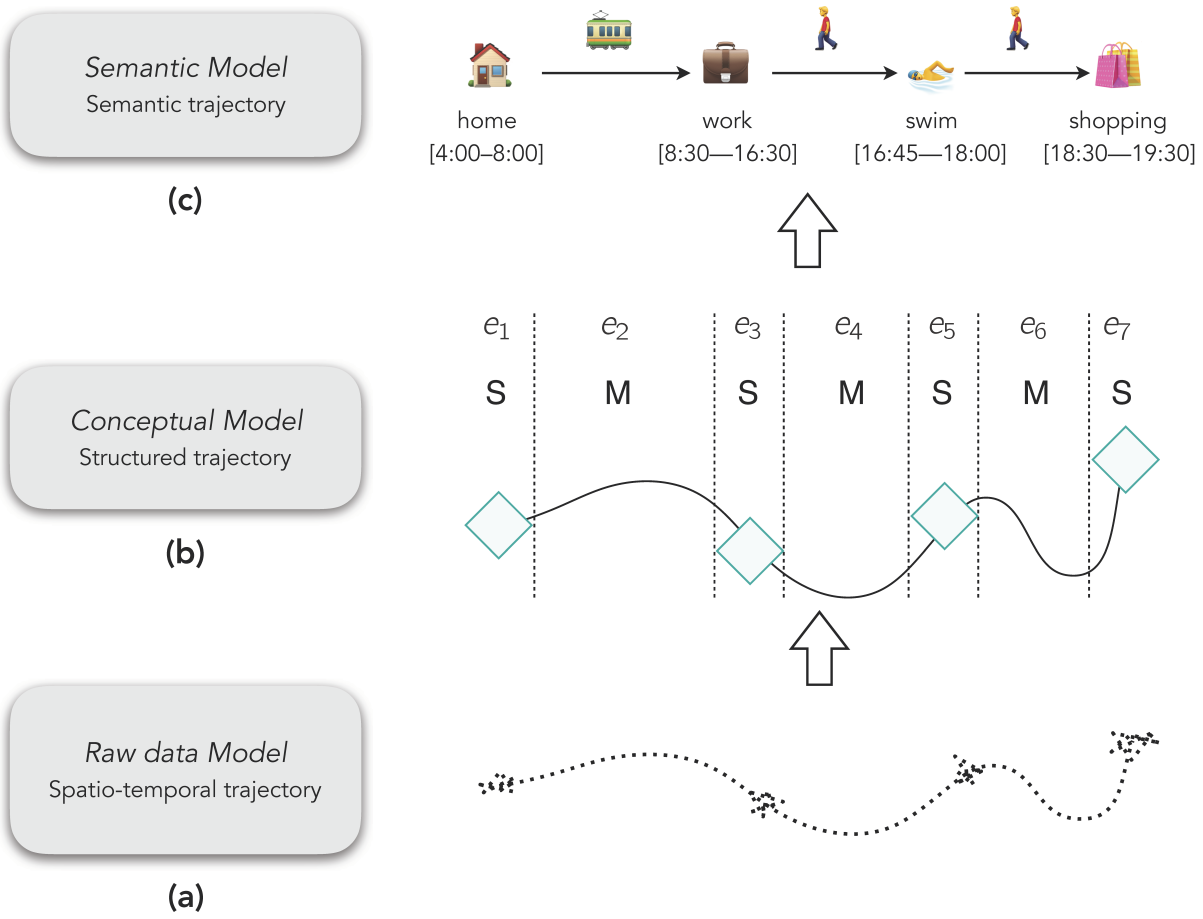


Figure 2.2 – Différentes représentations de trajectoires selon [**48**] : (a) Trace brute points GPS (b) Séquences d'épisodes *stop-move* (c) Séquences d'épisodes avec annotations sémantiques

de mobilité compréhensibles.

Certains objets connectés fournissent également des informations concernant le contexte dans lequel les objets mobiles évoluent (température, pollution de l'air, niveau sonore...). La combinaison des données de positionnement et de contexte aide à mieux comprendre les comportements de mobilité en réalisant un enrichissement sémantique des trajectoires.

Pour réaliser cet enrichissement sémantique, il est nécessaire de mettre en place une chaîne complète de traitement des données brutes de position, illustrée sur la figure 2.2, afin d'obtenir une trajectoire sémantique enrichie [**48**].

La première étape de cette chaîne de traitement consiste a étudier la qualité des données de positions obtenues afin de détecter des données aberrantes et réaliser un apurement [**Re3**]. Dans certains cas, le nombre de positions constituant les traces brutes est

trop important par rapport à l'échelle d'étude ou lorsque les algorithmes utilisés ont une complexité trop élevée. Il est alors nécessaire de réaliser une étape de compression des données [**124**, **13**]. Une fois les traces brutes filtrées, une étape de segmentation peut être réalisée [**70**, **72**, **33**]. Cette étape, détaillée dans le chapitre 3, découpe les traces brutes en une séquence d'arrêts (stops) et de déplacements (moves). Enfin, il existe de nombreuses sources de données contextuelles qui peuvent venir enrichir les séquences de mobilité. Les algorithmes de map-matching visent à associer les déplacements réalisés à un réseau routier [**74**, **39**, **18**]. Des points d'intérêts Point Of Interest (POI) peuvent également être associés aux arrêts [**68**, **55**, **20**] (voir chapitre 3). La météo rencontrée tout au long du parcours peut également être intégrée comme un nouvel enrichissement sémantique. Dans certains cas, l'utilisateur partage volontairement des informations concernant les activités réalisées via les réseaux sociaux [**47**, **40**, **27**] mais également dans son agenda personnel. Enfin, il est possible d'essayer d'inférer les activités réalisées par l'utilisateur à partir de ses traces [**24**] (voir chapitre 3) ou de les annoter à l'aide d'un outil de saisie comme nous l'avons réalisé dans le projet ANR MOBIKIDS.

Le modèle Conceptual model of semantic trajectories (CONSTAnT) proposé par Bogorny et al. dans [**41**] définit la notion de trajectoire sémantique. Le modèle conceptuel de CONSTAnT, présenté dans la figure 2.3, considère les différentes dimensions abordées dans les sections précédentes (intervalle temporel, représentation spatiale, description sémantique). Les travaux de recherche présentés dans les prochains chapitres de ce manuscrit visent à enrichir ces modèles en fournissant de nouveaux outils de synthèse et de comparaison de trajectoires tenant compte de ces 3 dimensions.

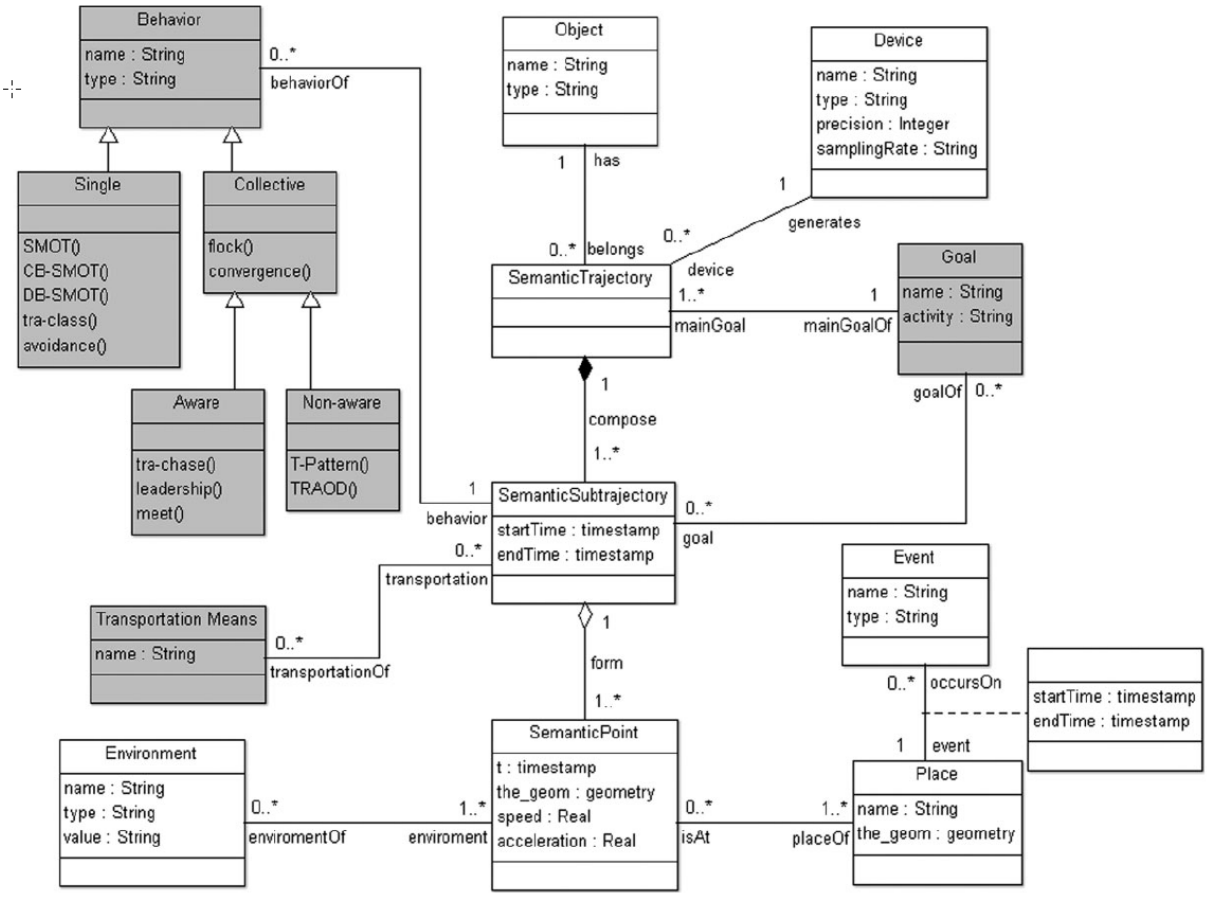


Figure 2.3 – Modèle conceptuel de trajectoire sémantique issu de [**41**]

Chapitre 3

APPORT DES OBJETS CONNECTÉS DANS L'ÉTUDE DES MOBILITÉS

Les objets connectés se développent de plus en plus et nous assistent dans nos tâches quotidiennes. Ces objets disposent de capteurs de plus en plus variés, sophistiqués et performants. Ils disposent de moyens de localisation, de stockage de l'information et de communication particulièrement intéressants dans le cadre de l'étude de la mobilité [**34**]. Ils collectent et transmettent en continu des informations concernant l'environnement dans lequel ils évoluent (bâtiment connecté, véhicule intelligent). Certains de ces objets, comme les smartphones et les montres connectées, restent toute la journée avec leur propriétaire qui les consulte régulièrement.

Les utilisateurs de ces objets connectés peuvent s'en servir pour diffuser volontairement leurs activités et leurs localisations sur des plateformes de partage (parcours sportif, photographie sur les réseaux sociaux, accident ou embouteillages sur la route). Les objets connectés génèrent des données qu'ils peuvent diffuser à d'autres objets interconnectés via différents réseaux (bluetooth, wifi, lorawan, sigfox, mobile, VHF, satellite...) au sein de l'Internet des Objets (IoT) [**36**].

L'étude de ces données produites par l'IoT offre des perspectives nouvelles pour la caractérisation de la mobilité [**16**, **17**]. Cependant, les masses de données produites par ces capteurs hétérogènes induisent différents verrous de recherches en lien avec l'étude des données massives (Big Data). En effet, le volume, la vélocité à laquelle ces données sont produites ainsi que la variété et l'hétérogénéité de celles-ci rendent leur manipulation et leur analyse complexe [**53**]. Enfin, l'utilisation de ces données pose également des problèmes d'éthique et de respect de la vie privée des utilisateurs [**32**]. L'étude de l'apport des objets connectés pour la compréhension fine de la mobilité est au cœur de ce chapitre.

3.1 Analyse de mobilité de véhicules connectés

Les véhicules connectés sont un bon exemple de système IoT générant des masses de données en lien avec la mobilité. En effet, les véhicules récents sont équipés de capteurs chargés de collecter de nombreuses informations concernant l'état du véhicule (vitesse, freins, éclairages, systèmes de sécurité, ceintures, ouverture des portes, système GPS, etc.). Tous ces capteurs transmettent des données à différents calculateurs via un réseau de communication interne au véhicule (Controller Area Network (CAN)). L'analyse en temps réel de certaines de ces données à l'aide de calculateurs permet d'optimiser le système et de détecter des anomalies. Ces données sont généralement utilisées en temps réel pour optimiser le système ou réagir immédiatement en cas d'anomalie (par exemple pour le déclenchement des airbags en cas de choc violent). Elles peuvent être sauvegardées à la manière d'une boîte noire en cas d'incident pour être analysées a posteriori.

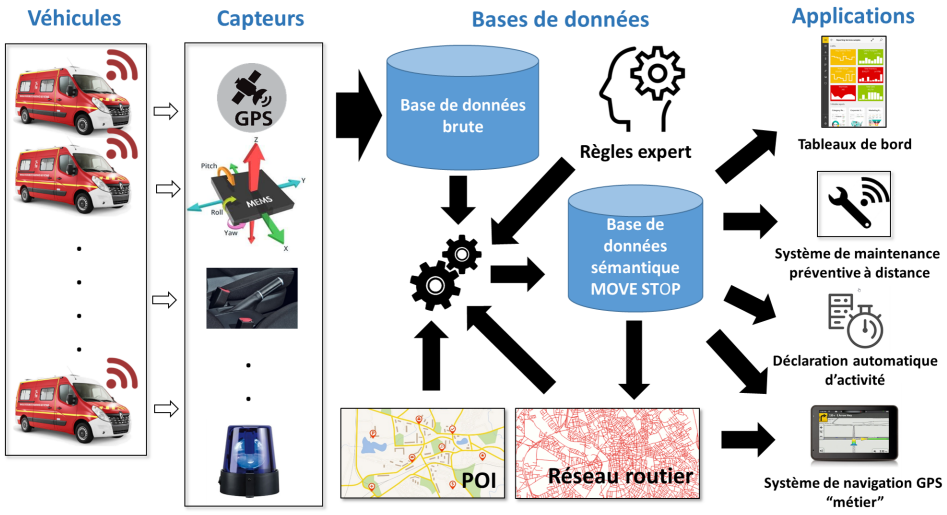


Figure 3.1 – Schéma de synthèse du projet SMART AMBULANCE.

Dans le cadre de la thèse CIFRE "Smart Ambulance" de Frédérick BISONE (1.4.2), réalisée en collaboration entre le Laboratoire d'Informatique Fondamentale et Appliquée de l'Université de Tours (LIFAT), l'entreprise Petit Picot by Gruau et le Service Départemental d'Incendie et de Secours d'Indre et Loire (SDIS 37), nous nous sommes intéressés à l'étude des flux de données générés par des véhicules de secours connectés [**2**]. L'objectif de cette thèse était de proposer des applications spécifiquement adaptées aux services de secours s'appuyant sur un corpus de données généré par une flotte de véhicules de secours

connectés (figure 3.1).

Différents types de données ont été collectés et sauvegardés dans une base de données de trajectoires lors de ce projet :

- Les données issues du châssis du véhicule ont été extraites du bus CAN (vitesse compteur, régime moteur, niveau de carburant, frein à main, portes ouvertes, clignotants, feux...),
- les données de positionnement du véhicule ont été récupérées depuis le système de navigation (GPS, accéléromètres),
- les données spécifiques au véhicule de secours ont été obtenues via le bus CAN auxiliaire (gyrophares, sirène, brancard...)

Le jeu de données a été collecté par deux ambulances connectées suivies pendant 18 jours. Ces ambulances ont généré 2,6M lignes de données structurées décrivant l'état des différents capteurs du véhicule tout au long de ces journées. L'emprise spatiale du jeu de données est présentée sur la figure 3.2.

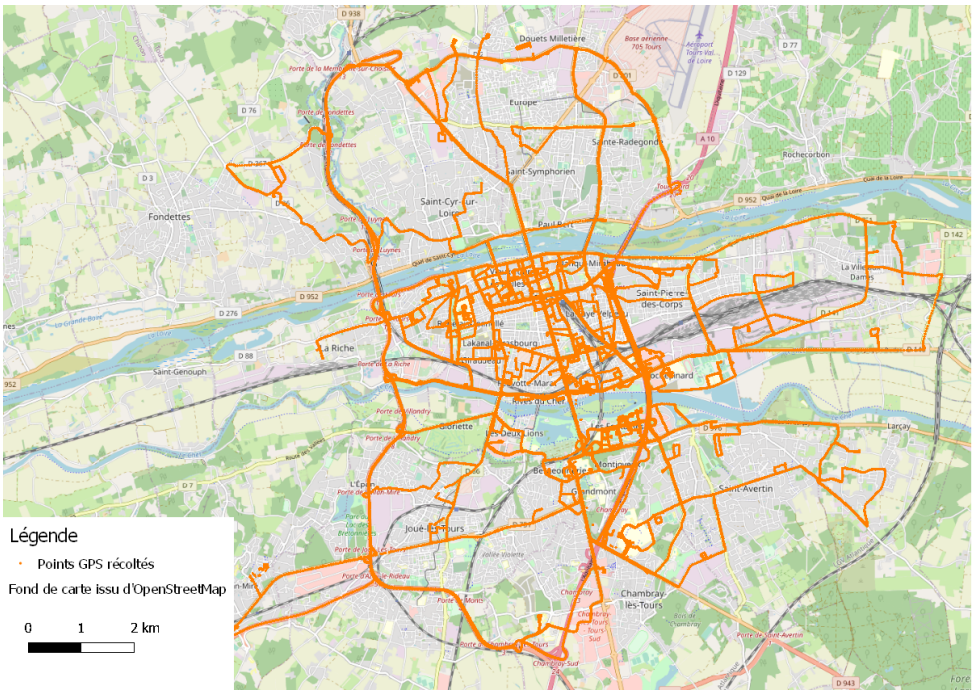


Figure 3.2 – Représentation géographique des données collectées dans le cadre de la thèse SMART AMBULANCE.

À partir de ces données collectées (figure 3.3.a), différents processus d'enrichissement

sémantique ont été réalisés. Le premier enrichissement consiste à détecter les moments ou le véhicule s'est arrêté (figure 3.3.b) puis d'y associer des lieux d'intérêt et en déduire des activités réalisées (figure 3.3.c).

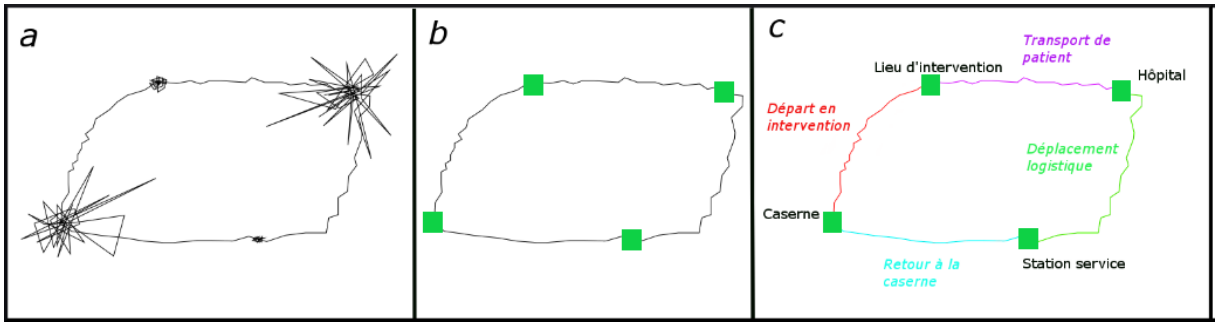


Figure 3.3 – Processus d'enrichissement sémantique.

3.2 Segmentation des trajectoires

La segmentation des trajectoires brutes, constituées principalement d'une suite de positions GPS, à pour objectif de différencier les épisodes de mouvements (move) des arrêts stationnaires (stop).

Il existe différentes méthodes pour réaliser cette segmentation [**21**, **22**, **30**]. Le principal verrou de cette étape du processus de segmentation consiste à définir les différents seuils (spatial et temporel) permettant de considérer que le véhicule est arrêté sachant que les systèmes de géolocalisation produisent des données parfois imparfaites [**14**].

3.2.1 Détection des stops à l'aide des positions GPS

Lorsqu'on dispose uniquement des positions GPS, le calcul de la vitesse moyenne entre deux positions successives est utilisé. Lorsque la vitesse est proche de 0, on peut alors considérer que l'objet ne se déplace plus. Cependant, le calcul de cette vitesse est particulièrement sensible aux variations induites par le système de positionnement GPS ainsi qu'à la fréquence d'échantillonnage utilisée [**23**]. Dans un environnement urbain, le signal GPS est perturbé par la présence d'obstacles (figure 3.4).

De plus, lorsqu'un véhicule entre dans un parking, la qualité du signal GPS est très fortement dégradée voire complètement perdue. Cette perte de qualité induit des sauts

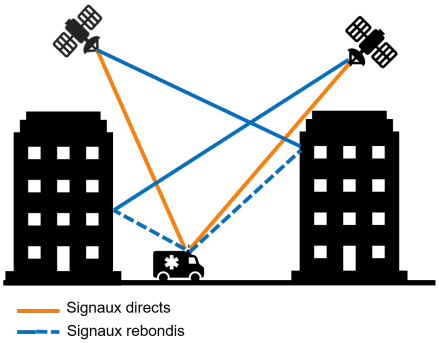


Figure 3.4 – Perturbation du signal GPS.

de positions générant un effet pelote illustré sur la figure 3.5. Dans certains cas, le signal GPS peut être perdu pendant un long moment (stationnement dans un garage). Lorsque le véhicule se remet en mouvement à l'extérieur, le système GPS n'est pas en mesure d'indiquer immédiatement la position du véhicule. En effet, le capteur doit recevoir un nombre minimal de signaux satellites avant de pouvoir calculer une position. Ce temps de latence au démarrage d'un GPS est appelé Time To First Fix (TTFF). Le système GPS dispose également d'un mécanisme permettant de spécifier la qualité du signal GPS reçu par le capteur. L'indice Dilution Of Precision (DOP) permet de calculer l'erreur potentielle de positionnement en fonction du nombre de signaux satellites collectés par le capteur GPS. Toutes ces contraintes rendent la détection des STOPs à l'aide des données GPS complexes et de nombreuses recherches portent sur cette problématique [**78**, **46**, **29**, **61**, **52**].

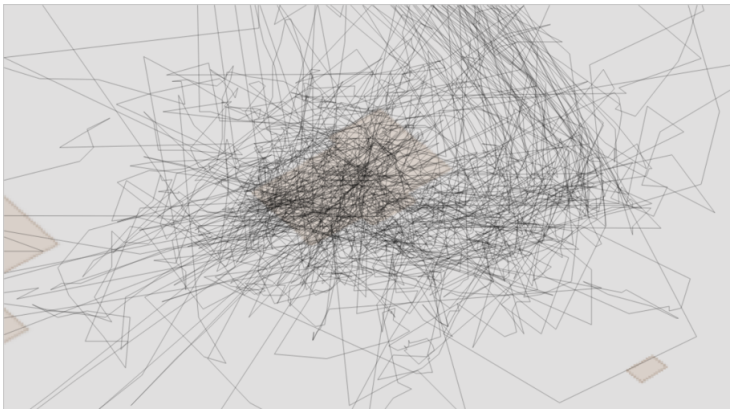


Figure 3.5 – Positions GPS dégradées dans un parking couvert induisant un effet pelote.

C'est pourquoi, lorsqu'on dispose uniquement des positions GPS, une étape de filtrage

et de préparation des données est nécessaire. La définition d'un seuil minimum de vitesse ϵ_v (ou d'une distance minimale entre positions successives) et d'une durée minimale des stops sont requis. La prise en compte de l'indice DOP est également importante pour détecter des données aberrantes.

Dans la littérature, la détection des stops peut être réalisée en recherchant des groupes de positions spatio-temporellement proches et denses. Les approches basées sur l'étude de la densité, telles que DBSCAN [**99**], cherchent à constituer des groupes (clusters) de positions à l'aide de deux paramètres, un seuil de distance minimal ϵ_d entre positions ainsi qu'un nombre minimal de positions $MinPts$ pour constituer un cluster. Encore une fois, ces deux paramètres sont difficiles à définir et dépendent fortement du contexte de l'étude. En effet, lorsqu'un véhicule reste plus ou moins longtemps au même endroit, le nombre de positions observées peut varier de manière significative entre deux clusters. De plus, les véhicules peuvent s'arrêter régulièrement au même endroit. C'est pourquoi il est important de prendre en compte l'aspect temporel des positions en ajoutant une contrainte de seuil temporel ϵ_t entre deux positions pour pouvoir les agréger au sein d'un même cluster comme proposé par dans les algorithmes ST-DBSCAN [**72**] et T-DBSCAN [**42**, **25**]. L'algorithme TrajDBSCAN [**54**] impose des contraintes temporelles encore plus strictes en restreignant la durée minimum $MinDur$ d'un cluster (au lieu d'un nombre minimum de positions $MinPts$) et en ne considérant comme voisin que les positions successives.

3.2.2 Apport des capteurs IoT pour la détection des arrêts

Dans le cas des véhicules connectés, de nombreux capteurs complémentaires au GPS sont disponibles. Le capteur de frein à main est particulièrement intéressant. En effet, lorsque le véhicule est stationné, le frein à main est presque systématiquement enclenché. De même, lorsque le véhicule est stationné dans un garage, il peut être branché sur le courant électrique ce qui permet de confirmer que celui-ci est bien remisé. Le capteur de vitesse permet également de connaître la vitesse du véhicule y compris lorsque le véhicule ne capte pas de signal GPS. Enfin, le capteur de contact moteur indique que le moteur du véhicule est coupé confirmant que le véhicule est bien arrêté et stationné.

La combinaison de tous ces capteurs apporte des informations pertinentes pour la détection plus fiable des arrêts du véhicule et la qualification des épisodes de STOP au sein des trajectoires.

Le modèle de trajectoire défini au chapitre 2 est étendu pour intégrer ces capteurs. Formellement, on considère alors un élément de trajectoire e_i^+ comme une extension d'une position p aux coordonnées c collectée à l'estampille temporelle t et disposant d'un ensemble C de valeurs provenant des capteurs.

$$e_i^+ . \begin{cases} p \begin{cases} t & \rightarrow & \text{estampille temporelle} \\ c & \rightarrow & \text{coordonnées} \end{cases} & \rightarrow & \text{position} \\ freinMain & \rightarrow & \text{état du frein à main à l'instant } t \end{cases} \tag{3.1}$$

L'algorithme de segmentation des trajectoires se focalise sur la détection de changement d'état du capteur de frein à main. Le début d'un potentiel arrêt du véhicule est détecté lorsque le frein à main est serré entre T_i et T_{i+1}. Respectivement, la fin d'un potentiel arrêt est détectée lorsque le frein à main est desserré entre T_i et T_{i+1}.

Le principal avantage de cet algorithme est sa faible complexité $O(n)$. Cependant, il ne permet pas de détecter les épisodes de STOP si le frein à main a été oublié, mais que le véhicule ne bouge pas. En revanche, cette méthode est fonctionnelle lorsque le système GPS ne capte pas (véhicule positionné dans un garage) et permet de segmenter les trajectoires en une succession d'arrêts et de déplacements.

Une fois les trajectoires segmentées, il est alors possible d'analyser plus finement les arrêts afin de les qualifier. Un seuil de durée minimum D_{min}^{arret} est défini en concertation avec des experts métier (120s). Les arrêts dont la durée est inférieure à D_{min}^{arret} sont ignorés et les sous-trajectoires précédant et suivant cet arrêt sont fusionnées. Les micro-arrêts comme les arrêts à un carrefour, feux rouges, démarrage en côte ne sont pas retenus. Les sous-trajectoires ou le capteur de frein à main est serré et dont la durée est supérieure à D_{min}^{arret} sont alors considérées comme des STOPs. De même un seuil de durée minimale de déplacement est défini D_{min}^{dep} (30m). Lorsqu'un déplacement dure très peu de temps entre deux arrêts, celui-ci est supprimé et les deux arrêts sont alors fusionnés.

3.3 Appariement des STOPs à des points d'intérêt

Une fois les épisodes de STOPs définis, il est intéressant de les qualifier plus précisément. La première étape consiste à spécifier leur représentation spatiale. Les données GPS collectées pendant le STOP peuvent varier légèrement en fonction de l'environnement de collecte (effet pelote). Une position centrale représentant l'ensemble du nuage de positions

GPS collectées pendant la durée d'un STOP doit alors être définie. La technique employée pour définir cette position centrale doit être robuste à l'effet pelote observé lorsque le signal GPS est perturbé (principalement lorsque le capteur GPS se trouve en intérieur comme dans un garage).

3.3.1 Représentation spatiale des STOPs

La médiane géométrique des positions GPS collectées durant l'épisode STOP est utilisée [**62**]. Cette médiane correspond aux coordonnées du point qui minimise la somme des distances à tous les autres points du nuage. De plus amples détails concernant la modélisation de patrons spatio-temporels de nuages de positions sont présentés dans le chapitre 6 à la section 6.1.1 de ce manuscrit.

Ainsi, les sous-trajectoires T_i de type STOP sont enrichies d'un nouvel attribut $T_i.MedGeom$ représentant la coordonnée centrale du nuage de positions constituant le STOP. Afin de qualifier la dispersion du nuage, on considère la plus grande distance observée entre les points du nuage et la médiane géométrique en retirant une marge de 5% des points les plus éloignés [**44**]. Cette distance R_{med} correspond au rayon d'un cercle centré sur la médiane géométrique représenté en rouge sur la figure 3.6.

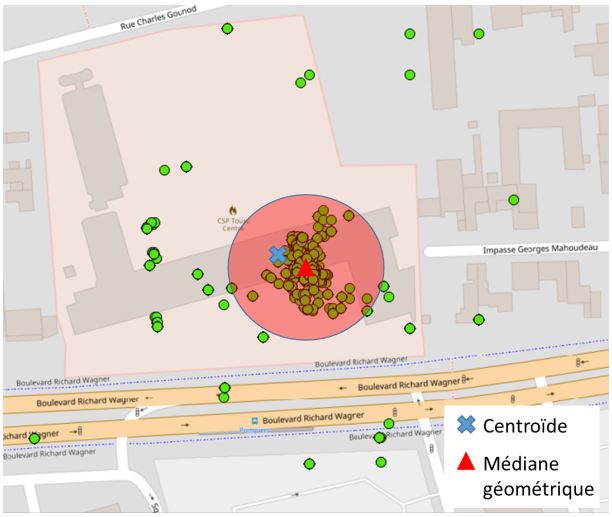


Figure 3.6 – Médiane géométrique et rayon de dispersion d'un nuage de positions correspondant à un STOP.

Disposant de la position centrale et de la dispersion du nuage de positions pour chaque STOPs, il est alors possible de rechercher un point d'intérêt (POI) en lien avec ce STOP.

3.3.2 Appariement des STOPs aux points d'intérêts

Une première étape consiste à extraire une liste de points d'intérêts, en lien avec les activités métiers du véhicule connecté, dans une base de données géographique telle que Open Street Map (OSM) ou la BD TOPO de l'Institut Géographique National (IGN) [**26**]. La base de données d'OSM a été filtrée en utilisant l'attribut *amenity* décrivant les usages des bâtiments [1] [**19**, **57**, **10**].

Les POI intéressants ont été extraits de la base de données OSM à l'aide de l'API Overpass turbo [2]. Ces POI disposent de représentations spatiales (points, lignes ou polygones) et d'attributs descriptifs. Le processus d'appariement utilisé entre les médianes géométriques des STOPs et les POI est décrit dans [**81**]. Dans notre cas d'étude, 92 POI ont été extraits et regroupés en 3 grandes catégories :

- Les casernes de pompiers ;
- Les hôpitaux ;
- Les lieux logistiques (garages, stations services ou de nettoyage).

La recherche du POI le plus proche d'un STOP est réalisée à l'aide d'un calcul de distance limité par un seuil de recherche D_{max}^{POI} (50m). Si aucun POI ne se trouve à une distance inférieure à D_{max}^{POI}, alors le STOP n'est pas associé. Cette situation se produit régulièrement lorsque le STOP correspond à un lieu d'intervention. Si un ou plusieurs POI se trouvent à une distance de moins de D_{max}^{POI}, le POI le plus proche du STOP est sélectionné. Cette solution simple est fonctionnelle, car les POI sélectionnés dans le cas d'étude sont tous relativement éloignés les uns des autres. Il existe d'autres techniques d'appariement se focalisant sur la résolution de ces problèmes d'ambiguïté spatiale [**81**, **66**, **38**].

Les POI attribués aux STOPs disposent de caractéristiques sémantiques particulièrement intéressantes pour qualifier les différents types d'activités potentiellement réalisées par le véhicule lors de ses arrêts. Les types de lieux visités sont également intéressants pour mieux comprendre la sémantique des déplacements réalisés par le véhicule connecté. L'enchaînement de ces séquences d'activités (STOP-MOVE) permet de reconstruire une description sémantique des activités réalisées.

1. `http://wiki.openstreetmap.org/wiki/Key:amenity`
2. `https://overpass-turbo.eu`

3.4 Étiquetage sémantique des activités réalisées par un véhicule connecté

Disposant d'un ensemble de STOPs enrichis sémantiquement à l'aide de POI proches, il est également possible d'étudier les déplacements entre ces STOPs (MOVEs). Comme pour les STOPs, les déplacements peuvent être qualifiés à l'aide d'ontologies décrivant les activités spécifiques à ce type de véhicule. Les travaux portant sur l'étude de similarité sémantique sont détaillés au chapitre 4 de ce manuscrit. Dans ce cas d'étude, la liste des activités statiques et mobiles a été construite avec des experts sapeurs-pompiers. Différentes icônes illustrant ces activités, listées dans la table 3.1, sont utilisées pour faciliter la description des séquences.

Une fois les listes d'activités déterminées, celles-ci sont associées aux MOVE et STOP précédemment extraits à l'aide de règles d'inférence [**59**]. Dans notre modèle, il n'existe pas de règles contradictoires et l'ordre de vérification des règles n'a pas d'importance. Une règle s'applique lorsqu'un ensemble de conditions sont vérifiées. L'étiquetage sémantique des activités est réalisé en plusieurs phases successives.

8 activités statiques

Prise en charge du patient sur lieu d'intervention	
Prise en charge du patient à l'hôpital	
Attente fin de prise en charge hôpital	
Attente hôpital	
Logistique	
Attente caserne	
Arrêt indéterminé	
Arrêt non positionnable	

14 activités mobiles

Départ vers intervention	
Départ vers intervention (reroutage)	
Départ vers intervention avant annulation	
Transport vers hôpital	
Trajet intra hôpital	
Trajet logistique hôpital	
Trajet logistique	
Retour caserne	
Trajet intra caserne	
Retour avant redirection	
Retour caserne après annulation	
Trajet indéterminé vers Hôpital	
Trajet indéterminé vers arrêt indéterminé	
Trajet indéterminé vers arrêt non positionnable	

Table 3.1 – Liste des activités statiques et mobiles et icônes associées.

3.4.1 Étiquetage sémantique des STOPs

En premier lieu, les activités sont associées aux STOPs à l'aide des règles suivantes issues de [**2**] :

- Si le STOP n'a pas de médiane géométrique, alors l'activité est "Arrêt non positionnable"
- Si le POI attaché au STOP est de type hôpital et que le POI du STOP précédent est également de type hôpital alors l'activité est "Attente fin de prise en charge hôpital"
- Si le POI attaché au STOP est de type hôpital et que les gyrophares sont activés plus de $gyro_{min}^{hopital}$ % du temps durant le MOVE précédent alors l'activité est "Prise en charge du patient à l'hôpital"
- Si le POI attaché au STOP est de type hôpital alors l'activité est "Attente hôpital"
- Si le POI attaché au STOP est de type logistique alors l'activité est "Logistique"
- Si le POI attaché au STOP est de type caserne alors l'activité est "Attente caserne"
- Si les gyrophares sont activés plus de $gyro_{min}^{inter}$ % du temps durant le STOP et qu'aucun POI n'a été associé au STOP alors l'activité est "Prise en charge du patient sur lieu d'intervention"
- Si aucune activité n'a été définie, alors l'activité est "Arrêt indéterminé"

Lorsque les stops sont étiquetés, il est alors possible de réaliser une seconde phase d'enrichissement sémantique appliquée aux MOVEs.

3.4.2 Détection des situations de reroutage et segmentation des MOVEs

Cette phase consiste à détecter des situations spécifiques de reroutage d'un véhicule en cours de déplacement. C'est une des spécificités du modèle proposé dans [**2**] qui autorise alors une succession de deux activités de type MOVE.

Dans notre cas d'étude, il existe deux situations de reroutage d'un véhicule. Le premier cas est l'annulation qui survient lorsqu'une intervention est annulée alors que le véhicule est déjà parti de la caserne (annulation). Pour détecter les situations d'annulation, une nouvelle règle a été introduite. Il existe une situation d'annulation dans un épisode MOVE si celui-ci commence et termine par une activité d'attente caserne et que les gyrophares

du véhicule ont été allumés selon certaines contraintes de temps ($duree_{debut}^{gyro}$ maximum après le début du MOVE et $duree_{min}^{gyro}$ pendant ce MOVE).

La position de segmentation du MOVE en deux sous-trajectoires correspond à la dernière position à laquelle les gyrophares ont été coupés lors du MOVE.

- Si activité du STOP précédent est "Attente caserne" et activité du STOP suivant est "Attente caserne" et le moment de première activation du gyrophare pendant le MOVE est inférieur à un seuil $duree_{debut}^{gyro}$ et la durée d'activation du gyrophare pendant le MOVE est supérieure à un seuil $duree_{min}^{gyro}$ alors c'est un re routage de type "Annulation"

Le second cas se produit lorsque le véhicule rentre de l'hôpital vers sa caserne d'affectation, mais est redirigé sur une nouvelle intervention pendant son trajet de retour (redirection). Pour détecter les redirections, la règle suivante est appliquée :

- Si activité du STOP précédent est différente de "Attente caserne" et activité du STOP suivant est "Prise en charge patient sur lieu d'intervention" alors c'est un re routage de type "Redirection"

Dans ce cas précis, l'ajout des informations provenant des capteurs IoT (gyrophares) est primordial pour réaliser une segmentation sémantique plus fine ainsi qu'une détection automatique des activités réalisées par le véhicule. Cette segmentation n'aurait pas pu être réalisée en utilisant uniquement les données en provenance du GPS.

3.4.3 Étiquetage sémantique des MOVEs

Les STOPs ayant été précédemment étiquetés. Chaque déplacement MOVE dispose des informations concernant son STOP précédent et suivant. De nouvelles règles peuvent alors être appliquées pour enrichir sémantiquement ces épisodes MOVE.

- Si un reroutage a été détecté et est de type "Début redirection", alors l'activité est "Retour avant redirection"
- Si un reroutage a été détecté et est de type "Fin redirection", alors l'activité est "Départ vers intervention (reroutage)"
- Si un reroutage a été détecté et est de type "Début annulation", alors l'activité est "Départ vers intervention avant annulation"

- Si un reroutage a été détecté et est de type "Fin annulation", alors l'activité est "Retour caserne après annulation"
- Si l'activité suivante est vide ou si elle est de type "Arrêt non positionnable", alors l'activité est "Trajet indéterminé - vers arrêt non positionnable"
- Si l'activité suivante est de type "Attente hôpital", alors l'activité est "Trajet logistique hôpital"
- Si l'activité suivante est de type "Attente fin de prise en charge hôpital", alors l'activité est "Trajet intra hôpital"
- Si l'activité suivante est de type "Prise en charge du patient à l'hôpital" et si l'activité précédente est vide ou différente de "Prise en charge patient sur lieu d'intervention", alors l'activité est "Trajet indéterminé - vers Hôpital"
- Si l'activité suivante est de type "Prise en charge du patient à l'hôpital" et si l'activité précédente est "Prise en charge patient sur lieu d'intervention", alors l'activité est "Transport vers hôpital"
- Si l'activité suivante est de type "Logistique", alors l'activité est "Trajet logistique"
- Si l'activité suivante est de type "Prise en charge patient sur lieu d'intervention" et si l'activité précédente est de type "Attente caserne", alors l'activité est "Départ vers intervention"
- Si l'activité suivante est de type "Prise en charge patient sur lieu d'intervention" et si l'activité précédente est différente de "Attente caserne", alors l'activité est "Départ vers intervention (reroutage)"
- Si l'activité suivante est de type "Attente caserne" et si l'activité précédente est différente de "Attente caserne", alors l'activité est "Retour caserne"
- Si l'activité suivante est de type "Attente caserne" et si l'activité précédente est vide ou de type "Attente caserne", alors l'activité est "Trajet intra caserne"
- Si aucune règle n'est valide, alors l'activité est "Trajet indéterminé - vers arrêt indéterminé"

3.4.4 Identification des lieux régulièrement fréquentés

Une fois les trajectoires enrichies sémantiquement, il est alors possible de s'intéresser à l'identification de lieux régulièrement fréquentés en regroupant ensemble les médianes

géométriques représentant les STOPs. Cette tâche de clustering a été réalisée à l'aide de l'algorithme DBSCAN [**100**] détaillé en section 3.2.1. Les seuils de l'algorithme ont été définis à 30m pour le seuil de distance et 3 positions minimum pour constituer un cluster.

14 clusters ont été détectés par l'algorithme DBSCAN. L'étude de ces clusters a permis de confirmer la détection des lieux fréquemment visités par les ambulances tels que le cluster 0 correspondant à la caserne, le cluster 1 pour l'hôpital et le cluster 2 pour la station logistique.

3.4.5 Étude des séquences fréquentes

Considérant une tournée comme une séquence d'activités réalisées entre deux passages à la caserne, une étude de fréquence et de durée des différents types de séquences a été réalisée.

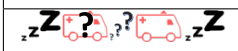
Motif de tournée	Nombre	Durée minimum	Durée médiane	Durée moyenne	Durée maximale
	46	12 :09	23 :06	27 :36	73 :04
	43	30 :26	61 :56	67 :44	166 :48
	13	53 :31	72 :28	76 :48	121 :35
	33	00 :06	01 :27	01 :55	16 :56
	11	18 :26	62 :05	57 :27	98 :02
	8	00 :46	04 :13	06 :20	20 :05
	7	12 :49	40 :59	33 :36	47 :33
	5	37 :14	56 :43	67 :00	130 :39
	5	45 :16	62 :55	74 :51	129 :57
	5	58 :50	63 :46	71 :13	87 :24

Table 3.2 – Motifs des tournées les plus fréquents et durées calculées, en (minutes :secondes)

Le tableau 3.2 présente les différents motifs de tournées observés une fois les STOPs et les MOVEs qualifiés sémantiquement.

La prise en compte de capteurs supplémentaires par rapport à l'approche initiale (GPS + frein à main) a permis d'améliorer la détection des activités réalisées ainsi que les motifs de tournées les plus fréquents. Ces résultats ont été présentés à des experts qui ont confirmé les nombres et durées des tournées détectées par notre algorithme.

3.5 Conclusion sur l'apport des objets connectés dans l'étude des trajectoires

Dans ce chapitre nous avons présenté un cas d'étude s'intéressant aux apports des objets connectés dans l'étude du mouvement. Les difficultés de segmentation des trajectoires à l'aide du GPS seul ont été présentées. Le GPS étant lui-même soumis à des problématiques de qualité (effet rebond, pelotes, pertes de signal) lorsqu'il est utilisé dans des environnements couverts ou lorsque l'objet mobile se trouve dans un bâtiment (garage, souterrain...). L'adjonction de capteurs supplémentaires a permis de mieux définir les emprises temporelles des épisodes de STOPs et de MOVEs. Dans certains cas, ces capteurs ont permis de qualifier un STOP alors même qu'aucune position GPS n'avait été collectée (cadre vert sur la figure 3.7).

Figure 3.7 – Processus de montée en sémantique proposé.

La montée en sémantique des activités réalisées pendant les STOP a été principalement induite par l'appariement à un POI préalablement sélectionné dans une liste d'intérêt couvrant des POI en lien avec les activités métier étudiées (casernes, hôpitaux, stations-service...) (cadre orange et vert sur la figure 3.7). La représentation spatiale des POI (points, lignes ou polygones) a un impact non négligeable sur ce processus d'appariement et peut parfois être source de confusion d'activité (cas d'une intervention proche d'une station-service par exemple qui risque d'être confondue avec un arrêt logistique).

Les règles du moteur d'inférence utilisé ont l'avantage d'être facilement explicables et sont directement issues des connaissances expertes (cadre bleu sur la figure 3.7). Cependant, une des difficultés de ce système réside dans la formulation de ces règles qui ne doivent être ni contradictoires, ni redondantes. Cette difficulté est encore accrue lorsque l'on dispose d'un nombre de capteurs plus conséquent. Une piste de résolution de ce problème peut être l'usage de système de résolution à base de cas comme proposé dans [**12**]. Il est alors nécessaire de définir des cas prototypiques représentant chaque situation ainsi

que des mesures de similarité entre les données observées et ces cas prototypiques.

Dans ce cas d'étude, nous ne disposions malheureusement pas d'un volume de données étiquetées suffisant pour appliquer des méthodes d'apprentissage comme les réseaux de neurones ou les forêts aléatoires (random forest) [**11**]. Il serait également intéressant de pouvoir utiliser une combinaison de ces deux techniques en mêlant apprentissage sur des jeux de données étiquetées et connaissances expertes à base de règles ou de cas prototypiques. L'explicabilité des résultats obtenus par ces processus de montée en sémantique est également un verrou de recherche actif en cas d'usage de mécanisme d'apprentissage.

La constitution de corpus de données de mobilité volumineux regroupant des données brutes issues de capteurs en provenance d'objets connectés ainsi que des étiquetages sémantiques décrivant finement les activités réalisées reste difficile à obtenir. Les problématiques de respect de la vie privée en lien avec ces corpus les rendent malheureusement difficilement partageables au sein de la communauté scientifique. Ce genre de corpus de données de mobilité enrichies sémantiquement a été produit et étudié dans le cadre du projet ANR MOBI'KIDS s'intéressant aux déplacements des enfants. Les apports portant sur l'étude de séquences sémantiques de mobilité sont détaillés dans le chapitre 4.

Chapitre 4

SIMILARITÉ SÉMANTIQUE ENTRE TRAJECTOIRES

Comme détaillé dans le chapitre 2, les trajectoires disposent d'une composante spatiale, temporelle, mais également sémantique. Ainsi, il est possible de décrire les activités réalisées lors d'un arrêt ou d'un déplacement plus ou moins finement [**43**]. La description des activités réalisées par une ambulance connectée a été présentée au chapitre 3. Une fois les activités décrites, il est alors intéressant de les comparer afin d'en déduire des similarités. C'est l'objet de ce chapitre dédié à l'étude de la similarité sémantique entre trajectoires étudiée dans la thèse de Clément Moreau [**4**] dans le cadre des projets MOBI'KIDS et SMART LOIRE (section 1.4.2).

4.1 Description des activités à l'aide d'ontologies

Dans cette section on s'intéresse à la notion de sens attribué, dans un contexte particulier, à un terme décrivant un élément constitutif d'une trajectoire. Ces éléments sont liés les uns aux autres via des concepts pouvant être représentés à l'aide de graphes de connaissances [**115**] décrivant des entités et leurs relations. Le modèle Resource Description Framework (RDF) est couramment utilisé pour représenter ces graphes de connaissances à l'aide de triplets (sujet, prédicat, objet).

Lorsqu'un graphe de connaissances regroupe un ensemble de concepts suffisants étendu pour décrire un domaine, il est alors possible de définir une Ontologie [**5**]. Dans une ontologie, les concepts sont reliés entre eux par des relations taxonomiques (*isA, partOf*) permettant une hiérarchisation des concepts [**60**]. Une taxonomie est une ontologie se limitant à la description de relations taxonomiques entre concepts (*isA*). Les ontologies permettent également de raisonner par transitivité sur les concepts. Le Web sémantique ou Linked Open Data (LOD) regroupe des millions d'objets référencés et structurés à l'aide

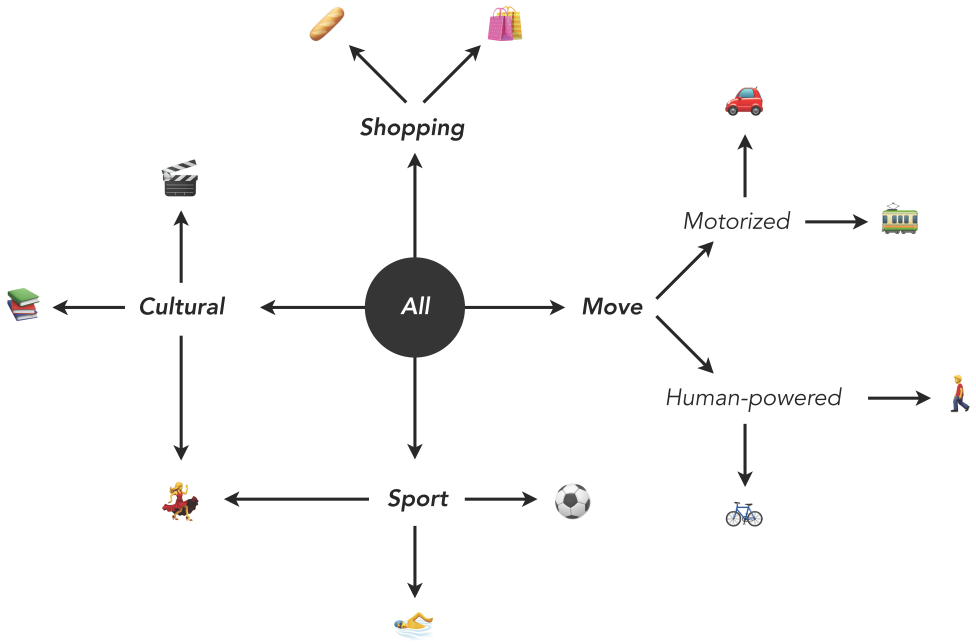


Figure 4.1 – Exemple de taxonomie d'activités quotidiennes

d'une ontologie cohérente. La figure 4.1 donne un exemple simplifié d'une taxonomie d'activités quotidiennes représentées à l'aide d'icônes qui serviront à illustrer les exemples de ce chapitre.

Certaines ontologies telles que les *Time Use Survey*[1] ont spécifiquement été créées pour décrire la mobilité quotidienne.

L'étude des relations transitives entre concepts au sein de ces ontologies est très utile pour définir la notion de similarité entre ces concepts.

4.1.1 Similarité entre concepts

La notion de similarité entre des concepts renvoie à la définition d'une distance entre deux objets issus d'un même ensemble satisfaisant les axiomes de symétrie, séparabilité et d'inégalité triangulaire. Les mesures de similarité, sont souvent normalisée dans $[0, 1]$. Elles sont très proches du concept de distance et respectent les mêmes axiomes. La similarité entre deux objets identiques vaut 1 lorsque leur distance est égale à 0.

Il existe de nombreuses approches pour comparer des concepts au sein d'une taxonomie

1. https://ec.europa.eu/eurostat/documents/3859598/11597606/KS-GQ-20-011-EN-N.pdf/2567be02-f395-f1d0-d64d-d375192d6f10?t=1607360062000

[**56**, **33**, **112**, **103**, **96**, **107**, **87**, **133**, **121**, **63**, **97**, **45**]. Ces approches se basent sur la topologie du graphe, les caractéristiques des concepts (traits), la quantité d'information portée par chaque concept (informationnelle) ou des résultats statistiques [**35**]. Ces mesures de similarité ont été étudiées par Clément Moreau dans [**4**] et sont synthétisées dans les tableaux 4.1 et 4.2.

Approche	**Nom**	**Type**	**Co-domaine**	**Utilise**				
				Plus grand ancêtre commun $LCA(x, y)$	Hyperonyme $\Gamma(x)$	Hyponyme $\mathcal{I}(x)$	Plus court chemin $d(x, y)$	Profondeur $d(x)$
Topologique	sim_{rada}	sim	$[0, 1]$				×	
	sim_{resnik}	sim	$\mathbb{R}^+$				×	×
	sim_{LC}	sim	$[0, 1]$				×	×
	sim_{wup}	sim	$[0, 1]$	×				×
	sim_{li}	sim	$[0, 1]$	×			×	
Par traits	$sim_{JaccardInt}$	sim	$[0, 1]$		×			
	$sim_{JaccardExt}$	sim	$[0, 1]$			×		
	$sim_{tversky}$	sim	$[0, 1]$	×				
	sim_{amato}	sim	$[0, 1]$	×		×		
Informationnelle	$sim_{resnikIC}$	sim	$\mathbb{R}^+, [0, 1]$	×				
	sim_{lin}	sim	$[0, 1]$	×				×
	sim_{cross}	sim	$[0, 1]$		×			
Statistique	NGD	dist.	$\mathbb{R}^+$		×	×		
	LSA	sim	$[0, 1]$		×	×		

Table 4.1 – Résumé synthétique des mesures de similarité sémantiques étudiées – 1 [4]

Approche	Nom	Description et commentaires
Topologique	sim_{rada}	Inverse du plus court chemin entre x et y.
	sim_{resnik}	Ajoute la notion de profondeur maximale de la taxonomie à la mesure de Rada.
	sim_{LC}	Ajoute un log afin de pénaliser davantage les valeurs importantes de plus court chemin.
	sim_{wup}	Prise en compte du plus grand ancêtre commun des concepts (le plus spécifique qui les subsume) et de la profondeur des concepts dans la taxonomie.
	sim_{li}	Similaire à Wu - Palmer mais utilise des fonctions non linéaires. Paramètres α, β contribuant respectivement au plus court chemin et au degré de parenté entre les concepts.
Par traits	$sim_{jaccardInt}$	Ratio du nombre de concepts ancêtres en commun dans la taxonomie.
	$sim_{jaccardExt}$	Ratio de nombre instances en commun dans la taxonomie.
	$sim_{tversky}$	Généralisation du modèle ensembliste de Jaccard. Paramètres α, β contribuant respectivement aux concepts uniques de y et x.
	sim_{amato}	Estime que deux concepts peuvent être similaires sans avoir aucune instances en commun.
Informationnelle	$sim_{resnikIC}$	IC du plus grand ancêtre commun. Ne respecte pas l'identité des indiscernables
	sim_{lin}	Équivalent de Wu - Palmer basé sur l'IC.
	sim_{cross}	Équivalent de Jaccard basé sur la somme des IC.
Statistique	NGD	Basée sur la fréquence des termes relatifs à x et y retourner par le moteur de recherche Google. Ne respecte pas les axiomes de la métrique.
	LSA	Utilisée avec la similarité du cosinus. Basée sur un ensemble de corpus pour représenter les concepts à partir des documents. S'appuie sur la notion de co-occurence de termes.

Table 4.2 – Résumé synthétique des mesures de similarité sémantiques étudiées – 2 [**4**]

L'approche Topologique de la mesure de Wu-Palmer [**107**] notée sim_{wup} a l'avantage d'être simple, normalisée, et tient compte des liens de parenté entre les concepts via la notion de plus petit ancêtre commun (LCA). La mesure de Wu-Palmer définit la similarité de deux concepts comme un ratio prenant en compte la profondeur d de chaque concept ainsi que la profondeur de leur LCA.

$$sim_{wup}(x, y) = \frac{2 \times d(LCA(x, y))}{d(x) + d(y)} \tag{4.1}$$

Lorsque l'on souhaite comparer un ensemble de concepts, il est nécessaire d'utiliser une fonction d'agrégation. La mesure de similarité définie par Halkidi et al. [**86**] est basée sur une combinaison de moyennes tenant compte du cardinal de chaque ensemble. Elle est moins sensible aux données aberrantes.

$$\zeta(X, Y) = \frac{1}{2}\left(\frac{1}{|X|}\sum_{x \in X}\max_{y \in Y}\{sim(x, y)\} + \frac{1}{|Y|}\sum_{y \in Y}\max_{x \in X}\{sim(x, y)\}\right) \tag{4.2}$$

4.1.2 Similarité entre séquences de concepts

Une fois la similarité entre concepts définis, il est alors possible de réaliser des comparaisons de séquences et séries temporelles sémantiques. Dans le chapitre 2, différents modèles formels ont été présentés. Ces modèles proposent une abstraction des activités humaines réalisées au cours du temps à l'aide de séquences d'éléments. Les relations entre ces activités peuvent être définies à l'aide des ontologies présentées à la section 4.1.

Une trajectoire symbolique est représentée par une séquence d'activités $S = \langle x_1, ..., x_n \rangle$ où $x_i \in \Sigma$ dont la taille est notée $|S| = n$. L'ensemble des activités réalisables est noté Σ.

Les différentes distances entre séries temporelles qualitatives sont listées ci-après [**49**, **4**, **Ra2**].

- Distance de Minkowski (ou distances ℓ_p)
- Distance de Hamming
- Distance d'édition (Edit distance)
- Longest Common Subsequence (LCSS)
- Dynamic Time Warping (DTW)

Distance de Minkowski

La distance de Minkowski d'ordre p est une distance entre deux points dans un espace à n dimensions.

$$\|S_1 - S_2\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}, \quad |S_1| = |S_2| = n \tag{4.3}$$

Différentes appellations de cette distance existent en fonction de la valeur du paramètre p. Cette distance est appelée distance de Manhattan pour $p = 1$ ou distance Euclidienne pour $p = 2$. Cette famille de distances compare des vecteurs de tailles égales.

Distance de Hamming

La distance de Hamming [**128**] s'intéresse au nombre de symboles à modifier permettant de transformer une séquence S_1 en S_2.

$$H(S_1, S_2) = \sum_{i=1}^{n} \delta(x_i, y_i), \quad |S_1| = |S_2| = n \tag{4.4}$$

où $\delta : \Sigma \times \Sigma \rightarrow [0, 1]$ est une distance sur Σ.

La distance de Hamming est très simple et rapide à calculer. Cependant, celle-ci est peu robuste aux décalages et distorsions temporelles.

Distance d'édition

La distance d'édition (Edit Distance (ED)) est une extension de la distance de Hamming pour des séquences de longueurs différentes. La distance d'édition réalise un appariement optimal des séquences en comptant le nombre minimum d'opérations d'édition nécessaires pour convertir une séquence S_1 en une séquence S_2 [**126**]. Trois opérations d'éditions de symboles sont réalisables (la suppression, l'ajout et la modification). Une fonction de coût est définie pour chaque opération d'édition.

Un chemin d'édition correspond à la composition successive des opérations d'édition permettant de transformer S_1 en S_2. Son coût total est défini par la somme des coûts de ses opérations d'édition successives.

La distance d'édition $ED : \Sigma^n \times \Sigma^m$ correspond coût cumulé minimal du meilleur chemin

d'édition trouvé.

$$ED(S_1, S_2) = \min_{(e_1,\ldots,e_N)\in\mathscr{P}(S_1,S_2)} \left\{ \sum_{i=1}^{N} \gamma(e_i) \right\} \tag{4.5}$$

La distance d'édition peut être calculée par programmation dynamique à l'aide d'un algorithme d'une complexité en $O(n \times m)$ [**123**].

Longuest Common Sub-Sequence

La Longuest Common Sub-Sequence (Longuest Common Sub-Sequence (LCSS)) est une extension de la distance d'édition. Cette distance s'intéresse à la longueur des sous-séquences d'éléments strictement identiques. Ces sous-séquences d'éléments identiques n'étant pas nécessairement consécutifs. LCSS permet donc d'ignorer certains éléments, mais ne permet pas de prendre en compte une fonction de coûts entre symboles (pas d'opération de modification).

LCSS peut également être calculée par programmation dynamique [**89**] selon l'équation :

$$C_{i,j} = \begin{cases} 0 & \text{Si } i = 0 \text{ ou } j = 0 \\ \begin{cases} C_{i-1,j-1} + 1 & \text{Si } x_i = y_j \\ \max\{C_{i,j-1}, C_{i-1,j}\} & \text{Sinon} \end{cases} & \text{Sinon} \end{cases} \tag{4.6}$$

La distance entre S_1 et S_2 est obtenue par $LCSS(S_1, S_2) = \max\{n, m\} - C_{n,m}$.

Dynamic Time Warping

La mesure Dynamic Time Warping (Dynamic Time Warping (DTW)) [**119**, **104**] est très utilisée pour la reconnaissance de motifs dans des séries temporelles. Cette mesure autorise la contraction ou dilatation temporelle de la série. Cependant, cette mesure ne respecte pas les axiomes de séparabilité et d'inégalité triangulaire [**83**].

Son calcul $DTW(S_1, S_2) = C_{n,m}$ peut être réalisé par programmation dynamique [**79**]

selon l'équation :

$$C_{i,j} = \begin{cases} \infty & \text{Si } i = 0 \text{ ou } j = 0 \\ 0 & \text{Si } i = 0 \text{ et } j = 0 \\ \delta(x_i, y_i) + \min \left\{ \begin{array}{c} C_{i-1,j-1} \\ C_{i-1,j} \\ C_{i,j-1} \end{array} \right\} & \text{Sinon} \end{cases} \tag{4.7}$$

Toutes ces distances prennent en compte l'ordre temporel via la notion de précédence entre les éléments d'une série. Cependant, elles ne sont pas en mesure de tenir compte du contexte lors des opérations d'édition ni de la durée des éléments de la série.

C'est pourquoi de nouvelles mesures de similarité sémantique entre séquences ont été proposées.

4.2 Mesure de similarité sémantique entre séquences tenant compte du contexte.

Cette section présente les travaux réalisés pour la définition de nouvelles mesures de similarité entre séquences sémantiques intégrant trois notions de proximité [**4**].

- La proximité sémantique ($ProxSem$) : Calculée en utilisant une mesure de similarité sémantique entre activités s'appuyant sur une ontologie et présentée à la section 4.1.1.
- La proximité temporelle ($ProxTemp$) : Deux activités sont considérées comme proches temporellement si elles sont réalisées à des périodes proches.
- La proximité contextuelle ($ProxContext$) : Deux activités sont proches contextuellement si elles ont à la fois une proximité sémantique et une proximité temporelle forte.

Découlant de ces trois notions, cinq spécificités sont recherchées :

1. *Homogénéité sémantique* : Deux séquences sémantiques regroupant des activités proches sémantiquement devraient être plus similaires que deux séquences d'activités regroupant des activités sémantiquement différentes.

2. *Temporalité d'activités* : Deux séquences sémantiques ayant les mêmes activités se déroulant à des temporalités proches devraient être plus similaires que deux séquences d'activités ayant les mêmes activités, mais se déroulant à des temporalités éloignées.
3. *Décalage temporel* : Deux séquences sémantiques ayant les mêmes activités à un décalage temporel Δt près (en termes d'ordre ou de durée) devraient avoir une dissimilarité d'autant plus faible que le décalage temporel est faible.
4. *Permutation d'activités* : Deux séquences sémantiques ayant les mêmes activités à une permutation près, devraient avoir une dissimilarité faible. Cette dissimilarité est d'autant plus faible que les activités permutées sont proches temporellement.
5. *Redondance d'activités* : Deux séquences sémantiques ayant les mêmes activités à des redondances (i.e., répétitions) près, devraient avoir une dissimilarité faible.

Les distances présentées dans la revue de littérature de la section 4.1.2 ne couvrent pas toutes ces spécificités. Cependant, parmi les distances étudiées, la distance d'édition est capable de comparer des séquences de tailles différentes, elle tient compte de la similarité entre symboles, possède la capacité d'effectuer des permutations et calcule un appariement optimal des séquences.

Les spécificités recherchées font référence à des qualificatifs imprécis comme "proche", "éloigné", "faible", "fort". La distance proposée est en mesure de s'adapter à cette situation en utilisant une approche à base de logique floue [**127**, **108**] pour la définition des fonctions de coût d'édition.

La distance d'édition contextuelle Contextual Edit Distance (CED) a été proposée par Clément Moreau dans le cadre de sa thèse et publiée dans [**7**].

Les opérateurs d'édition ont été modifiés pour y intégrer l'indice k_{edit} auquel l'édition du symbole $\mathbf{x}$ est réalisée dans la séquence S_i.

Une opération d'édition contextuelle $e = (op, \mathbf{x}, k_{edit}, S_i)$ est représentée sous la forme d'un quadruplet ou op désigne l'opération réalisée (add, del, mod).

Le nombre de symboles d'écart ($\Delta k = k - k_{edit}$) dans la séquence par rapport à l'indice k_{edit} du symbole à éditer est utilisé pour définir la relation de proximité temporelle. Cette proximité temporelle est représentée à l'aide d'une fonction d'appartenance floue $\mu : \mathbb{R} \to [0, 1]$ centrée sur 0 comme illustrée sur la figure 4.2.

Pour une opération d'édition $e = (op, x, k_{edit}, S_i)$ sur une séquence de taille n, on associe un vecteur temporel $\nu_e \in [0, 1]^n$ décrivant la proximité temporelle de chaque indice

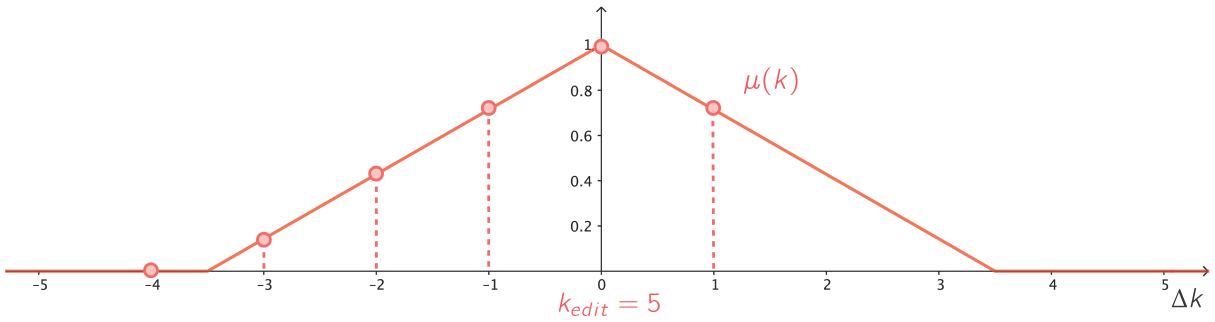


Figure 4.2 – Exemple de fonction floue μ pour l'encodage de l'opérateur `mod`

k de la séquence S_i. Plus l'indice k est proche de k_{edit}, plus la valeur de $\nu_{e,k}$ tend vers 1 et plus le symbole devra être pris en compte dans l'opération d'édition.

Cependant, les opérations d'ajout et de suppression de symboles impactant la taille n des séquences, il est nécessaire de prendre en compte ces changements d'indices dans la définition des vecteurs d'édition. Lors d'une opération d'ajout, les indices des éléments $k > k_{edit}$ sont tous incrémentés. L'opération de suppression d'un symbole est réalisée par un forçage du vecteur temporel à 0 pour l'élément positionné à l'indice k_{edit} de la séquence.

$$\nu_{e,k} = \begin{cases} \nu_{e,k} = \mu(\Delta k) & \text{Si } op = \texttt{mod} \\ \begin{cases} \mu(-1) & \text{Si } \Delta k = 0 \\ \mu(\Delta k + 1) & \text{Si } \Delta k \geq 1 \\ \mu(\Delta k) & \text{Si } \Delta k \leq -1 \end{cases} & \text{Si } op = \texttt{add} \\ \begin{cases} 0 & \text{Si } \Delta k = 0 \\ \mu(\Delta k) & \text{Sinon} \end{cases} & \text{Si } op = \texttt{del} \end{cases}$$

Disposant du vecteur temporel il est alors possible de définir la fonction de coût $\gamma : E \rightarrow [0, 1]$ d'une opération d'édition contextuelle e ou $sim(x_{ik}, \mathbf{x})$ est l'une des mesures de similarité sémantique (par exemple celle de Wu-Palmer sim_{wup}) proposée à la section 4.1.1.

$$\gamma(e) = 1 - \max_{k \in [\![1,n]\!]} \{sim(x_{ik}, \mathbf{x}) \times \nu_{e,k}\} \tag{4.8}$$

Enfin, à l'aide de l'équation 4.8, il est désormais possible de calculer un coût global

requis pour transformer la séquence $S_1 = \langle x_{1,1}, \ldots, x_{1,n} \rangle$ en $S_2 = \langle x_{2,1}, \ldots, x_{2,p} \rangle$.

$$CED_{S_1 \to S_2} = \min_{(e_1,\ldots,e_N) \in \mathscr{P}(S_1,S_2)} \left\{ \sum_{k=1}^{N} \gamma(e_k) \right\} \tag{4.9}$$

Afin de conserver la symétrie, la distance CED est définie telle que :

$$CED(S_1, S_2) = \max \{ CED_{S_1 \to S_2}, CED_{S_2 \to S_1} \} \tag{4.10}$$

4.2.1 Exemple d'application de la distance d'édition contextuelle.

Dans cette section, un exemple simple [**4**], constitué de 8 séquences d'activité sémantiques de mobilité, est présenté afin d'illustrer les capacités de notre distance CED comparée à la distance d'édition classique. Les activités sont décrites à l'aide de la taxonomie simplifiée introduite en début de chapitre sur la figure 4.1.

- $S_1 = \langle$🚲$\rangle$
- $S_2 = \langle$🚲, 🥖, 🚲$\rangle$
- $S_3 = \langle$🚶, 🛍️, 🚶$\rangle$
- $S_4 = \langle$⚽, 🚶, 🚃, 🎬$\rangle$
- $S_5 = \langle$🚃, 🎬, 🚶, ⚽$\rangle$
- $S_6 = \langle$🚶, 🚤, 🚃, 🎬$\rangle$
- $S_7 = \langle$🚶, 📚, 🚶, 🎬$\rangle$
- $S_8 = \langle$🚶, 🐓, 📚$\rangle$

Les couleurs des numéros de séquences correspondent à 3 classes définies par des experts :

1. La classe orange regroupe des séquences comportant des achats (🥖, 🛍️) combinés avec des modes de déplacement actifs (🚶,🚲). Ces séquences disposent également d'une importante homogénéité sémantique ainsi que de répétitions d'activités.
2. La classe bleu est composée de séquences comportant à la fois des modes de déplacements motorisés 🚃et actifs 🚶. Différentes activités sportives (🚤,⚽) et culturelles 🎬y sont réalisées. Les séquences S_4 et S_5 sont composées des mêmes activités, mais réalisées dans un ordre différent afin de tester les capacités de prise en compte de la temporalité et des permutations de notre distance.

3. La classe rouge se focalise sur des séquences comportant des activités artistiques ( , ) avec des déplacements exclusivement à pied  . L'homogénéité sémantique y est importante ainsi que la présence de répétitions.

Les matrices de distances entre toutes les séquences ont été calculées pour la distance d'édition (ED) et la distance contextuelle d'édition (CED). La pente de la fonction triangulaire floue utilisée pour le calcul du vecteur temporel de la distance d'édition contextuelle a été fixée à $1 - \frac{|k_{edit} - i|}{4}$ afin de prendre en compte tous les symboles des séquences. La mesure de similarité sémantique retenue entre les concepts est la mesure de Wu-Palmer (équation 4.1).

À partir de ces matrices de distances, deux dendrogrammes, présentés sur la figure 4.3, ont été calculés selon l'algorithme Linkage et le critère d'agrégation de Ward.

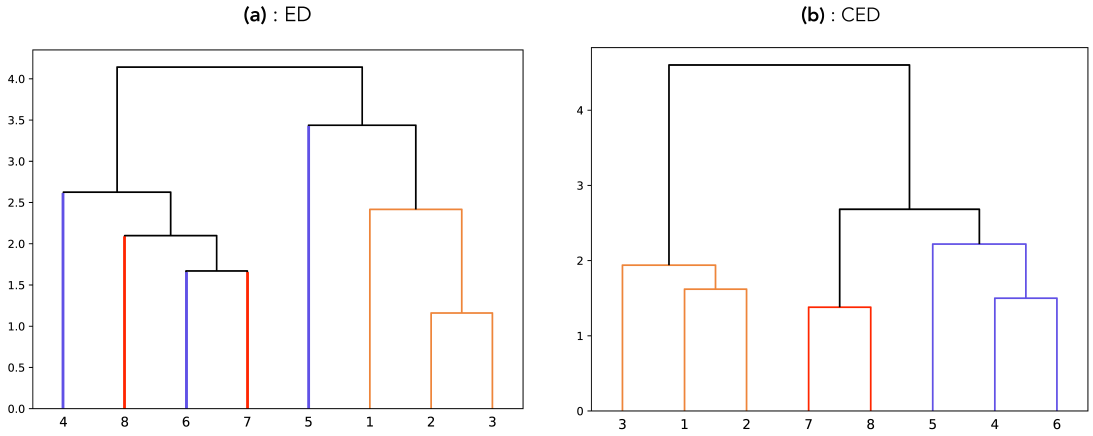


Figure 4.3 – Dendrogrammes des séquences sémantiques pour les mesures (a) Edit Distance (b) Contextual Edit Distance

La figure 4.3 illustre les différences entre la distance d'édition et notre distance contextuelle d'édition. Les 5 spécificités introduites en début de section 4.2 sont respectées dans cet exemple. La distance CED est en mesure de retrouver les classes experts ainsi que l'ordre hiérarchique de regroupement des séquences. La distance d'édition classique ne parvient pas à regrouper correctement les séquences des classes rouges et bleues et ne respecte pas non plus l'ordre de regroupement de la classe orange.

La distance CED a également été appliquée avec succès sur différents jeux de données sémantiques tels que des traces d'exploration de bases de données [**9**], des trajectoires d'enfants annotées sémantiquement (projet MOBIKIDS) et des visites touristiques (projet SMART LOIRE).

Les codes python réalisés pour le calcul de la distance CED sont accessibles sur le dépôt GitHub de Clément Moreau [2].

4.3 Extension de la distance de Hamming aux trajectoires sémantiques.

La distance CED, présentée dans la section précédente, est en mesure de prendre en compte le contexte temporel (symboles précédents et suivants) lors des opérations d'édition d'une séquence. Cependant, la durée des activités n'est pas prise en compte dans cette distance. Dans CED, la prise en compte de la dimension temporelle est considérée uniquement via la notion d'ordre de précédence des symboles. De plus, l'ordre de complexité de CED, pour deux séquences de tailles n et p, est relativement important ($O(n \times p \times \max\{n, p\})$).

La distance de Hamming, présentée à la section 4.1.2, dispose d'un ordre de complexité linéaire. Cette distance est prévue pour comparer des séquences de tailles identiques. La distance de Hamming n'est pas robuste aux décalages et distorsions temporelles.

Cette section présente les travaux réalisés dans la thèse de Clément Moreau [**4**] pour l'extension de la distance de Hamming aux trajectoires sémantiques intégrant les notions de durées et robuste aux décalages temporels et publiés dans [**CI1**].

4.3.1 Séquence sémantique temporelle

La définition des séquences sémantiques proposées dans la section 4.1.2 est étendue pour y intégrer des durées δ attribuées à chaque symbole. La séquence sémantique temporelle est désormais notée :

$$S_i = \langle (x_{i1}, \delta_{i1}), \ldots, (x_{in}, \delta_{in}) \rangle$$

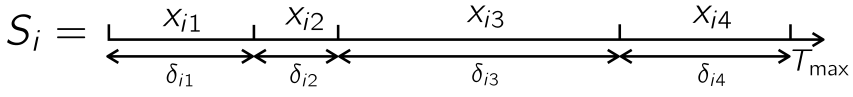


Figure 4.4 – Illustration d'une séquence sémantique temporelle

2. `https://github.com/Clement-Moreau-Info/CED`

Tous les symboles sont comparables à l'aide d'une des mesures de similarité introduites en section 4.1.1 ($sim : \Sigma \times \Sigma \rightarrow [0, 1]$). La durée de l'activité x_{ik} est notée δ_{ik}. La répétition des symboles dans une séquence est interdite. La somme cumulée des durées des activités d'une séquence doit correspondre à la durée $T_{\max}$ de l'intervalle d'étude choisi.

Dans la suite de ce chapitre, l'unité de temps choisie est la minute et la durée d'étude $T_{\max}$ est fixée à une journée soit 1440 minutes.

4.3.2 Approche floue de la distance de Hamming

En s'inspirant des travaux réalisés pour la prise en compte du contexte dans CED, cette section s'intéresse à la prise en compte du voisinage temporel dans la comparaison des séquences tout en y considérant la notion de durée.

Les opérations d'éditions (e) définies à la section 4.2 sont alors modifiées pour y intégrer la notion d'instant d'édition (t_{edit}) et de durée de l'opération d'édition (δ).

$$e = (\mathbf{x}, \delta, t_{edit}, S_i)$$

L'opération d'édition e signifie que tous les symboles de S_i sont remplacés par le symbole $\mathbf{x}$ pour une durée de δ unités de temps à partir de l'instant t_{edit} comme illustré en rouge sur la figure 4.5.

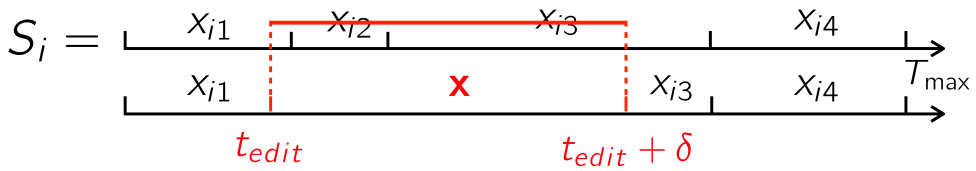


Figure 4.5 – Exemple d'opération d'édition sur une séquence sémantique temporelle.

Tout comme pour CED, une fonction floue $\mu_e : I \rightarrow [0, 1]$, illustrée en rouge sur l'exemple de la figure 4.6, est proposée pour prendre en compte le contexte temporel des séquences lors des opérations d'édition. La fonction temporelle est définie en fonction de l'opération d'édition e.

Le noyau $C(\mu_e)$ de la fonction flou, représenté en vert sur la figure 4.6, est défini entre $[t_{edit}, t_{edit} + \delta[$ avec un degré d'appartenance fixé à 1. Une frontière floue décroissante $B(\mu_e)$, représenté en orange sur la figure 4.6, est définie selon un paramètre β de part et d'autre du noyau.

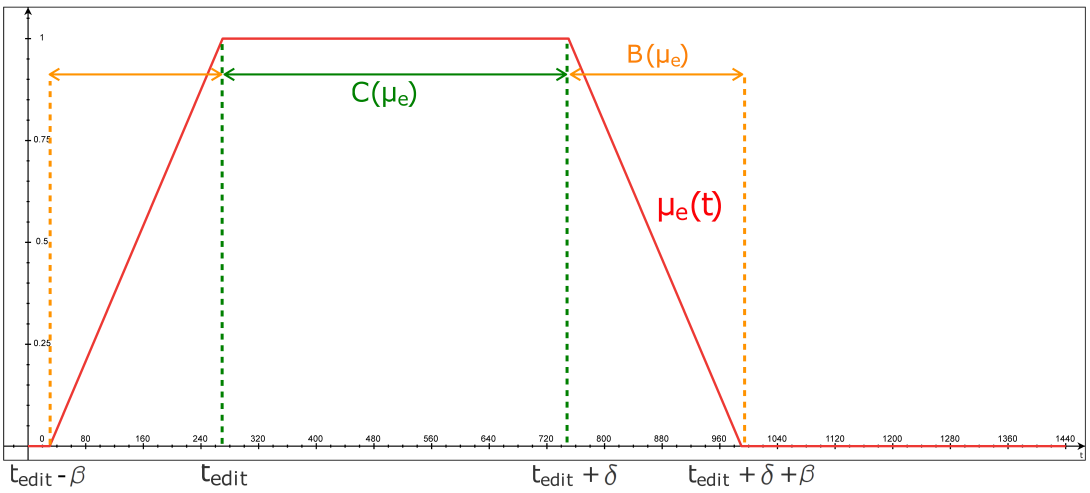


Figure 4.6 – Exemple de fonction μ_e

Une seconde fonction notée $sim_e(t)$ compare le symbole x de l'opération d'édition e avec tous les symboles de la séquence sur l'intervalle de temps $[O, T_{max}[$ (Figure 4.7).

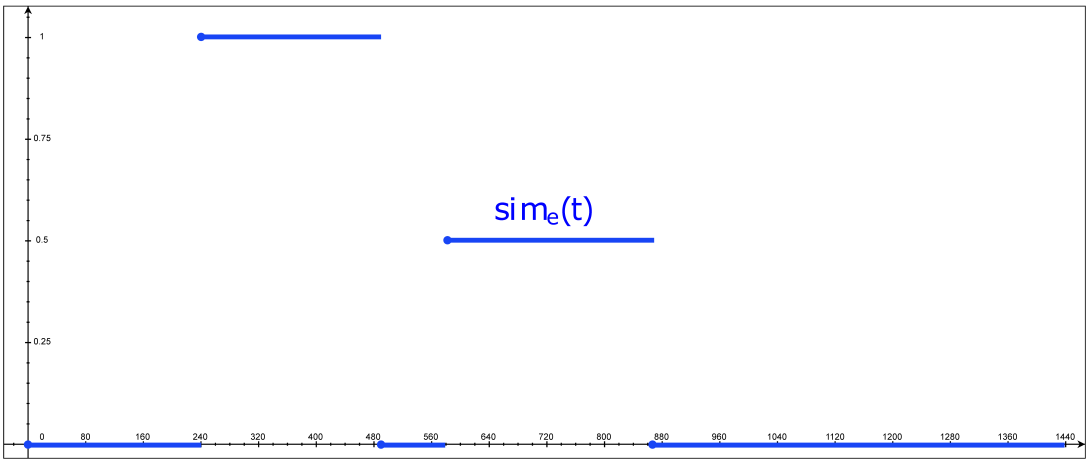


Figure 4.7 – Fonction étagée de similarité sim_e d'une opération d'édition e avec les symboles d'une séquence sémantique temporelle

Enfin, en associant ces deux fonctions, il est alors possible de définir une fonction de coût normalisée $\gamma : \mathbb{E} \rightarrow [0, 1]$ qui combine la similarité sémantique d'une opération d'édition (sim_e) en tenant compte de l'emprise temporelle floue (μ_e).

$$\gamma(e) = 1 - \sup_{\tau \in I} \left\{ \frac{1}{\delta} \int_{\tau}^{\tau+\delta} sim_e(t) \times \mu_e(t) dt \right\} \tag{4.11}$$

L'objectif de cette fonction consiste à trouver le segment temporel $[\tau, \tau + \delta[$ qui maximise à la fois la similarité du symbole édité $\mathbf{x}$ et la fonction temporelle. Ce segment temporel est illustré en gris sur la figure 4.8. L'algorithme de transformation de Fourier rapide (Fast Fourier Transform (FFT)) est utilisé pour réaliser ce calcul [**122**].

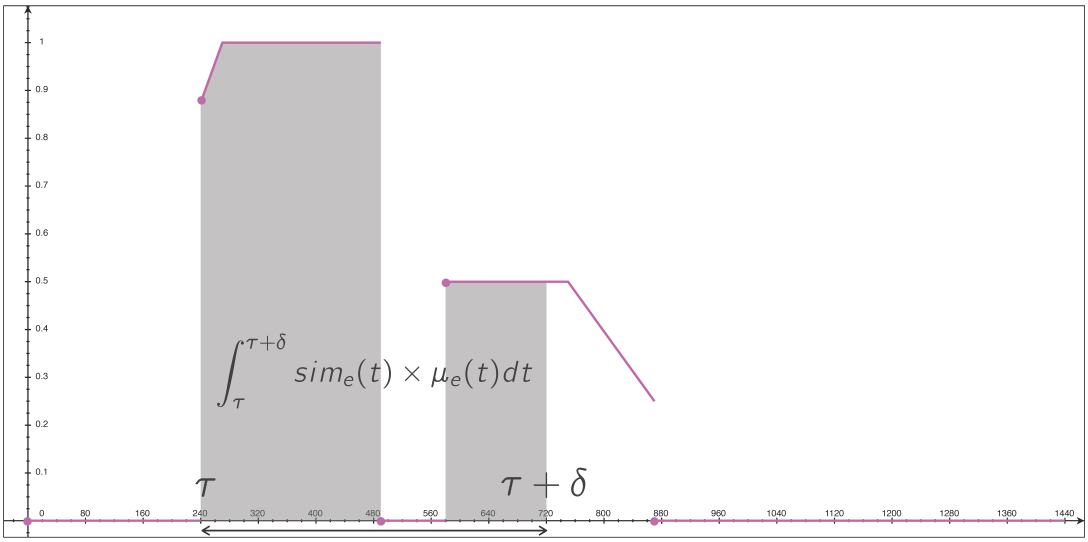


Figure 4.8 – Exemple d'application de la fonction de coût normalisée $\gamma(e)$ pour une opération d'édition e

La fonction de coût normalisée $\gamma(e)$ proposée à l'avantage de permettre de prendre en compte de manière plus importante des symboles dont la durée est courte. C'est intéressant dans l'étude des mobilités quotidiennes car certaines activités, comme être à la maison ou travailler, se réalisent sur une majeure partie de la journée. Les activités de plus courte durée sont alors saillantes dans l'étude et la comparaison des mobilités quotidiennes. Cependant, il est tout à fait possible de modifier cette fonction de coût pour la rendre proportionnelle à la durée du symbole édité comme proposé dans la fonction de coût $\Delta(e)$ suivante :

$$\Delta(e) = \delta \times \gamma(e) \tag{4.12}$$

Enfin, les fonctions de coût $\gamma(e)$ et $\Delta(e)$ sont utilisables pour calculer la similarité de deux séquences. Pour cela on s'inspire du calcul la distance de Hamming pour définir la dissimilarité unilatérale Fuzzy Temporal Hamming (équation 4.13. La fonction de coût $f(e)$ utilisée pouvant être au choix $\gamma(e)$ ou $\Delta(e)$ en fonction du poids alloué à la durée

des symboles dans la séquence.

$$\mathrm{FTH}_{S_1 \to S_2} = \sum_{k=1}^{n} f(e_k) \tag{4.13}$$

Tout comme pour CED, l'usage du maximum permet de rendre cette mesure symétrique. L'équation 4.14 permet d'obtenir la Fuzzy Temporal Hamming distance (Fuzzy Temporal Hamming distance (FTH)) entre les deux séquences. On note $FTH\gamma$ la distance dont la fonction de coût est γ_e et $FTH\Delta$ la distance dont la fonction de coût est Δ_e.

$$\mathrm{FTH}(S_1, S_2) = \max\{\mathrm{FTH}_{S_1 \to S_2}, \mathrm{FTH}_{S_2 \to S_1}\} \tag{4.14}$$

4.3.3 Exemple d'application de la Fuzzy Temporal Hamming distance.

De même que pour la distance CED, un exemple illustratif est présenté dans cette section issu de [**4**]. Cet exemple propose 6 séquences d'activité sémantiques disposant de caractéristiques intéressantes permettant de mettre en évidence les propriétés attendues des différentes mesures proposées. Les différentes activités sont représentées par un code couleur sur la figure 4.9. La durée en minutes des activités est représentée sur l'axe des abscisses. La durée T_{max} des séquences est fixée à une journée soit 1440 minutes.

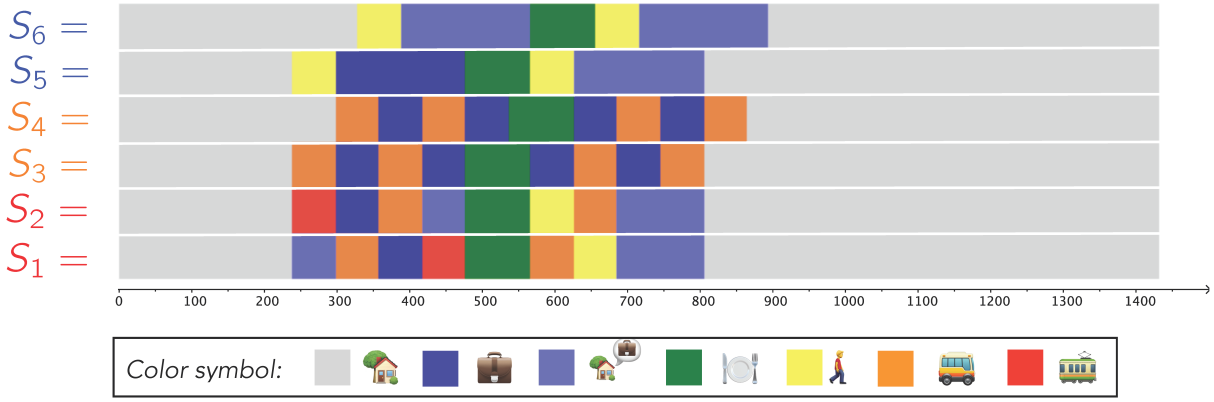


Figure 4.9 – Exemple illustratif comprenant 6 séquences sémantiques temporelles.

Les séquences de la classe rouge ($S1, S2$) sont composées d'activités identiques, mais mélangées avec de nombreuses permutations. La classe orange (S_3, S_4) présente un simple décalage temporel de 60 minutes des activités. Enfin, la classe bleue (S_5, S_6) comporte un

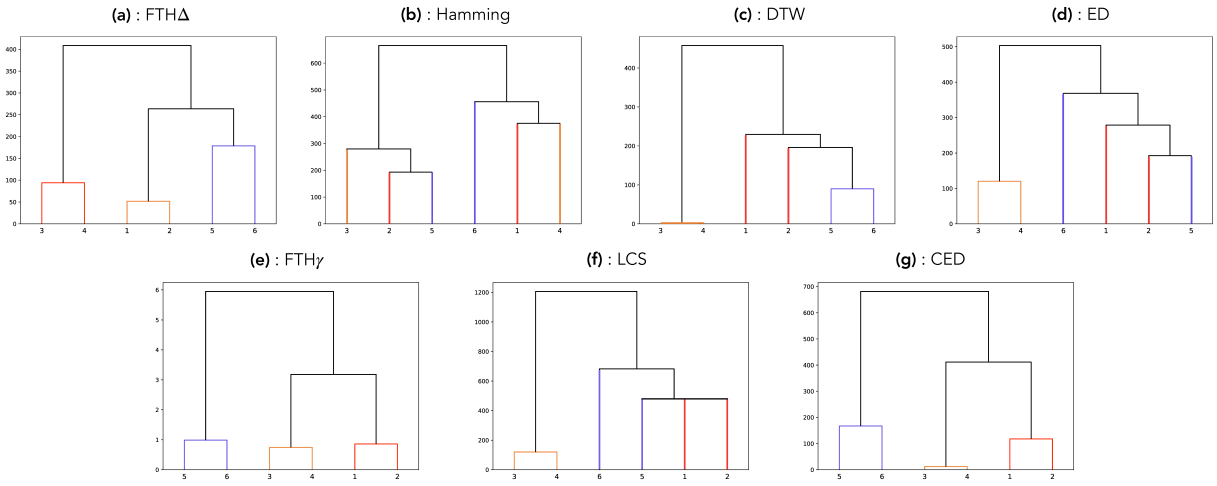


Figure 4.10 – Dendrogrammes des séquences sémantiques temporelles de la figure 4.9 pour différentes mesures. Les couleurs indiquent la classe experte de la paire d'origine.

décalage temporel plus important ainsi qu'une différence d'activité le matin réalisée sur des concepts proches ("Travail" vs "Travail à la maison"). Ces trois classes sont considérées comme les classes expertes que nous souhaitons retrouver à l'aide des différentes mesures de similarité.

La figure 4.10 présente les différents dendrogrammes générés à partir des matrices entre toutes les trajectoires sémantiques temporelles pour les principales mesures de similarité définies à la section 4.1.2. Ces dendrogrammes ont tous été générés en utilisant l'algorithme linkage et le critère d'agrégation de Ward. La fenêtre temporelle utilisée pour le calcul de FTH a été définie avec un paramètre de frontière floue β de 240 minutes.

L'étude de ces dendrogrammes montre que les séquences de la classe orange disposant d'un simple léger décalage temporel sont bien regroupées par la majorité des mesures de similarité à l'exception de la distance de Hamming. Les séquences de la classe bleue, comportant un décalage temporel important ainsi qu'un changement d'activité sémantiquement proche sont correctement regroupées avec CED, FTH mais également DTW. Ce résultat confirme bien que ces 3 mesures sont bien robustes à la présence de décalages temporels. Enfin, seules les mesures que nous avons proposées (CED et FTH) regroupent correctement les séquences de la classe rouge ou les symboles des séquences sont identiques, mais avec des permutations temporelles importantes.

Afin d'éprouver ces résultats, la distance FTH a également été testée sur des jeux de données de mobilité conséquents (1200 séquences sémantiques temporelles) issus de

l'enquête Enquête Ménages-Déplacements (Enquête Ménages-Déplacements (EMD)) de 2018 pour la ville de Rennes. Les résultats de cette étude sont exposés dans [**8**].

Dans cet article, les séquences sémantiques temporelles de mobilité ont été comparées à l'aide de 3 mesures de similarité (Hamming, $FTH\gamma$ et $FTH\Delta$) et regroupées dans 5 clusters pour chaque mesure. Le paramètre β de la frontière temporelle utilisé pour le calcul des distances FTH a été fixé à 12h. Le diagramme de Sankey de la figure 4.11 présente les ventilations et transitions des séquences entre les clusters des différentes distances étudiées.

Seulement 10% des séquences changent de cluster entre les clustering de Hamming et de $FTH\Delta$. La prise en compte de la durée des activités dans $FTH\Delta$ rapproche les résultats de ceux obtenus par la distance de Hamming par rapport à $FTH\gamma$. L'étude d'un sous-ensemble des 46 séquences réarrangées entre Hamming (C1) et $FTH\Delta$ (C2), surlignées en violet sur la figure 4.11, montre que ces réarrangements de séquences sont dus à la capacité de FTH à tolérer des décalages temporels plus importants que Hamming. En effet, ces séquences concernent des étudiants qui passent plus ou moins de temps à l'école puis réalisent des activités de loisirs. On note sur ce diagramme de Sankey, un nombre important de réarrangements entre les clusters obtenus par $FTH\Delta$ et ceux obtenus en appliquant $FTH\gamma$. La distance $FTH\gamma$ n'est pas pondérée par la durée des activités éditées. Le paradigme de coût est donc clairement différent. Ces changements

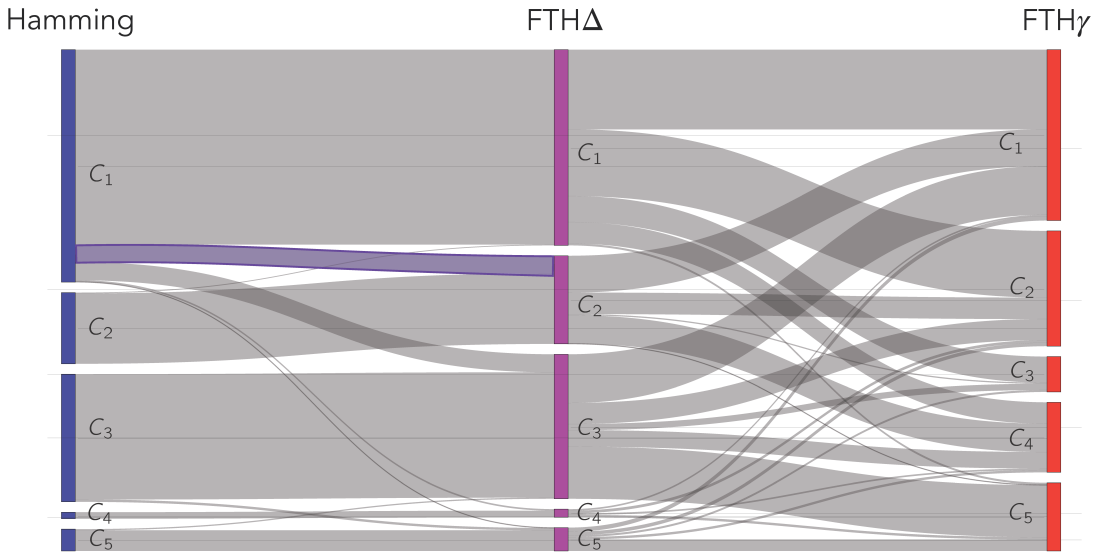


Figure 4.11 – Diagramme de Sankey illustrant les flux entre les 5 clusters obtenus par les mesures de Hamming, FTHΔ et FTHγ

indiquent que ces deux mesures offrent des outils de comparaisons différents qui doivent être choisis avec pertinence en fonction des besoins exprimés par les experts métiers.

Ces expérimentations et le code de calcul de la distance FTH sont disponibles sur le dépôt Github de Clément Moreau [3].

3. https://github.com/Clement-Moreau-Info/FTH

Chapitre 5

SIMILARITÉ SPATIALE ENTRE TRAJECTOIRES

Ce chapitre s'intéresse plus particulièrement à l'étude de la similarité spatiale entre trajectoires. Grâce à l'essor des objets connectés et au développement des systèmes de géolocalisation GNSS, les positions (p) de nombreux objets mobiles peuvent être capturées et sauvegardées dans de volumineuses bases de données. Les séries de positions sont généralement connectées par des lignes droites pour former la trace spatiale de l'objet mobile comme présenté dans la section 2.3 du chapitre 2. La notion de similarité spatiale entre ces traces repose sur la définition de distances entre ces traces et les positions qui les composent.

5.1 Distance entre positions

La distance entre deux positions est souvent exprimée à l'aide de la distance euclidienne qui découle de la distance de Minkowski présentée à la section 4.1.2.

Considérant deux positions $p_1 = (x_1, y_1, z_1, t_1)$ et $p_2 = (x_2, y_2, z_2, t_2)$ la distance spatiale euclidienne entre ces deux positions est définie par l'équation 5.1.

$$d_E(p1, p2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{5.1}$$

5.2 Distances spatio-temporelles entre trajectoires

Les trajectoires peuvent être vues comme des séries de données (positions) ordonnées dans le temps. Ainsi, les différentes mesures de similarité de séries de données temporelles (ED, LCSS, DTW) exposées en section 4.1.2 s'appliquent également à l'étude de la composante spatiale des séquences de positions.

Outre les mesures de similarité basées sur l'analyse de séries de données temporelles, les trajectoires peuvent également être comparées sur leurs formes géométriques.

5.2.1 Distance moyenne

La distance moyenne, illustrée sur la figure 5.1, est calculée à l'aide de la surface entre deux lignes divisée par la longueur de la ligne de référence [**116**]. Cette mesure n'est par conséquent pas symétrique.

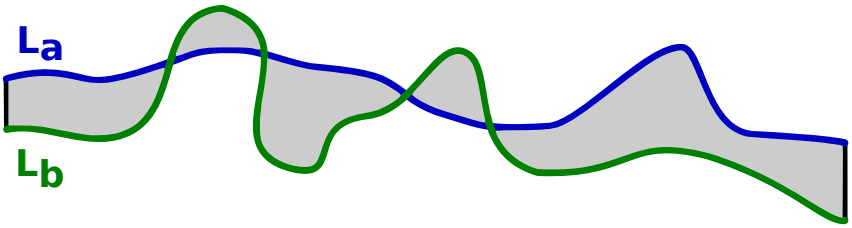


Figure 5.1 – Illustration du calcul de la surface entre deux lignes L_a et L_b

5.2.2 Distance de Hausdorff

La d_H correspond à l'écart maximal existant entre deux lignes L_a et L_b [**131**]. Elle est définie formellement par l'équation 5.2.

$$d_H(L_a, L_b) = Max\left(\underset{p_a \in L_a}{Max}\left(\underset{p_b \in L_b}{Min}\left(d(p_a, p_b)\right)\right), \underset{p_b \in L_b}{Max}\left(\underset{p_a \in L_a}{Min}\left(d(p_a, p_b)\right)\right)\right) \tag{5.2}$$

La distance de Hausdorff est la plus grande des deux distances entre :

- $d1$ qui est la plus grande distance parmi les distances minimales entre les points de L_a et le point le plus proche de L_b,
- $d2$ qui est la plus grande distance parmi les distances minimales entre les points de L_b et le point le plus proche de L_a.

Ces deux distances $d1$ et $d2$ sont illustrées sur la figure 5.2.

La distance de Hausdorff ne tient pas compte de l'ordre temporel des positions [**58**, **98**, **91**, **76**].

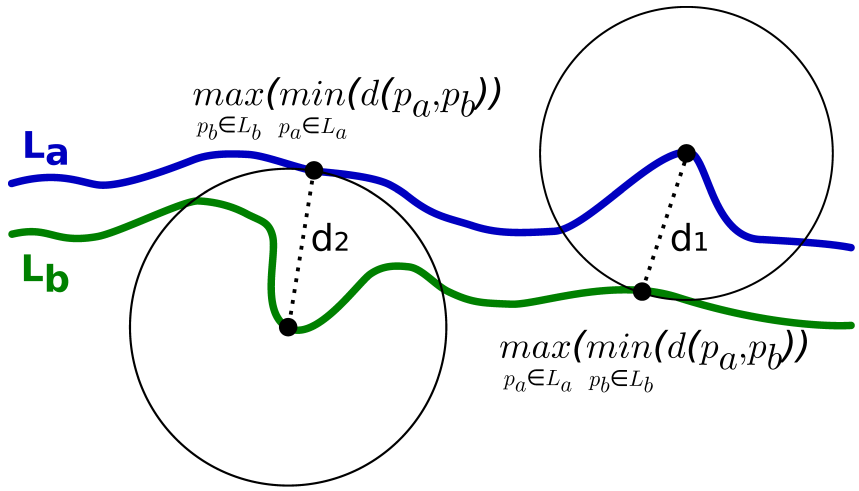


Figure 5.2 – Illustration de la distance de Hausdorff entre deux lignes L_a et L_b

5.2.3 Distance de Fréchet

La distance de Fréchet (d_F) [**132**] est une distance permettant de calculer la distance maximale entre deux lignes [**109**, **76**].

Cette distance est généralement illustrée en prenant l'exemple d'un maître promenant son chien en laisse. Chacun suivant son propre chemin, s'arrêtant et avançant, mais ne pouvant jamais revenir en arrière. La distance de Fréchet entre les chemins du maître et du chien peut être représentée par la longueur minimale de la laisse permettant au maître et au chien de se promener ensemble. Cependant, la complexité de calcul de la version continue de cette distance est relativement importante ($O(N_a\ N_b\ log^2(N_a\ N_b))$ ou N_a et N_b représentant le nombre de segments des lignes L_a et L_b [**101**, **76**].

5.2.4 Distance de Fréchet discrète

La distance de Fréchet peut être approximée à l'aide de l'algorithme proposé par [**105**] dont la complexité algorithmique est réduite à $O(N_a\ N_b)$. Cette distance est alors nommée distance de Fréchet discrète (d_{Fd}).

La distance de Fréchet discrète est particulièrement adaptée à l'étude des mobilités qui sont généralement composées de suites ordonnées temporellement de positions discrètes.

Cette distance est calculée par programmation dynamique en recherchant les couples successifs de positions minimisant la longueur de la laisse permettant au maître et à son chien de se déplacer sur un même chemin en partant du premier couple de points (L_{a_1}, L_{b_1})

jusqu'au couple de points (L_{a_N}, L_{b_M})

Le déplacement du maître et de son chien est régi par les trois actions suivantes :

- cas 1 : le maître et le chien avancent en même temps $(L_{a_{i+1}}, L_{b_{i+1}})$,
- cas 2 : seul le maître se déplace $(L_{a_{i+1}}, L_{b_i})$,
- cas 3 : seul le chien se déplace $(L_{a_i}, L_{b_{i+1}})$.

La distance de Fréchet discrète entre L_a et L_b est calculée de façon récursive en utilisant la formule 5.3.

$$d_{Fd}(L_a, L_b) = max \left(\begin{array}{l} d_E(L_{a_n}, L_{b_m}) \\ min \left(\begin{array}{ll} d_{Fd}(\{L_{a_1} \ldots L_{a_{n-1}}\}, \{L_{b_1} \ldots L_{b_m}\}) & \forall n > 1 \\ d_{Fd}(\{L_{a_1} \ldots L_{a_n}\}, \{L_{b_1} \ldots L_{b_{m-1}}\}) & \forall m > 1 \\ d_{Fd}(\{L_{a_1} \ldots L_{a_{n-1}}\}, \{L_{b_1} \ldots L_{b_{m-1}}\}) & \begin{array}{l} \forall n > 1 \\ \forall m > 1 \end{array} \end{array} \right) \end{array} \right) \quad (5.3)$$

Cependant, cette distance reste une approximation de la distance de Fréchet continue. L'erreur d'approximation est directement lié à la discrétisation utilisée et à la longueur du plus grand segment entre deux positions des trajectoires comparées. L'optimisation du pas de discrétisation des trajectoires est un axe de recherche que nous avons étudié dans la section 5.3.

Un exemple de calcul de la distance de Fréchet est présenté sur la figure 5.3.

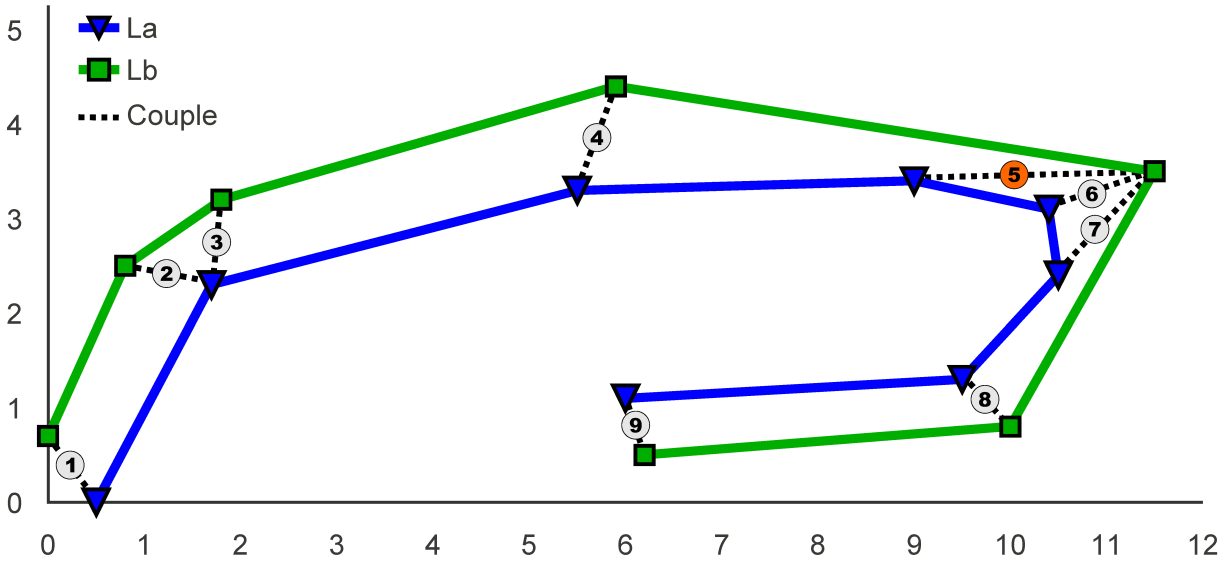


Figure 5.3 – Couples de points appariés de deux polylignes L_a et L_b

Pour chaque couple de points de L_a et L_b, la distance euclidienne d_E entre les points est calculée et sauvegardée dans la matrice de distances (MD) de taille ($N \times M$).

	N°	1	2	3	4	5	6	7	8
Ligne a	Xa	0,5	1,7	5,5	9	10,4	10,5	9,5	6
	Ya	0	2,3	3,3	3,4	3,1	2,4	1,3	1,1

Ligne b

N°	Xb	Yb
1	0	0,7
2	0,8	2,5
3	1,8	3,2
4	5,9	4,4
5	12	3,5
6	10	0,8
7	6,2	0,5

Matrice de distances

MD	1	2	3	4	5	6	7	8
1	0,86	2,33	6,08	9,40	10,67	10,64	9,52	6,01
2	2,52	0,92	4,77	8,25	9,62	9,70	8,78	5,39
3	3,45	0,91	3,70	7,20	8,60	8,74	7,93	4,70
4	6,97	4,70	1,17	3,26	4,68	5,02	4,75	3,30
5	11,54	9,87	6,00	2,50	1,17	1,49	2,97	6,00
6	9,53	8,43	5,15	2,79	2,33	1,68	0,71	4,01
7	5,72	4,85	2,89	4,03	4,94	4,70	3,40	0,63

Matrice de Fréchet

MF	1	2	3	4	5	6	7	8
1	0,86	2,33	6,08	9,40	10,67	10,67	10,67	10,67
2	2,52	0,92	4,77	8,25	9,62	9,70	9,70	9,70
3	3,45	0,92	3,70	7,20	8,60	8,74	8,74	8,74
4	6,97	4,70	1,17	3,26	4,68	5,02	5,02	5,02
5	11,54	9,87	6,00	2,50	2,50	2,50	2,97	6,00
6	11,54	9,87	6,00	2,79	2,50	2,50	2,50	4,01
7	11,54	9,87	6,00	4,03	4,94	4,70	3,40	2,50

Figure 5.4 – Matrices de distances et de Fréchet de deux polylignes L_a et L_b

La matrice de Fréchet (MF) est calculée par programmation dynamique en utilisant la formule 5.3. On remarque que le calcul de la distance de Fréchet est très proche de la mesure du DTW.

La distance de Fréchet discrète entre les deux lignes L_a et L_b est contenue dans la dernière cellule de la matrice de Fréchet ($MF_{[N,M]}$) valant $2,50$ pour notre exemple. Cette distance représente l'écartement maximal entre deux points homologues des deux lignes (le couple de points n°5 dans l'exemple de la figure 5.3 écartés de $2,50$m).

La distance de Fréchet discrète, contrairement à la distance de Hausdorff, respecte l'ordonnancement temporel des positions des trajectoires comparées.

5.2.5 Distance de Fréchet discrète moyenne

En utilisant les résultats obtenus par le calcul de la distance de Fréchet discrète, il est possible de calculer la distance moyenne entre tous les couples homogènes de positions

des deux trajectoires qui composent le chemin minimum (lignes en pointillés sur la figure 5.3 correspondant aux cases surlignées en vert dans les matrices de la figure 5.4).

Ce chemin minimum peut être obtenu à partir de la matrice de Fréchet par une analyse des valeurs des cellules des matrices de Fréchet et de distance en partant de la dernière cellule de la matrice de Fréchet et en remontant dans la matrice (back tracking) jusqu'à obtenir le couple des points de départ.

La moyenne des distances entre ces couples de points homogènes (cellules vertes dans la matrice de distances euclidiennes) peut être calculée. Pour notre exemple, la distance de Fréchet discrète moyenne d_{Fdm} vaut $(d_{Fdm}(L_a, L_b)) = ((0,63 + 0,71 + 1,49 + 1,17 + 2,50 + 1,17 + 0,91 + 0,92 + 0,86)/9) = 1,15$

5.2.6 Distance de Fréchet discrète partielle

Dans certains cas, les lignes à comparer disposent d'une emprise spatiale différente. On s'intéresse alors à trouver, au sein de la ligne la plus grande (L_a), la sous-partie L_c de la ligne L_a minimisant la distance de Fréchet discrète entre L_b et L_c. Devogele a proposé un algorithme de calcul de cette distance de Fréchet discrète partielle dans [**92**, **88**]. Lorsque les points de départ de L_a et L_b sont connus et appariés ensemble, l'inconnue à trouver est le premier point de L_{an} apparié avec le dernier point de L_b. L'algorithme se base sur l'étude des matrices de distances et de Fréchet pour détecter la plus faible valeur de distance de Fréchet dans la ligne correspondant à la dernière position de la plus petite ligne (L_b). Cet algorithme est également détaillé dans la section 2.2.6 de [**Ra2**].

5.3 Optimisation du calcul de la distance de Fréchet discrète

Tout au long des projets de recherche réalisés, différents corpus volumineux de données de mobilités ont été générés (Navires, Véhicules de secours, Pigeons voyageurs, Enfants allant à l'école, Parcours de touristes...).

La taille de ces corpus rend parfois difficile l'application de certaines mesures dont les complexités algorithmiques sont élevées.

D'un point de vue spatial, la distance de Fréchet discrète est une mesure intéressante permettant de comparer des trajectoires. Cependant, comme présenté en section 5.2.4,

son ordre de complexité $O(N * M)$, en lien direct avec le nombre de positions N et M des trajectoires à comparer, rend son usage coûteux.

C'est pourquoi il est intéressant de réduire au maximum le nombre de positions nécessaires pour représenter une trajectoire tout en conservant une précision acceptable.

Le filtrage de Douglas et Peucker spatio-temporel est particulièrement efficace pour représenter des trajectoires de mobilité. Ce filtrage induit un échantillonnage irrégulier des trajectoires avec de potentiels segments très longs lorsque les trajectoires réalisent de longues lignes droites à vitesse constante et une densification du nombre de positions en cas de changements de comportements importants.

Cependant, lorsque l'on souhaite comparer deux trajectoires filtrées à l'aide de la distance de Fréchet discrète, cette différence d'échantillonnage pose des problèmes.

La distance de Fréchet discrète se base uniquement sur les distances entre les positions discrètes des trajectoires. L'erreur maximale de mesure entre la distance de Fréchet discrète et la distance de Fréchet continue est alors directement liée à la longueur des plus grands segments. Il est alors nécessaire de détecter les cas particuliers induisant ces erreurs d'approximation.

Cette section présente les apports réalisés dans le cadre de l'optimisation du calcul de la distance de Fréchet discrète introduite dans la section 5.2.4. Ce verrou de recherche international a fait l'objet d'une compétition (GIS CUP 2017) lors de la conférence ACM SIGSPATIAL 2017 dans laquelle nous avons publié un article dans le Workshop Analytics for Big Geospatial Data [**CI7**].

En collaboration avec le Laboratoire d'Informatique Fondamentale d'Orléans (LIFO), nous avons proposé de nouveaux algorithmes de calcul de la distance de Fréchet discrète optimisée entre trajectoires diminuant les temps de calcul.

Dans cet article, nous avons calculé les ratios entre temps de calcul et l'écart de mesure avec de la distance de Fréchet discrète pour un jeu de 495 trajectoires de pigeons voyageurs composé de plus de 1,5 million de positions (soit environ 3000 positions par trajectoire).

Le calcul de la distance de Fréchet entre chaque couple de trajectoires est utilisé pour réaliser une matrice de distance entre trajectoires. Grâce à cette matrice de distance, un dendrogramme est généré afin de constituer des clusters de trajectoires (Figure 5.5).

L'algorithme non optimisé de la distance de Fréchet discrète nécessite de calculer une matrice de distance euclidienne entre chaque position des deux trajectoires à comparer. La complexité algorithmique de ce calcul est donc directement liée au nombre de positions

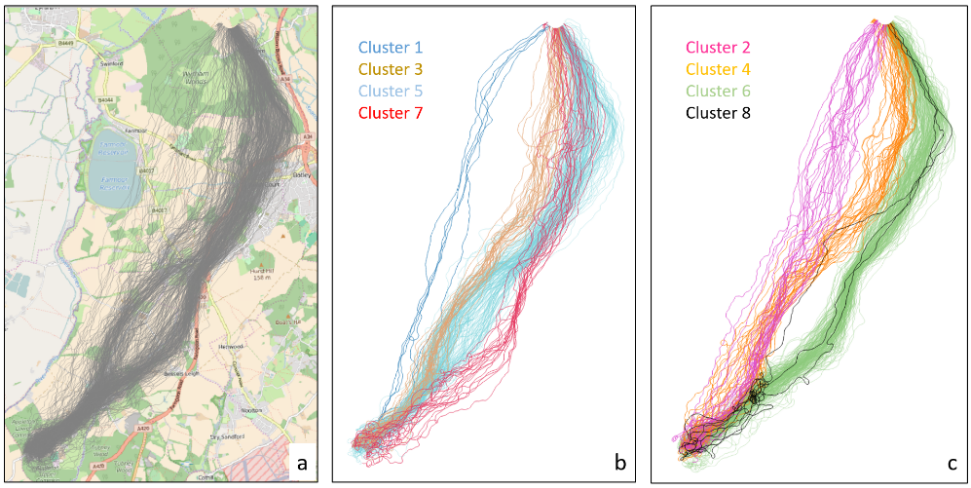


Figure 5.5 – Clusters des 495 trajectoires de pigeons voyageurs

de chaque trajectoire à comparer (3000_2) multiplié par le nombre total de couples de trajectoires soit 1096200000000 calculs de distances euclidiennes entre positions.

Afin de limiter le temps de calcul tout en conservant une approximation de la distance de Fréchet discrète acceptable, différentes optimisations ont été proposées.

La première consiste à limiter le nombre de points requis pour représenter chaque trajectoire en filtrant celle-ci à l'aide d'un filtrage de Douglas et Peucker. La seconde optimisation retenue consiste à limiter les calculs de distances entre positions des deux trajectoires à comparer grâce â une heuristique basée sur le calcul de la diagonale de la matrice de distance. Enfin, la troisième optimisation proposée vise à recréer certains points par projection sur les longs segments de trajectoires afin de limiter l'erreur de calcul de la distance de Fréchet discrète.

Les résultats obtenus en faisant varier le paramètre de filtrage de Douglas et Peucker sont présentés en bleu sur la figure 5.6 [**CN1**]. L'ajout des points projetés, en orange sur la figure 5.6, permet de minimiser l'erreur de calcul de distance. Cependant, le temps de calcul induit par l'ajout de ces points est relativement coûteux.

Une étude approfondie des techniques de résolution par projections de nouveaux points sur les segments doit être réalisée.

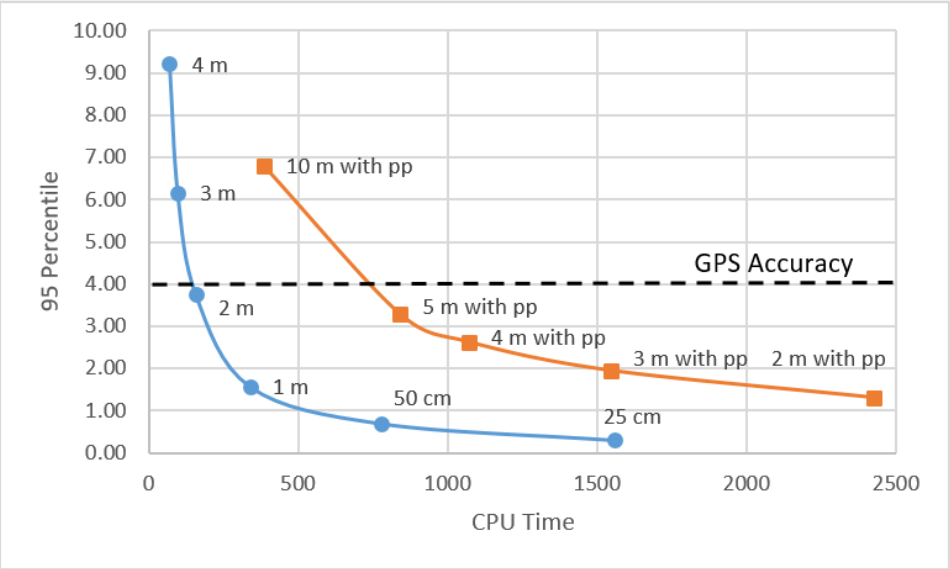


Figure 5.6 – Relation entre le temps de calcul CPU, l'erreur de mesure de la distance de Fréchet discrète optimisée et le paramètre de filtrage de Douglas et Peucker pour l'ensemble des 495 trajectoires de pigeons

Chapitre 6

PATRONS SPATIO-TEMPORELS

Dans les chapitres 4 et 5, différentes mesures de similarité entre trajectoires ont été proposées. Ces mesures se focalisent sur la représentation spatiale des trajectoires ou sur leur aspect sémantique. Le chapitre 3 quant à lui s'est intéressé à l'apport des objets connectés dans l'étude des mobilités ainsi que les limites liées à l'usage de système de positionnement par satellite (GPS).

L'axe de recherche présenté dans ce chapitre porte sur l'usage de ces mesures de similarité pour la définition de groupes homogènes de trajectoires (clustering) ainsi que la découverte de patrons (patterns) synthétisant ces groupes.

Lorsque l'on dispose d'un nombre important de données de mobilité, il est parfois difficile d'arriver à bien comprendre le comportement des objets mobiles ayant réalisé ces trajectoires. C'est pourquoi il est intéressant de synthétiser ce comportement à l'aide de patron en complément de règles et connaissances provenant d'experts [**6**]. Dans cette section, on s'intéresse à l'extension de concepts de statistiques descriptives à l'étude des trajectoires.

Le schéma fonctionnel de la figure 6.1, issu de [**Ra2**] présente les différentes étapes nécessaires à la génération et à la visualisation de ces patrons dans un contexte applicatif maritime.

6.1 Patrons de nuages de positions

Dans [**CI10**], nous avons présenté une étude des différentes techniques de synthèse et de visualisation de patrons de nuages de positions. Ces patrons sont en général composés d'un élément central et de bordures décrivant l'emprise spatiale, la densité du nuage de positions et sa dispersion. Ils sont particulièrement intéressants pour comparer des nuages de positions par analyse visuelle, détecter des positions inhabituelles (outliers), comprendre les évolutions de ces nuages dans le temps et dans l'espace.

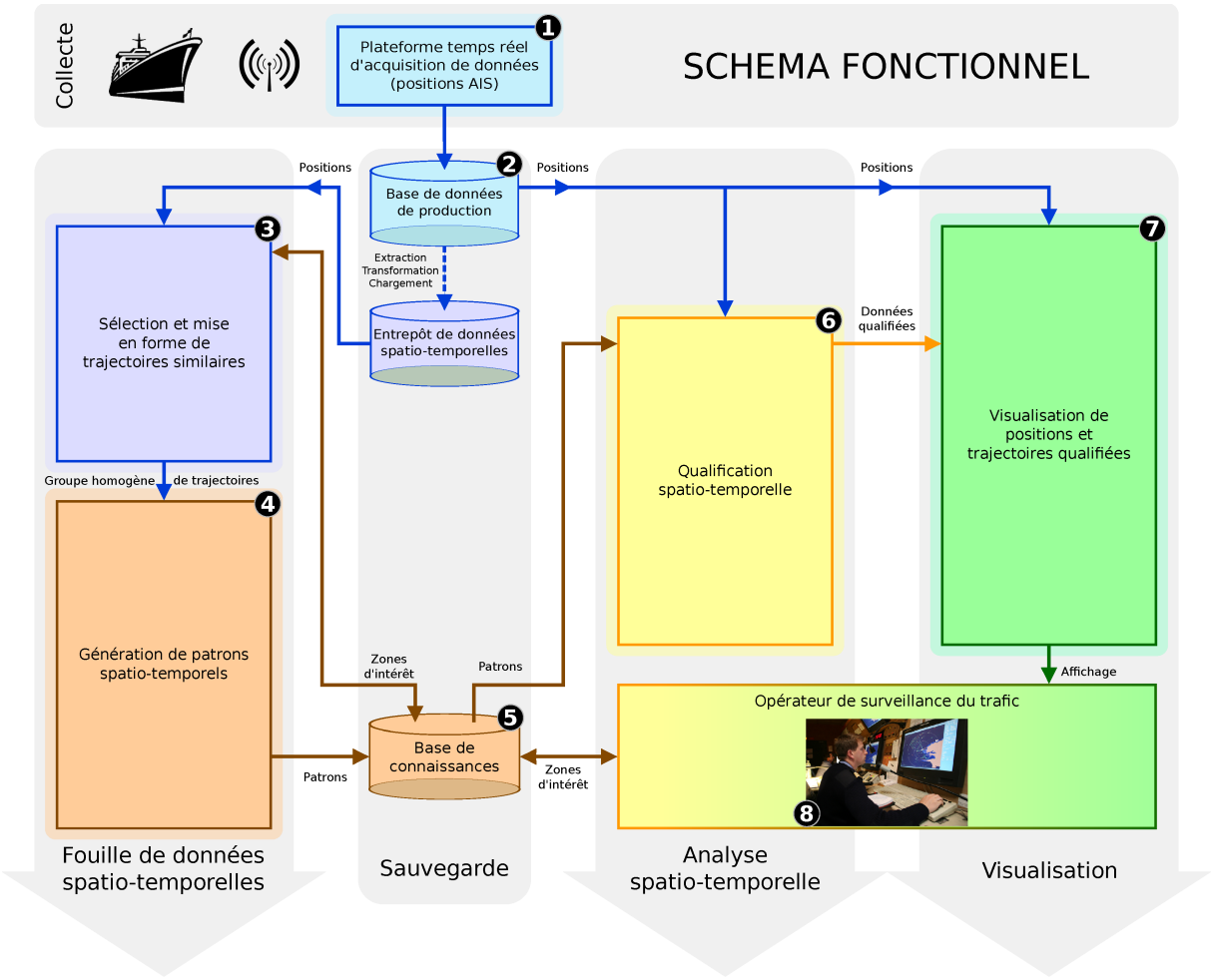


Figure 6.1 – Schéma fonctionnel du processus de génération de patrons de groupes de trajectoires

6.1.1 La position centrale d'un nuage de positions

Le premier concept s'intéresse à la représentation de la tendance centrale d'un nuage de positions. L'élément central d'un ensemble de données est souvent calculée à l'aide d'une valeur moyenne ou médiane. Il existe différentes généralisations de ces concepts pour des données multidimensionnelles telles que le barycentre, la médiane géométrique ou le médoïde [**111**, **84**]. Ces différentes représentations ont été présentées dans [**CI10**] et [**Re6**]. La figure 6.2 présente un nuage de points 2D ainsi que les différentes représentations des tendances centrales étudiées.

Le barycentre, figurant en rouge sur la figure 6.2, est calculé à l'aide de la moyenne arithmétique des différentes composantes des coordonnées (x,y et potentiellement z) de

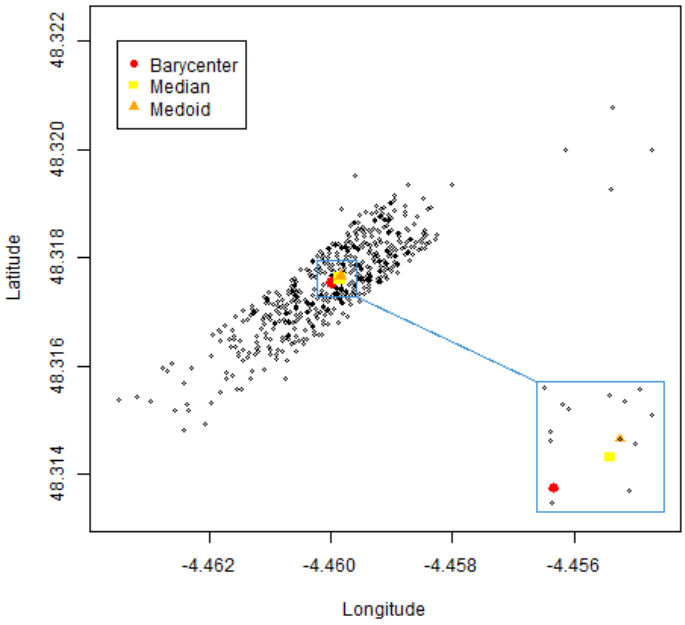


Figure 6.2 – Barycentre, médoïde et médiane géométrique d'un nuage de points.

chacune des positions du nuage. Les coordonnées du barycentre sont calculées et ne correspondent pas forcément à une des positions du nuage. De plus, le barycentre est relativement sensible à la présence de données aberrantes (outliers).

Une autre approche calcule les valeurs médianes des composantes des coordonnées des positions du nuage pour former la médiane géométrique. La médiane géométrique, présentée en jaune sur la figure 6.2, est moins sensible à la présence de données aberrantes, mais ne correspond pas non plus à une position réelle du nuage.

Enfin, il est parfois intéressant de s'assurer que la représentation centrale du nuage correspond bien à une des positions du cluster. Le médoïde, illustré en orange sur la figure 6.2, correspond à la position du nuage de points qui minimise la distance à tous les autres points du nuage.

6.1.2 Les bordures d'un nuage de positions

Une fois la représentation centrale d'un nuage de positions définie, il est également intéressant de qualifier la dispersion de ce nuage autour de cette tendance centrale.

Cette dispersion est classiquement représentée, pour des données unidimensionnelles, à l'aide de la notion d'écart type lorsqu'on étudie la moyenne ou de boites à moustaches (percentiles) dans la cadre d'une approche médiane.

Dans le cas de données multivariées pour une approche multinormale, il est important d'étudier la corrélation entre ces variables. Cette corrélation peut être caractérisée à l'aide d'une matrice de co-variance présentée à l'équation 6.1. Sachant que la covariance de deux variables aléatoires indépendantes est nulle, la matrice de covariance est une matrice diagonale s'il n'existe pas de relations statistiques entre les variables.

$$\begin{pmatrix} var(X) & covar(X,Y) \\ covar(X,Y) & var(Y) \end{pmatrix} \tag{6.1}$$

Le coefficient de corrélation de Pearson décrit la relation linéaire qui existe entre deux variables [**134**]. Considérant un nuage de n positions, l'équation 6.2 permet de calculer le coefficient de corrélation de Pearson pour ce nuage.

Ce coefficient est positif $]0,1]$ lorsque les deux variables évoluent dans la même direction, nul lorsqu'il n'y a pas de corrélation entre les variables et négatif $[-1,0[$ lorsque les variables évoluent dans des directions opposées.

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{6.2}$$

Le coefficient de corrélation de Pearson des variables X et Y correspondant aux coordonnées des points du nuage de la figure 6.2 est de 0,8493. Ce coefficient indique une importante corrélation positive entre les deux composantes X et Y du nuage. En effet, on observe bien que les points dont les coordonnées x sont élevées ont également tendance à avoir une coordonnée y élevée.

Lorsque les variables observées suivent une loi normale, il est alors possible d'utiliser cette corrélation entre les variables X et Y pour définir une ellipse représentant la dispersion des positions au sein du nuage autour de la moyenne. La Standard Deviational Ellipse (SDE) et la Cross of Dispersion définissent l'orientation de l'ellipse en fonction de la corrélation entre X et Y. Les largeurs des deux axes de l'ellipse sont définies selon l'écart type. La figure 6.3 présente l'ellipse obtenue pour le nuage de positions d'exemple.

Malheureusement, la SDE n'est pas adaptée à des variables dont les distributions ne suivent pas une loi normale et ne sont pas symétriques.

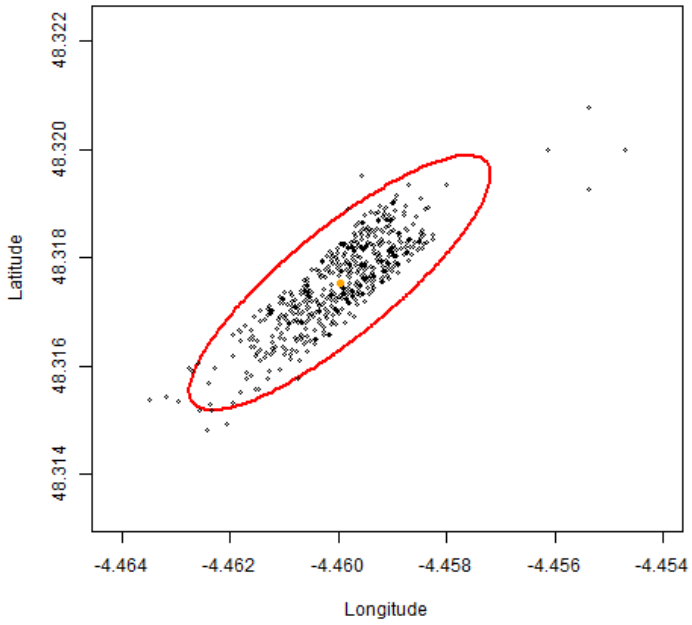


Figure 6.3 – Standard Deviational Ellipse (SDE) du nuage de positions.

Le skewness ([**77**]) est une mesure qui permet de qualifier la symétrie d'une distribution. La figure 6.4 illustre différentes distributions dont les symétries varient. Le skewness d'une distribution symétrique est nul (figure 6.4.b). Le skewness est négatif lorsque la densité est plus forte du côté droit de la distribution et que la queue de la distribution est plus longue du côté gauche (figure 6.4.a). Inversement, le skewness est positif lorsque la densité est plus forte du côté gauche de la distribution et que la queue de la distribution est plus longue du côté droit (figure 6.4.c).

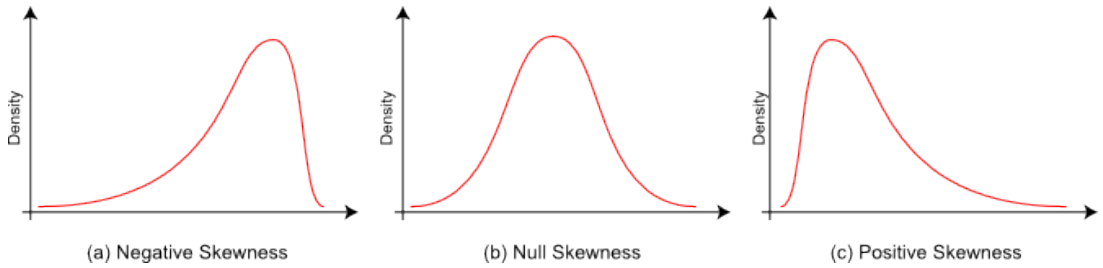


Figure 6.4 – Skewness de différentes distributions.

Le skewness décrivant la distribution des latitudes (y) du nuage de positions vaut

-0,149 alors que celui des longitudes (x) vaut -0.818. Ces deux valeurs indiquent que les distributions ne sont pas symétriques. C'est pourquoi il est important de disposer de patrons en mesure de synthétiser des nuages de positions dont les distributions ne sont pas forcément symétriques.

Le Minimum Convex Polygon (MCP) défini par Mohr dans [**129**] représente le plus petit polygone convexe entourant le nuage de positions (figure 6.5). Afin d'éviter les perturbations liées aux données aberrantes, il est possible de définir le MCP en excluant 5% des positions les plus éloignées de la tendance centrale (telle que définie dans la section 6.1.1) [**114**, **94**, **82**].

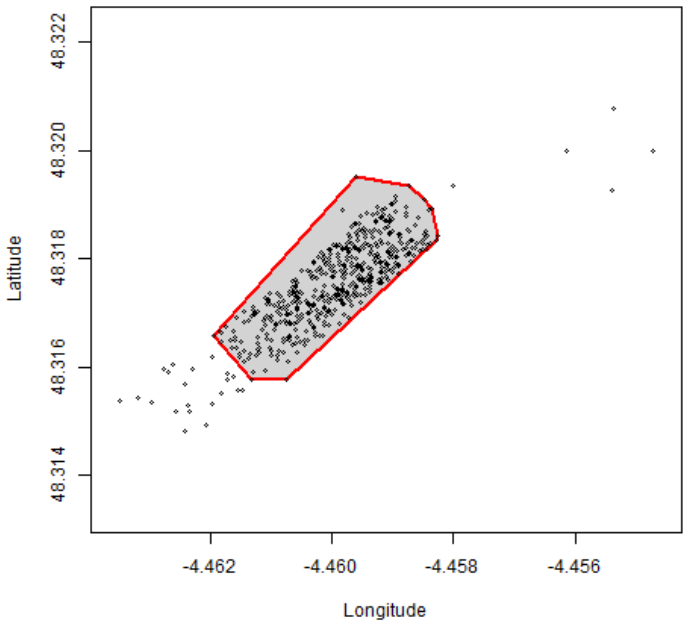


Figure 6.5 – Polygone convexe minimum du nuage de positions englobant 95% des positions.

Pour définir l'enveloppe du MCP, il est parfois nécessaire d'utiliser de nombreux points lorsque le nuage est conséquent. De plus, les enveloppes sont relativement complexes et coûteuses à calculer.

Les boites à moustaches (boxplot), introduites par Tukey dans [**120**], sont la représentation graphique classique utilisée en statistiques descriptives. Elles permettent de visualiser graphiquement différents paramètres statistiques de la distribution en se basant sur l'étude

des percentiles :

- Le minimum ou la limite basse (5%) ;
- Le premier quartile $Q1$ (25%) ;
- La médiane (50%) ;
- Le troisième quartile $Q3$ (75%) ;
- Le maximum ou la limite haute (95%) ;

La figure 6.6 correspond aux boites à moustaches des longitudes et latitudes du nuage de positions d'exemple. La médiane est représentée en orange, la boite verte représente l'espace interquartile entre le premier et troisième quartile. Enfin, les moustaches représentées par les barres rouges indiquent les limites au-delà desquelles on considère une donnée comme aberrante (5%).

Il existe différentes façons de définir les limites des moustaches. Dans certains cas, on considère un percentile (5%) comme étant la limite (on considère alors que 5% des données les plus fortes et les plus faibles sont considérées comme aberrantes). L'autre méthode utilise l'Espace inter-quartile (EIQ), correspondant à la taille de la boite verte soit $Q3 - Q1$, comme taille maximale au-delà de laquelle une donnée est considérée comme aberrante. La limite inférieure de la boite à moustache est alors fixée à la plus grande des valeurs entre le minimum et $(Q1 - 1,5 * EIQ)$. La limite supérieure est fixée à la plus petite des valeurs entre le maximum et $(Q3 + 1,5 * EIQ)$.

Les boites à moustaches sont particulièrement intéressantes pour visualiser les différences de symétrie (skewness) des distributions, mais également pour observer les différents étalements de ces distributions.

Le kurtosis caractérise cet étalement comme illustré dans la figure 6.7. Les distributions dont le kurtosis est supérieur à 1 (figure 6.7.c) sont très pointues au niveau de la médiane et disposent d'une queue plus plate et longue résultant en une boite interquartile très courte avec des moustaches plus éloignées. A contrario, les distributions dont le kurtosis est inférieur à 1 sont plutôt aplaties avec une queue plus courte (figure 6.7.a). Leurs boites à moustaches ont alors des boites interquartiles plus grandes et des moustaches plus rapprochées.

Lorsque les dimensions étudiées sont décorrélées, il est possible d'étendre le concept des boites à moustache à plusieurs dimensions comme proposé par Becketti et Gould qui définit les Rangefinder Box Plot dans [**113**]. Cependant, lorsque les coordonnées x et y

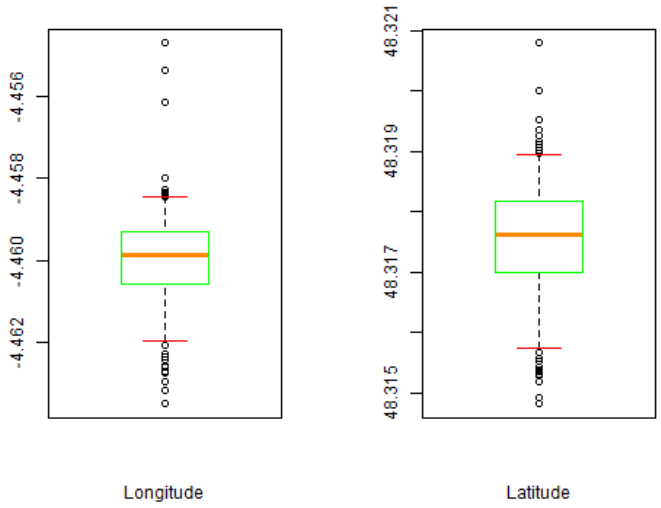


Figure 6.6 – Boites à moustaches des longitudes et latitudes du nuage de positions.

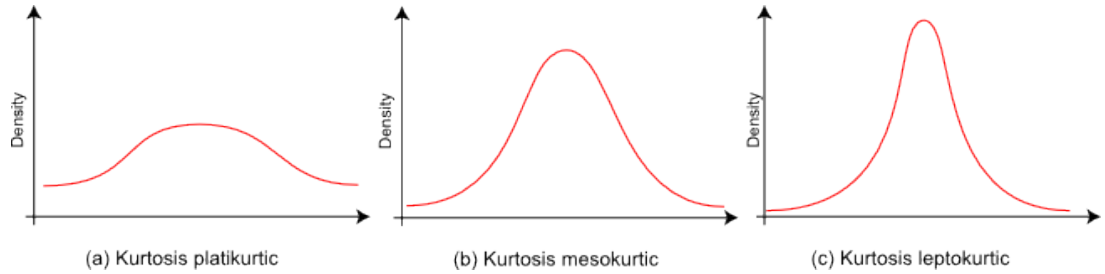


Figure 6.7 – Illustration des kurtosis de distributions de formes différentes.

d'un nuage de positions sont corrélées, les résultats obtenus ne sont pas concluants comme l'illustre la figure 6.8.

Des solutions hybrides ont été proposées pour étendre les MCP en intégrant les concepts des boites à moustaches. Rousseeuw, Ruts et Tukey dans [**93**] a proposé de représenter deux polygones emboîtés contenant différents sous-ensembles du nuage de positions. Cette approche nommée BagPlot consiste à définir un premier polygone (inner bag) englobant 50% des positions les plus proches de la position centrale puis de définir un second polygone de bordure (outer bag) englobant 95% des positions du nuage.

Le BagPlot est particulièrement intéressant, car il permet de visualiser la symétrie, la forme, la dispersion du nuage de positions tout en s'appliquant à des données corrélées. Cependant, cette technique requiert l'usage de nombreux point pour définir les différentes bordures. Ceci rend le calcul et la comparaison des BagPlot complexe.

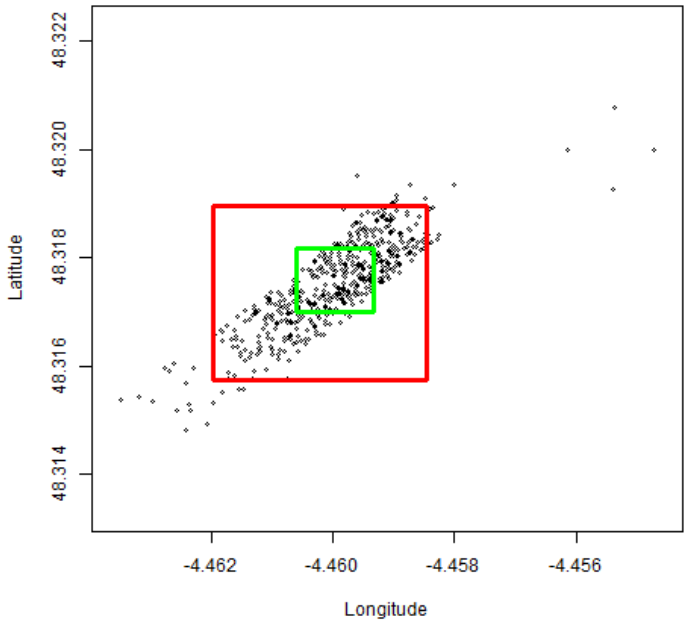


Figure 6.8 – Rangefinder boxplot.

Le Quelplot (Quarter elliptic plot) proposé par Goldberg et Iglewicz dans [**110**] étend les concepts de la SDE en utilisant différentes ellipses combinées sur leurs grands et petits axes. Le quelplot est en mesure de représenter des distributions asymétriques en réalisant des estimations statistiques pour modéliser la distribution sous la forme d'une déformation complexe d'une distribution normale. Les différentes limites générées par les ellipses sont présentées dans la figure 6.10.

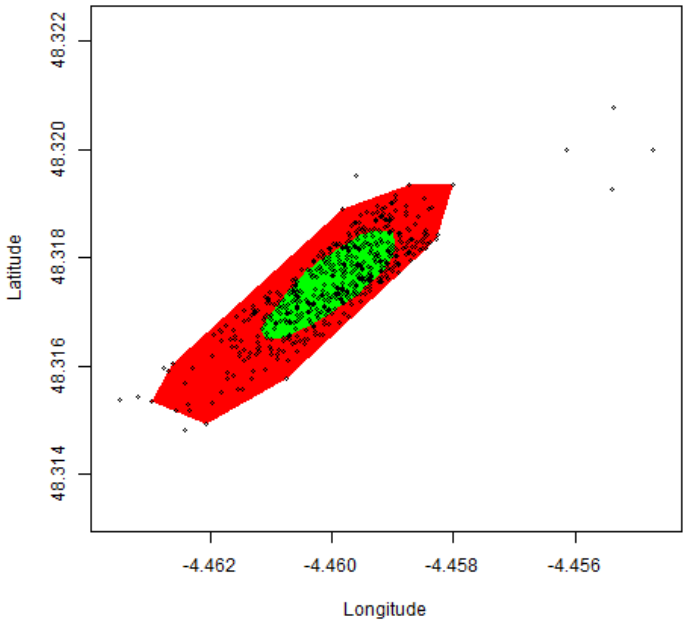


Figure 6.9 – BagPlot.

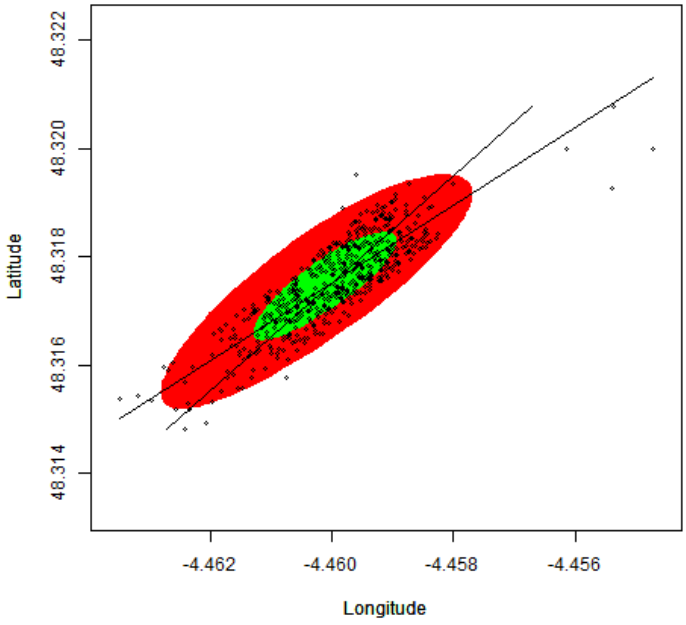


Figure 6.10 – Quelplot.

6.1.3 Oriented Spatial Box Plot

Notre proposition nommée OSBP, publiée dans [**CI10**] et [**Re6**] est une extension des travaux de Tongkumchum sur les boxplot en 2 dimensions [**80**].

Le patron proposé doit être en mesure de représenter l'élément central d'un nuage de positions ainsi que deux zones s'inspirant des boxplots pour la définition d'un espace interquartile 2D ainsi qu'une limite au-delà de laquelle les positions sont considérées comme anormales. La figure 6.11 décrit les différentes étapes du processus de génération de ce patron.

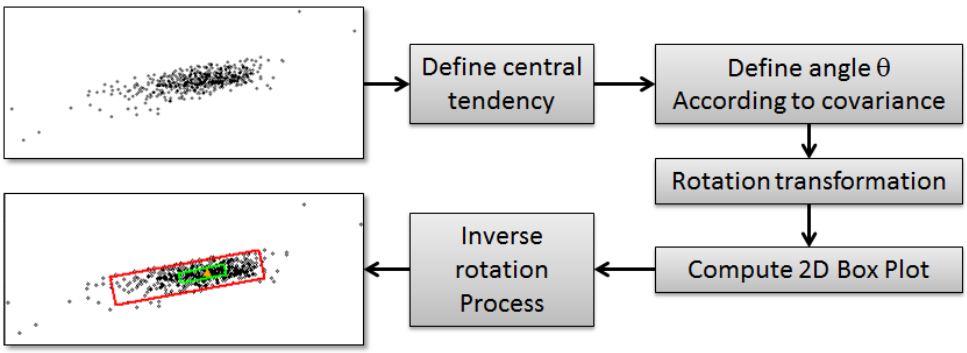


Figure 6.11 – Processus de calcul d'un OSBP.

La première étape consiste à calculer l'élément central du nuage de positions en s'inspirant des techniques présentées à la section 6.1.1. Les boites à moustaches sont basées sur une approche médiane/percentiles. Aussi, la médiane géométrique ou le médoïde sont préférables au barycentre. Dans la suite de cette section, le médoïde est choisi, car il a l'avantage de correspondre à une position réelle du nuage.

La seconde étape de processus de calcul de l'OSBP consiste à découvrir la relation entre les coordonnées X et Y des positions du nuage. L'objectif recherché est de minimiser la taille des boites englobantes afin d'éviter la situation représentée par la figure 6.8 pour le Range Finder Plot.

C'est pourquoi il est nécessaire de calculer la relation linéaire entre les coordonnées du nuage à l'aide du coefficient de corrélation de Pearson présenté à l' équation 6.2. Lorsque les composantes X et Y des coordonnées du nuage de positions sont corrélées, il est alors possible de réaliser une Analyse en Composantes Principales (ACP) [**134**, **130**]. La première composante de cette ACP permet de calculer et de représenter la droite de régression linéaire illustrée à la figure 6.12. L'angle θ entre cette droite et l'axe des

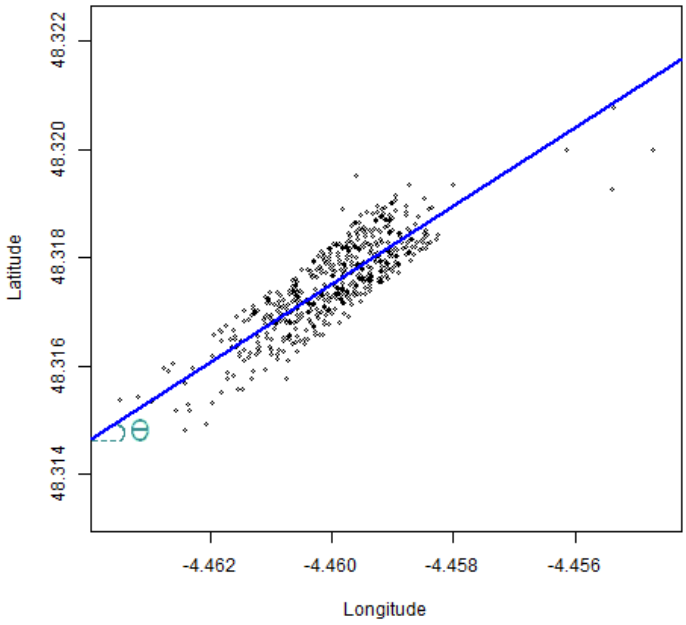


Figure 6.12 – Droite de régression linéaire.

abscisses, représenté en vert sur la figure 6.12, est alors donné par l'équation 6.3. Cet angle vaut 11, 56 pour le nuage de positions d'exemple.

$$\theta = tan^{-1}\left(\frac{cov(X,Y)}{var(X)}\right) \tag{6.3}$$

Disposant de cet angle θ, il est alors simple d'effectuer un changement de repères afin de reprojeter toutes les positions du nuage selon ce nouvel axe de référence (figure 6.13). On note (x', y') les coordonnées des points projetés par rotation à l'aide des équations 6.4 et 6.5. Le nouvel axe de direction principal des abscisses est alors nommé X', et Y' pour le nouvel axe des ordonnées (figure 6.13).

$$x' = xcos(\theta) - ysin(\theta) \tag{6.4}$$

$$y' = xsin(\theta) + ycos(\theta) \tag{6.5}$$

Les positions ayant été projetées dans un nouvel espace, il est alors possible de calculer le Range Finder Plot pour le nuage de positions projetées. Les limites pour considérer une position comme aberrante sur chaque axe X' et Y' sont définies à 2,5% par axe (1.25% des

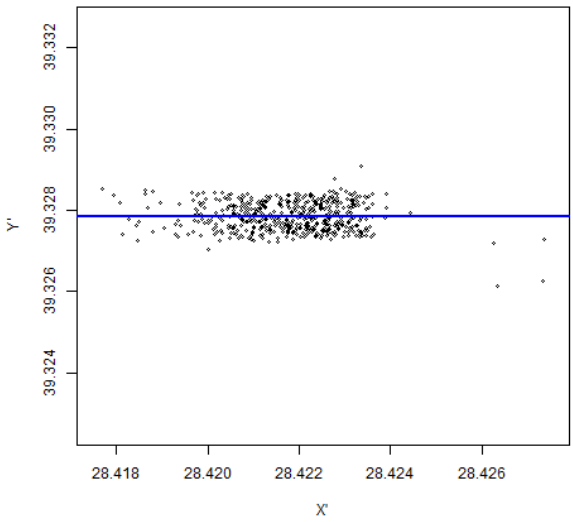


Figure 6.13 – Nuage de positions réorienté selon l'axe principal de l'ACP.

valeurs les plus faibles et 1.25% des valeurs les plus fortes). Les percentiles (1,25%, 25%, 75%, 98,75%) des coordonnées projetées du nuage sur X' et Y' sont utilisés pour définir les limites des rectangles englobants internes (vert) et externes (rouge). Ces rectangles sont tracés sur la figure 6.14. Les boxplot 1D de chaque axe sont dessinés sur les bords de la figure 6.14.

Finalement, le processus de rotation inverse ($-\theta$) est appliqué aux coordonnées définissant les angles des deux rectangles englobants en utilisant les équations 6.4 et 6.5. Le patron OSBP est alors constitué de :

- Les coordonnées du point central du nuage ;
- Les 4 coordonnées du rectangle interne ;
- Les 4 coordonnées du rectangle externe.

La figure 6.15 présente l'OSBP final obtenu pour le nuage de points d'exemple. Les positions qui se trouvent en dehors du rectangle rouge sont alors considérées comme des données aberrantes (outlier). L'OSBP est en mesure de représenter efficacement la distribution d'un nuage de points 2D. Sur l'exemple de la figure 6.15, la différence de densité au sein du nuage est visible à l'aide du positionnement relatif des rectangles englobants par rapport à la position centrale représentée par le triangle orange.

Le positionnement relatif de ces rectangles par rapport au point central est un indicateur

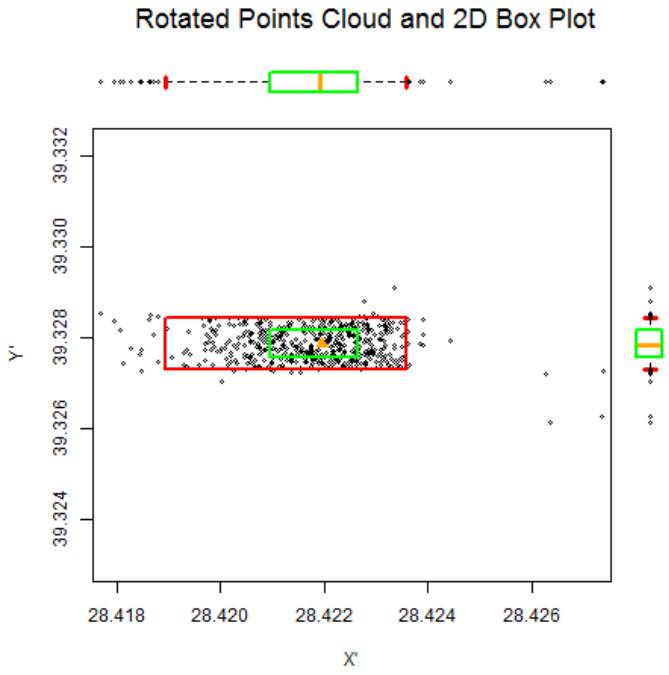


Figure 6.14 – BoxPlot 2D du nuage de positions reprojeté sur la composante principale de l'ACP.

visuel intéressant pour définir la symétrie et l'aplatissement de la distribution. Dans [**Re6**], nous avons proposé deux indicateurs décrivant la symétrie et l'aplatissement du nuage de positions. Ces indicateurs se basent sur des mesures de distances normalisées entre le point central et les différentes bordures des rectangles internes (a,b,c,d) et externes (A,B,C,D) comme présenté sur la figure 6.16.

Les distances entre les bordures du rectangle externe et le médoïde sont utilisées pour définir le coefficient de symétrie (s) à l'aide de l'équation 6.6. Ce coefficient tend vers 0 lorsque la distribution 2D du nuage de positions est symétrique (Si $A = B$ et $C = D$). Deux exemples d'OSBP de distributions asymétriques sont présentés sur la figure 6.17.

$$s = \frac{(A+B)}{(A+B+C+D)} * \frac{|A-B|}{(A+B)} + \frac{(C+D)}{(A+B+C+D)} * \frac{|C-D|}{(C+D)} \tag{6.6}$$

Les distances entre les bordures des deux rectangles interne et externe avec le médoïde sont utilisées pour définir le coefficient d'aplatissement (k) à l'aide de l'équation 6.7. Le ratio des surfaces des rectangles internes et externes est calculé. Plus ce coefficient est faible, plus les positions sont regroupées autour de la position centrale. Deux exemples

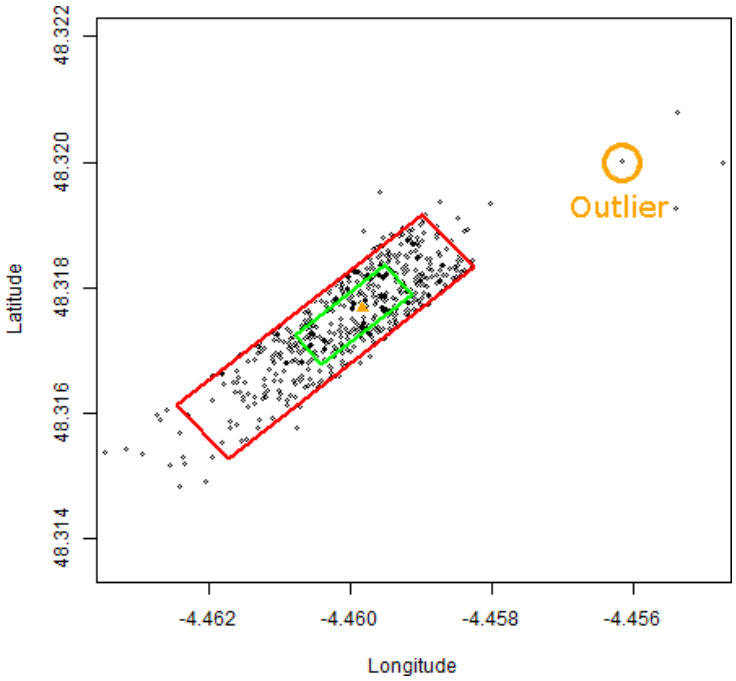


Figure 6.15 – OSBP du nuage de positions.

d'OSBP de distributions disposant de coefficients d'aplatissement différents sont présentés sur la figure 6.18.

$$k = \frac{((a+b)*(c+d))}{((A+B)*(C+D))} \tag{6.7}$$

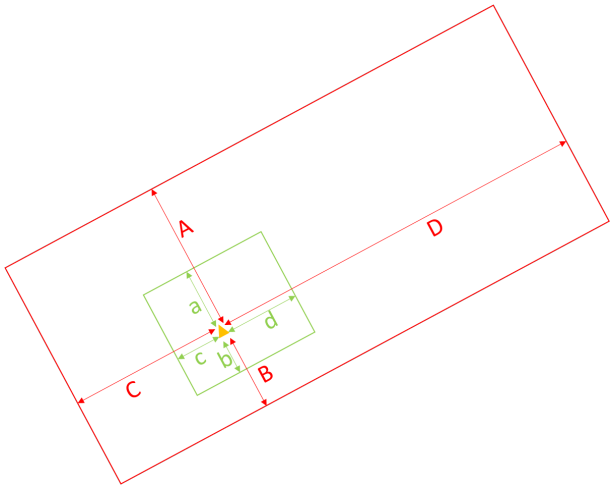


Figure 6.16 – Distances utilisées pour mesurer la symétrie et l'aplatissement du nuage de positions.

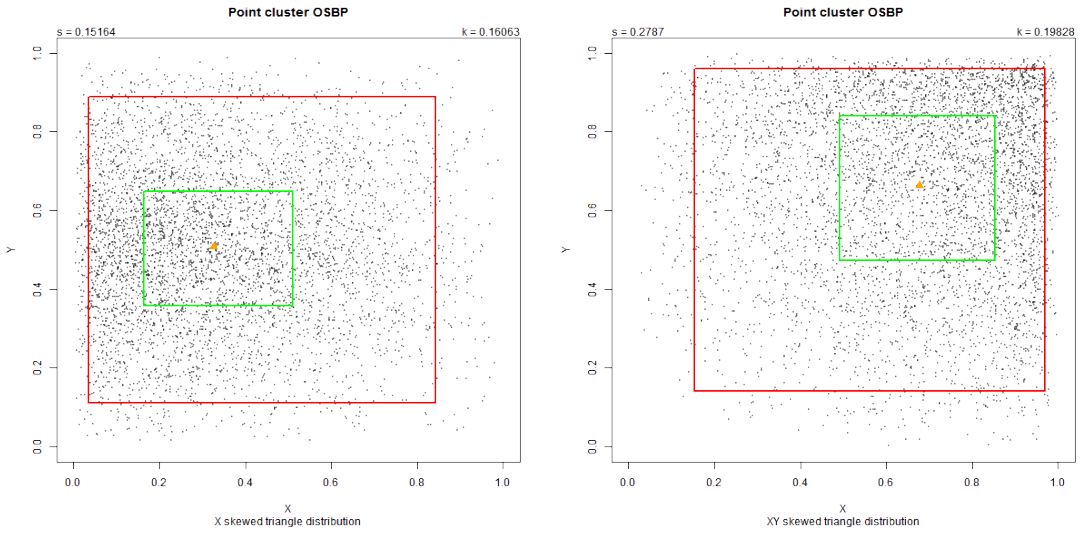


Figure 6.17 – Illustration de deux nuages de points disposant d'une symétrie différente.

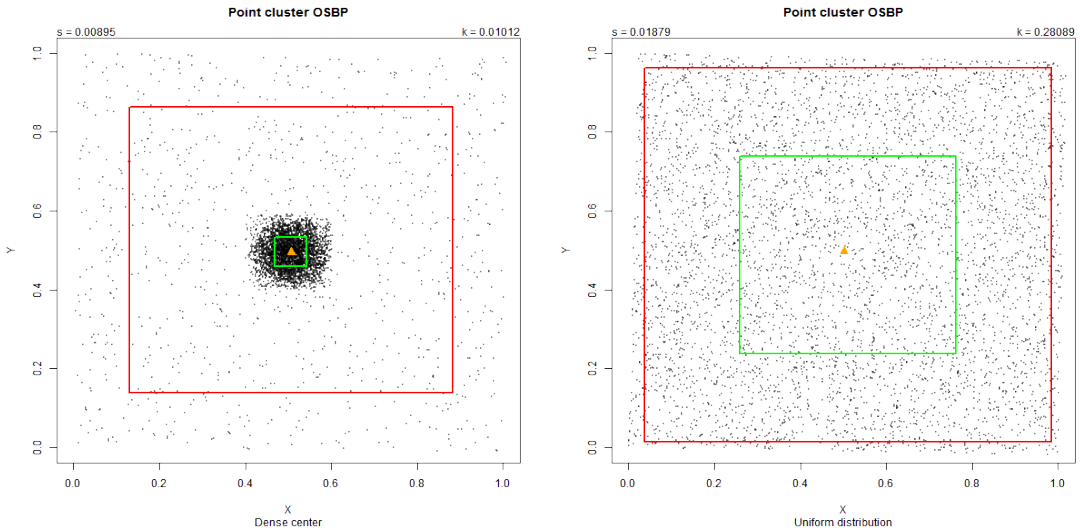


Figure 6.18 – Illustration de deux nuages de points disposant d'un coefficient d'aplatissement différent.

6.1.4 Oriented Spatio-Temporal Box Plot

L'OSBP présenté à la section 6.1.3 ne tient pas compte de la composante temporelle des nuages de positions. Dans [**Re5**], nous avons proposé une extension du modèle OSBP à l'aide de cube 3D imbriqués. La dimension temporelle est cependant traitée séparément des composantes spatiales.

Toutes les estampilles temporelles du nuage de positions sont triées afin de calculer la boite à moustache 1D de cette composante. Puis les différents percentiles (1,25%, 25%, 50%, 75%, 98,75%) sont utilisés pour définir les limites temporelles de chaque cube 3D. Le temps étant alors représenté sur la troisième dimension (Z). Une nouvelle position centrale est générée en remplaçant l'attribut temporel du médoïde du nuage par la valeur de la médiane des temps du nuage. Le rectangle interne est extrudé sur l'axe Z en utilisant les limites basses et hautes de l'espace interquartile temporel (25% et 25%). Enfin le rectangle externe est également extrudé sur l'axe Z en utilisant les limites temporelles basses et hautes (1, 25% et 98, 75%).

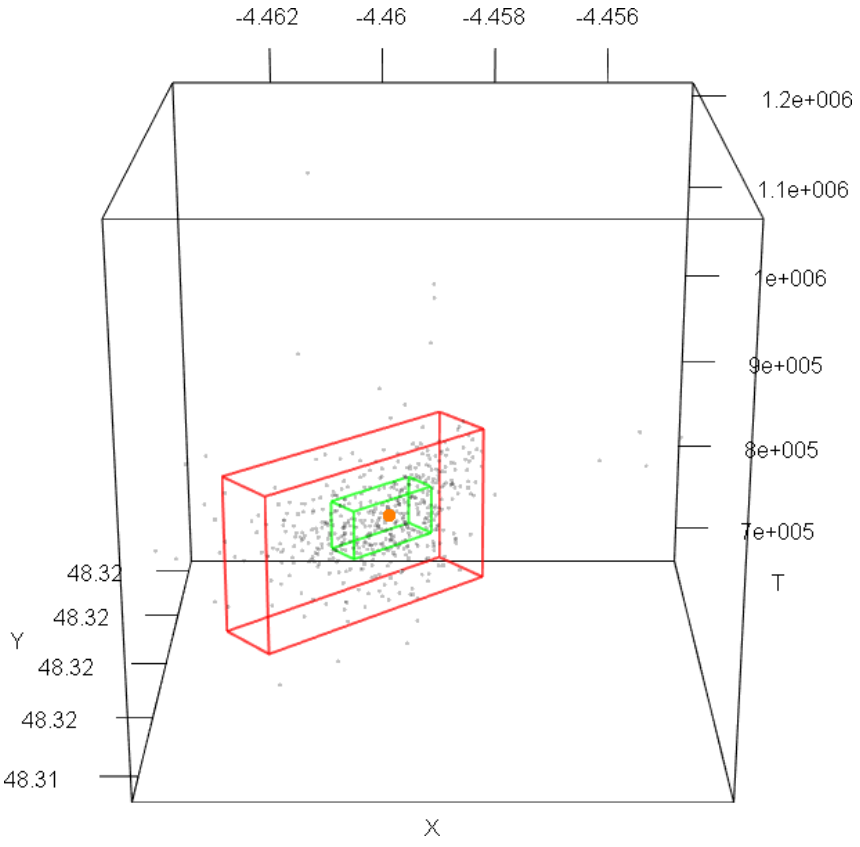


Figure 6.19 – OSTBP du nuage de positions d'exemple.

La figure 6.19 illustre les capacités de représentation visuelle de symétrie et d'étalement d'un nuage de positions en 3D. Le positionnement relatif ainsi que les différences de volumes des cubes offrent une représentation synthétique de la densité du nuage de positions. De plus, les différents espaces 3D permettent de qualifier le comportement d'un objet mobile comparé aux autres positions du nuage. Un élément qui se trouve à l'intérieur du cube interne vert peut alors être qualifié de très similaire aux autres éléments du nuage. Les éléments contenus en dehors du cube rouge sont des éléments anormaux (spatialement et/ou temporellement). Enfin l'espace contenu dans le cube rouge en dehors du cube vert est un espace de transition.

6.2 Patrons de trajectoires

Dans la section 6.1, nous nous sommes focalisés sur les patrons synthétisant le comportement d'un nuage de positions. Cependant, dans le cas des trajectoires, ces nuages de positions sont connectés les uns aux autres pour former des ensembles de trajectoires.

La section 5 a introduit différentes mesures de similarité entre trajectoires ainsi que nos apports pour un calcul efficient de la distance de Fréchet discrète optimisée (voir section 5.3) ainsi que la capacité de cette distance à générer des couples homogènes de positions via un algorithme d'alignement des positions par backtracking (voir section 5.2.5).

Cependant, cette mesure de similarité est relativement coûteuse à calculer. C'est pourquoi nous avons proposé des optimisations de calcul (section 5.3) se basant sur une réduction du nombre de positions nécessaire pour représenter les trajectoires à l'aide d'un filtre de Douglas et Peucker spatio-temporel [**124**, **71**].

Aussi, l'alignement des couples de positions homogènes entre deux trajectoires peut être fortement impacté par ce filtrage initial.

Cette section présente les travaux réalisés pour la génération de patrons de trajectoires synthétisant le comportement central d'un groupe de trajectoires homogènes ainsi que différents paramètres qualifiant l'évolution de la densité du groupe au cours du temps (étalement, symétrie) tout comme pour un nuage de positions.

6.2.1 La trajectoire médiane

Tout comme le médoïde représente l'élément central d'un nuage de positions, il est intéressant de définir la trajectoire dite "centrale" d'un groupe de trajectoires homogènes. Dans [**CN2**], nous avons proposé une méthode de calcul de cette trajectoire centrale.

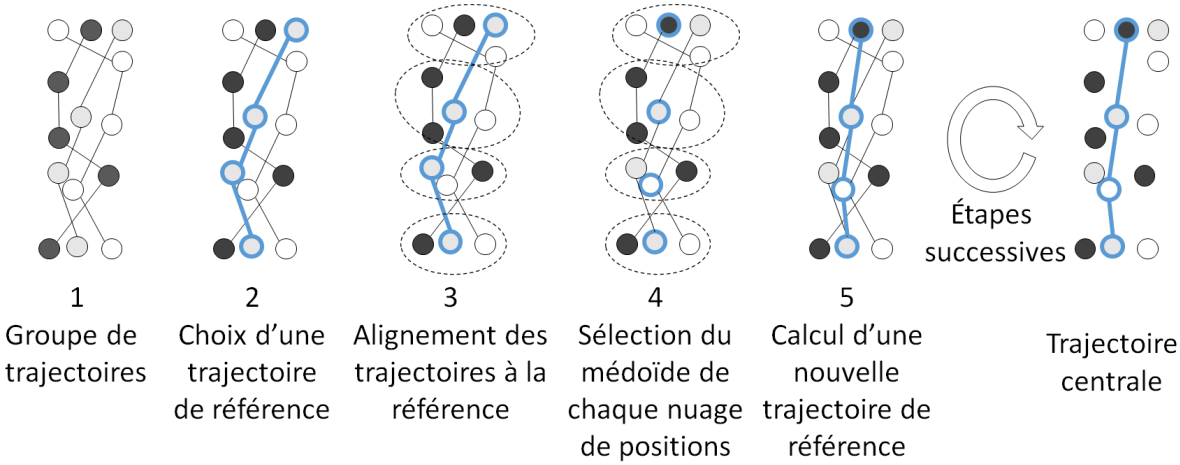


Figure 6.20 – Processus de calcul de la trajectoire centrale.

En premier lieu, un groupe homogène de trajectoires (cluster) est sélectionné (figure 6.20 étape 1). Toutes les trajectoires de ce groupe réalisent le même trajet d'un lieu A vers un lieu B. Les trajectoires de ce groupe sont toutes filtrées à l'aide d'un filtre de Douglas et Peucker Spatio-Temporel [**124**, **71**].

Une première trajectoire est sélectionnée comme référence de départ (figure 6.20 étape 2). Cette trajectoire est choisie judicieusement à l'aide d'heuristiques (trajectoire ayant une durée et une longueur proche des durées et longueurs médianes des trajectoires du groupe).

Toutes les trajectoires du groupe sont alors comparées à la trajectoire de référence à l'aide de la distance de Fréchet discrète. L'alignement des positions des trajectoires du groupe est réalisé avec la trajectoire de référence. Ainsi, pour chaque position de la trajectoire de référence, on dispose d'un nuage de toutes les positions appariées des autres trajectoires du groupe. Ces nuages de positions homologues sont illustrés par des cercles pointillés sur la figure 6.20 étape 3).

L'élément central (médoïde présenté en section 6.1.1) de ces nuages est alors calculé (figure 6.20 étape 2).

Une nouvelle trajectoire de référence est calculée en connectant les médoïdes de chaque

nuage. Cette nouvelle trajectoire de référence est comparée à la trajectoire de référence précédente à l'aide de la distance de Fréchet discrète afin de vérifier un critère de convergence.

Les opérations 3,4 et 5 du processus de la figure 6.20 sont répétées successivement jusqu'à ce que la trajectoire de référence converge (plus de différence observée entre la nouvelle trajectoire et la précédente).

La trajectoire de référence est alors considérée comme la trajectoire centrale. Il est également possible de choisir une trajectoire réelle dont la distance de Fréchet discrète est la plus proche de la trajectoire de référence finale afin d'obtenir une trajectoire réelle plutôt qu'une trajectoire centrale composée de positions réelles, mais issues potentiellement de différentes trajectoires du groupe.

Le calcul de la distance de Fréchet discrète est assez coûteux. C'est pourquoi cet algorithme itératif cherche à éviter le calcul de la matrice de distances de Fréchet discrètes entre toutes les trajectoires du groupe deux à deux.

Cependant, le filtrage de Douglas et Peucker initial des trajectoires implique un échantillonnage irrégulier de celles-ci. Aussi, certaines positions d'une même trajectoire peuvent être appariées à des positions différentes de la trajectoire de référence et se retrouver ainsi dans plusieurs nuages de positions homologues. Cette situation peut engendrer un phénomène d'oscillation de positions d'un nuage à un autre empêchant la convergence de l'algorithme.

C'est pourquoi il est nécessaire d'ajouter un paramètre définissant le nombre maximum d'itérations de l'algorithme et/ou un seuil minimum de distance entre deux trajectoires de référence consécutives garantissant l'arrêt de l'algorithme.

Afin d'illustrer le fonctionnement de cet algorithme, un groupe de 100 trajectoires homogènes a été généré par ajout d'un bruit gaussien sur les coordonnées d'une trajectoire de référence dessinée en vert sur la figure 6.21.

La figure 6.22 présente les résultats du calcul de la trajectoire de référence au bout de 1 et de 23 itérations de l'algorithme. La trajectoire représentée en jaune indique la première trajectoire de référence, la trajectoire en rouge représente la nouvelle trajectoire de référence obtenue à l'issue de l'itération de l'algorithme. La trajectoire verte montre la trajectoire utilisée pour générer le groupe homogène de trajectoires.

Une fois la trajectoire centrale obtenue, celle-ci est comparée à toutes les trajectoires du groupe à l'aide de la distance de Fréchet discrète et de la distance DTW. On remarque sur

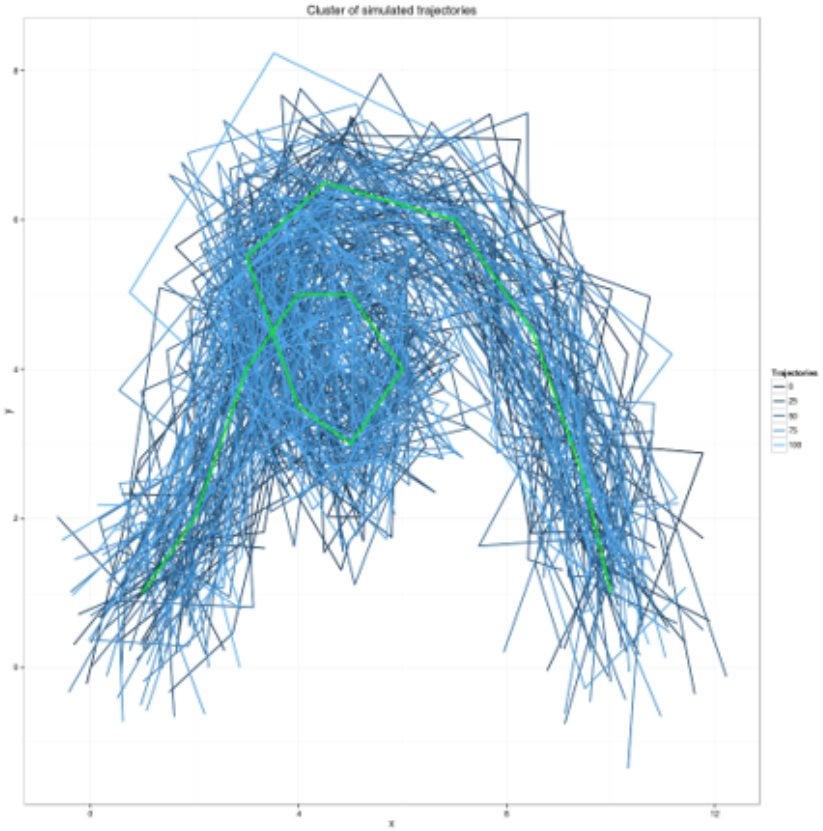


Figure 6.21 – Groupe de 100 trajectoires homogènes générées par bruit gaussien autour d'une trajectoire de référence (verte).

la figure 6.23 que la trajectoire centrale "virtuelle" composée des médoïdes des nuages de positions est plus proche visuellement de la trajectoire initiale (verte) utilisée pour générer le groupe de trajectoires. Selon la distance choisie (DTW ou Fréchet) présentée en jaune sur la figure 6.23, la trajectoire la plus proche de la trajectoire centrale du groupe n'est pas forcément la même.

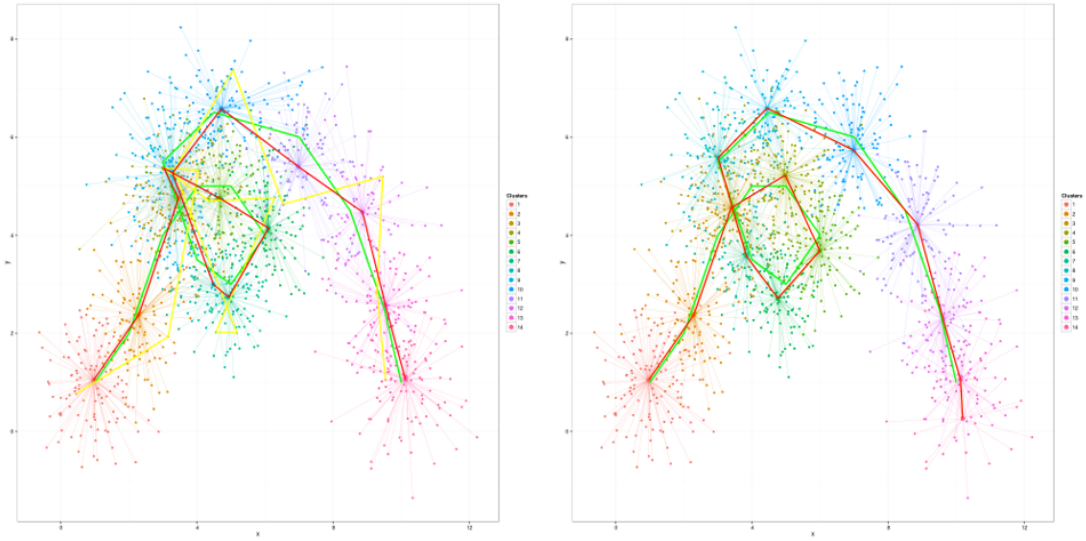


Figure 6.22 – Trajectoires de référence obtenues au bout de 1 et 23 itérations de l'algorithme.

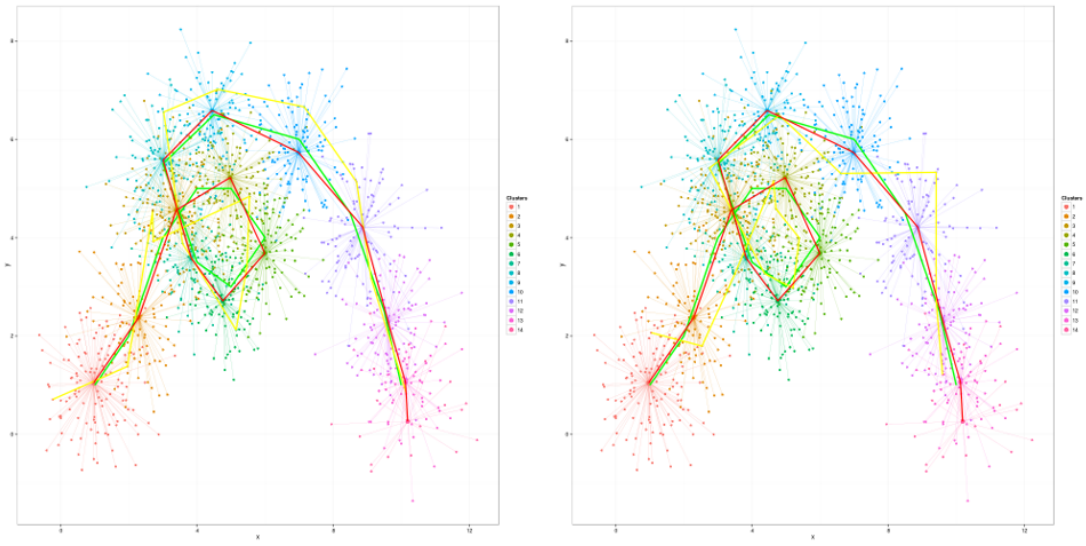


Figure 6.23 – Trajectoires du groupe les plus proches de la trajectoire de référence selon la distance DTW ou la distance de Fréchet discrète.

6.2.2 Le couloir spatio-temporel

Les objets mobiles évoluent dans des environnements divers, parfois contraints par des réseaux en fonction de leurs modes de transports, parfois plus libres de leurs mouvements dans des espaces plus ouverts.

Lorsque l'on étudie des groupes de trajectoires évoluant dans des espaces ouverts, il est alors important de décrire la manière dont les trajectoires s'étalent et se répartissent autour de la trajectoire centrale. On retrouve ici la notion de description de la densité des trajectoires autour de la trajectoire centrale. Tout comme pour les nuages de positions, nous avons proposé une approche basée sur les boites à moustaches en étendant les travaux réalisés sur les OSTBP aux trajectoires. Ces travaux détaillés dans cette section ont été publiés dans [**Re5**].

Le processus itératif permettant de générer la trajectoire centrale (figure 6.20) génère des nuages de positions homogènes rattachés à une position centrale médoïde. La composante temporelle utilisée pour définir le temps médian de la position centrale est exprimée en temps relatif depuis le départ de la trajectoire. Pour chaque nuage de positions, une OSTBP est calculée. La succession de boites 3D permet de générer une sorte de couloir de navigation 3D, nommée TBP, au sein duquel les trajectoires du groupe évoluent. La trajectoire centrale virtuelle, composée des médoïdes des nuages, est représentée en connectant ces médoïdes les uns aux autres (ligne orange sur la figure 6.24). Cette visualisation permet de représenter efficacement la trajectoire centrale ainsi que l'évolution des densités des nuages de points au cours du temps. La corrélation entre les composantes X et Y des positions est représentée par l'orientation des boites, les positionnements relatifs des boites vertes et rouge par rapport à la position centrale indiquent l'évolution de l'étalement et de la symétrie du groupe de trajectoires au cours du temps. De plus, le vecteur connectant les positions centrales des nuages permet de définir l'orientation principale du nuage en représentant une synthèse du cap et de la vitesse médiane du nuage de positions. Cette orientation est importante, car elle permet alors de définir les notions suivantes par rapport à la position centrale :

- Devant ;
- Derrière ;
- À droite ;
- À gauche.

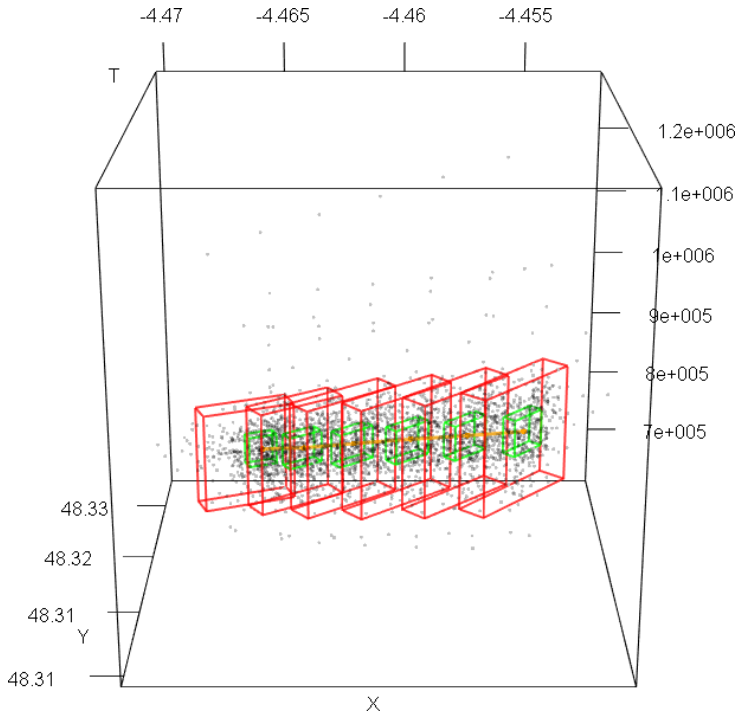


Figure 6.24 – Suite d'OSTBP des nuages de positions homologues appariés aux positions de la trajectoire centrale.

Une des problématiques principales de cette approche réside dans l'usage de l'ACP pour le calcul de l'axe principal du boxplot qui considère que les variables étudiées sont distribuées selon une loi normale. Dans notre cas d'étude, nous nous sommes intéressés à des nuages de positions dont la densité n'est pas forcément symétrique. C'est pourquoi il serait intéressant d'étudier d'autres méthodes de réductions de dimensions pour la génération du patron de trajectoires. De plus, les composantes spatiales et temporelles sont traitées séparément pour générer les OSTBP et la TBP. Une piste intéressante de recherche consiste à étudier ces dimensions de manière conjointe ce qui devrait permettre de mieux visualiser la densité temporelle du nuage. En effet, les positions des trajectoires ayant pris du retard sont censées se trouver regroupées en début de nuage alors que les positions des navires allant plus vite devraient être regroupées en fin de nuage. Cependant, l'alignement de Fréchet discret autorise l'alignement multiple entre positions. Certaines positions peuvent alors être incluses dans plusieurs nuages et potentiellement perturber l'analyse de densité temporelle.

6.3 Mesures de similarité entre patrons et trajectoires

Une fois le patron d'un groupe de trajectoires calculé, celui-ci peut être utilisé pour visualiser un couloir de navigation au sein duquel une trajectoire normale réalisant le même voyage devrait se trouver.

Lorsqu'un objet se déplace, sa trajectoire partielle depuis son point de départ peut être comparée et alignée avec la trajectoire centrale du patron à l'aide de la distance de Fréchet discrète partielle introduite dans la section 5.2.6. Chaque position de la trajectoire partielle est alors alignée avec un des OSBP de la TBP.

6.3.1 Similarité spatiale normalisée entre une position et un OSBP

Chaque OSBP représente la densité de distribution d'un nuage de positions et défini un repère X', Y' dont l'origine correspond à la position centrale (médoïde) du nuage et l'axe Y' est orienté vers le prochain OSBP du TBP. La position p de coordonnées (x, y) est alors projetée dans le repère de l'OSBP. On note $p'(x', y')$ les coordonnées de la position projetée.

L'écart entre la composante x' de la position projetée et le médoïde est alors négatif si la position se trouve à gauche du médoïde et positif si la position se trouve à sa droite. L'écart entre la composante y' de la position projetée et le médoïde (y') est négatif si la position se trouve en dessous du médoïde et positif si la position se trouve au-dessus.

Considérant que l'écart spatial maximal toléré est défini par la limite externe de l'OSBP (boite rouge), cette limite asymétrique est utilisable pour normaliser les écarts avec le médoïde (voir figure 6.16 pour le rappel des notations).

On note dS_X la distance spatiale normalisée sur l'axe X'. Et dS_Y la distance spatiale normalisée sur l'axe Y'.

$$dS_X(p') = \begin{cases} \frac{x'}{C} & si \quad x' \leq 0 \\ \frac{x'}{D} & si \quad x' > 0 \end{cases} \tag{6.8}$$

$$dS_Y(p') = \begin{cases} \frac{y'}{A} & si \quad y' > 0 \\ \frac{y'}{B} & si \quad y' \leq 0 \end{cases} \tag{6.9}$$

$$dS(p') = \frac{dS_X(p') + dS_Y(p')}{2} \tag{6.10}$$

Les distances normalisées dS_X et dS_Y tiennent compte de l'asymétrie potentielle du nuage ainsi que de son aplatissement. La figure 6.25 donne un exemple de distances spatiales normalisées entre différentes positions d'une trajectoire et un TBP.

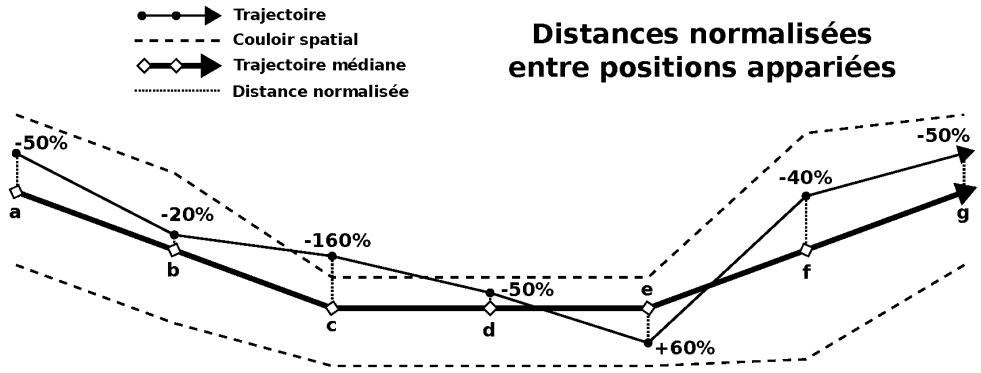


Figure 6.25 – Distances normalisées entre positions appariées à une TBP exprimées en pourcentages par rapport à la limite externe du TBP.

6.3.2 Similarité temporelle normalisée entre une position et un OSTBP

La similarité temporelle d'une position p comparée à un OSTBP est définie de manière similaire à la similarité temporelle. Cependant, l'axe Z représentant la composante temporelle du boxplot 3D ayant été traité séparément des composantes spatiales, il n'est pas nécessaire de réaliser un changement de repère.

La similarité temporelle dT est alors directement comparée au temps relatif médian du nuage de positions de l'OSTBP. Tout comme pour la similarité spatiale, l'écart entre l'estampille temporelle relative de la position p et le temps relatif médian de l'OSTBP est normalisé à l'aide des limites temporelles de la boite externe du boxplot 3D.

Cette similarité temporelle permet de définir si la position est en avance $dT < -1$ ou en retard $dT > 1$ par rapport aux limites habituellement observées pour le nuage de positions ayant permis de générer l'OSTBP.

6.3.3 Similarité spatio-temporelle entre une trajectoire et un TBP

Les différentes positions d'une trajectoire T peuvent être qualifiées à l'aide des mesures de similarité spatiale et temporelle définies aux sections 6.3.1 et 6.3.2.

Ces mesures de similarité évoluent à chaque position de la trajectoire à comparer avec le patron $\tilde{T}$. Différentes mesures agrégées basées sur ces distances ont été proposées dans [**Ra2**, **Re7**].

Les valeurs moyennes et maximales des distances normalisées spatiales et temporelles sont calculées. La moyenne des variations des distances normalisées (spatiale et temporelle) entre deux OSTBP successifs est également calculée.

- La distance spatiale maximale : DSM

$$DSM(T, \tilde{T}) = \max_{1 \leq i < n} (|dS_N(p_i, \tilde{p}_j)|) \tag{6.11}$$

- La distance spatiale moyenne : DSm

$$DSm(T, \tilde{T}) = \frac{\sum_{i=1}^{n} |dS_N(p_i, \tilde{p}_j)|}{n} \tag{6.12}$$

- La moyenne des variations spatiales : δSm

$$\delta Sm(T, \tilde{T}) = \frac{\sum_{i=1}^{n-1} |dS_N(p_i, \tilde{p}_j) - dS_N(p_{i+1}, \tilde{p}_{j+1})|}{n-1} \tag{6.13}$$

- La distance temporelle maximale : DTM

$$DTM(T, \tilde{T}) = \max_{1 \leq i < n} (|dT_N(p_i, \tilde{p}_j)|) \tag{6.14}$$

- La distance temporelle moyenne : DTm

$$DTm(T, \tilde{T}) = \frac{\sum_{i=1}^{n} |dT_N(p_i, \tilde{p}_j)|}{n} \tag{6.15}$$

- La moyenne des variations temporelles : δTm

$$\delta Tm(T, \tilde{T}) = \frac{\sum_{i=1}^{n-1} |dT_N(p_i, \tilde{p}_j) - dT_N(p_{i+1}, \tilde{p}_{j+1})|}{n-1} \tag{6.16}$$

Ces nouvelles mesures sont calculées pour toutes les trajectoires du groupe comparées au TBP. On dispose ainsi d'un ensemble conséquent de mesures décrivant les caractéris-

tiques spatiales et temporelles des trajectoires comparées au patron.

Des statistiques sont réalisées sur ces données afin de définir différentes bornes définissant les notions de similarité plus ou moins forte entre une trajectoire et un patron. La logique floue ([**127**, **102**]) est alors employée pour définir des règles floues décrivant la similarité spatio-temporelle globale entre une trajectoire et un patron.

- R1 : si l'écart spatial maximum est faible alors la similarité spatiale est forte.
  $(DSM = Faible) \Rightarrow (SIM_S = Fort)$
- R2 : si la forme de la trajectoire est très différente de celle de la trajectoire médiane, alors la similarité est faible en d'autres termes si la δSm est grande alors la similarité spatiale est faible.
  $(\delta Sm = Fort) \Rightarrow (SIM_S = Faible)$
- R3 : si l'écart spatial moyen est faible et l'écart spatial maximum est faible et la moyenne des deltas spatiaux est faible alors la similarité spatiale est très forte.
  $((DSm = Faible) \wedge (DSM = Faible) \wedge (\delta Sm = Faible)) \Rightarrow (SIM_S = Tfort)$
- R5 : si l'écart temporel maximum est faible alors la similarité temporelle est forte.
  $(DTM = Faible) \Rightarrow (SIM_T = Fort)$
- R6 : si les écarts temporels sont très différents de ceux de la trajectoire médiane, alors la similarité temporelle est faible (si δTm est grande alors la similarité temporelle est faible).
  $(\delta Tm = Fort) \Rightarrow (SIM_T = Faible)$
- R7 : si l'écart temporel moyen est faible et l'écart temporel maximum est faible et la moyenne des variations temporelles est faible alors la similarité temporelle est très forte.
  $((DTm = Faible) \wedge (DTM = Faible) \wedge (\delta Tm = Faible)) \Rightarrow (SIM_T = Tfort)$

Dans ces règles, des descriptions qualitatives sont utilisées (fort, faible, très fort). Afin de définir ces notions, des ensembles flous sont utilisés. Ces ensembles sont définis à l'aide de fonctions d'appartenance floues.

Les ensembles flous définissant les termes "Faible", "Moyen" et "Fort" pour chaque mesure de similarité sont générés à l'aide des statistiques produites par l'ensemble des trajectoires comparées au patron.

Une fonction d'appartenance μ est une fonction qui définit, pour toute entrée numérique x, le degré d'appartenance compris entre 0 et 1 de x à l'ensemble flou correspondant.

Par exemple, la fonction d'appartenance $\mu^{DSM}_{Faible}(x)$ définit le degré d'appartenance de x à l'ensemble flou $Faible$ pour la mesure DSM. Les différentes fonctions d'appartenance utilisées sont listées dans le tableau 6.1.

DSM	DSm	δSm	DTM	DTm	δTm
$\mu^{DSM}_{Faible}(x)$	$\mu^{DSm}_{Faible}(x)$	$\mu^{\delta Sm}_{Faible}(x)$	$\mu^{DTM}_{Faible}(x)$	$\mu^{DTm}_{Faible}(x)$	$\mu^{\delta Tm}_{Faible}(x)$
$\mu^{DSM}_{Moyen}(x)$	$\mu^{DSm}_{Moyen}(x)$	$\mu^{\delta Sm}_{Moyen}(x)$	$\mu^{DTM}_{Moyen}(x)$	$\mu^{DTm}_{Moyen}(x)$	$\mu^{\delta Tm}_{Moyen}(x)$
$\mu^{DSM}_{Fort}(x)$	$\mu^{DSm}_{Fort}(x)$	$\mu^{\delta Sm}_{Fort}(x)$	$\mu^{DTM}_{Fort}(x)$	$\mu^{DTm}_{Fort}(x)$	$\mu^{\delta Tm}_{Fort}(x)$

Table 6.1 – Fonctions d'appartenance des différentes mesures de similarité

Les fonctions d'appartenance floue utilisée dans [**Re7**] sont de type linéaire par morceaux. Les différentes bornes de ces fonctions sont définies à l'aide des valeurs statistiques obtenues lors des comparaisons de toutes les trajectoires du groupe avec le TBP. La figure 6.26 présente les ensembles flous "Faible", "Moyen" et "Fort" et fonctions d'appartenance correspondantes pour la mesure DSM. Les limites de ces ensembles sont toutes fixées à 20, 40, 60 et 80% des valeurs observées de la DSM pour l'ensemble des trajectoires du cluster. On remarque que ces ensembles se chevauchent entre 20 et 40% et entre 60 et 80%. C'est pourquoi, une valeur x peut appartenir à plusieurs ensembles flous à la fois avec différents degrés d'appartenance.

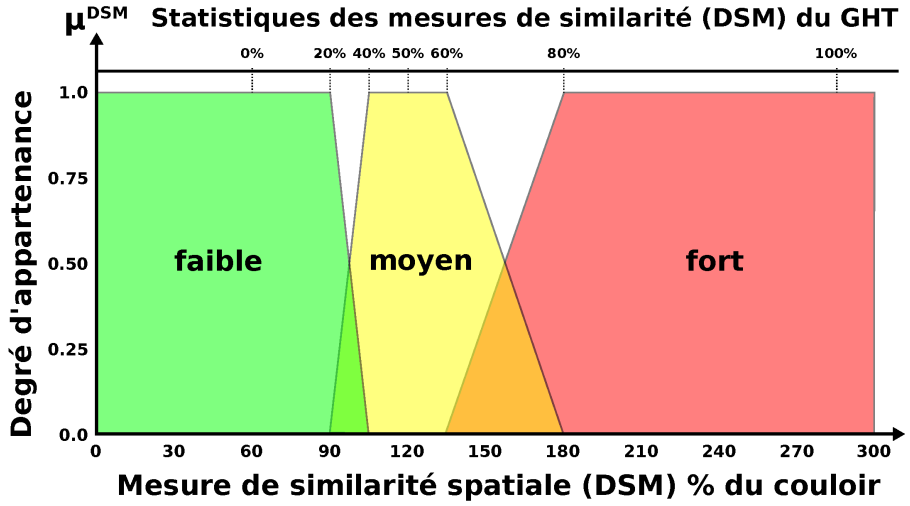


Figure 6.26 – Fonctions d'appartenance associées à la mesure de similarité DSM.

Une fois toutes les fonctions d'appartenance et ensembles flous correspondants définis

pour toutes les mesures de similarité spatiales et temporelles observées, il est alors possible de "fuzzifier" une valeur observée pour une nouvelle trajectoire à comparer au patron.

Par exemple, la valeur de la DSM d'une nouvelle trajectoire comparée au patron a été calculée et vaut 145%. Pour savoir si cette valeur de DSM est "Faible", "Moyenne" ou "Forte", on utilise les ensembles flous et les fonctions d'appartenance dont les résultats sont illustrés sur la figure 6.27.

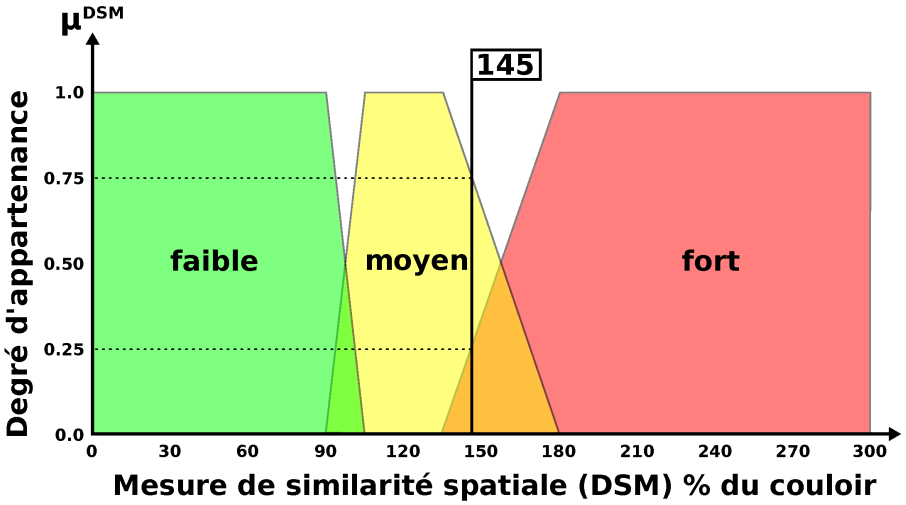


Figure 6.27 – Valeur floue de la DSM d'une trajectoire appariée à un TBP.

Après évaluation des fonctions d'appartenance, les résultats suivants sont obtenus : $\mu^{DSM}_{Faible}(x) = 0.00 \quad \mu^{DSM}_{Moyen}(x) = 0.75 \quad \mu^{DSM}_{Fort}(x) = 0.25$

Ces valeurs indiquent que la trajectoire observée dispose d'un DSM plutôt "Moyen" voir légèrement "Fort" en comparaison de toutes les autres trajectoires utilisées pour générer le patron.

Pour obtenir un indice de similarité spatiale et temporelle combinant toutes les mesures précédemment citées, des règles floues sont définies.

Ces règles floues combinent les différents ensembles flous des 6 mesures de similarité afin de décrire le niveau de similarité global d'une trajectoire avec un patron. Les 27 règles combinant tous les ensembles flous des 3 mesures spatiales sont illustrées sur l'arbre de décision de la figure 6.28. Le même arbre est défini pour les 3 mesures de similarité temporelle.

Un exemple d'activation des règles floues spatiales est proposé sur la figure 6.28. Les

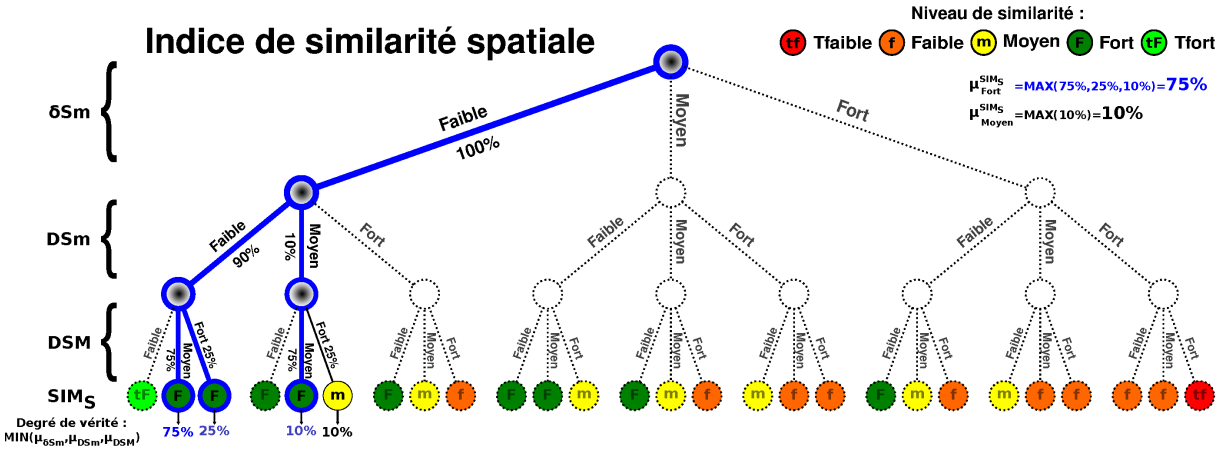


Figure 6.28 – Exemple d'activation d'un arbre de décision de règles floues définissant la similarité spatiale globale.

3 mesures de similarité de la trajectoire d'exemple comparée au patron ont activé les différentes fonctions d'appartenance suivantes :

$$\mu^{\delta Sm}_{Faible}(x) = 1.00 \quad \mu^{\delta Sm}_{Moyen}(x) = 0.00 \quad \mu^{\delta Sm}_{Fort}(x) = 0.00$$
$$\mu^{DSm}_{Faible}(x) = 0.90 \quad \mu^{DSm}_{Moyen}(x) = 0.10 \quad \mu^{DSm}_{Fort}(x) = 0.00$$
$$\mu^{DSM}_{Faible}(x) = 0.00 \quad \mu^{DSM}_{Moyen}(x) = 0.75 \quad \mu^{DSM}_{Fort}(x) = 0.25$$

Dans cet exemple, il existe plusieurs branches qui activent une similarité spatiale considérée comme "Forte" à 75% (branches surlignées en bleu dans la figure 6.28). Il existe également une branche qui considère que la similarité spatiale est plutôt moyenne. Cependant, cette règle n'est activée qu'à 10% en raison de la similarité DSm dont la fonction d'appartenance à l'ensemble flou "Moyen" n'est activée qu'à 10%.

Un premier niveau de description qualitatif de la similarité spatiale est obtenu. La similarité est considérée comme Forte à 75% et Moyenne à 10%. Cependant, cette description n'est pas forcément pratique lorsque l'on souhaite comparer des trajectoires deux à deux pour définir laquelle ressemble le plus au patron.

Pour obtenir une valeur numérique normalisée décrivant la similarité spatiale entre une trajectoire et un patron, il faut alors "défuzzifier" les degrés d'appartenance aux ensembles flous. La méthode du centre de gravité des surfaces des ensembles flous activés permet de calculer cette valeur numérique (Figure 6.29).

Un exemple complet de calcul de la similarité spatio-temporelle détaillant toutes les valeurs des mesures de similarité, les niveaux d'activation des fonctions d'appartenance, les différentes règles floues activées ainsi que la "défuzzification" des mesures de similarité

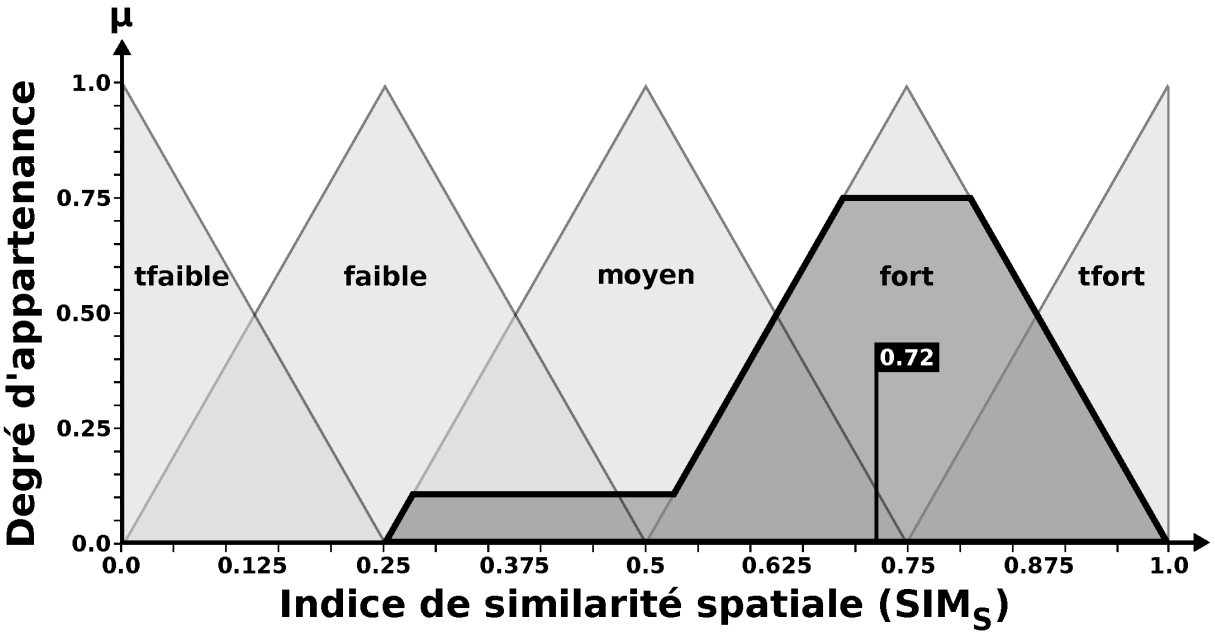


Figure 6.29 – Défuzzification de l'indice de similarité spatiale ($SIMS$) par la méthode du centre de gravité sur les surfaces.

globales est présenté dans [**Re7**].

Chapitre 7

PRISE EN COMPTE DU CONTEXTE ENVIRONNEMENTAL

Dans les chapitres 4 et 5, différentes mesures de similarité entre trajectoires ont été présentées. Ces mesures étudient l'aspect spatial, temporel et sémantique des trajectoires. Cependant, l'environnement dans lequel les objets mobiles évoluent influence également grandement les trajectoires observées.

Par exemple, les embouteillages sur certains axes routiers ont lieu à certains horaires de la journée. Certaines trajectoires peuvent ainsi varier en fonction de l'horaire et du contexte dans lequel le déplacement est réalisé. Dans le cadre d'un réseau routier, les véhicules doivent alors choisir une route secondaire en cherchant à optimiser un critère évoluant au cours du temps. Ce problème de recherche opérationnelle est complexe, il s'apparente à la recherche classique d'un plus court chemin ou les valeurs des coûts de déplacements sur les arcs d'un graphe évoluent en fonction des horaires de départ d'un lieu (noeud du graphe) [**64**, **15**, **3**].

Lorsque les objets mobiles étudiés se déplacent dans un environnement ouvert, il est alors nécessaire de représenter l'impact de l'environnement sur les trajectoires en fonction du temps.

Dans ce chapitre, les résultats de recherche sur l'étude des déplacements des navires en zone polaire sont présentés.

7.1 Modélisation de l'environnement

Les déplacements des navires sont fortement impactés par l'environnement marin dans lequel ils évoluent tels que les courants marins ainsi que les vents. Dans le contexte polaire, la présence de glaces de mer introduit un risque supplémentaire limitant fortement les capacités de navigation des navires.

La plateforme européenne COPERNICUS dispose d'un espace dédié au partage de données marines [1] à différentes échelles spatiales et temporelles. Le modèle TOPAZ retrace et simule l'évolution de nombreuses variables décrivant l'environnement marin Arctique [**51**]. Les données journalières produites par ce modèle sont téléchargeables au format NetCDF sur la plateforme COPERNICUS [2].

Les différentes variables fournies par ce modèle sont les suivantes :

- Températures de surface et de fond marin
- Salinité de l'eau
- Concentration de glace (SIC)
- Épaisseur de glace (SIT)
- Épaisseur de neige (SNOW)
- Vitesse de déplacement de la glace (SIUV)
- Courant marin
- Bathymétrie

Les deux variables décrivant la concentration (SIC) et l'épaisseur (SIT) de glace sont particulièrement intéressantes pour l'étude des capacités de déplacement des navires dans la glace.

Une autre source de données basée sur ce modèle TOPAZ est également en mesure d'établir des prévisions l'état de l'Océan Arctique sous 10 jours [**50**]. Ces prévisions contiennent des données horaires plus détaillées. [3].

Ces sources de données décrivent l'état de l'océan Arctique à une résolution spatiale d'environ $12,5x12,5$ km et une résolution temporelle journalière depuis 1991.

Chaque donnée journalière comporte environ 140 000 cellules raster, 10 variables descriptives, sur une période couvrant 1991 à 2021 soit 10 950 journées collectées soit environ 15 milliards de données à prendre en compte. La figure 7.1 présente l'épaisseur de glace issue des données du modèle TOPAZ pour une journée (01/09/2000).

Ces données ont été collectées et intégrées dans une base de données spatio-temporelle PostgreSQL à l'aide du plugin PostGIS. Une analyse géostatistique a été réalisée sur

1. https://marine.copernicus.eu/
2. https://resources.marine.copernicus.eu/product-detail/ARCTIC_MULTIYEAR_PHY_002_003/INFORMATION
3. https://resources.marine.copernicus.eu/product-detail/ARCTIC_ANALYSIS_FORECAST_PHYS_002_001_a/INFORMATION

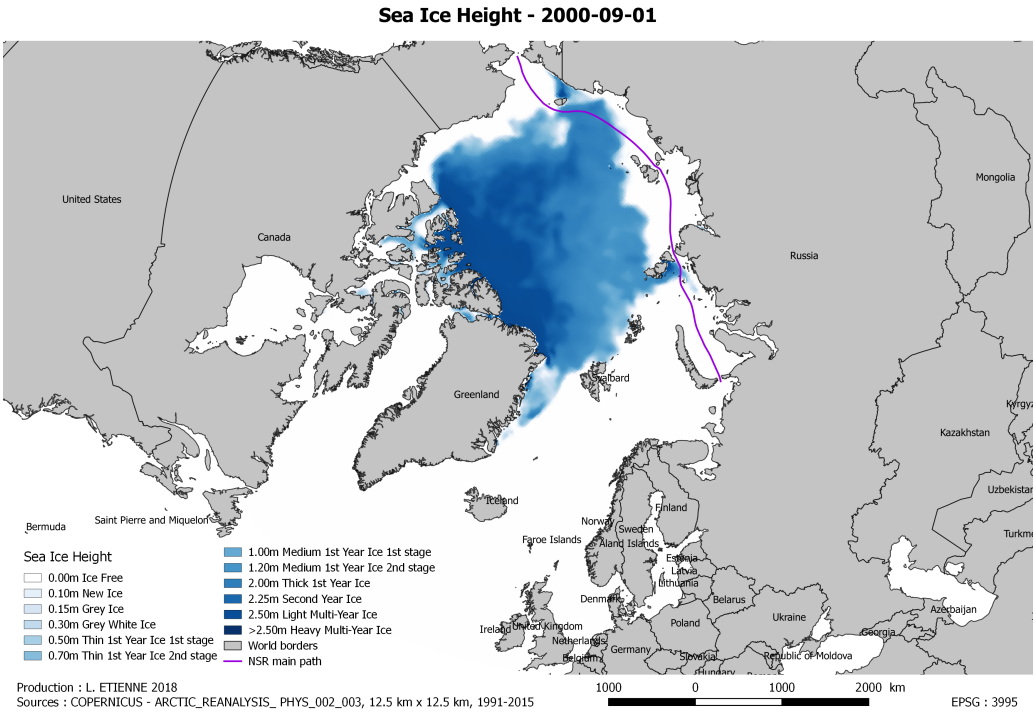


Figure 7.1 – Carte d'épaisseur de glace au 01/09/2000 issue des données COPERNICUS.

ces données afin de produire pour chaque cellule raster, une agrégation statistique des différentes variables observées sur les 30 dernières années.

Les différents percentiles des variables observées ont été calculés afin de définir des boîtes à moustaches décrivant les valeurs d'épaisseur et de concentration de glace observées. Ces éléments de statistique descriptive sont ensuite utilisés pour définir les conditions de glaces observées au cours des 30 dernières années et établir différents scénarios possibles de navigation. La taille des boîtes (espace interquartile) et l'étendue moustaches pour chaque cellule raster pour un jour spécifique de l'année donne alors une information complémentaire concernant la variabilité interannuelle du paramètre étudié. Cette variabilité permet alors de représenter et de visualiser l'incertitude spatio-temporelle de ce paramètre pour la cellule raster et le jour de l'année concerné. Les figures 7.2 et 7.3 présentent les boîtes à moustaches des concentrations et des épaisseurs de glaces agrégées par semaine au niveau du passage d'Ostrova Anzhu.

Ces données sont ensuite utilisées pour définir les capacités de navigation des navires en fonction de l'épaisseur et de la concentration de glace de mer. Les vitesses des navires varient en fonction de ces deux paramètres. Dans certains cas, les navires ne sont pas

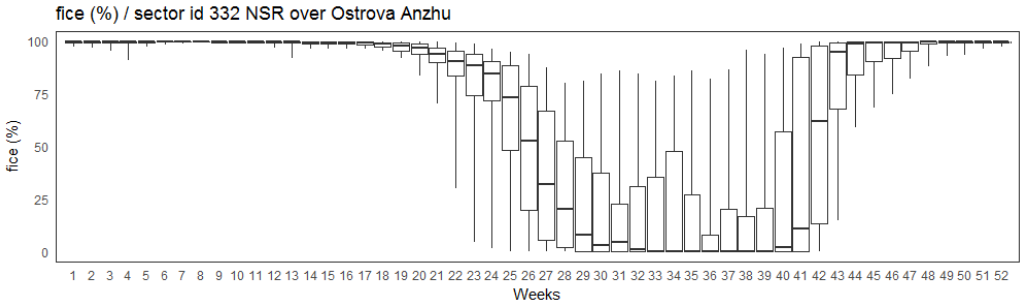


Figure 7.2 – Boîtes à moustaches des concentrations de glaces agrégées par semaine au niveau du passage d'Ostrova Anzhu.

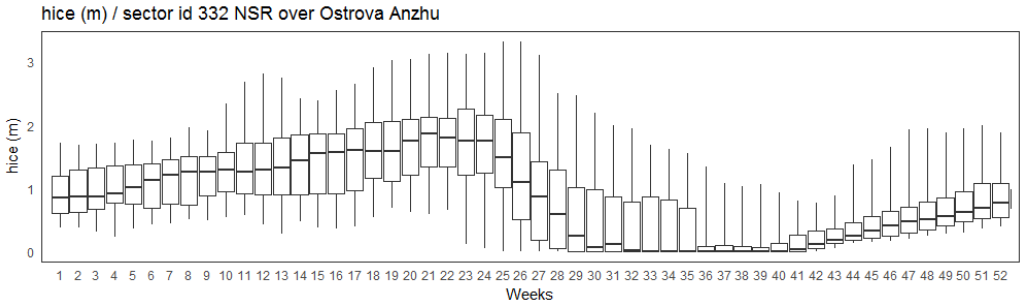


Figure 7.3 – Boîtes à moustaches des épaisseurs de glaces agrégées par semaine au niveau du passage d'Ostrova Anzhu.

capables de franchir certaines zones lorsque les glaces de mer sont trop épaisses et doivent faire appel à une escorte de brise-glace.

7.2 Modélisation de l'évolution du trafic Maritime dans l'Arctique Canadien

Dans le cadre de mon post-doctorat au Canada, je me suis intéressé à la modélisation et à l'analyse du trafic maritime dans l'Arctique Canadien [**Ra1**]. La problématique liée à ces travaux provient de la grande incertitude relative aux différents facteurs impactant le trafic maritime.

Dans un premier temps, l'étude des facteurs socio-économiques ainsi que la modélisation de différents scénarios de projection pour 2020 ont été réalisées. L'évolution temporelle des conditions météorologiques dans l'Arctique Canadien et plus particulièrement l'impact des différents types de glaces sur les déplacements des navires a été étudiée.

Cette modélisation a été couplée avec l'étude du trafic actuel dans l'Arctique en utilisant des bases de données de positions de navires utilisant le système Long Range Identification and Tracking (LRIT). Une méthodologie détaillée reposant sur trois grandes parties a été proposée.

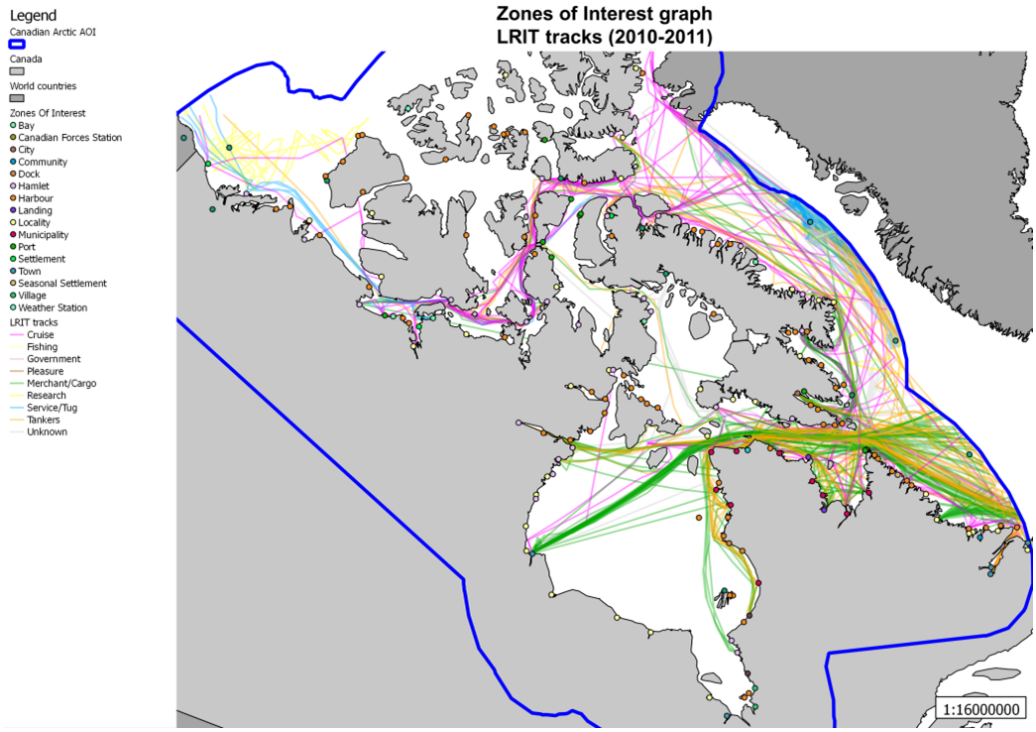
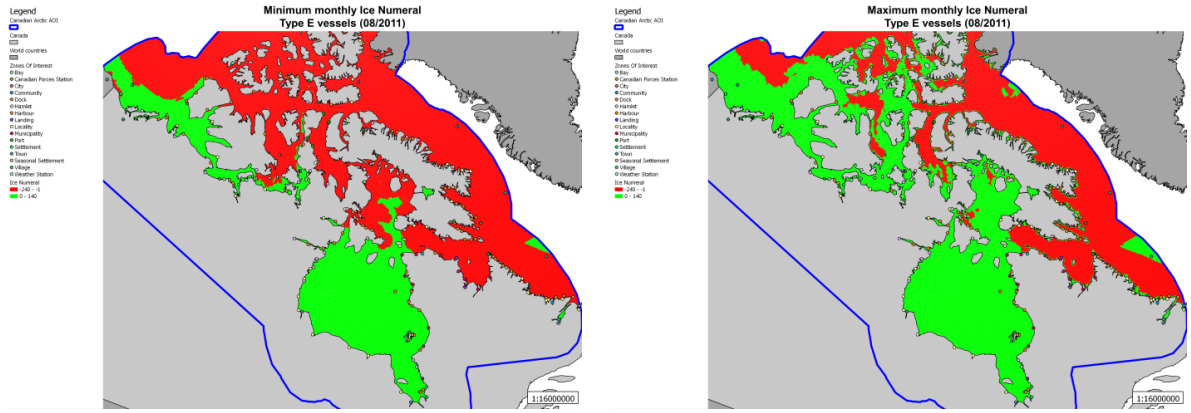


Figure 7.4 – Graphe des Zones d'Intérêt de l'Arctique Canadien.

La première partie de cette méthodologie consiste à modéliser le trafic maritime actuel dans l'Arctique [**CI11**]. Une analyse des données de positions d'une année complète de trafic a été réalisée. L'objectif de cette étape consiste à découvrir les Zones d'Intérêt (ZOI) générant du trafic maritime, les types de navires fréquentant ces zones, et comment sont-elles interconnectées. Une analyse spatiale et temporelle des trajectoires a permis de construire un graphe spatio-temporel pondéré du trafic maritime entre les différentes ZOI (Figure 7.4). Ce graphe permet de connaître le niveau de trafic mensuel par catégorie de navires entre deux zones d'intérêt.

La seconde partie de la méthodologie s'intéresse à la prise en compte de l'évolution temporelle de la glace dans les eaux arctiques. Différents types de navires sont spécialement conçus pour naviguer dans certaines catégories de glaces plus ou moins denses. Ces navires sont classés selon la classification Canadienne dans la catégorie des brise-glaces (CAC1-



(a) Minimum Ice Numeral (b) Maximum Ice Numeral

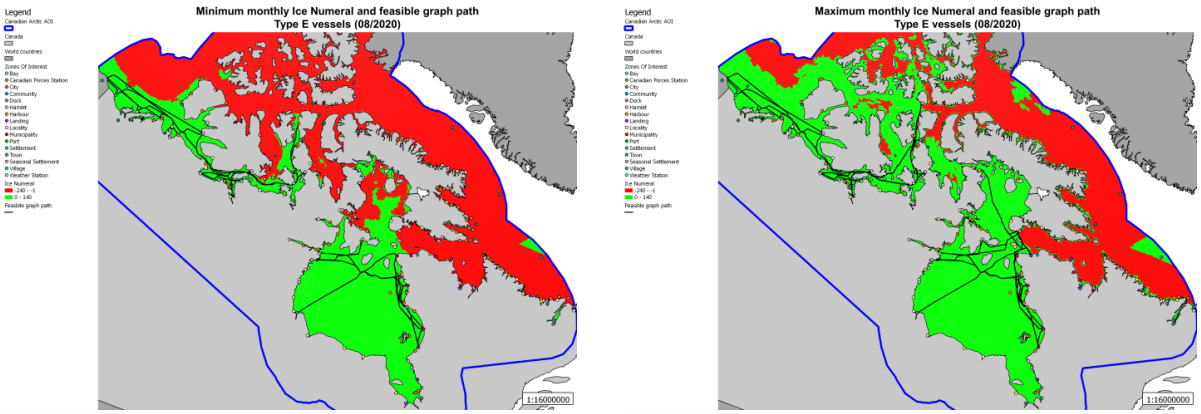
Figure 7.5 – Carte de risques pour les navires de type E au mois d'août 2011.

CAC4) ou des navires à coque renforcée (Type A-Type E) pouvant naviguer uniquement dans des zones où la glace est de faible épaisseur. Les navires de la catégorie Type E sont les plus limités, car ils ne peuvent naviguer que dans la glace ayant une épaisseur réduite (inférieure à 15 cm) alors que ceux de la catégorie CAC1 peuvent se déplacer dans n'importe quel type de glace rencontrée dans l'Arctique.

Les cartes climatiques de concentration des différentes catégories de glaces de mer du Service Canadien des Glaces ont été utilisées afin de définir des cartes mensuelles de risques par catégorie de navires. L'indice de risque utilisé est nommé "Ice Numeral" (Arctic Ice Regime Shipping System (AIRSS) défini par Transport Canada dans [**1**]), il est calculé à partir d'une combinaison entre la catégorie du navire, la concentration et le type de glace rencontrée. Lorsque l'indice de risque est négatif, cela signifie que la glace est trop dense pour que cette catégorie de navires puisse la traverser (représenté en rouge sur les cartes de risques). Une analyse raster a été réalisée pour chaque mois de l'année 2011. Cette analyse a permis de calculer l'indice de risque minimal et maximal mensuel par catégorie de navires [**CI11**]. Les cartes de risques pour le mois d'août 2011 sont présentées sur la Figure 7.5a (minimum) et la Figure 7.5b (maximum).

Une fois les cartes de risques établies, nous avons utilisé un modèle de prédiction de changement climatique pour simuler l'évolution du type de glace présent mensuellement dans l'Arctique en 2020. Puis nous avons calculé les plus courts chemins entre les zones d'intérêt du graphe afin d'en déduire quels arcs du graphe seront navigables pour chaque mois en 2020. Ces routes navigables simulées pour le mois d'août 2020 sont présentées

sur l'exemple de la Figure 7.6a (minimum) et de la Figure 7.6b (maximum).



(a) Minimum Ice Numeral (b) Maximum Ice Numeral

Figure 7.6 – Routes navigables pour les navires de type E au mois d'août 2020.

À partir de l'analyse de trafic réalisée lors de la première étape et des cartes de routes navigables en 2020, la dernière étape de cette méthodologie consiste à modéliser les changements de trafic liés à des facteurs spatio-temporels socio-économiques. Un étalement temporel du trafic maritime dû à la fonte des glaces a été simulé. Nous sommes partis du postulat que le trafic pouvait s'étendre sur un arc du graphe de zones d'intérêt si ce même arc est également navigable les mois précédents et suivants. Cet étalement temporel a été modélisé à l'aide d'une simulation Monte-Carlo. Nous avons également pris en compte les différents facteurs socio-économiques tels que l'activité liée à l'exploitation des ressources (mines, pétrole, gaz, pêche), au changement de population et croissance du produit intérieur brut, au tourisme... Chacun de ces facteurs a été modélisé par une distribution triangulaire permettant de définir un scénario minimum, maximum et médian ainsi qu'une emprise spatiale et temporelle. Le trafic des arcs du graphe impactés spatialement et temporellement par les différents facteurs est modifié de façon multiple et combinée grâce à la simulation Monte-Carlo. Le résultat final de cette étude est fusionné dans des cartes mensuelles de densités de trafic par catégorie de navires pour l'année 2020.

Ainsi, grâce à ces cartes, la densité de trafic minimale, maximale et moyenne dans différentes zones de l'Arctique Canadien peut être visualisée. Ces informations sont ensuite intégrées dans un logiciel d'aide à la décision utilisé par la Direction de la Recherche et du développement pour la Défense du Canada (DRDC) afin d'optimiser l'allocation des ressources de la Marine Nationale Canadienne et des garde-côtes. Un exemple de carte de

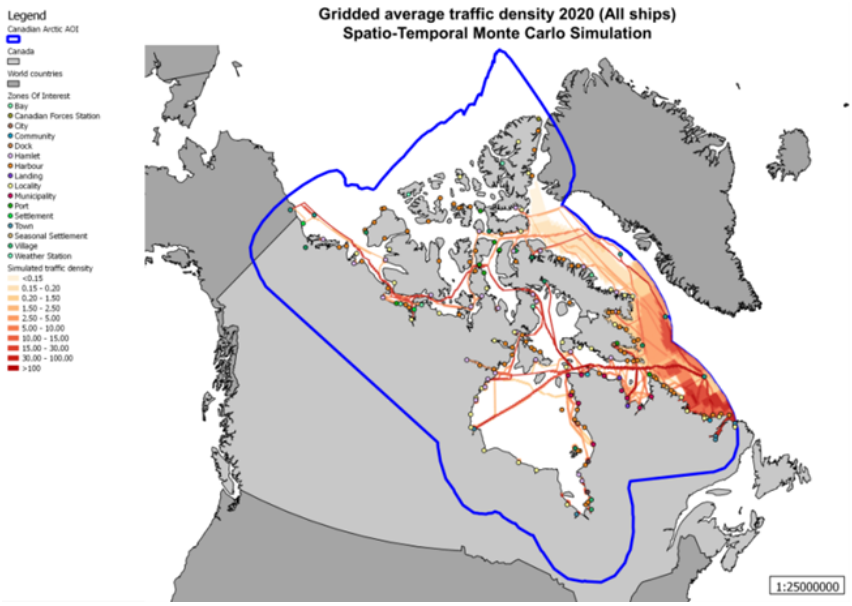


Figure 7.7 – Carte de densité moyenne de trafic simulé pour l'année 2020 dans l'Arctique Canadien.

densité moyenne de trafic pour l'année 2020 est présenté sur la Figure 7.7.

7.3 Évaluation du risque lié à la présence de glaces de mer

Cette section porte sur la modélisation spatio-temporelle d'aléas ainsi que l'analyse de risques associée. Ces travaux sont en lien avec la thèse de Marc STODDARD réalisée en collaboration avec l'Université Dalhousie au Canada. On s'intéresse ici à la représentation du concept d'isolement des navires voyageant dans l'Arctique et la prise en compte de ce facteur dans l'analyse des risques liés aux déplacements des navires dans des régions polaires éloignées (présence de glaces de mer, bathymétrie incertaine, isolement, conditions climatiques extrêmes...) [**Ch3**, **CI9**, **CI8**].

7.3.1 Système d'évaluation des risques et Code polaire

Suite à l'adoption par l'IMO du Code polaire [**37**] mis en place au 1er janvier 2017, j'ai travaillé sur l'analyse de risques liés à la navigation de différents types de navires dans la glace en me basant sur l'index Polar Operational Limit Assessment Risk Indexing System (POLARIS) [**28**]. Dans ce document, les navires sont classés en 12 différentes catégories en fonction de leurs capacités à naviguer dans 12 différents types et épaisseurs de glaces (Figure 7.8).

Ice Thickness upper limit (cm)		0	10	15	30	50	70	100	120	200	225	250	>250
Category	**Ice Class**	Ice Free	New Ice	Grey Ice	Grey White Ice	Thin First Year Ice, 1st	Thin First Year ice, 2nd	Medium First Year Ice	Medium First Year Ice 2nd	Thick First Year Ice	Second Year Ice	Light Multi Year Ice	Heavy Multi- Year Ice
Polar Class (Cat A)	PC1	3	3	3	3	2	2	2	2	2	2	1	1
	PC2	3	3	3	3	2	2	2	2	2	1	1	0
	PC3	3	3	3	3	2	2	2	2	2	1	0	-1
	PC4	3	3	3	3	2	2	2	2	1	0	-1	-2
	PC5	3	3	3	3	2	2	1	1	0	-1	-2	-2
Polar Class (Cat B)	PC6	3	2	2	2	2	1	1	0	-1	-2	-3	-3
	PC7	3	2	2	2	1	1	0	-1	-2	-3	-3	-3
Ice Class	1AS	3	2	2	2	2	1	0	-1	-2	-3	-4	-4
	1A	3	2	2	2	1	0	-1	-2	-3	-4	-5	-5
	1B	3	2	2	1	0	-1	-2	-3	-4	-5	-6	-6
	1C	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8
Cat II	Not IS	3	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-8
		Operation permitted		Low speed		Ice Breaker escort		Ice Breaker escort at low speed		Operation not permitted			

Figure 7.8 – Tableau des indices de risques par types de navires en fonction de l'épaisseur de glace.

Catégorie de navire	Vitesse limite recommandée
PC1	11 noeuds
PC2	8 noeuds
PC3-PC5	5 noeuds
Below PC5	3 noeuds

Table 7.1 – Vitesses limites recommandées en cas d'opération à risque élevé.

L'indice de risque (Risk Index Outcome (RIO)) est un raffinement du concept de "Ice Numeral" défini dans le système Canadien AIRSS [**1**]. Cet indice est calculé à l'aide d'un tableau d'indices de risques (RV) de navigation dans différentes épaisseurs de glaces présenté Figure 7.8. Cet indice tient également compte de la concentration de glace (IC). L'équation 7.1 donne la formule de calcul du RIO.

$$RIO = \sum_{i=1}^{n} (IC_i x RV_i) \tag{7.1}$$

En fonction de la valeur du RIO et du type de navire, différentes recommandations sont formulées dans [**28**].

Lorsque le RIO est positif, le navire peut se déplacer sans contraintes particulières (Normal Operation). En revanche, lorsque le RIO est négatif ($-10 <= RIO < 0$), celui-ci doit faire appel à une escorte de brise-glace et limiter sa vitesse (Elevated Operational Risk). Lorsqu'un navire est escorté par un brise-glace, son RIO est diminué ($RIO + 10$). Enfin, lorsque le RIO est inférieur à -10, le navire devrait éviter de naviguer dans cette zone (Operation subject to special consideration).

Adaptation des vitesses des navires en fonction de l'indice de risque POLARIS

L'IMO donne des recommandations concernant les vitesses des navires transitant dans des zones polaires. Le tableau 7.1, issu de [**28**], indique les vitesses recommandées pour les différents types de navires en cas de navigation dans une zone de risques élevés.

Dans [**Re2**, **Re1**], la vitesse d'un navire en fonction de l'indice RIO POLARIS a été définie de manière à respecter ces différentes vitesses conseillées. La vitesse optimale du navire (design speed) est également prise en compte dans la formule.

Les formules utilisées dans [**Re1**] sont linéaires par morceaux (figure 7.9). Les vitesses optimales des navires sont obtenues pour l'indice POLARIS le moins risqué ($RIO = 30$). La vitesse de référence de 8 noeuds est obtenue lorsque les navires atteignent un indice

RIO nul. Enfin, la vitesse minimale de 3 noeuds est attribuée aux navires (Not Ice Class et 1A) escortés par un brise-glace pour un RIO atteignant -10. Le navire de classe glace PC3 est en mesure de naviguer à 5 noeuds jusqu'à un RIO de -20 s'il est lui-même escorté par un brise-glace de catégorie supérieure (PC1 ou PC2).

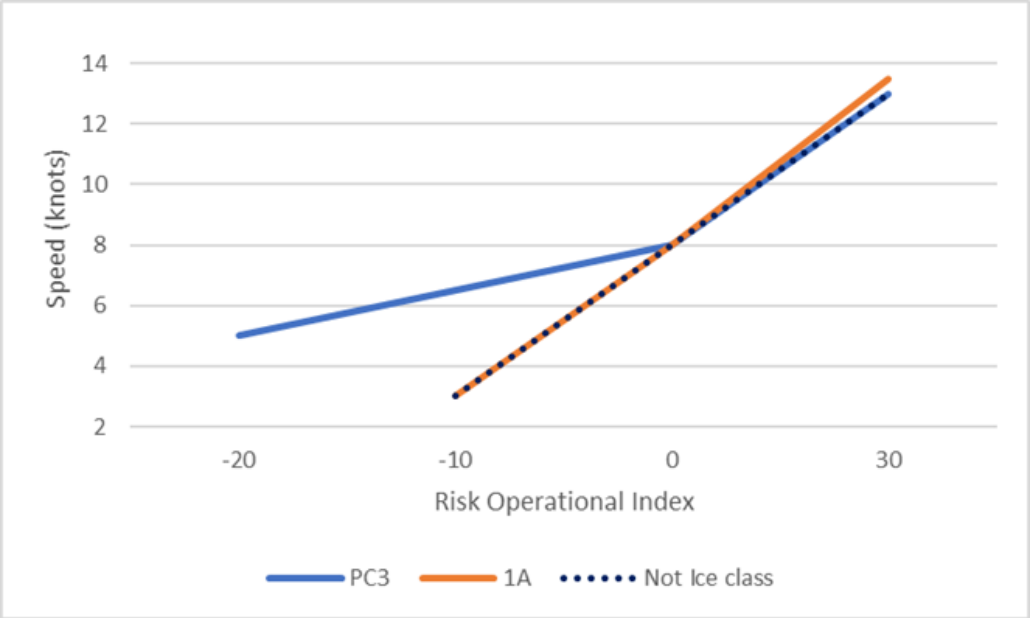


Figure 7.9 – Courbe de vitesses de 3 navires de classes glace différentes (Not Ice Class, 1A, PC3) en fonction de l'indice de risque POLARIS.

Dans un autre exemple publié dans [**Re1**], le navire étudié est le porte-conteneur "Venta Maersk". Ce navire dispose d'une certification de classe glace 1A et d'une vitesse optimale de 19 noeuds. Sa vitesse minimale recommandée de manoeuvrabilité dans la glace, sous escorte d'un brise-glace, a été déduite du minimum proposé dans la table 7.1 soit 3 knt.

La formule utilisée dans cet article pour le calcul de la vitesse V en fonction du RIO POLARIS est donnée par l'équation 7.2 dont la courbe est illustrée figure 7.10.

$$V = \frac{-1}{300} RIO^2 + \frac{7}{15} RIO + 8 \tag{7.2}$$

Ces formulations mathématiques de la vitesse du navire en fonction de l'indice de risque pourraient être améliorées en les confrontant à une approche statistique combinant les données météorologiques (et les indices de risques correspondants) avec les données de positionnement Automatic Identification System (AIS) des navires. De premières expérimentations ont été menées sur ces croisements de données soulevant des difficultés

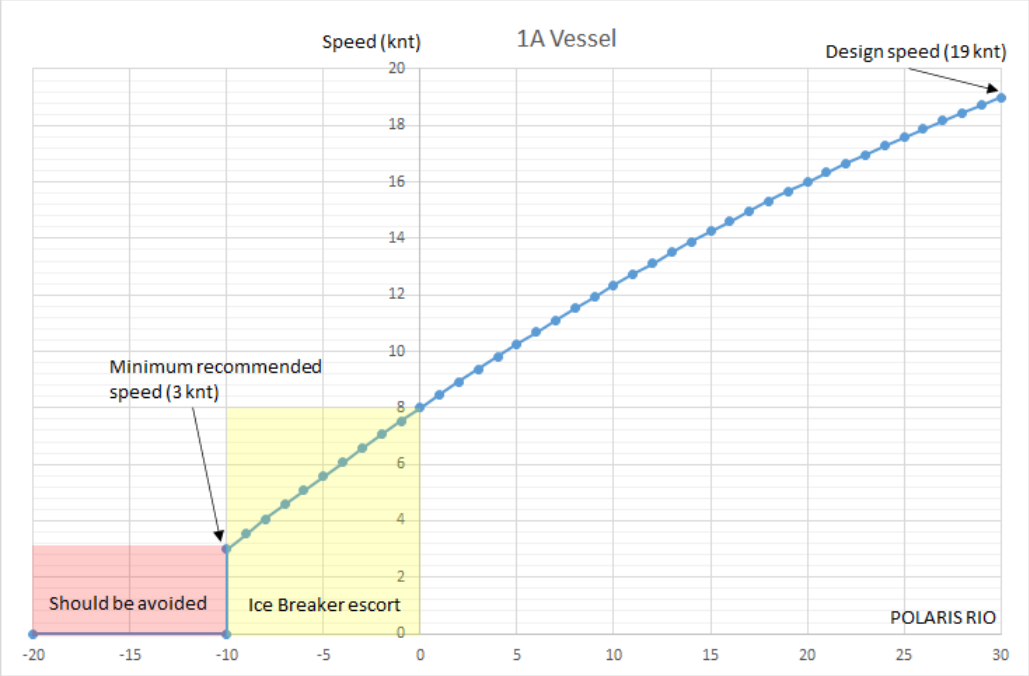


Figure 7.10 – Courbe des vitesses d'un navire 1A dont la vitesse optimale est de 19 noeuds en fonction de l'indice de risque POLARIS.

insoupçonnées. Dans certains cas, le navire réduit sa vitesse pour d'autres raisons que la présence de glaces de mer. De plus, lorsqu'un navire est escorté par un brise-glace, sa vitesse est également plus rapide et son indice de risque est minimisé ($RIO + 10$). Ces analyses statistiques nécessitent donc de connaître de nombreuses informations contextuelles concernant la certification de classe glace des navires (convertie dans le système POLARIS) ainsi que la présence d'autres navires pouvant potentiellement les escorter. Actuellement le système AIS ne diffuse pas d'information concernant les certifications de classe glace des navires. Il est donc nécessaire de s'appuyer sur des bases de données d'organismes privés référençant les caractéristiques des navires.

7.3.2 Sources de données environnementales historiques utilisées

Différentes sources de données historiques des observations des glaces de mer pour l'Arctique Canadien (National Snow & Ice Data Center, Canadian Ice Service Arctic Regional Sea Ice Charts in SIGRID-3 Format, 2006-2016) ainsi que les données fournies par la plateforme européenne COPERNICUS (Arctic Ocean Physical Reanalysis Product, ARCTIC_REANALYSIS_PHYS_002_003) ont été utilisées [**CI9**, **Ch2**].

Ces données ont été intégrées dans une base de données spatio-temporelle (voir section 7.1). Les indices POLARIS indiquant les capacités de navigation des 12 différentes classes de navires ont été calculés pour chaque jour de chaque année et chaque cellule raster en utilisant les variables de concentration de glace (SIC) et d'épaisseur (SIT). Le tableau de la figure 7.8 indique les épaisseurs de glace maximales utilisées pour définir les différents types de glaces et obtenir le coefficient de risque RV associé.

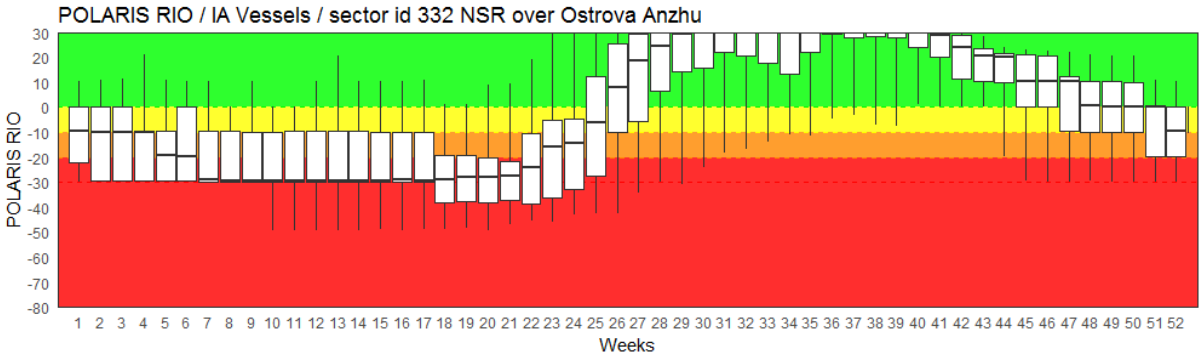


Figure 7.11 – Boîtes à moustaches obtenues par agrégation des indices de risque POLARIS du secteur Ostrova Anzhu regroupés par semaines.

Ainsi, pour chaque jour de l'année, il est possible d'observer différents scénarios correspondant aux percentiles (boxplot) des indices de risques POLARIS observés au cours des 30 dernières années. Un exemple préliminaire de ces travaux est détaillé sur les cartes du Tableau 7.2 pour un navire de classe 1A et les jours de l'année 130, 170, 260 (mai, juin, septembre). On peut y observer la grande variabilité spatiale et temporelle des zones accessibles pour ce type de navire (1A). La figure 7.11 présente les boîtes à moustaches regroupées par semaines pour le secteur de Ostrova Anzhu. Cette visualisation est particulièrement intéressante, car elle permet d'observer et de différencier les périodes de l'année où le régime de risque est stable (boîte entièrement située dans la zone verte ou dans la zone rouge) des périodes où il existe une plus grande incertitude concernant les capacités de navigation d'un navire dans cette zone (boîte étalée entre le rouge et le vert).

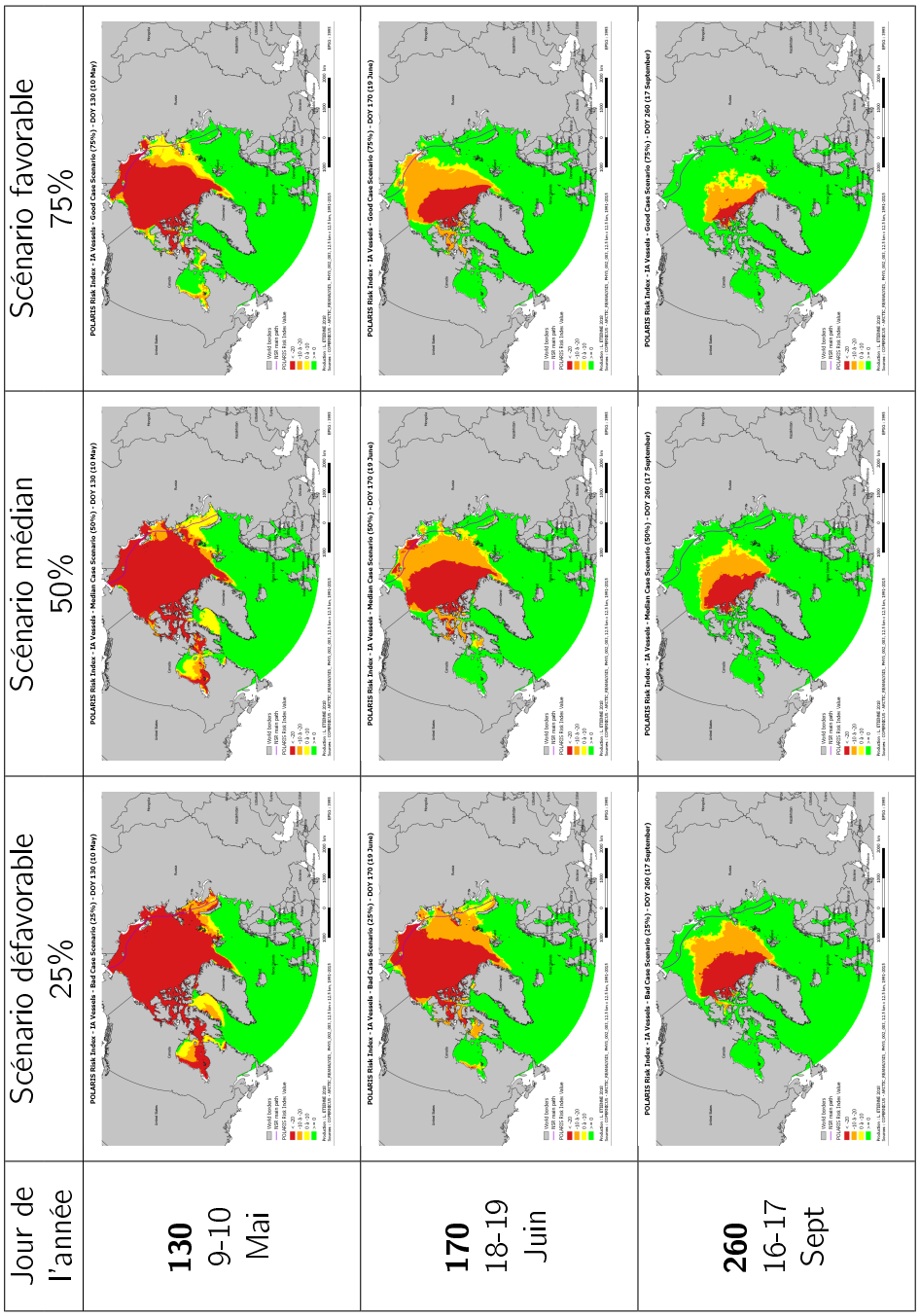


Table 7.2 – Cartes spatio-temporelles des indices POLARIS pour un navire de type IA (1991-2015)

Ces travaux sont préliminaires à une étude sur la prise de décision lors de la planification des voyages dans l'Arctique et la prise en compte de l'incertitude dans ce processus de décision. Pour l'Arctique Canadien, les résultats de cette analyse POLARIS ont été présentés et publiés dans différentes conférences et revues [**Ch3**, **CI9**, **CI8**].

7.4 Étude des incidents de navigation en zone polaire.

Le changement climatique impacte fortement les zones polaires. Le réchauffement climatique est plus rapide que la moyenne en Arctique et la fonte des glaces de mer y est de plus en plus prononcée. Cette situation a pour conséquence un intérêt accru pour les capacités de navigation le long de la route maritime du Nord (NSR).

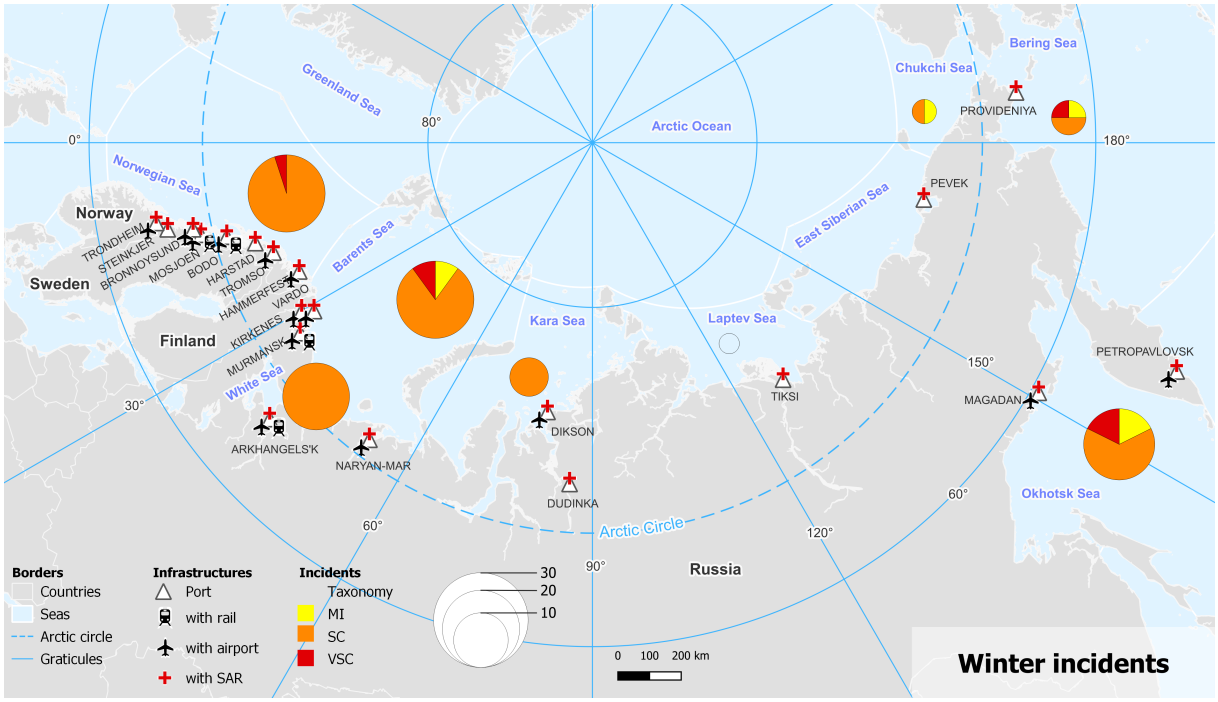


Figure 7.12 – Carte représentant l'analyse des incidents de navigation en zone Arctique.

Cependant, le risque lié aux glaces de mer est toujours présent et cette zone maritime, bien que moins fréquentée, reste particulièrement accidentogène. Une étude des rapports d'incidents nous a permis de constituer une base de données ainsi qu'une analyse de ces incidents classés à l'aide d'une taxonomie et présentés sur la figure 7.12. Les travaux en lien avec ces analyses d'incidents ont été publiés dans [**CI5**, **CI3**, **Re4**]. Une étude comparative

des indices POLARIS observés lors de certains incidents a été réalisée dans [**Re4**]. Cette étude a souligné que certains incidents surviennent dans des zones où l'indice POLARIS reste favorable au navire. Cependant, certains cas étudiés ont montré l'intérêt de cet indice qui prédisait des conditions de glace défavorables au navire étudié lors de son accident.

7.5 Optimisation du transit maritime en zone polaire

Les travaux de recherche présentés à la section 7.3 sont particulièrement utiles pour définir les limites des capacités de navigation de différentes catégories de navires dans l'Océan Arctique.

Un modèle à base de graphe a été généré à partir de ces données statistiques agrégées. Chaque cellule raster couvrant l'Arctique est alors considérée comme un nœud du graphe. Ce nœud est accessible à différentes périodes de l'année en fonction des conditions de glace et de l'indice POLARIS calculé pour chaque classe de navire. Chaque nœud est ensuite connecté topologiquement aux 8 cellules voisines pour former les arcs du graphe de navigation.

Ces arcs disposent d'attributs (indice RIO POLARIS, capacité et vitesse de déplacement, escorte de brise-glaces) variant au cours du temps. La vitesse du navire étant directement reliée à l'indice POLARIS comme détaillé à la section 7.3.1. Le résultat obtenu est un graphe dynamique dépendant du temps (Time-Dependent Graph).

Ce graphe a ensuite été exploité pour calculer différents scénarios de plus courts chemins entre différents ports de l'Arctique [**CI4**, **Ch1**]. Ces scénarios (favorable, médian, défavorable) se basent sur les statistiques descriptives calculées dans le modèle de la section 7.3.

La recherche de plus court chemin (en termes de temps) sur des graphes dépendant du temps est une tâche complexe nécessitant l'usage d'heuristiques et d'optimisations de calculs. Pour chaque catégorie de navires, nous avons calculé les plus courts chemins des 3 scénarios pour chaque jour de l'année avec ou sans escorte de brise-glace. Les courbes des temps de parcours des différents scénarios mettent en lumière des fenêtres de navigation plus ou moins larges en fonction des navires employés [**Re2**, **CI4**].

La figure 7.14 extraite de [**CI4**] illustre les temps nécessaires à un navire de catégorie 1A pour réaliser le transit entre Mourmansk et Bering pour chaque jour de départ de l'année en fonction des 3 scénarios statistiques retenus (25% en rouge, 50% en bleu et

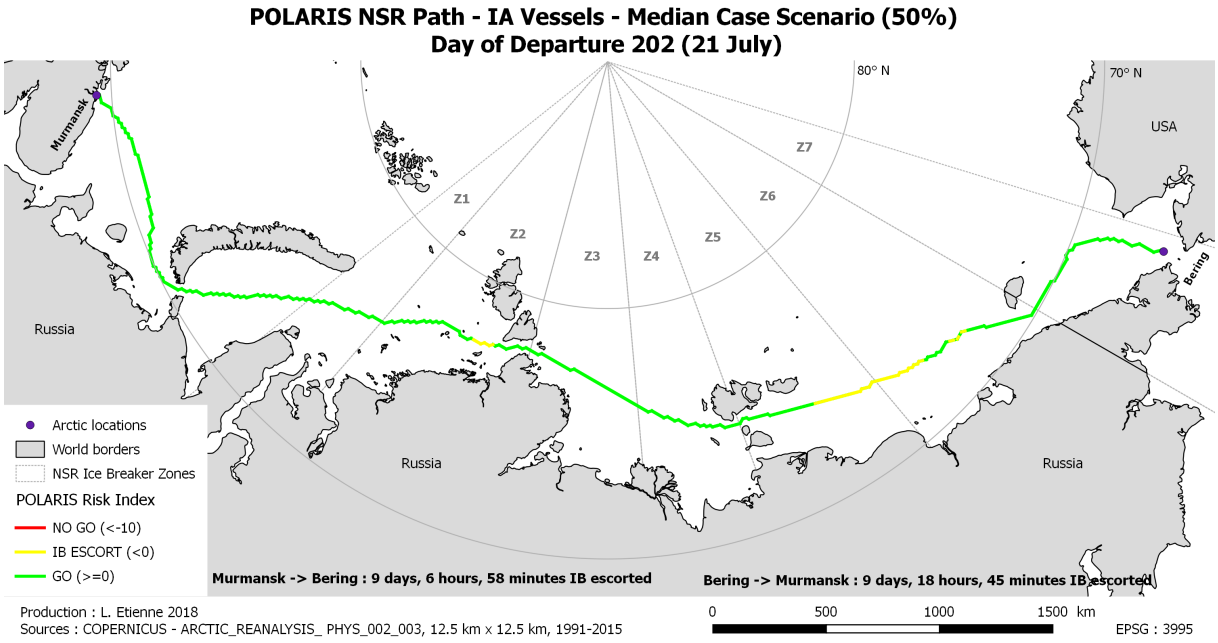


Figure 7.13 – Chemin et temps de transit réalisé par un navire de type 1A entre Mourmansk et Bering selon le scénario médian.

75% en vert). Dans cet exemple, une version simplifiée du graphe a été utilisée. Seuls les arcs situés le long d'une route de référence fixée ont été utilisés. La figure 7.13 montre le chemin utilisé ainsi que les zones en jaune nécessitant une escorte de brise-glace.

Les résultats de ces calculs de temps de parcours sont ensuite intégrés dans un modèle d'optimisation de trafic maritime. Ce modèle s'intéresse à l'étude des combinaisons de flottes de navires, à l'ordonnancement des navires et au choix de la route (Suez ou NSR) pour le transport de modules servant à l'assemblage de l'usine de gaz naturel liquéfié de Yamal en Russie [**CI6**, **CI2**].

Une version modifiée de l'algorithme A* tenant compte du risque POLARIS a été étudiée dans le cadre de la thèse de Marc STODDARD en collaboration avec l'université de Dalhousie et le DRDC au Canada. En s'appuyant sur les 3 différents scénarios du graphe spatio-temporel, différents plus courts chemins ont été calculés entre différents lieux importants situés dans l'arctique Canadien. Ces résultats préliminaires ne sont plus contraints à une zone spatiale restreinte. La figure 7.15 montre les 3 différents chemins obtenus pour aller à Resolute à l'aide d'un navire de type PC5 en partant le 90ème jour de l'année. La route verte, d'une durée de 4,82 jours, correspond à la meilleure route possible en cas de scénario favorable (25%). La route bleue, d'une durée de 6,3 jours, correspond

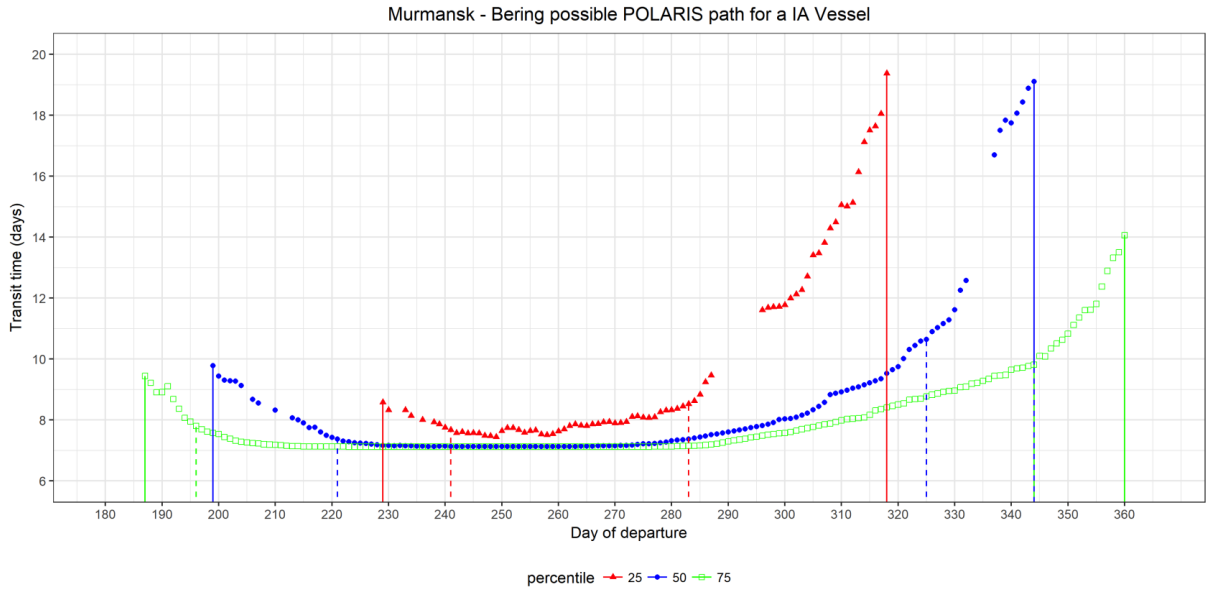


Figure 7.14 – Temps de transit requis entre Mourmansk et Bering pour un navire de type IA selon les 3 différents scénarios statistiques.

à la route la plus courte en cas de scénario médian (50%). Enfin, la route rouge, d'une durée de 7,17 jours, illustre la route réalisable en cas de scénario défavorable (75%). Le fond de carte utilisé présente l'épaisseur de glace médiane observée le jour de navigation (jour 93) les portions des routes parcourues lors du troisième jour de navigation (jour 93) sont surlignées en jaune.

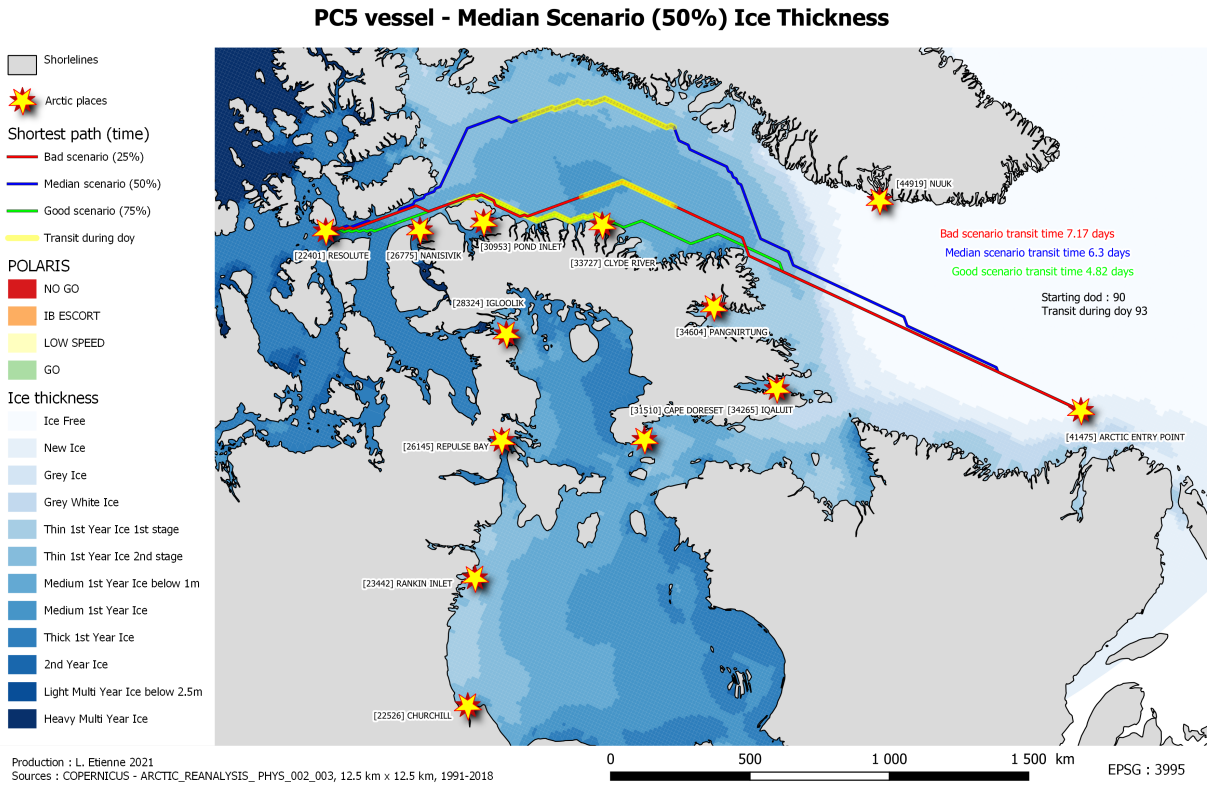


Figure 7.15 – Plus courts chemins pour un navire de type PC5 s'appuyant sur les 3 scénarios du graphe spatio-temporel.

Chapitre 8

CONCLUSION ET PERSPECTIVES DE RECHERCHE

Les différents chapitres de ce mémoire ont présenté une synthèse des travaux et des contributions réalisés dans le cadre de l'étude des trajectoires. Ce chapitre présente une synthèse des contributions significatives ainsi que des pistes de recherche associées.

8.1 Optimisation du calcul de la distance de Fréchet discrète

Dans le chapitre 5, de premières contributions visant à optimiser le temps de calcul de la distance de Fréchet discrète ont été présentées. Les optimisations proposées se basent sur une première étape de simplification des trajectoires à comparer à l'aide d'un filtre de Douglas et Peucker spatio-temporel. Néanmoins, ce filtrage induit un échantillonnage irrégulier des trajectoires qui complexifie la comparaison entre trajectoires et augmente considérablement la marge d'erreur comme indiqué dans la section 5.3. Le filtre de Douglas et Peucker spatio-temporel dispose d'un paramètre de seuil de filtrage qui spécifie la marge d'erreur maximale tolérée pour la simplification de la trajectoire.

Cependant, l'erreur maximale de mesure entre la distance de Fréchet discrète et la distance de Fréchet continue est directement liée à la longueur des plus grands segments. Dans le cas où certaines trajectoires sont composées de longues lignes droites, de très longs segments peuvent être gérés. La longueur de ces segments étant alors très supérieure à la marge d'erreur tolérée pour le filtrage de Douglas et Peucker.

Certains cas particuliers induisant ces erreurs d'approximation de la distance de Fréchet discrète ont été étudiés dans [**CI7**, **CN1**]. Il reste néanmoins des cas de figures pour lesquels aucune solution n'a été proposée.

L'étude de ces cas particuliers nécessite la prise en compte du positionnement relatif de segments successifs des deux trajectoires à comparer. Cette piste de recherche nécessite une formalisation mathématique du problème qui devra ensuite être implémentée dans les algorithmes de ré-échantillonnage des trajectoires. Des points supplémentaires étant alors ajoutés sur les segments de trajectoires afin de limiter l'erreur d'approximation liée aux segments trop longs.

L'objectif de ces travaux vise à optimiser le temps de calcul de la distance de Fréchet discrète tout en conservant une erreur d'approximation bornée. C'est pourquoi il sera important de qualifier précisément l'impact de ces optimisations sur les temps de calcul. Une autre piste de recherche consiste à étudier l'impact de la technique de filtrage des trajectoires sur l'erreur de calcul de la distance de Fréchet discrète.

8.2 Intégration de la composante spatiale dans les mesures de similarité sémantique

Dans le chapitre 4, deux nouvelles mesures de similarité sémantique ont été proposées. La mesure de similarité CED introduite à la section 4.2 est capable de comparer des séquences sémantiques d'activités tout en tenant compte du contexte exprimé dans ces séquences. Les séquences sémantiques d'activités peuvent être de tailles différentes. CED tient compte de l'ordre dans lequel les activités sont réalisées, mais pas de leurs durées respectives.

Une extension floue de la distance de Hamming, nommée FTH a été présentée à la section 4.3.1. Cette mesure tient également compte du contexte des séquences sémantiques tout en incorporant la dimension temporelle à l'aide de fonctions floues. Cependant, les séquences sémantiques étudiées doivent alors avoir la même durée totale.

Dans ces travaux, la composante spatiale des trajectoires a été volontairement mise de côté. Une perspective de recherche prometteuse consiste à étendre les mesures de similarité CED et FTH pour la prise en compte de cette dimension spatiale.

La première piste de recherche consiste à convertir les attributs décrivant la composante spatiale de la trajectoire en éléments sémantiques. Ainsi, la localisation plus ou moins précise et détaillée d'une activité réalisée dans une séquence sémantique pourrait être convertie en un sous-ensemble fini d'éléments sémantiques décrivant l'espace. Cette

représentation de l'espace pouvant être conceptualisée sous forme de graphes à différentes échelles. Par exemple, une activité de shopping réalisée dans un centre commercial pourrait être rattachée à un lieu lui-même connecté à un quartier situé dans une ville d'un même département. Cependant, dans le cadre spécifique de l'étude des trajectoires, les activités statiques (STOPS) et de déplacement (MOVES) disposent de caractéristiques spatiales très différentes. Les travaux de recherche présentés dans le chapitre 6 se sont focalisés sur la définition de patrons caractérisant les variations spatiales et temporelles d'un ensemble d'objets mobiles.

Le concept d'OSBP a été introduit à la section 6.1.3 pour caractériser les variations spatiales d'un nuage de positions observées. Ce concept a ensuite été étendu section 6.1.4 à l'OSTBP pour des suites de nuages de positions homogènes obtenus par appariements successifs à une trajectoire médiane définie à la section 6.2.1 afin d'obtenir des couloirs spatio-temporels décrivant l'évolution des objets mobiles dans le temps et dans l'espace (Concept de TBP défini section 6.2.2).

La seconde piste de recherche proposée vise à combiner les mesures de similarité de séquences sémantiques CED et FTH avec les mesures de similarité spatio-temporelles s'appuyant sur l'OSTBP pour les STOPS et le TBP pour les MOVES.

8.3 Définition de nouvelles mesures de similarité entre patrons spatio-temporels

Différents patrons spatio-temporels ont été présentés au chapitre 6. En complément de ces patrons, des mesures de similarité entre un patron et une trajectoire ont été proposées à la section 6.3. Ces mesures de similarité reposent sur un ensemble de règles floues calculées à l'aide de descripteurs portant sur des mesures de distances spatiales et temporelles normalisées. La similarité entre un patron TBP et une trajectoire peut être mesurée uniquement lorsque la trajectoire à comparer est entièrement réalisée. La distance de Fréchet discrète, utilisée pour aligner les positions d'une trajectoire avec la trajectoire médiane du TBP, est applicable à des alignements de trajectoires incomplètes de tailles différentes. Une première piste de recherche porte sur l'adaptation des mesures de similarité entre un patron (TBP) et une trajectoire partielle.

La seconde piste de recherche s'intéresse à la définition de nouvelles mesures per-

mettant de comparer deux patrons l'un à l'autre. Dans ce projet de recherche, différents groupes de trajectoires réalisant le même itinéraire (ayant la même origine et même destination) pourraient permettre de générer des patrons spatio-temporels distincts en fonction de critères de sélection (différentes dates, différents types d'objets mobiles). La comparaison de ces patrons ouvre des perspectives de recherche intéressantes concernant l'étude des différences entre ces groupes de trajectoires. Ainsi, au lieu de devoir comparer toutes les trajectoires composant les clusters deux à deux, il est envisageable de ne comparer que les éléments caractérisant le patron TBP tel que sa trajectoire médiane et les couloirs spatio-temporels. Outre la définition d'une nouvelle mesure de similarité entre patrons permettant d'indiquer globalement quels patrons sont les plus similaires les uns par rapport aux autres, il devrait être possible d'indiquer précisément où se trouvent ces similarités et différences tant sur le plan spatial que temporel. De plus, la définition de nouveaux mécanismes de comparaison de patrons de trajectoires devrait également permettre de faciliter l'étude de gros volumes de données de trajectoires.

8.4 Étude des mécanismes d'agrégation ou de division de patrons spatio-temporels

L'algorithme itératif de calcul des patrons de trajectoires TBP repose sur un appariement des trajectoires basé sur le calcul de la distance de Fréchet discrète. La complexité de cet algorithme, présentée au chapitre 5, est relativement élevée. Dans certains cas, les trajectoires constituant le groupe ayant permis de générer le TBP ne se comportent pas de façon homogène. Il est alors possible d'observer des changements notables dans la symétrie et l'étalement des OSTBP constituants le TBP. L'étude de l'évolution de ces paramètres (symétrie, étalement) au sein d'un TBP devrait permettre de détecter des portions spécifiques du TBP où les objets mobiles font preuve d'un comportement multimodal. Ainsi, les sous-parties homogènes d'un TBP pourraient être extraites afin de constituer de nouveaux sous-groupes de trajectoires pour lesquels il serait alors possible de générer des sous-patrons. Les sous-patrons de ces sous-ensembles de trajectoires pourraient alors être regroupés (ou divisés) en appliquant une classification ascendante hiérarchique s'appuyant sur les résultats des mesures de similarité entre patrons présentés dans la section précédente.

Enfin, une étude de potentiels mécanismes d'agrégation d'OSTBP pourrait permettre d'éviter de devoir recalculer un OSTBP constitué de l'ensemble des positions des deux sous-clusters. Ce mécanisme d'agrégation pourrait être basé sur une modélisation mathématique des distributions des deux clusters de positions initiaux.

8.5 Génération de trajectoires fictives à partir de patrons spatio-temporels

Disposant de patrons spatio-temporels décrivant l'espace spatio-temporel dans lequel évolue un ensemble de trajectoires, il est alors envisageable d'utiliser ces patrons afin de générer des trajectoires fictives simulant de manière plausible une trajectoire suivant ce même itinéraire.

Les TBP sont actuellement constituées d'une suite ordonnée d'OSTBP définissant l'espace 3D dans lequel il est plus ou moins probable qu'une position d'une trajectoire suivant cet itinéraire puisse se trouver. Des mesures complètent le TBP et donnent des informations concernant les variations successives observées entre deux OSTBP consécutifs au sein d'un TBP.

Ainsi, il est possible de générer un ensemble positions dont le tirage aléatoire des coordonnées respecte la probabilité de distribution spécifiée par un OSTBP. Cependant, il est également important de prendre en compte la façon dont deux positions successives d'une trajectoire sont connectées l'une à l'autre. Les résultats du premier tirage aléatoire limitant alors l'espace des possibles pour le tirage aléatoire au sein de l'OSTBP suivant à la manière d'un bruit de perlin. Cette piste de recherche nécessite de qualifier plus précisément les transitions entre les OSTBP. Une extension du concept de OSTBP aux segments pourrait alors être envisagée dans cet axe de recherche.

8.6 Prise en compte du contexte environnemental

Le chapitre 7 de ce manuscrit s'est focalisé sur la prise en compte du contexte environnemental dans l'étude des mobilités. La section 7.1 détaille un cadre d'application maritime pour lequel des bases de données environnementales historiques volumineuses ont été intégrées afin de définir des probabilités de présence de glaces de mers ainsi qu'une

caractérisation de l'incertitude concernant les risques associés pour la navigation maritime en zones polaires.

L'axe de recherche présenté dans cette section s'intéresse à l'étude du risque lié à la navigation maritime dans l'Océan Arctique. Lors du projet PASSAGES, nous avons réalisé des enquêtes auprès de compagnies de transport maritime travaillant dans l'Arctique Canadien. Différents risques ont été évoqués par ces compagnies lors des entretiens (Figure 8.1). Parmi ces risques, le plus saillant est celui lié à la présence de glaces et aux conditions climatiques difficiles (météo, vents, températures basses, brouillards). Le second facteur de risque mentionné est temporel, la saisonnalité et variabilité de ces conditions génèrent des difficultés d'anticipation et de planification des opérations.

Figure 8.1 – Carte heuristique des risques liés au transport maritime en Arctique.

Ces variabilités et incertitudes sont fortement liées à l'évolution climatique de l'environ-

nement arctique comme décrit dans la section 7.3 du chapitre 7. L'Organisation Maritime Internationale (IMO), consciente des dangers et spécificités liés à la navigation en zones polaires, a adopté une réglementation spécifique « International Code for Ships Operating in Polar Waters » (Polar Code) mise en application le 1er janvier 2017. Ce code polaire dispose d'un volet dédié à l'analyse des risques liés à la navigation dans la glace. POLARIS (Polar Operational Limit Assessment Risk Indexing System), présenté en section 7.3.1, est un système d'analyse de risque combinant différents paramètres de glaces (concentration, épaisseur) et capacités de différentes classes de navires. Disposant des données environnementales collectées et préalablement traitées dans le chapitre 7, je propose d'appliquer le système d'évaluation du risque POLARIS à ces données de glaces pour modéliser et visualiser les variations spatio-temporelles du risque pour chaque classe de navires. Ainsi, la saisonnalité et la variabilité des conditions climatiques pourront être observées et analysées sous forme de cartes des zones de l'Océan Arctique disposant des plus fortes variations statistiques du niveau de risque POLARIS tout au long de l'année (agrégation par jour, par semaine et/ou par zones) ou sur des scénarios de projections décennales.

8.7 Prise en compte de l'incertitude dans les processus de décision

L'autre risque que je propose de traiter dans ce projet porte sur l'incertitude concernant la bathymétrie et les cartes maritimes de cette zone immense et peu explorée. Un challenge de recherche en géomatique consiste à combiner le risque lié à la glace de mer (indice POLARIS) avec le risque d'échouement (Bathymétrie). L'équipe de recherche KLaIM du LabISEN dispose de compétences dans le domaine des simulations à base d'agents. Je souhaite intégrer ces données environnementales complexes au sein d'une simulation dans laquelle des agents (navires), disposant de caractéristiques différentes (type et catégorie de navires) et de connaissances plus ou moins précises de leur environnement, sont en mesure de prendre des décisions afin d'optimiser leur déplacement à travers l'océan Arctique. Les Systèmes Multi-Agents (SMA) permettent de simuler de très nombreuses situations complexes dans lesquels des agents doivent interagir et prendre des décisions qui peuvent mener dans certains cas à des situations d'échec dans lesquelles les navires se retrouvent dans des situations catastrophiques (bloqué dans la glace, échouage, collisions, naufrage).

L'étude des bases de données historiques d'accidents (section 7.4) ainsi que les données issues de capteurs de position AIS permettront également d'enrichir les modèles utilisés dans la simulation.

8.8 Définition d'un nouvel indice d'isolement

Enfin, je souhaite développer le concept d'indice d'isolement (Remoteness index) comme troisième indicateur de risque de navigation. Inspiré de l'indice de Nordicité proposé par Louis-Edmond Hamelin en 1976 combinant 10 variables qualifiant des phénomènes naturels et liés à l'activité humaine, l'indice d'isolement représente le temps nécessaire à un navire pour obtenir de l'aide dans un espace vaste et inhospitalier limitant ses capacités de déplacements. Ces travaux de recherche sur le concept d'indice d'isolement font actuellement l'objet d'un co-encadrement de thèse de doctorant avec l'Université de Dalhousie au Canada (section 7.5). L'indice d'isolement dynamique combinera les capacités de navigation du navire étudié, son environnement immédiat et futur (POLARIS, Bathymétrie) ainsi que les infrastructures avoisinantes (bases, communautés, autres navires) et leurs capacités d'interventions dans l'environnement.

Cet indice est particulièrement intéressant, car il n'est pas symétrique. En effet, dans les zones polaires, deux navires peuvent être proches spatialement les uns des autres, mais n'ont pas forcément les mêmes capacités de navigation dans des zones infestées de glaces de mer. Ainsi, en cas d'avarie sur l'un des deux navires, il n'est possible d'obtenir de l'assistance que si l'autre navire est en mesure de traverser la glace qui les sépare.

8.9 Modélisation du trafic maritime arctique

Le dernier volet de ce projet porte sur l'étude de l'évolution des activités humaines dans l'Arctique et leur impact sur le trafic maritime. Les récentes évolutions réglementaires (Polar Code) combinées au réchauffement climatique et à la fonte des glaces ouvrent peu à peu la voie au développement de nouvelles activités humaines dans l'Arctique. Ce projet vise à caractériser spatialement et temporellement les dynamiques environnementales du milieu marin arctique ainsi que l'impact sur les usages de ce territoire particulier en pleine mutation. La diversité des usages de cet espace maritime est en pleine expansion (raccourci de navigation mondiale, exploitation des ressources (pétrole, gaz, minerais, pêche...), tou-

risme d'aventure, campagnes de recherche scientifique, activités traditionnelles des communautés du Grand Nord (chasse, pêche, déplacements)). Ces usages, parfois concurrents, se développent actuellement en fonction de divers facteurs socio-politico-économiques (investissement économique, fluctuation du prix des matières premières, guerre en Ukraine...).

Les bases de données historiques des déplacements de navires (RADAR, LRIT, AIS, S-AIS) sont des sources de données précieuses pour l'analyse de ces activités maritimes et la définition de zones d'activité et de routes de navigation entre ces zones. Je dispose de connaissances approfondies dans ce domaine ainsi que de collaborations de recherche actives sur ce sujet. À l'aide de ces données, un graphe représentant les interconnexions entre les zones d'activités dans l'Arctique peut être défini. Différents scenarios de développement de ces usages ainsi que leurs incidences sur la navigation maritime arctique pourront être proposés. Je souhaite étendre le modèle "A spatio-temporal simulation model for forecasting marine traffic in the Canadian Arctic in 2020" à l'ensemble de l'Arctique en me basant sur les résultats produits à la seconde phase de ce projet de recherche. Ce projet bénéficiera des collaborations de recherche avec Olivier Faury (EM Normandie) et Patrick Rigot-Müller (National University of Ireland Maynooth) experts en logistique et transports. Ces collaborations visent à analyser et optimiser les activités de transport maritime intra et transarctique en fonction des résultats produits par la seconde phase de ce projet (analyse de risques et capacités de navigation dans la glace).

En conclusion, ces projets de recherche abordent des problématiques de géographie de l'environnement d'un point de vue numérique. Le sujet proposé contribue au développement de connaissances relatives aux interactions nature/société (étude de l'impact du réchauffement climatique sur le développement des activités marines dans l'Océan Arctique). Ils portent sur un terrain d'étude privilégié de l'équipe KLaIM du LabISEN et proposent une démarche modélisatrice ainsi qu'une approche spatio-temporelle. Ce projet vise à développer des collaborations de recherche multiples et interdisciplinaires au sein du LabISEN et de l'écosystème de recherche local, national et international via la participation et la contribution au GDR MAGIS dans l'action de recherche MISE (Mobilité et impacts socio-environnementaux) que je co-anime.

J'ai travaillé avec des collègues issus de nombreux champs disciplinaires (Géographie, Aménagement, Urbanisme, Environnement, Écologie, Histoire, Archéologie, Sociologie, Informatique, Mathématique, Économie, Droit ...) dans le cadre des Humanités numériques. Je souhaite vivement entretenir et conserver ces interactions pluridisciplinaires en-

richissantes dans mes futures activités de recherche.

Ces travaux de recherche portant sur des problématiques liées à l'environnement littoral, marin et à l'analyse de risques peuvent être appliqués à l'aide à la décision pour l'aménagement du territoire et sont directement transposables aux enseignements du domaine professionnel Numérique Environnement et Développement Durable que je gère à l'ISEN.

BIBLIOGRAPHIE

[**Ch1**] Faury, Olivier, **Etienne**, Laurent, Fedi, Laurent, Rigot-Müller, Patrick, Cheaitou, Ali et Stephenson, Scott. **sept. 2019**, « L'impact de la gestion du risque sur l'attractivité du passage du nord-est », in : *Baltic Arctic – Strategic Perspective*, Les collections Océanides, Editions EMS, p. 169-190, url : `https://hal.archives-ouvertes.fr/hal-02885901` (cf. p. 134).

[**Ch2**] Stoddard, Mark, **Etienne**, Laurent, Pelot, Ronald, Fournier, Melanie et Beveridge, Leah. **août 2018**, « From Sensing to Sense-Making : Assessing and Visualizing Ship Operational Limitations in the Canadian Arctic Using Open-Access Ice Data », in : *Sustainable Shipping in a Changing Arctic*, url : `https://hal.archives-ouvertes.fr/hal-01897714` (cf. p. 130).

[**Ch3**] **Etienne**, Laurent, Fournier, Melanie, Beveridge, Leah, Stoddard, Mark et Pelot, Ronald. **2017c**, « Vessel navigation constraints in Canadian Arctic waters », in : *Advances in Shipping Data Analysis and Modeling Tracking and Mapping Maritime Flows in the Age of Big Data*, url : `https://hal.archives-ouvertes.fr/hal-01627377` (cf. p. 127, 133).

[**Re1**] Cheaitou, Ali, Faury, Olivier, **Etienne**, Laurent, Fedi, Laurent, Rigot-Müller, Patrick et Stephenson, Scott. **nov. 2022**, « Impact of CO2 emission taxation and fuel types on Arctic shipping attractiveness », in : *Transportation Research Part D : Transport and Environment* 112, p. 103491, doi : `10.1016/j.trd.2022.103491`, url : `https://hal.archives-ouvertes.fr/hal-03912714` (cf. p. 128, 129).

[**Re2**] Rigot-Müller, Patrick, Cheaitou, Ali, **Etienne**, Laurent, Faury, Olivier et Fedi, Laurent. **jan. 2022**, « The role of polarseaworthiness in shipping planning for infrastructure projects in the Arctic : The case of Yamal LNG plant », in : *Transportation Research Part A : Policy and Practice* 155, p. 330-353, doi : `10.1016/j.tra.2021.11.009`, url : `https://hal.archives-ouvertes.fr/hal-03549995` (cf. p. 128, 134).

[**Re3**] Duroudier, Sylvestre, Chardonnel, Sonia, Mericskay, Boris, Andre-Poyaud, Isabelle, Bedel, Olivier, Depeau, Sandrine, Devogele, Thomas, **Etienne**, Laurent, Lepetit, Arnaud, Moreau, Clément, Pelletier, Nicolas, Ployon, Estelle et Tabaka, Kamila. **2020b**, « Diagnostic qualité et apurement des données de mobilité quotidienne issues de l'enquête mixte et longitudinale Mobi'Kids », in : *Revue Internationale de Géomatique* 30.*1-2*, p. 127-148, doi : 10.3166/rig.2020.00105, url : https://hal.archives-ouvertes.fr/hal-03254291 (cf. p. 32).

[**Re4**] Fedi, Laurent, Faury, Olivier et **Etienne**, Laurent. **avr. 2020**, « Mapping and analysis of maritime accidents in the Russian Arctic through the lens of the Polar Code and POLARIS system », in : *Marine Policy* 118, p. 103984, doi : 10.1016/j.marpol.2020.103984, url : https://hal.archives-ouvertes.fr/hal-02885952 (cf. p. 133, 134).

[**Re5**] **Etienne**, Laurent, Devogele, Thomas, Buckin, Maike et Mcardle, Gavin. **2016c**, « Trajectory Box Plot : a new pattern to summarize movements », in : *International Journal of Geographical Information Science*, Analysis of Movement Data 30.*5*, p. 835-853, doi : 10.1080/13658816.2015.1081205, url : https://hal.archives-ouvertes.fr/hal-01215945 (cf. p. 102, 108).

[**Re6**] **Etienne**, Laurent, Devogele, Thomas et Mcardle, Gavin. **2014d**, « Oriented spatial box plot, a new pattern for points clusters », in : *International Journal of Business Intelligence and Data Mining* 9.*3*, http://www.inderscience.com/info/inarticle.php?artid=68367, doi : 10.1504/IJBIDM.2014.068367, url : https://hal.archives-ouvertes.fr/hal-01172635 (cf. p. 86, 95, 98).

[**Re7**] Devogele, Thomas et **Etienne**, Laurent. **2012a**, « Mesures de similarité de trajectoires basées sur l'utilisation de patrons spatio-temporels », in : *Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes d'Information* 17.*1*, p. 11-34, url : https://hal.archives-ouvertes.fr/hal-01171533 (cf. p. 112, 114, 117).

[**CI1**] Moreau, Clément, Devogele, Thomas, De Runz, Cyril, Peralta, Veronika, MOREAU, Evelyne et **Etienne**, Laurent. **juill. 2021**, « A Fuzzy Generalisation of the Hamming Distance for Temporal Sequences », in : *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Luxembourg, France : IEEE, p. 1-8, doi : 10.1109/FUZZ45933.2021.9494445, url : https://hal.archives-ouvertes.fr/hal-03319236 (cf. p. 66).

[**CI2**] Faury, Olivier, Cheaitou, Ali, **Etienne**, Laurent, Fedi, Laurent et Rigot-Muller, Patrick. **juin 2020**, « The loading capacity of convoy for the transit of container along the Northeast Passage », in : *28th Annual Conference of the International Association of Maritime Economists*, Hong Kong, China, url : `https://hal.archives-ouvertes.fr/hal-03355136` (cf. p. 135).

[**CI3**] Fedi, Laurent, Faury, Olivier, Cheaitou, Ali, **Etienne**, Laurent et Rigot-Muller, Patrick. **juin 2020**, « Application and analysis of the IMO taxonomy on casualty investigation over 20 years of marine events in the Russian and Norwegian Arctic », in : *28th Annual Conference of the International Association of Maritime Economists*, Hong Kong, China, url : `https://hal.archives-ouvertes.fr/hal-03355149` (cf. p. 133).

[**CI4**] Faury, Olivier, Cheaitou, Ali, **Etienne**, Laurent, Fedi, Laurent, Rigot-Muller, Patrick et Stephenson, Scott. **juin 2019**, « How attractive is the Northern Sea Route for container shipping ? An economic model », in : *27th Annual Conference of the International Association of Maritime Economists*, Athens, Greece, url : `https://hal.archives-ouvertes.fr/hal-03354313` (cf. p. 134).

[**CI5**] Fedi, Laurent, Faury, Olivier, **Etienne**, Laurent et de FERRIERE le VAYER, Amaury. **juin 2019**, « Mapping and analysis of maritime claims in the Russian Arctic based on POLARIS System », in : *27th Annual Conference of the International Association of Maritime Economists*, Athens, Greece, url : `https://hal.archives-ouvertes.fr/hal-03354308` (cf. p. 133).

[**CI6**] Rigot-Müller, Patrick, **Etienne**, Laurent, Faury, Olivier et Stephenson, Scott. **juin 2018**, « Ship routing and scheduling for the assembly of a LNG plant in the arctic : a decision support system », in : *25th EUROMA Conference*, Budapest, Hungary, url : `https://hal.archives-ouvertes.fr/hal-03355195` (cf. p. 135).

[**CI7**] Devogele, Thomas, **Etienne**, Laurent, Esnault, Maxence et Lardy, Florian. **nov. 2017**, « Optimized Discrete Fréchet Distance between trajectories », in : *6th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, Redondo Beach, United States, 9 pages, doi : `10.1145/3150919.3150924`, url : `https://hal.archives-ouvertes.fr/hal-01631192` (cf. p. 81, 139).

[**CI8**] Stoddard, M A, **Etienne**, Laurent, Fournier, Melanie, Pelot, R et Beveridge, L. **2015g**, « Making sense of Arctic maritime traffic using the Polar Operational

Limits Assessment Risk Indexing System (POLARIS) », in : *9th Symposium of the International Society for Digital Earth (ISDE)*, t. 34, IOP Conf. Series : Earth and Environmental Science 1, Halifax, Canada, p. 012034, doi : 10.1088/1755-1315/34/1/012034, url : https://hal.archives-ouvertes.fr/hal-01512696 (cf. p. 127, 133).

[**CI9**] Stoddard, Mark, **Etienne**, Laurent et Pelot, Ronald. **2015h**, « From Sensing to Sense-Making : Assessing and visualizing ship operational limitations in the Canadian Arctic using open-access ice data », in : *Arctic Council International Conference on Safe and Sustainable Shipping in a Changing Arctic Environment (ShipArc 2015)*, Malmö, Sweden, url : https://hal.archives-ouvertes.fr/hal-01627369 (cf. p. 127, 130, 133).

[**CI10**] **Etienne**, Laurent, Devogele, Thomas et Mcardle, Gavin. **juin 2014**, « State of the Art in Patterns for Point Cluster Analysis », in : *Computational Science and Its Applications–ICCSA*, Guimaraes, Portugal, doi : 10.1007/978-3-319-09144-0_18, url : https://hal.archives-ouvertes.fr/hal-01118544 (cf. p. 85, 86, 95).

[**CI11**] **Etienne**, Laurent et Pelot, Ronald. **2013c**, « Simulation of maritime paths taking into account ice conditions in the Arctic », in : *11th International Symposium for GIS and Computer Cartography for Coastal Zone Management (CoastGIS)*, Victoria, Canada, url : https://hal.archives-ouvertes.fr/hal-01740751 (cf. p. 123, 124).

[**CN1**] Devogele, Thomas, Esnault, Maxence et **Etienne**, Laurent. **nov. 2016**, « Distance discrète de Fréchet optimisée », in : *Spatial Analysis and Geomatics (SAGEO)*, Nice, France, url : https://hal.archives-ouvertes.fr/hal-02110055 (cf. p. 82, 139).

[**CN2**] **Etienne**, Laurent et Devogele, Thomas. **jan. 2014**, « Trajectoires médianes », in : *14 ème conférence Extraction et Gestion des Connaissances, Ateliers fouille de données spatiales et temporelles & construction, enrichissement et exploitation de ressources géographiques pour l'analyse de données*, Rennes, France, url : https://hal.archives-ouvertes.fr/hal-02110086 (cf. p. 104).

[**Ra1**] **Etienne**, Laurent, Pelot, Ronald et Cecilia, Engler. **2013d**, *Analysis of Marine Traffic along Canada's Coasts : Phase 2 - Part 2 : A spatio-temporal simulation model for forecasting marine traffic in the Canadian Arctic in 2020*, Rapport de

recherche, Defense R&D Canada, Centre for Operational Research & Analysis, 182 p. (cf. p. 122).

[**Ra2**] **Etienne**, Laurent. **déc. 2011**, « Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime. », Thèse de doctorat, Université de Bretagne occidentale - Brest, 209 p., url : https://tel.archives-ouvertes.fr/tel-00667953 (cf. p. 58, 80, 85, 112).

[**1**] Guard, Canadian Coast. **2022b**, *Ice Navigation in Canadian Waters* (cf. p. 124, 128).

[**2**] Bisone, Frédérick. **mai 2021**, « Extraction de trajectoires sémantiques à partir de donnéesmulti-capteurs : application à des véhicules de secours », Theses, Université de tours, url : https://hal.archives-ouvertes.fr/tel-03365137 (cf. p. 36, 46).

[**3**] Jaballah, Rabie, Veenstra, Marjolein, Coelho, Leandro C. et Renaud, Jacques. **2021b**, « The time-dependent shortest path and vehicle routing problem », in : *INFOR : Information Systems and Operational Research* 59.4, p. 592-622, doi : 10.1080/03155986.2021.1973785, eprint : https://doi.org/10.1080/03155986.2021.1973785, url : https://doi.org/10.1080/03155986.2021.1973785 (cf. p. 119).

[**4**] Moreau, Clément. **nov. 2021**, « Fouille de séquences de mobilité sémantique : sur l'élaboration de mesures pour la comparaison, l'analyse et la découverte de comportements », Theses, Université de Tours, url : https://hal.archives-ouvertes.fr/tel-03607421 (cf. p. 28, 53, 55-58, 61, 64, 66, 70).

[**5**] Claramunt, Christophe. **2020a**, « Ontologies for Geospatial information : progress and challenges ahead », in : *Journal of Spatial Information Science 20*, p. 35-41, doi : 10.5311/JOSIS.2020.20.666, url : https://hal.science/hal-03200202 (cf. p. 53).

[**6**] Iphar, Clément, Napoli, Aldo et Ray, Cyril. **nov. 2020**, « An expert-based method for the risk assessment of anomalous maritime transportation data », in : *Applied Ocean Research* 104, p. 102337, doi : 10.1016/j.apor.2020.102337, url : https://hal-mines-paristech.archives-ouvertes.fr/hal-03405527 (cf. p. 85).

[7] Moreau, Clement, Devogele, Thomas, Peralta, Veronika et Etienne, Laurent. **2020g**, « Contextual Edit Distance for Semantic Trajectories », in : *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, p. 635-637 (cf. p. 62).

[8] Moreau, Clement, Devogele, Thomas, Peralta, Veronika, Etienne, Laurent et de Runz, Cyril. **2020h**, « Methodology for Mining, Discovering and Analyzing Semantic Human Mobility Behaviors », in : *arXiv preprint arXiv :2012.04767* (cf. p. 72).

[9] Moreau, Clement, Peralta, Veronika, Marcel, Patrick, Chanson, Alexandre et Devogele, Thomas. **2020i**, « Learning Analysis Patterns using a Contextual Edit Distance », in : *DOLAP 2020, EDBT/ICDT*, t. 2572, p. 46-55 (cf. p. 65).

[10] Raifer, Martin, Auer, Michael, Loos, Lukas, Troilo, Rafael, Kowatsch, Fabian, Visintini, Johannes et Zipf, Alexander. **2020j**, « Ohsome – OpenStreetMap data quality analysis », in : *3rd International Workshop on Spatial Data Quality*, Valletta, Malta, url : `https://eurogeographics.org/wp-content/uploads/2019/06/3-SDQ2020-Ohsome-OpenStreetMapData-Quality-Analysis.pdf` (cf. p. 43).

[11] Rehrl, Karl, Gröchenig, Simon et Kranzinger, Stefan. **2020k**, « Why did a vehicle stop ? A methodology for detection and classification of stops in vehicle trajectories », in : *International Journal of Geographical Information Science* 34.*10*, p. 1953-1979 (cf. p. 51).

[12] Bisone, Frédérick, Devogele, Thomas et Etienne, Laurent. **2019a**, « From raw sensor data to semantic trajectories », in : *Proceedings of the 5th ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management*, p. 1-4 (cf. p. 50).

[13] Chen, Chao, Ding, Yan, Xie, Xuefeng, Zhang, Shu, Wang, Zhu et Feng, Liang. **2019b**, « TrajCompressor : An online map-matching-based trajectory compression framework leveraging vehicle heading direction and change », in : *IEEE Transactions on Intelligent Transportation Systems* 21.*5*, p. 2012-2028 (cf. p. 33).

[14] Follin, Jean-Michel, Girres, Jean-François, Olteanu-Raimond, Ana-Maria et Sheeren, David. **sept. 2019**, « The Origins of Imperfection in Geographic Data », in : *Geographic Data Imperfection 1 : From Theory to Applications*, sous la dir. de Mireille Batton-Hubert, Eric Desjardin et François Pinet, Chapter 3, Wiley, doi :

10.1002/9781119507284.ch3, url : https://hal.science/hal-03792971 (cf. p. 38).

[15] Heni, Hamza, Coelho, Leandro C et Renaud, Jacques. **2019g**, « Determining time-dependent minimum cost paths under several objectives », in : *Computers & Operations Research* 105, p. 102-117 (cf. p. 119).

[16] Ivanovic, Stefan, Olteanu-Raimond, Ana-Maria, Mustière, Sébastien et Devogele, Thomas. **sept. 2019**, « A Filtering-Based Approach for Improving Crowdsourced GNSS Traces in a Data Update Context », in : *ISPRS International Journal of Geo-Information* 8.*9*, p. 380, doi : 10.3390/ijgi8090380, url : https://hal.science/hal-02276043 (cf. p. 35).

[17] Ivanovic, Stefan, Olteanu-Raimond, Ana-Maria, Mustière, Sébastien et Devogele, Thomas. **2019i**, « Potential of Crowdsourced Traces for Detecting Updates in Authoritative Geographic Data », in : *The Annual International Conference on Geographic Information Science*, p. 205-221, url : https://hal.science/hal-02269233 (cf. p. 35).

[18] Jepsen, Tobias Skovgaard, Jensen, Christian S et Nielsen, Thomas Dyhre. **2019j**, « Graph convolutional networks for road networks », in : *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, p. 460-463 (cf. p. 33).

[19] Truong, Quy Thy, de Runz, Cyril et Touya, Guillaume. **2019k**, « Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining », in : *International Journal of Geographical Information Science* 33.*8*, p. 1651-1682, url : https://www.tandfonline.com/doi/full/10.1080/13658816.2018.1556395 (cf. p. 43).

[20] Yan, Xiongfeng, Ai, Tinghua, Yang, Min et Yin, Hongmei. **2019l**, « A graph convolutional neural network for classification of building patterns using spatial vector data », in : *ISPRS journal of photogrammetry and remote sensing* 150, p. 259-273 (cf. p. 33).

[21] Zhang, Yue et Lin, Yaping. **2019m**, « An interactive method for identifying the stay points of the trajectory of moving objects », in : *Journal of Visual Communication and Image Representation* 59, p. 387-392 (cf. p. 38).

[22] Bermingham, Luke et Lee, Ickjai. **2018a**, « A probabilistic stop and move classifier for noisy GPS trajectories », in : *Data Mining and Knowledge Discovery* 32.*6*, p. 1634-1662 (cf. p. 38).

[23] Karaim, Malek, Elsheikh, Mohamed, Noureldin, Aboelmagd et Rustamov, RB. **2018b**, « GNSS error sources », in : *Multifunctional Operation and Application of GPS ; Rustamov, RB, Hashimov, AM, Eds*, p. 69-85 (cf. p. 38).

[24] Beber, Marco Aurelio. **2017a**, « Individual and group activity recognition in moving object trajectories », thèse de doct., Universidade federal de Santa Catarina (cf. p. 33).

[25] Luo, Ting, Zheng, Xinwei, Xu, Guangluan, Fu, Kun et Ren, Wenjuan. **2017d**, « An Improved DBSCAN Algorithm to Detect Stops in Individual Trajectories », in : *ISPRS Int. J. Geo-Information* 6.*3*, p. 63, doi : 10.3390/ijgi6030063, url : https://doi.org/10.3390/ijgi6030063 (cf. p. 40).

[26] Mooney, Peter et Minghini, Marco. **2017e**, « A Review of OpenStreetMap Data », in : *Citizen Sensor*, p. 37-59 (cf. p. 43).

[27] Tu, Wei, Cao, Jinzhou, Yue, Yang, Shaw, Shih-Lung, Zhou, Meng, Wang, Zhensheng, Chang, Xiaomeng, Xu, Yang et Li, Qingquan. **2017f**, « Coupling mobile phone and social media data : A new approach to understanding urban functions and diurnal patterns », in : *International Journal of Geographical Information Science* 31.*12*, p. 2331-2358 (cf. p. 33).

[28] Committee, Maritime Safety. **2016a**, *Guidance on methodologies for assessing operational capabilities and limitations in ice*, rapp. tech., Tech. Rep. MSC. 1/Circ. 1519, International Maritime Organization, London (cf. p. 127, 128).

[29] Ivanović, Stefan S, Raimond, Ana-Maria Olteanu, Mustière, Sébastien et Devogele, Thomas. **2016d**, « Detection of outliers in crowdsourced GPS traces », in : *12th international symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences* (cf. p. 39).

[30] Longgang, XIANG et Xiaotian, SHAO. **2016e**, « Visualization and extraction of trajectory stops based on kernel-density », in : *Acta Geodaetica et Cartographica Sinica* 45.*9*, p. 1122 (cf. p. 38).

[31] Miller, Harvey J. **2016f**, « Time geography and space–time prism », in : *International encyclopedia of geography : People, the earth, environment and technology*, p. 1-19 (cf. p. 29, 30).

[**32**] Mittelstadt, Brent Daniel et Floridi, Luciano. **2016g**, « The ethics of big data : current and foreseeable issues in biomedical contexts », in : *Science and engineering ethics* 22.*2*, p. 303-341 (cf. p. 35).

[**33**] Zhu, Ganggao et Iglesias, Carlos A. **2016h**, « Computing semantic similarity of concepts in knowledge graphs », in : *IEEE Trans. on Knowledge and Data Engineering* 29.*1*, p. 72-85 (cf. p. 33, 55).

[**34**] Bonnel, Patrick, Hombourger, Etienne, Olteanu-Raimond, Ana-Maria et Smoreda, Zbigniew. **2015a**, « Passive Mobile Phone Dataset to Construct Origin-destination Matrix : Potentials and Limitations », in : *Transportation Research Procedia* 11, p. 381-398, doi : 10.1016/j.trpro.2015.12.032, url : https://shs.hal.science/halshs-01664219 (cf. p. 35).

[**35**] Harispe, Sébastien, Ranwez, Sylvie, Janaqi, Stefan et Montmain, Jacky. **2015b**, « Semantic similarity from natural language and ontology analysis », in : *Synthesis Lectures on Human Language Technologies* 8.*1*, p. 1-254 (cf. p. 55).

[**36**] Helbing, Dirk et Pournaras, Evangelos. **2015c**, « Society : Build digital democracy », in : *Nature* 527.*7576*, p. 33-34 (cf. p. 35).

[**37**] IMO. **2015d**, *International code for ships operating in polar waters (Polar Code)* (cf. p. 127).

[**38**] Olteanu-Raimond, Ana-Maria, Mustiere, Sebastien et Ruas, Anne. **2015e**, « Knowledge formalization for vector data matching using belief theory », in : *Journal of Spatial Information Science* 2015.*10*, p. 21-46 (cf. p. 43).

[**39**] Popa, Iulian Sandu, Zeitouni, Karine, Oria, Vincent et Kharrat, Ahmed. **2015f**, « Spatio-temporal compression of trajectories in road networks », in : *GeoInformatica* 19.*1*, p. 117-145 (cf. p. 33).

[**40**] Wu, Fei, Li, Zhenhui, Lee, Wang-Chien, Wang, Hongjian et Huang, Zhuojie. **2015i**, « Semantic annotation of mobility data using social media », in : *Proceedings of the 24th International Conference on World Wide Web*, p. 1253-1263 (cf. p. 33).

[**41**] Bogorny, Vania, Renso, Chiara, de Aquino, Artur Ribeiro, de Lucca Siqueira, Fernando et Alvares, Luis Otavio. **2014a**, « Constant–a conceptual data model for semantic trajectories of moving objects », in : *Transactions in GIS* 18.*1*, p. 66-88 (cf. p. 33, 34).

[**42**] Chen, Wen, Ji, MH et Wang, JM. **2014b**, « T-DBSCAN : A Spatiotemporal Density Clustering for GPS Trajectory Segmentation. », in : *International Journal of Online Engineering* 10.*6* (cf. p. 40).

[**43**] Vandecasteele, Arnaud, Devillers, Rodolphe et Napoli, Aldo. **2014f**, « From Movement Data to Objects Behavior Using Semantic Trajectory and Semantic Events », in : *Marine Geodesy* 37.*2 - Special Issue : Special Issue on Coastal and Marine Geographic Information Systems*, p. 126-144, doi : 10.1080/01490419.2014.902885, url : https://hal-mines-paristech.archives-ouvertes.fr/hal-01002799 (cf. p. 53).

[**44**] William, J. Hughes, NSTB/WAAS et Airport, Atlantic City International. **2014g**, « Global positioning system (gps) standard positioning service (sps) performance analysis report », in : *GPS Product Team : Washington, DC, USA* (cf. p. 42).

[**45**] Cross, Valerie, Yu, Xinran et Hu, Xueheng. **2013a**, « Unifying ontological similarity measures : A theoretical and empirical investigation », in : *International Journal of Approximate Reasoning* 54.*7*, p. 861-875 (cf. p. 55).

[**46**] Duran, Adam et Earleywine, Matthew. **2013b**, *GPS data filtration method for drive cycle analysis applications*, rapp. tech., National Renewable Energy Lab.(NREL), Golden, CO (United States) (cf. p. 39).

[**47**] Hasan, Samiul, Zhan, Xianyuan et Ukkusuri, Satish V. **2013e**, « Understanding urban human activity and mobility patterns using large-scale location-based data from online social media », in : *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*, p. 1-8 (cf. p. 33).

[**48**] Yan, Zhixian, Chakraborty, Dipanjan, Parent, Christine, Spaccapietra, Stefano et Aberer, Karl. **2013f**, « Semantic trajectories : Mobility data computation and annotation », in : *ACM Transactions on Intelligent Systems and Technology (TIST)* 4.*3*, p. 1-38 (cf. p. 32).

[**49**] Esling, Philippe et Agon, Carlos. **2012b**, « Time-series data mining », in : *ACM Computing Surveys* 45.*1*, p. 1-34 (cf. p. 58).

[**50**] Melsom, Arne, Counillon, François, LaCasce, Joseph Henry et Bertino, Laurent. **2012c**, « Forecasting search areas using ensemble ocean circulation modeling », in : *Ocean Dynamics* 62.*8*, p. 1245-1257 (cf. p. 120).

[**51**] Sakov, Pavel, Counillon, F, Bertino, L, Lisæter, KA, Oke, PR et Korablev, A. **2012d**, « TOPAZ4 : an ocean-sea ice data assimilation system for the North Atlantic and Arctic », in : *Ocean Science* 8.*4*, p. 633-656 (cf. p. 120).

[**52**] Ordonez, Celestino, Martinez, J, Rodriguez-Pérez, JR et Reyes, Andria. **2011b**, « Detection of outliers in GPS measurements by using functional-data analysis », in : *Journal of Surveying Engineering* 137.*4*, p. 150-155 (cf. p. 39).

[**53**] Rajaraman, Anand et Ullman, Jeffrey David. **2011c**, *Mining of massive datasets*, Cambridge University Press (cf. p. 35).

[**54**] Tran, Le Hung, Nguyen, Quoc Viet Hung, Do, Ngoc Hoan et Yan, Zhixian. **2011d**, *Robust and hierarchical stop discovery in sparse and diverse trajectories*, rapp. tech. (cf. p. 40).

[**55**] Yan, Zhixian, Chakraborty, Dipanjan, Parent, Christine, Spaccapietra, Stefano et Aberer, Karl. **2011e**, « SeMiTri : a framework for semantic annotation of heterogeneous trajectories », in : *Proc. of EDBT*, p. 259-270 (cf. p. 33).

[**56**] Choi, Seung-Seok, Cha, Sung-Hyuk et Tappert, Charles C. **2010a**, « A survey of binary similarity and distance measures », in : *Journal of systemics, cybernetics and informatics* 8.*1*, p. 43-48 (cf. p. 55).

[**57**] Girres, Jean-François et Touya, Guillaume. **2010b**, « Quality Assessment of the French OpenStreetMap Dataset », in : *Trans. GIS* 14.*4*, p. 435-459, doi : 10.1111/j.1467-9671.2010.01203.x, url : https://doi.org/10.1111/j.1467-9671.2010.01203.x (cf. p. 43).

[**58**] Mascret, A. **2010c**, « D'eveloppement d'une approche SIG pour l'int'egration de donn'ees Terre/Mer », thèse de doct., 'ecole Nationale Sup'erieure d'Arts et M'etiers (cf. p. 76).

[**59**] Browne, Paul. **2009a**, *JBoss Drools business rules*, Packt Publishing Ltd (cf. p. 44).

[**60**] Guarino, Nicola, Oberle, Daniel et Staab, Steffen. **2009b**, « What Is an Ontology ? », in : *Handbook on Ontologies*, p. 1-17 (cf. p. 53).

[**61**] Knight, Nathan L et Wang, Jinling. **2009c**, « A comparison of outlier detection procedures and robust estimation methods in GPS positioning », in : *The Journal of Navigation* 62.*4*, p. 699-709 (cf. p. 39).

[**62**] Weiszfeld, Endre et Plastria, Frank. **2009d**, « On the point for which the sum of the distances to n given points is minimum », in : *Annals of Operations Research* 167.*1*, p. 7-41 (cf. p. 42).

[**63**] d'Amato, Claudia, Staab, Steffen et Fanizzi, Nicola. **2008a**, « On the influence of description logics ontologies on conceptual similarity », in : *International Conference on Knowledge Engineering and Knowledge Management*, Springer, p. 48-63 (cf. p. 55).

[**64**] Ding, Bolin, Yu, Jeffrey Xu et Qin, Lu. **2008b**, « Finding Time-dependent Shortest Paths over Large Graphs », in : *Proceedings of the 11th International Conference on Extending Database Technology : Advances in Database Technology*, EDBT '08, Nantes, France : ACM, p. 205-216, isbn : 978-1-59593-926-5, doi : 10.1145/1353343.1353371, url : http://doi.acm.org/10.1145/1353343.1353371 (cf. p. 119).

[**65**] Giannotti, F. et Pedreschi, D. **2008c**, *Mobility, Data Mining and Privacy : Geographic Knowledge Discovery*, Springer Publishing Company, Incorporated, isbn : 9783540751762 (cf. p. 31).

[**66**] Mustière, Sébastien et Devogele, Thomas. **2008d**, « Matching networks with different levels of detail », in : *GeoInformatica* 12.*4*, p. 435-453 (cf. p. 43).

[**67**] Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. et Smouse, P.E. **2008e**, « A Movement Ecology Paradigm for Unifying Organismal Movement Research », in : *Proceedings of the National Academy of Sciences* 105.*49*, p. 19052-19059 (cf. p. 27).

[**68**] Palma, Andrey Luis Tietbohl. **2008f**, « A Clustering-Based Approach for Discovering Interesting Places in Trajectories », mém. de mast., Universidade Federal Do Rio Grande Do Sul (cf. p. 33).

[**69**] Spaccapietra, S., Parent, C., Damiani, M.L., de Macedo, J.A., Porto, F. et Vangenot, C. **2008g**, « A Conceptual View on Trajectories », in : *Data & Knowledge Engineering* 65.*1*, p. 126-146 (cf. p. 31).

[**70**] Alvares, Luis Otavio, Bogorny, Vania, Kuijpers, Bart, de Macedo, Jose Antonio Fernandes, Moelans, Bart et Vaisman, Alejandro. **2007a**, « A model for enriching trajectories with semantic geographical information », in : *GIS '07 : Proceedings of the 15th annual ACM international symposium on Advances in geographic*

information systems, Seattle, Washington : ACM, p. 1-8, isbn : 978-1-59593-914-2, doi : http://doi.acm.org/10.1145/1341012.1341041 (cf. p. 33).

[**71**] Bertrand, F., Bouju, A., Claramunt, C., Devogele, T. et Ray, C. **2007b**, « Web and Wireless Geographical Information Systems », in : sous la dir. de Springer Berlin / Heidelberg, t. 4857, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, chap. Web Architecture for Monitoring and Visualizing Mobile Objects in Maritime Contexts, p. 94-105, doi : 10.1007/978-3-540-76925-5_7 (cf. p. 31, 103, 104).

[**72**] Birant, Derya et Kut, Alp. **2007c**, « ST-DBSCAN : An algorithm for clustering spatial–temporal data », in : *Data & knowledge engineering* 60.*1*, p. 208-221 (cf. p. 33, 40).

[**73**] Laube, P., Dennis, T., Forer, P. et Walker, M. **2007d**, « Movement Beyond the Snapshot-Dynamic Analysis of Geospatial Lifelines », in : *Computers, Environment and Urban Systems* 31.*5*, p. 481-501 (cf. p. 31).

[**74**] Quddus, Mohammed A, Ochieng, Washington Y et Noland, Robert B. **2007e**, « Current map-matching algorithms for transport applications : State-of-the art and future research directions », in : *Transportation research part c : Emerging technologies* 15.*5*, p. 312-328 (cf. p. 33).

[**75**] Wu, Y. et Pelot, R. **2007f**, « Geomatics Solutions for Disaster Management », in : sous la dir. de Springer Berlin Heidelberg, Jonathan Li, Sisi Zlatanova et Andrea G. Fabbri, chap. Comparison of Simplifying Line Algorithms for Recreational Boating Trajectory Dedensification, p. 321-334, doi : 10.1007/978-3-540-72108-6 (cf. p. 31).

[**76**] Aronov, B., Har-Peled, S., Knauer, C., Wang, Y. et Wenk, C. **2006a**, « Fr'echet Distance for Curves, Revisited », in : *Algorithms–ESA 2006*, p. 52-63 (cf. p. 76, 77).

[**77**] Ferreira, José TAS et Steel, Mark FJ. **2006b**, « On describing multivariate skewed distributions : a directional approach », in : *Canadian Journal of Statistics* 34.*3*, p. 411-429 (cf. p. 89).

[**78**] Cain III, James W, Krausman, Paul R, Jansen, Brian D et Morgart, John R. **2005a**, « Influence of topography and GPS fix interval on GPS collar performance », in : *Wildlife Society Bulletin* 33.*3*, p. 926-934 (cf. p. 39).

[**79**] Keogh, E. et Ratanamahatana, C.A. **2005b**, « Exact Indexing of Dynamic Time Warping », in : *Knowledge and Information Systems* 7.*3*, p. 358-386 (cf. p. 60).

[**80**] Tongkumchum, Phattrawan. **2005c**, « Two-dimensional box plot », in : *Songklanakarin Journal of Science and Technology* 27.*4*, p. 859-866 (cf. p. 95).

[**81**] Beeri, Catriel, Kanza, Yaron, Safra, Eliyahu et Sagiv, Yehoshua. **2004a**, « Object fusion in geographic information systems », in : *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, p. 816-827 (cf. p. 43).

[**82**] Getz, Wayne M et Wilmers, Christopher C. **2004b**, « A local nearest-neighbor convex-hull construction of home ranges and utilization distributions », in : *Ecography* 27.*4*, p. 489-505 (cf. p. 90).

[**83**] Kim, S.W., Park, S. et Chu, W.W. **2004c**, « Efficient Processing of Similarity Search Under Time Warping in Sequence Databases : An Index-based Approach », in : *Information Systems* 29.*5*, p. 405-420 (cf. p. 60).

[**84**] Bhadury, J., Eiselt, H.A. et Jaramillo, J.H. **2003a**, « An alternating heuristic for medianoid and centroid problems in the plane », in : *Computers & Operations Research* 30.*4*, p. 553-565, issn : 0305-0548, doi : `https://doi.org/10.1016/S0305-0548(02)00024-2`, url : `https://www.sciencedirect.com/science/article/pii/S0305054802000242` (cf. p. 86).

[**85**] Dykes, J.A. et Mountain, D.M. **2003b**, « Seeking Structure in Records of Spatio-Temporal Behaviour : Visualization Issues, Efforts and Applications », in : *Computational Statistics & Data Analysis* 43.*4*, p. 581-603 (cf. p. 31).

[**86**] Halkidi, M., Nguyen, B., Varlamis, Iraklis et Vazirgiannis, M. **2003c**, « THESUS : Organizing Web document collections based on link semantics », in : *The VLDB Journal* 12, p. 320-332 (cf. p. 58).

[**87**] Li, Yuhua, Bandar, Zuhair A et McLean, David. **2003d**, « An approach for measuring semantic similarity between words using multiple information sources », in : *IEEE Transactions on knowledge and data engineering* 15.*4*, p. 871-882 (cf. p. 55).

[**88**] Devogele, Thomas. **2002a**, « A New Merging Process for Data Integration Based on the Discrete Fr'echet Distance », English, in : *Advances in Spatial Data Handling*, sous la dir. de DianneE. Richardson et Peter van Oosterom, Springer Berlin Heidelberg, p. 167-181, isbn : 978-3-642-62859-7, doi : 10.1007/978-3-642-

56094-1_13, url : http://dx.doi.org/10.1007/978-3-642-56094-1_13 (cf. p. 80).

[89] Vlachos, M., Kollios, G. et Gunopulos, D. **2002b**, « Discovering similar multidimensional trajectories », in : *Proceedings 18th International Conference on Data Engineering*, p. 673-684 (cf. p. 60).

[90] Claramunt, Christophe. **juin 2001**, « Vers une intégration de la temporalité dans les SIG », Habilitation à diriger des recherches, Université de Rouen, url : https://hal.science/tel-01211264 (cf. p. 28).

[91] Alt, H. et Guibas, L. J. **2000a**, « Discrete Geometric Shapes : Matching, Interpolation, and Approximation », in : *Handbook of Computational Geometry*, sous la dir. de J.-R. Sack et J. Urrutia, Amsterdam : North-Holland, p. 121-153, isbn : 978-0-44-482537-7, doi : DOI : 10.1016/B978-044482537-7/50004-8, url : http://www.sciencedirect.com/science/article/pii/B9780444825377500048 (cf. p. 76).

[92] Devogele, T. **2000b**, « Mesure d'exactitude et processus de fusion 'a l'aide de la distance de Fr'echet discr'ete », in : *Revue internationale de G'eomatique* 10, p. 359-381 (cf. p. 80).

[93] Rousseeuw, Peter J, Ruts, Ida et Tukey, John W. **1999a**, « The bagplot : a bivariate boxplot », in : *The American Statistician* 53.*4*, p. 382-387 (cf. p. 92).

[94] Thériault, Marius, Claramunt, Christophe et Villeneuve, Paul Y. **1999b**, « A spatio-temporal taxonomy for the representation of spatial set behaviours », English, in : *Spatio-Temporal Database Management*, sous la dir. de MichaelH. Böhlen, ChristianS. Jensen et MichelO. Scholl, t. 1678, Lecture Notes in Computer Science, Springer, Springer Berlin Heidelberg, p. 1-18, isbn : 978-3-540-66401-7, doi : 10.1007/3-540-48344-6_1, url : http://dx.doi.org/10.1007/3-540-48344-6_1 (cf. p. 90).

[95] Claramunt, Christophe, Parent, Christine et Thériault, Marius. **1998a**, « Design patterns for spatio-temporal processes », in : *Data Mining and Reverse Engineering*, Springer, p. 455-475 (cf. p. 31).

[96] Leacock, Claudia et Chodorow, Martin. **1998b**, « Combining local context and WordNet similarity for word sense identification », in : *WordNet : An electronic lexical database* 49.*2*, p. 265-283 (cf. p. 55).

[**97**] Lin, Dekang. **1998c**, « An information-theoretic definition of similarity. », in : *International Conference on Machine Learning*, t. 98, 1998, p. 296-304 (cf. p. 55).

[**98**] Alt, H. et Guibas, L.J. **1996a**, *Discrete Geometric Shapes : Matching, Interpolation, and Approximation. A survey.* Rapp. tech., Handbook of Computational Geometry (cf. p. 76).

[**99**] Ester, Martin, Kriegel, Hans-Peter, Sander, Jørg et Xu, Xiaowei. **1996b**, « A density-based algorithm for discovering clusters in large spatial databases with noise. », in : *Kdd*, t. 96, 34, p. 226-231 (cf. p. 40).

[**100**] Ester, Martin, Kriegel, Hans-Peter, Sander, Jörg et Xu, Xiaowei. **1996c**, « A density-based algorithm for discovering clusters in large spatial databases with noise. », in : *Kdd*, t. 96, 34, p. 226-231 (cf. p. 49).

[**101**] Alt, H. et Godau, M. **1995a**, « Computing the Fr'echet Distance Between two Polygonal Curves », in : *International Journal of Computational Geometry and Applications* 5.*1*, p. 75-91 (cf. p. 77).

[**102**] Bouchon-Meunier, B. **1995b**, *La logique floue et ses applications*, Addison-Wesley France, isbn : 2879080738 (cf. p. 113).

[**103**] Resnik, Philip. **1995c**, « Using information content to evaluate semantic similarity in a taxonomy », in : *arXiv preprint cmp-lg/9511007* (cf. p. 55).

[**104**] Berndt, Donald J et Clifford, James. **1994a**, « Using dynamic time warping to find patterns in time series. », in : *ACM SIGKDD* 10.*16*, p. 359-370 (cf. p. 60).

[**105**] Eiter, T. et Mannila, H. **1994b**, *Computing Discrete Fr'echet Distance*, rapp. tech., Technische Universitat Wien, p. 64 (cf. p. 77).

[**106**] Peuquet, Donna J. **1994c**, « It's about time : A conceptual framework for the representation of temporal dynamics in geographic information systems », in : *Annals of the Association of american Geographers* 84.*3*, p. 441-461 (cf. p. 31).

[**107**] Wu, Zhibiao et Palmer, Martha. **1994d**, « Verb semantics and lexical selection », in : *Association for Computational Linguistics*, p. 133-138 (cf. p. 55, 58).

[**108**] Bouchon-Meunier, Bernadette. **1993**, *La logique floue*, QUE SAIS-JE ?, PUF (cf. p. 62).

[**109**] Alt, H. et Godau, M. **1992a**, « Measuring the Resemblance of Polygonal Curves », in : *Proceedings of the eighth annual symposium on Computational geometry*, ACM, p. 102-109 (cf. p. 77).

[**110**] Goldberg, Kenneth M et Iglewicz, Boris. **1992b**, « Bivariate extensions of the boxplot », in : *Technometrics* 34.*3*, p. 307-320 (cf. p. 93).

[**111**] Small, Christopher G. **1990**, « A survey of multidimensional medians », in : *International Statistical Review/Revue Internationale de Statistique*, p. 263-277 (cf. p. 86).

[**112**] Rada, Roy, Mili, Hafedh, Bicknell, Ellen et Blettner, Maria. **1989**, « Development and application of a metric on semantic nets », in : *IEEE transactions on systems, man, and cybernetics* 19.*1*, p. 17-30 (cf. p. 55).

[**113**] Becketti, Sean et Gould, William. **1987a**, « Rangefinder box plots : A note », in : *The American Statistician* 41.*2*, p. 149-149 (cf. p. 91).

[**114**] Kenward, Robert. **1987b**, *Wildlife radio tagging : equipment, field techniques and data analysis*, Academic Press London (cf. p. 90).

[**115**] Sowa, John F. **1987c**, « Semantic networks », in : *Encyclopedia of Artificial Intelligence* (cf. p. 53).

[**116**] McMaster, R.B. **1986**, « A Statistical Analysis of Mathematical Measures for Linear Simplification », in : *Cartography and Geographic Information Science* 13.*2*, p. 103-116, issn : 1523-0406 (cf. p. 76).

[**117**] Allen, James F. **1983**, « Maintaining knowledge about temporal intervals », in : *Communications of the ACM* 26.*11*, p. 832-843 (cf. p. 27).

[**118**] Lamport, Leslie. **1978a**, « Time, clocks, and the ordering of events in a distributed system », in : *Communications of the ACM*, p. 179-196 (cf. p. 27).

[**119**] Sakoe, H. et Chiba, S. **1978b**, « Dynamic Programming Algorithm Optimization for Spoken Word Recognition », in : *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, p. 43-49 (cf. p. 60).

[**120**] Tukey, J.W. **1977a**, *Exploratory Data Analysis*, Addison-Wesley Series in Behavioral Science - Quantitative Methods, Addison-Wesley (cf. p. 90).

[**121**] Tversky, Amos. **1977b**, « Features of similarity. », in : *Psychological review* 84.*4*, p. 327 (cf. p. 55).

[**122**] Knuth, Donald E. **1975**, « Son of seminumerical algorithms », in : *ACM SIGSAM Bulletin* 9.*4*, p. 10-11 (cf. p. 69).

[**123**] Wagner, Robert A. et Fischer, Michael J. **1974**, « The String-to-String Correction Problem », in : *Journal of the ACM* 21.*1*, p. 168-173, issn : 0004-5411, doi :

10.1145/321796.321811, url : https://doi.org/10.1145/321796.321811 (cf. p. 60).

[**124**] Douglas, D.H. et Peucker, T.K. **1973**, « Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature », in : *Cartographica : The International Journal for Geographic Information and Geovisualization* 10, p. 112-122 (cf. p. 31, 33, 103, 104).

[**125**] Hägerstrand, Torsten. **1970**, « What about people in regional science ? », in : *Papers of the Regional Science Association*, t. 24 (cf. p. 29).

[**126**] Levenshtein, V. I. **1966**, « Binary codes capable of correcting deletions, insertions, and reversals », in : *Soviet physics doklady* 10.*8*, p. 707-710 (cf. p. 59).

[**127**] Zadeh, L.A. **1965**, « Fuzzy Sets », in : *Information and control* 8.*3*, p. 338-353, issn : 0019-9958 (cf. p. 62, 113).

[**128**] Hamming, Richard W. **1950**, « Error detecting and error correcting codes », in : *The Bell system technical journal* 29.*2*, p. 147-160 (cf. p. 59).

[**129**] Mohr, Carl O. **1947**, « Table of Equivalent Populations of North American Small Mammals », English, in : *American Midland Naturalist* 37.*1*, pp. 223-249, issn : 00030031, url : http://www.jstor.org/stable/2421652 (cf. p. 90).

[**130**] Hotelling, Harold. **1933**, « Analysis of a complex of statistical variables into principal components. », in : *Journal of Educational Psychology* 24, p. 498-520 (cf. p. 95).

[**131**] Hausdorff, F. **1918**, « Dimension und äußeres Maß », in : *Mathematische Annalen* 79.*1*, p. 157-179, issn : 0025-5831 (cf. p. 76).

[**132**] Fréchet, M. **1905**, « Sur l'écart de deux courbes et sur les courbes limites », in : *Transactions of the American Mathematical Society* 6.*4*, p. 435-449, issn : 00029947 (cf. p. 77).

[**133**] Jaccard, Paul. **1902**, « Distribution comparée de la flore alpine dans quelques régions des Alpes occidentales et orientales », in : *Bulletin de la Société Vaudoise de Sciences Naturelle 31*, p. 81-92 (cf. p. 55).

[**134**] Pearson, Karl. **1901**, « On lines and planes of closest fit to systems of points in space », in : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.*11*, p. 559-572 (cf. p. 88, 95).

GLOSSAIRE

ACP Analyse en Composantes Principales. 95, 97, 98, 109

AIRSS Arctic Ice Regime Shipping System. 124, 128

AIS Automatic Identification System. 129, 130, 146

CAN Controller Area Network. 36, 37

CED Contextual Edit Distance. 62, 64–67, 70, 71, 140, 141

CONSTAnT Conceptual model of semantic trajectories. 33

DOP Dilution Of Precision. 39, 40

DRDC Direction de la Recherche et du développement pour la Défense du Canada. 125, 135

DTW Dynamic Time Warping. 60, 71, 75, 79, 105–107

ED Edit Distance. 59, 65, 75

EIQ Espace inter-quartile. 91

EMD Enquête Ménages-Déplacements. 72

FFT Fast Fourier Transform. 69

FTH Fuzzy Temporal Hamming distance. 70–73, 140, 141

GNSS Global Navigation Satellite System. 30, 75

GPS Global Positionning System. 30, 41, 42

IGN Institut Géographique National. 43

LCSS Longuest Common Sub-Sequence. 60, 75

LOD Linked Open Data. 53

LRIT Long Range Identification and Tracking. 123

MCP Minimum Convex Polygon. 90, 92

Ontologie Une ontologie est un outil permettant de structurer, représenter et partager des corpus de connaissances rattachés à un domaine, utilisable par un ordinateur et permettant de raisonner par transitivité sur les concepts.. 53

OSBP Oriented Spatial BoxPlot. 5, 95, 97–99, 102, 110, 141

OSM Open Street Map. 43

OSTBP Oriented Spatio-Temporal BoxPlot. 5, 102, 108, 109, 111, 112, 141–143

POI Point Of Interest. 33, 42, 43

POLARIS Polar Operational Limit Assessment Risk Indexing System. 127–131, 133–135

RDF Resource Description Framework. 53

RIO Risk Index Outcome. 128

SCR Système de Coordonnées de Référence. 30

SDE Standard Deviational Ellipse. 88, 93

SMA Systèmes Multi-Agents. 145

TBP Trajectory BoxPlot. 5, 108–112, 114, 115, 141–143

temps absolu t_A faisant référence à un instant temporel absolu exprimé au format Coordinated Universal Time (UTC). 27

temps relatif t_R faisant référence à un instant temporel relatif par rapport à une référence temporelle, c'est le temps écoulé entre deux temps absolus.. 27

trace spatiale Suite temporellement ordonnée de positions discrètes connectées entre elles par des segments de droites. 31

TTFF Time To First Fix. 39

UTC Coordinated Universal Time. 167

WGS84 World Geodesic System 1984. 30

List of Abbreviations and Symbols

a^+ estampille temporelle de fin de l'activité a. 28

a^- estampille temporelle de début de l'activité a. 28

c coordonnée $c = (x, y, z)$. 30, 31, 41

d_E Distance Euclidienne (d_E). 79

d_F Distance de Fréchet (d_F). 77

d_{Fd} Distance de Fréchet discrète (d_{Fd}). 77

d_{Fdm} Distance de Fréchet discrète moyenne (d_{Fdm}). 80

d_H Distance de Hausdorff (d_H). 76

D_{max}^{POI} Distance maximale de recherche d'un POI D_{max}^{POI}. 43

D_{min}^{arret} Durée minimale d'un arrêt D_{min}^{arret}. 41

D_{min}^{dep} Durée minimale d'un déplacement D_{min}^{dep}. 41

DSM distance spatiale maximale (DSM). 112, 114, 115

DSm distance spatiale moyenne (DSm). 112

δSm variation spatiale moyenne (δSm). 112

DTM distance temporelle maximale (DTM). 112

DTm distance temporelle moyenne (DTm). 112

δTm variation temporelle moyenne (δTm). 112

δ durée d'un intervalle temporel $\delta = a^+ - a^-$. 28

$FTH\Delta$ Fuzzy Temporal Hamming Distance pondérée par la durée ($FTH\Delta$). 70, 72

$FTH\gamma$ Fuzzy Temporal Hamming Distance non pondérée par la durée ($FTH\gamma$). 70, 72

$[a^-, a^+]$ intervalle temporel de l'activité A. 28

LCA Plus petit ancêtre commun entre deux concept d'une taxonomie. Least Common Ancestor (LCA). 58

MD Matrice de distances (MD). 79

MF Matrice de Fréchet (MF). 79

$MinDur$ Durée minimale d'un cluster de positions TrajDBSCAN $MinDur$. 40

$MinPts$ Nombre minimal de positions formant un cluster DBSCAN $MinPts$. 40

p position $p = (x, y, z, t)$. 30, 41, 75

ϵ_d Seuil de distance entre positions utilisé pour agréger une position au sein d'un même cluster ϵ_d. 40

ϵ_t Seuil de temps maximum entre positions utilisé pour agréger une position au sein d'un même cluster ϵ_t. 40

ϵ_v Seuil de vitesse utilisé pour considérer un objet comme immobile ϵ_v. 39

sim_{wup} Mesure de similarité de Wu-Palmer sim_{wup}. 58, 63

t estampille temporelle t. 31, 41

DOCTORAT SCIENCES BRETAGNE DE LA MER LOIRE ET DU LITTORAL

UBO Université de Bretagne Occidentale

Titre : Fouille de trajectoires, des données aux connaissances

Mot clés : Trajectoires, mobilité, fouille de données volumineuses

Résumé : La communauté scientifique s'intéressant à l'étude des déplacements est une communauté riche et multi-disciplinaire étudiant ce thème de recherche sous de nombreux prismes. Grâce à l'essor des objets connectés et au développement des technologies de géo-localisation, nous disposons aujourd'hui de masses de données décrivant, de manière plus ou moins détaillée, nos activités quotidiennes. La combinaison et la fusion de ces données permettent d'enrichir la représentation des mouvements réalisés en y ajoutant des descriptions sémantiques porteuses de sens en lien avec la thématique d'étude. Cependant, la multiplicité des attributs décrivant ces déplacements et leur hétérogénéité rend leur analyse particulièrement complexe. Pour extraire des connaissances de ces corpus, il est nécessaire de proposer de nouvelles mesures permettant de comparer la mobilité selon différents axes (spatial/temporel/sémantique). A l'aide de ces mesures de similarité, il est alors possible de constituer des ensembles de trajectoires (clusters) et d'en déduire des comportements types à l'aide de techniques de fouille de données. Enfin le contexte dans lequel les personnes évoluent impacte leur mobilité. Couplées aux données de mobilité, les données environnementales viennent enrichir et décrire le contexte des trajectoires réalisées. L'impact du contexte environnemental sur la mobilité peut alors être pris en compte pour la planification et l'optimisation des déplacements.

Title: Trajectory mining, from data to knowledge

Keywords: Trajectories, mobility, data mining, big data

Abstract: Mobility analysis is an active research area covered by a rich and multi-disciplinary community. The rise of the Internet of Things (IoT) and the wide spreading of geo-location services and connected devices produce an enormous amount of data describing, more or less precisely, our daily activities. Once these data are fused together, users trips can be described using semantic attributes. However the analysis of such heterogeneous dataset are complicated. New measures must be designed to take into account the different trajectories features (spatial/temporal/semantic). Using these similarity measures, cluster of trajectories (clusters) can be defined and typical behaviors can be mined. Finally, environmental data is uselfull to describes the context of theses trajectories. The impact of the environmental context on mobility can then be taken into account for travel planning and optimization.