HAL
archives-ouvertes.fr

Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime.

Laurent Etienne

▶ **To cite this version:**

Laurent Etienne. Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime.. Modeling and Simulation. Université de Bretagne occidentale - Brest, 2011. French. <tel-00667953>

HAL Id: tel-00667953

https://tel.archives-ouvertes.fr/tel-00667953

Submitted on 8 Feb 2012

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

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



THÈSE / UNIVERSITÉ DE BRETAGNE OCCIDENTALE
sous le sceau de l'Université européenne de Bretagne

pour obtenir le titre de
DOCTEUR DE L'UNIVERSITÉ DE BRETAGNE OCCIDENTALE
Mention : géomatique
École Doctorale des Sciences de la Mer

présentée par

Laurent ETIENNE

Préparée à l'Institut de Recherche de l'École Navale (EA 3634)

Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime.

Thèse soutenue le jeudi 08 décembre 2011
devant le jury composé de :

Alain BOUJU, Maître de conférences HDR
Université de La Rochelle / *Directeur de thèse*

Christophe CLARAMUNT, Professeur des universités
Institut de Recherche de l'École Navale / *Examinateur*

Thomas DEVOGELE, Professeur des universités
Université de Tours / *Co-Directeur de thèse*

Hervé MARTIN, Professeur des universités
Université de Grenoble / *Rapporteur*

Pascal PONCELET, Professeur des universités
Université de Montpellier / *Examinateur*

Jacques TISSEAU, Professeur des universités
Université Européenne de Bretagne / *Président*

Karine ZEITOUNI, Professeur des universités
Université de Versailles / *Rapporteur*

UNIVERSITÉ DE BRETAGNE OCCIDENTALE
ÉCOLE DOCTORALE DES SCIENCES DE LA MER

T H È S E

pour obtenir le titre de

Docteur en Sciences

de l'Université de Bretagne Occidentale

Mention : Géomatique

Présentée et soutenue par

Laurent ETIENNE

Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels.

Application au trafic maritime.

Thèse dirigée par Alain BOUJU

préparée à l'Institut de Recherche de l'École Navale,
Groupe Systèmes d'Information Géographique (EA 3634)

soutenue le 08 décembre 2011

Jury :

Président :	Pr. Jacques TISSEAU	-	Université Européenne de Bretagne
Rapporteurs :	Pr. Hervé MARTIN	-	Université de Grenoble
	Pr. Karine ZEITOUNI	-	Université de Versailles
Examinateurs :	Pr. Christophe CLARAMUNT	-	École Navale, Lanvéoc-Poulmic
	Pr. Pascal PONCELET	-	Université de Montpellier
Directeur :	Dr. Alain BOUJU	-	Université de La Rochelle
Co-Directeur :	Pr. Thomas DEVOGELE	-	Université de Tours

Remerciements

En premier lieu, je souhaite remercier mes parents qui m'ont permis de réaliser ce long parcours universitaire menant à la thèse de doctorat, à ma sœur pour sa présence rassurante lors des phases difficiles de rédaction de ce mémoire ainsi qu'à mon oncle et ma tante qui m'ont inspiré lors de mes études en informatique.

Cette thèse n'aurait pu aboutir sans les échanges amicaux et fructueux avec mes directeurs de thèse Alain Bouju (Université de La Rochelle) et Thomas Devogele (Université de Tours), je les en remercie grandement.

Je remercie Karine Zeitouni (Université de Versailles) et Hervé Martin (Université de Grenoble) qui m'ont fait l'honneur d'accepter d'être rapporteurs de ma thèse, Jacques Tisseau d'avoir présidé le jury ainsi que Pascal Poncelet pour avoir bien voulu examiner cette thèse.

J'exprime ma gratitude au ministère de la Défense pour avoir financé cette thèse et en particulier à Christophe Claramunt, directeur de l'institut de recherche de l'École Navale, pour ses échanges éclairés sur diverses thématiques de recherche et le métier d'enseignant chercheur ainsi que pour sa participation à mon jury de thèse. Je remercie également Rémy Thibaud, responsable du groupe Systèmes d'Information Géographique (SIG) de l'IRENAV pour son accueil chaleureux au sein du groupe.

Merci à Cyril Ray qui m'a permis de découvrir cette thématique de recherche ainsi qu'aux collègues du groupe SIG (David, Eric, Joseph).

Un grand merci également à l'ensemble du personnel de l'École Navale que j'ai côtoyé avec énormément de plaisir lors de ces trois années de thèse en qualité d'assistant d'enseignement et de recherche. J'adresse mes sincères remerciements à Marie Coz et Christiane Minssen pour avoir relu et corrigé ce mémoire de thèse, aux drôles de dames des secrétariats et bibliothèque (Elise, Magalie, Marie-France, Christine, Josiane)

Je pense également à tous mes collègues enseignants-chercheurs (doctorants et pour certains désormais docteurs) avec lesquels j'ai passé de très bons moments de convivialité, mes collègues de bureau successifs (Thierry, Imad, Géraldine et Mathieu D), ceux du groupe SIG (Dieudonné, Jean-Marie, Mathieu P, Meriam), mes amis presqu'îliens (Ariane, Marie-Laure, Sébastien, Thomas), l'équipe du carré (Emerry, Georges, Pierre-Luc, Jingya, Zhibin)...

Je remercie également Gabriel Ahtune pour la qualité de son travail lors de son stage au sein de l'IRENAV.

Je tiens également à remercier les civils et militaires (SAGA, SEFER...) avec lesquels j'ai travaillé au sein de l'École Navale ainsi que lors de mes périodes de réserve, et tout particulièrement Erwan Alincourt pour m'avoir permis de tisser un lien fort entre les mondes de la recherche et de la défense.

Table des matières

Introduction **1**
L'étude du mouvement . . . 1
Contexte . . . 3
Contexte scientifique : l'informatique décisionnelle . . . 3
Contexte applicatif maritime . . . 7
Problématique et objectifs de recherche . . . 15
Plan de thèse . . . 18

1 Modélisation des déplacements d'objets mobiles **19**
1.1 Modèle conceptuel . . . 20
1.1.1 Représentation d'un objet mobile . . . 21
1.1.2 Modélisation de positions dans l'espace . . . 22
1.1.3 Représentation du mouvement . . . 22
1.1.4 Modélisation des trajectoires . . . 23
1.1.5 Définition d'un itinéraire . . . 24
1.1.6 Spécification du concept de graphe de zones . . . 24
1.1.7 Regroupement de trajectoires similaires . . . 26
1.1.8 Modélisation d'une route-type . . . 27
1.1.9 Synthèse du modèle conceptuel . . . 28
1.2 Présentation du schéma fonctionnel . . . 29
1.2.1 Acquisition de données de positions . . . 31
1.2.2 Fouille de données spatio-temporelles . . . 31
1.2.3 Analyse spatio-temporelle . . . 34
1.2.4 Visualisation . . . 35

2 Concepts de distances et similarité entre trajectoires **37**
2.1 Mesures de similarité basées sur l'analyse de séries temporelles . . . 39
2.1.1 Distance de Minkowski ($L_p - norm$) . . . 39
2.1.2 Déformation temporelle dynamique ("Dynamic Time Warping") 41
2.1.3 Plus longue sous-séquence commune ("Longest common subsequence") . . . 45
2.1.4 Distance d'édition . . . 46
2.1.5 Comparatif des mesures de similarité basées sur l'analyse de séries temporelles . . . 47
2.2 Mesures de similarité basées sur l'analyse de la géométrie . . . 47
2.2.1 Distance moyenne . . . 48

2.2.2 Distance de Hausdorff . . . 49
2.2.3 Distance de Fréchet . . . 50
2.2.4 Distance de Fréchet discrète . . . 51
2.2.5 Distance de Fréchet discrète moyenne . . . 53
2.2.6 Distance de Fréchet discrète partielle . . . 55
2.3 Mesures de similarité appliquées aux trajectoires . . . 58
2.4 Synthèse . . . 61

3 Clusters et motifs de trajectoires 65
3.1 Sélection et extraction d'un groupe homogène de trajectoires . . . 66
3.2 Préparation des données . . . 72
3.2.1 Filtrage des trajectoires erronées . . . 72
3.2.2 Recalage spatial . . . 73
3.2.3 Filtrage de Douglas et Peucker spatio-temporel . . . 74
3.2.4 Échantillonnage d'un groupe homogène de trajectoires . . . 76
3.2.5 Normalisation temporelle d'un groupe homogène de trajectoires 76
3.3 Génération d'un patron spatio-temporel . . . 77
3.3.1 Le concept des boîtes à moustaches . . . 78
3.3.2 Calcul d'une trajectoire-type . . . 80
3.3.3 Calcul d'un couloir spatial . . . 90
3.3.4 Calcul d'un couloir temporel . . . 99
3.3.5 Synthèse d'un couloir spatio-temporel et d'une route-type . . . 100
3.4 Distribution des distances spatiales et temporelles des nuages de positions . . . 103
3.5 Sauvegarde de la route-type . . . 112

4 Qualification spatio-temporelle de trajectoires 113
4.1 Qualification d'une position . . . 114
4.2 Mesures de similarité entre une trajectoire et un patron . . . 116
4.2.1 Similarité spatiale d'une trajectoire comparée à un patron . . . 117
4.2.2 Similarité temporelle de trajectoires comparées à un patron . . . 119
4.3 Indice de similarité spatio-temporelle entre une trajectoire et un patron 120
4.3.1 Définition des ensembles flous et des fonctions d'appartenance 122
4.3.2 Fuzzification des variables linguistiques . . . 126
4.3.3 Règles floues et indice de similarité spatio-temporelle . . . 127
4.3.4 Calcul de l'indice de similarité spatio-temporelle . . . 131

Conclusion 139
Apport de la thèse . . . 139
Perspectives . . . 143
Perspectives à court terme . . . 143

Perspectives applicatives 144
Perspectives à plus long terme 145

Bibliographie **149**

Liste des symboles **167**

A Annexes **173**
A.1 Tableau des différents fréquences de transmission AIS 173
A.2 Tableau des différents types de trames AIS 174
A.3 Protocole de communication de l'AIS 175
A.4 Histogrammes de densité des distances des nuages de positions . . . 176
A.5 Histogrammes de densité des temps des nuages de positions 177
A.6 Histogrammes de densité des caps des nuages de positions 178
A.7 Histogrammes de densité des vitesses des nuages de positions 179
A.8 Route type de l'itinéraire "Brest-Lanvéoc" 180
A.9 Route type de l'itinéraire "Calais-Douvres" 182
A.10 Route type de l'itinéraire "Pas de calais" 184
A.11 Route type 3D de l'itinéraire "Brest-Lanvéoc" 186
A.12 Histogrammes des mesures de similarité 187
A.13 Probabilités cumulées des mesures de similarité 188
A.14 Publications 189

Table des figures

1 Informatique décisionnelle 3
2 Les différentes étapes menant à l'aide à la décision 5
3 Carte du réseau de surveillance du littoral maritime Français 8
4 Opérateur de surveillance du trafic 9
5 Carte de trafic dense dans la Manche 10
6 Console *ECDIS* 11
7 Exemple de visualisation 3D du port de Milford Haven 11
8 Illustration du fonctionnement complémentaire d'un RADAR et d'un *AIS* 12
9 Signaux *AIS* reçus par la station spatiale internationale 15
10 Carte de densité des trajectoires de navires circulant en Europe sur une durée d'un mois (décembre 2010) 16
11 Plan de thèse 18

1.1 Exemple de geospatial lifeline 23
1.2 Positions de navires dans la région de Brest 25
1.3 Graphe de zones de la région de Brest 25
1.4 Exemple de différents groupes homogènes de trajectoires dans la région de Brest 27
1.5 Extrait du schéma UML du modèle conceptuel 28
1.6 Processus d'extraction de connaissances dans des données 29
1.7 Schéma fonctionnel simplifié 30
1.8 Processus d'extraction de motifs à partir de clusters 31
1.9 Schéma fonctionnel détaillé de la phase d'extraction du groupe homogène de trajectoires 32
1.10 Schéma fonctionnel détaillé de la phase de génération des patrons spatio-temporels 33
1.11 Base de données inductive 34
1.12 Schéma fonctionnel détaillé de la phase de qualification spatio-temporelle 35
1.13 Schéma fonctionnel détaillé de la phase de visualisation 36

2.1 Illustration de la déformation temporelle dynamique entre deux lignes L_a et L_b 43
2.2 Pondération de parcours symétrique (min) 45
2.3 Pondération de parcours symétrique (max) 45
2.4 Pondération de parcours asymétrique 45

2.5 Pondération de parcours de Rabiner et Juang 45
2.6 Illustration du calcul de la surface entre deux lignes L_a et L_b 48
2.7 Illustration de la distance de Hausdorff entre deux lignes L_a et L_b . 49
2.8 Distance de Hausdorff entre deux lignes L_a et L_b d'emprise différente 49
2.9 Distance de Hausdorff faible entre deux lignes L_a et L_b sinueuses . . 50
2.10 Exemple de deux polylignes L_a et L_b 52
2.11 Matrices de distances et de Fréchet de deux polylignes L_a et L_b . . . 52
2.12 Parcours des cases dans la matrice de Fréchet 53
2.13 Couples de points appariés de deux polylignes L_a et L_b 55
2.14 Exemple de deux polylignes L_a et L_b d'emprise différente 57
2.15 Matrices de distances et de Fréchet de deux polylignes L_a et L_b d'emprise différente 58

3.1 Schéma fonctionnel de l'étape de préparation des données 66
3.2 Différents groupes homogènes de trajectoires obtenus par application de l'outil $T - Clustering$ du logiciel $M - Atlas$ 67
3.3 Graphe de zones de la région de Brest 69
3.4 Groupe homogène de trajectoires de navires de type "navire à passager" suivant l'itinéraire "Brest Arsenal → Lanvéoc Ecole Navale" . . 71
3.5 Lissage des trajectoires par filtre de Douglas et Peucker spatio-temporel 72
3.6 Suppression des données erronées 73
3.7 Recalage spatial des trajectoires 74
3.8 Illustration de l'algorithme de Douglas et Peucker spatio-temporel . 75
3.9 Carte de densité des trajectoires du GHT (transparence 1%) 77
3.10 Schéma fonctionnel de l'étape de génération du patron spatio-temporel 78
3.11 Exemple de représentation d'une distribution gaussienne sous la forme d'une boite à moustache 79
3.12 Position médiane agrégée d'un nuage de positions homologues de GHT 81
3.13 Trajectoire médiane du GHT de l'itinéraire A-F (Brest → Lanvéoc) 82
3.14 Comparatif des techniques d'appariement de positions de trajectoires 83
3.15 Nuages de positions homologues et trajectoires de référence à différentes itérations de l'algorithme de génération de la trajectoire-type par appariement de Fréchet 85
3.16 Évolution de la position médiane d'un nuage de positions homologues en fonction du nombre d'itérations de l'algorithme de génération de la trajectoire-type par appariement de Fréchet 86
3.17 Impact d'une trajectoire marginale disposant d'un nombre important de positions appariées à un même nuage 88
3.18 Correction de l'impact d'une trajectoire disposant d'un nombre important de positions appariées à un même nuage par agrégation . . . 88

3.19 Comparatif des trajectoires types obtenues par approche moyenne, médiane ou appariement 89
3.20 Carte de densité et trajectoire médiane du Groupe homogène des trajectoires de navires à passagers effectuant l'itinéraire Brest Arsenal → Lanvéoc Ecole Navale 91
3.21 Classement des positions à droite ou à gauche en fonction du cap de la position médiane 92
3.22 Calcul du couloir spatial par la technique de la droite perpendiculaire au cap 94
3.23 Intersection de deux lignes perpendiculaires aux caps de positions de la trajectoire médiane 95
3.24 Bornes du couloir spatial du *GHT* obtenu par analyse statistique des distances entre positions homogènes de nuages et position de la trajectoire médiane obtenue par appariement de Fréchet 97
3.25 Trajectoire type et couloir spatial du *GHT* "Brest-Lanvéoc" 98
3.26 Bornes du couloir spatial du *GHT* obtenu par analyse statistique des distances entre positions homogènes de nuages et position de la trajectoire médiane obtenue par appariement de Fréchet 101
3.27 Représentation 3D du *GHT* 101
3.28 Représentation 3D de la trajectoire médiane et du couloir spatio-temporel 102
3.29 Représentation 3D du couloir spatio-temporel et d'une trajectoire sortant du couloir à droite 103
3.30 Représentation 3D du couloir spatio-temporel et d'une trajectoire prenant du retard 103
3.31 Histogramme de densité des écarts de distances des positions du nuage 60/120 104
3.32 Histogramme de densité des écarts de temps des positions du nuage 60/120 105
3.33 Comparatif des distances moyennes entre les trajectoires du *GHT* et la trajectoire médiane en fonction de la technique d'appariement utilisée 106
3.34 Exemples de distributions ayant un coefficient d'asymétrie non nul . 107
3.35 Exemples de distributions ayant un coefficient d'aplatissement différent de 3 107
3.36 Diagrammes en boîte de trois nuages de distances 108
3.37 Comparatif des courbes d'évolution des valeurs statistiques en fonction des nuages de positions appariées 110
3.38 Comparatif des courbes d'évolution des coefficients d'aplatissement et d'asymétrie en fonction des nuages de positions appariées 111

4.1 Schéma fonctionnel détaillé de l'étape de qualification spatio-temporelle 114
4.2 Appariement d'une trajectoire à une route-type 115
4.3 Illustration d'une trajectoire inhabituelle contenue dans un couloir spatial 116
4.4 Couples des positions appariées d'une trajectoire avec une trajectoire médiane, les distances spatiales normalisées sont exprimées en pourcentage de ΔS 117
4.5 Exemple de trois trajectoires associées à une même trajectoire médiane 119
4.6 Processus détaillé de l'étape d'inférence floue 122
4.7 Différents types de fonctions d'appartenance 123
4.8 Histogramme et courbe de probabilité cumulée de la DSM du GHT 125
4.9 Fonctions d'appartenance associées à la mesure de similarité DSM . 125
4.10 Valeur floue de la DSM d'une trajectoire appariée à une route-type 126
4.11 Exemple de règle floue 127
4.12 Fonctions d'appartenance des indices de similarité 129
4.13 Arbre de décision de la matrice associative floue de SIM_S 130
4.14 Arbre de décision de la matrice associative floue de SIM_T 130
4.15 Exemple d'activation d'une règle floue 131
4.16 Exemple de valeurs des mesures de similarité spatio-temporelle d'une trajectoire comparée à une route-type 132
4.17 Degrés de vérité des branches des arbres de décision de SIM_S et SIM_T 133
4.18 Défuzzification de SIM_S et SIM_T par la méthode du centre de gravité sur les surfaces 134
4.19 Courbe de l'indice de similarité spatiale du GHT 135
4.20 Courbe de l'indice de similarité temporelle du GHT 135
4.21 Courbe de l'indice de similarité spatio-temporelle du GHT 136

A.1 Illustration du fonctionnement du protocole de communication AIS. 175
A.2 Histogrammes de densité des distances du nuage 30/120 de positions homologues du GHT 176
A.3 Histogrammes de densité des distances du nuage 60/120 de positions homologues du GHT 176
A.4 Histogrammes de densité des distances du nuage 90/120 de positions homologues du GHT 176
A.5 Histogrammes de densité des temps du nuage 30/120 de positions homologues du GHT 177
A.6 Histogrammes de densité des temps du nuage 60/120 de positions homologues du GHT 177
A.7 Histogrammes de densité des temps du nuage 90/120 de positions homologues du GHT 177

A.8 Histogrammes de densité des caps du nuage 30/120 de positions homologues du GHT . . . 178
A.9 Histogrammes de densité des caps du nuage 60/120 de positions homologues du GHT . . . 178
A.10 Histogrammes de densité des caps du nuage 90/120 de positions homologues du GHT . . . 178
A.11 Histogrammes de densité des vitesses du nuage 30/120 de positions homologues du GHT . . . 179
A.12 Histogrammes de densité des vitesses du nuage 60/120 de positions homologues du GHT . . . 179
A.13 Histogrammes de densité des vitesses du nuage 90/120 de positions homologues du GHT . . . 179
A.14 Route type de l'itinéraire "Brest-Lanvéoc" (navires à passagers) . . . 180
A.15 Route type de l'itinéraire "Lanvéoc-Brest" (navires à passagers) . . . 180
A.16 Routes types de l'itinéraire "Brest-Lanvéoc" et "Lanvéoc-Brest" (navires à passagers) . . . 181
A.17 Route type de l'itinéraire "Calais-Douvres" (navires à passagers) . . . 182
A.18 Route type de l'itinéraire "Douvres-Calais" (navires à passagers) . . . 182
A.19 Routes types de l'itinéraire "Calais-Douvres" et "Douvres-Calais" (navires à passagers) . . . 183
A.20 Route type de l'itinéraire "Manche-Mer du nord" (tous navires) . . . 184
A.21 Route type de l'itinéraire "Mer du nord-Manche" (tous navires) . . . 184
A.22 Routes types de l'itinéraire "Manche-Mer du nord" et "Mer du nord-Manche" (tous navires) . . . 185
A.23 Représentation 3D de la route type de l'itinéraire "Brest-Lanvéoc" (navires à passagers) . . . 186
A.24 Histogramme des valeurs de la DSM . . . 187
A.25 Histogramme des valeurs de la DTM . . . 187
A.26 Histogramme des valeurs de la DSm . . . 187
A.27 Histogramme des valeurs de la DTm . . . 187
A.28 Histogramme des valeurs de la δSm . . . 187
A.29 Histogramme des valeurs de la δTm . . . 187
A.30 Probabilités cumulées des valeurs de la DSM . . . 188
A.31 Probabilités cumulées des valeurs de la DTM . . . 188
A.32 Probabilités cumulées des valeurs de la DSm . . . 188
A.33 Probabilités cumulées des valeurs de la DTm . . . 188
A.34 Probabilités cumulées des valeurs de la δSm . . . 188
A.35 Probabilités cumulées des valeurs de la δTm . . . 188

Liste des tableaux

2.1 Comparatif des mesures de similarité 64

3.1 Pourcentage de la population contenue dans l'intervalle $\mu \pm n * \sigma$ pour une distribution gaussienne 80
3.2 Comparaison des moyennes des distances spatiales entre le GHT et la trajectoire médiane en fonction des techniques d'appariement . . . 105

4.1 Expressions de différents types de fonctions d'appartenance 123
4.2 Statistiques des différentes mesures de similarité du GHT 124
4.3 Indices statistiques des différentes mesures de similarité du GHT . . 124
4.4 Fonctions d'appartenance des différentes mesures de similarité 126
4.5 Opérateurs de Mamdani 128
4.6 Conclusions des règles floues en fonction des prédicats 129
4.7 Exemple de mesures de similarité et degrés d'appartenance associés . 132
4.8 Degrés de vérité des fonctions d'appartenance de SIM_S et SIM_T . . 133

A.1 Fréquences de transmission des différentes informations AIS. 173
A.2 Liste des différents types de messages AIS. 174

Liste des Algorithmes

1 Distance de déformation temporelle dynamique 43
2 Distance de Fréchet discrète partielle 56
3 Distance de Fréchet discrète semi-partielle 57
4 Extraction d'un groupe homogène de trajectoires 70
5 Filtrage de Douglas et Peucker spatio-temporel 75
6 Calcul d'une trajectoire type par normalisation temporelle 82
7 Calcul d'une trajectoire-type par appariement 87
8 Calcul d'un couloir spatial par calcul de DPC 95
9 Calcul d'un couloir spatial par appariement de positions homologues 98
10 Calcul d'un couloir temporel 100

Introduction

L'étude du mouvement

Le mouvement fait partie de notre quotidien. C'est un composant élémentaire de processus nécessaire à la réalisation de certaines activités. Aujourd'hui, un très grand nombre d'objets mobiles sont équipés de systèmes permettant de transmettre de manière plus ou moins précise des informations concernant leurs mouvements. Ces données peuvent provenir d'une multitude de sources différentes. Ces capteurs sont de plus en plus miniaturisés et leurs coûts de production ont fortement diminué, ce qui a facilité leur intégration dans de nombreux objets mobiles. Tous ces capteurs permettent d'obtenir différents renseignements concernant les mouvements des objets. Ces renseignements sont parfois combinés pour augmenter la précision avec laquelle les mouvements des objets mobiles sont mesurés.

Grâce aux technologies des systèmes de positionnement mondial (Global Navigation Satellite System : $GNSS$) tels que le Global Positioning System (GPS), ou encore Galileo, tous deux composés d'une constellation de satellites, il est possible de connaître de façon précise la position à laquelle un objet se situe à tout moment sur le globe terrestre à condition de recevoir les signaux des satellites.

D'autres technologies se focalisent sur la cinématique de l'objet mobile en mesurant son accélération linéaire sur trois axes orthogonaux (accéléromètre). De même, les gyromètres cherchent à détecter une rotation ou vitesse angulaire de l'objet mobile. Ces capteurs sont parfois combinés au sein de centrales inertielles, elles permettent de calculer en temps réel l'évolution du vecteur vitesse ainsi que de son orientation à partir de ces mesures. Cependant, les accéléromètres et les gyroscopes composant ces centrales sont sensibles à l'environnement et sont soumis à un phénomène de dérive. C'est pourquoi, certaines centrales inertielles actuelles disposent d'un GPS pour recaler leur position. Ce genre de centrale est désormais embarqué dans de nombreux objets tels que les avions, les navires, les sous-marins, les navettes spatiales, les missiles, les manettes de jeu vidéo, les téléphones portables.

De nombreux autres systèmes fonctionnent par triangulation pour déterminer la localisation d'un objet mobile non équipé de capteurs de position. C'est le cas des systèmes de vidéo-surveillance, la triangulation grâce aux ondes des systèmes GSM (Global System for Mobile Communications) ou des réseaux sans fil $WLAN$ (Wireless Local Area Network). Dans d'autres applications, les objets sont équipés de puces $RFID$ (Radio Frequency IDentification) et sont détectés automatiquement lorsqu'ils se trouvent dans une zone spatiale équipée d'un lecteur $RFID$.

L'étude du mouvement est un thème commun à de nombreux domaines tels que l'écologie [Turchin 1998, Nathan *et al.* 2008], la météorologie [Sinha & Mark 2005], la sociologie, les études comportementales, les transports [Mountain & Raper 2001, du Mouza & Rigaux 2005, Gonzalez *et al.* 2008], la surveillance [Morel *et al.* 2008] et les renseignements.

De nombreux exemples d'application existent comme la gestion d'une flotte de véhicules en temps réel, la prédiction des embouteillages, la surveillance du trafic maritime, la prédiction des horaires de transports en commun, le suivi de migration d'animaux (oiseaux).

Tous ces capteurs génèrent un volume important de données à des fréquences parfois très élevées (plusieurs centaines de mesures par seconde et par capteur). Grâce aux récentes évolutions des capacités de stockage et de traitement informatisé, ces flux de données peuvent être stockés dans des bases de données spatio-temporelles ($BDST$) en vue d'une analyse hors ligne.

Cependant, ce stockage de l'historique des données de positions des objets mobiles n'est pas sans poser de problèmes. En effet, les données de positions peuvent provenir de capteurs différents, comporter des incohérences, disposer de caractéristiques différentes (précision des données, fréquence d'échantillonnage). De plus, les requêtes dans des gros volumes de données engendrent des problèmes d'optimisation.

La sauvegarde des données de mouvement ainsi que leur analyse en vue de réaliser des prédictions pose également des problèmes de respect de la vie privée. Certaines études visent à extraire des informations sur les mouvements contenus dans des bases de données sans pour autant exposer la vie privée des personnes. En effet, en fonction des données traitées, des informations stratégiques peuvent être exposées comme, par exemple, la position d'objets sensibles (avions, transports de fonds), certaines zones d'activités confidentielles (zones de pêche).

Le projet de recherche européen "Mobility, Data Mining and Privacy" ($MODAP$) s'intéresse plus particulièrement à la problématique du respect de la vie privée dans l'étude des bases d'objets mobiles. Ce projet fait suite à un précédent projet nommé "Geographic Privacy-aware Knowledge Discovery and Delivery" ($GeoPKDD$) dont l'objectif est de développer des techniques et des systèmes pour extraire des connaissances dans des bases de données de mouvements d'objets mobiles tout en garantissant le respect de la vie privée.

Contexte

Actuellement, les données de positions sont utilisées principalement à des fins de gestion courante et d'affichage de la position des objets mobiles. Grâce à ces données, il est possible de suivre en temps réel l'évolution d'objets mobiles. Le stockage des données historiques offre des perspectives encourageantes en termes d'analyse de trajectoires et d'enrichissement de ces systèmes par des connaissances sur les déplacements et les comportements habituels liés aux usages. L'application des techniques de l'informatique décisionnelle à ce type de données de mouvement est un challenge intéressant. Cet axe de recherche sur le mouvement est d'ailleurs au centre de travaux visant à concentrer les efforts de recherche sur cette thématique au niveau européen [COST 2009]. La fouille de données dans ces $BDST$ offre de nouvelles perspectives concernant l'analyse des trajectoires de ces objets mobiles et la détection de comportements inhabituels.

Contexte scientifique : l'informatique décisionnelle

Les systèmes d'information sont très souvent utilisés pour manipuler et gérer les données des entreprises et ainsi répondre aux besoins quotidiens liés à leur métier. Au fil du temps, ces données gérées par ces bases de données de production représentent des masses de données volumineuses. Ces dernières sont des sources d'information primordiales pour aider les décideurs à définir la stratégie de l'entreprise à l'aide de système d'information d'aide à la décision ($SIAD$). Dans ce cadre, l'informatique décisionnelle (Business Intelligence, BI) désigne l'ensemble des moyens, outils et méthodes employés en vue d'offrir cette aide indispensable aux responsables.

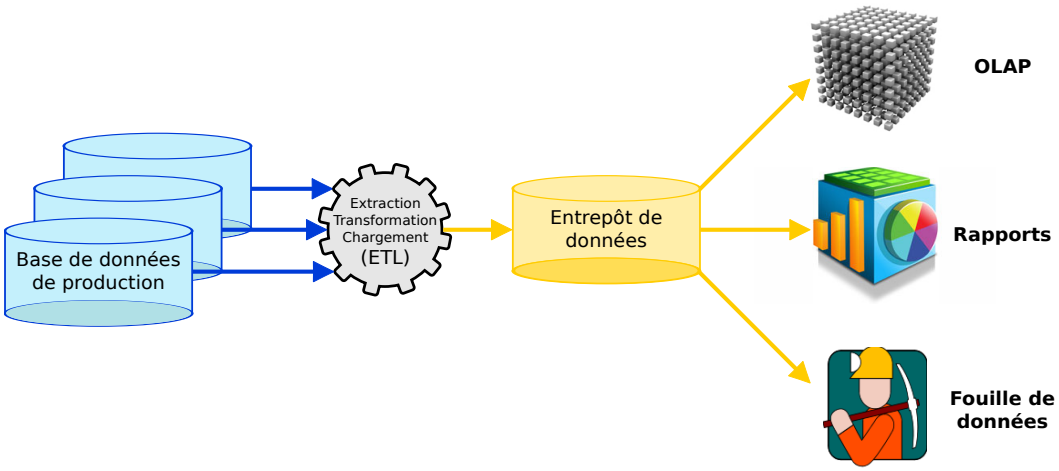


FIGURE 1 – Informatique décisionnelle

Ces derniers disposent alors d'une vue globale de l'activité traitée. La figure 1, retrace les différentes étapes de l'informatique décisionnelle. Les données des bases de production sont extraites, transformées et chargées de manière automatique dans un entrepôt de données généralement par un outil de type ETL (Extract-Transform-Load). L'entrepôt de données (Data Warehouse) est la base de données qui va regrouper l'ensemble des informations consolidées de l'entreprise. Dans l'entrepôt, les données vont s'accumuler et être historisées (ajout de date, de numéro de version) afin de pouvoir suivre leur évolution. Ce principe va à l'inverse de celui des bases de production où les données peuvent être mises à jour, voir supprimées. L'entrepôt de données a pour but de fournir les données de référence utilisées pour la prise de décisions. Ces données sont analysées afin de répondre à des questions du type :

- pourquoi cet évènement est-il arrivé ?
- quels sont les évènements liés dans le temps ?
- quels sont les comportements typiques ou inhabituels ?

Le premier type d'exploitation de l'entrepôt consiste à générer des rapports (reporting). Cette activité permet de sélectionner, trier, regrouper ou répartir ces données selon des critères de choix (période, type de production, type de client, etc.). Elle fournit aussi des résumés des données à l'aide de calculs (totaux, moyennes, écarts, comparatif d'une période à l'autre, etc.). Finalement des graphiques et des tableaux de bord sont produits pour représenter ces résumés suivant différents niveaux de détails selon les besoins des décideurs.

Le deuxième type d'utilisation s'appuie sur les cubes $OLAP$ (Online Analytical Processing). L'analyse en ligne ou l'$OLAP$ est une technique essentielle dans les applications décisionnelles. Elle se base sur la modélisation multidimensionnelle des entrepôts de données. Cette structure facilite l'analyse des données selon différentes dimensions (temporelle, géographique, gamme de produits, etc.) et à différents niveaux d'agrégation. Par exemple pour la dimension temporelle les niveaux d'agrégation peuvent être organisés selon la hiérarchie suivante :
année > trimestre > mois > semaine > jour.

Un langage de requêtes spécifiques aux cubes OLAP nommé "Multidimensional Expressions" (MDX) permet de manipuler les données multidimensionnelles et par exemple synthétiser les informations selon une dimension ou de changer de niveau d'agrégat. Son application aux données spatio-temporelles soulève des problèmes dus à la complexité des dimensions spatiales et temporelles. De plus en plus de travaux portent sur les techniques $OLAP$ pour les données spatiales ($SOLAP$) et/ou spatio-temporelles [Stefanovic *et al.* 2000, Bimonte 2007, Leonardi *et al.* 2010, Gomez *et al.* 2011b, Gomez *et al.* 2011a]. Quelques travaux récents ont traité particulièrement le cas des objets mobiles [Tao *et al.* 2004, Wan & Zeitouni 2005] pour lesquels il est nécessaire d'étendre les modèles $OLAP$

pour le support des applications spatio-temporelles décisionnelles [Bimonte 2007].

Le troisième type d'emploi consiste à extraire des connaissances nouvelles à partir de la masse d'information contenue dans l'entrepôt. L'ensemble des méthodes d'extraction de connaissances non triviales, inconnues jusqu'à maintenant et potentiellement intéressantes est regroupé sous l'expression fouille de données (Data mining ou encore Knowledge Discovery from Data en anglais : KDD). La fouille de données peut être définie comme un "processus non trivial d'extraction d'informations potentiellement utiles, implicites, et inconnues auparavant à partir d'un ensemble de données" [Piatetsky-Shapiro & Frawley 1991]. Ces méthodes sont issues de disciplines scientifiques diverses telles que les statistiques, l'intelligence artificielle et sont souvent utilisées conjointement. Elles facilitent la description des données. Par exemple, les données peuvent être regroupées selon des notions de similarité (clustering).

De même, les comportements fréquents peuvent être représentés à l'aide de motifs encore appelés patron ou pattern en anglais. Ces motifs fréquents sont employés afin d'expliquer le comportement d'objets ou de prévoir les comportements futurs. Des recommandations sont aussi envisageables en fonction des comportements d'objets proches. Inversement, des comportements statistiquement aberrants (inhabituels) sont identifiables. Contrairement à des domaines tels l'analyse de la consommation ou les banques, les systèmes d'information d'aide à la décision sont encore très peu utilisés dans le domaine du suivi d'objet mobile et les outils employés sont mal adaptés à la fouille sur des positions ou des trajectoires d'objets mobiles.

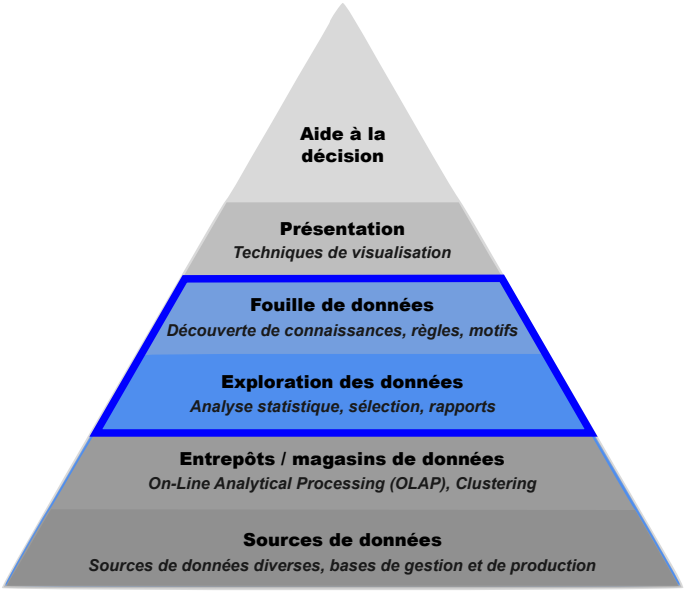


FIGURE 2 – Les différentes étapes menant à l'aide à la décision

La figure 2 illustre le principe de l'informatique décisionnelle ainsi que les étapes traitées dans cette thèse (représentées en bleu). La première étape consiste à récupérer des flux de données provenant de différentes sources (bases de données de gestion et de production). Ces données brutes et d'un volume important ne sont pas forcément adaptées à une analyse. C'est pourquoi elles sont formatées et sauvegardées dans des entrepôts de données et regroupées sous la forme de magasins de données. L'exploration de ces données permet de calculer certaines informations telles que des statistiques et produire des rapports. L'étape de fouille de données a pour objectif d'extraire des connaissances telles que des règles d'association, des motifs. Enfin ces connaissances doivent être restituées à l'utilisateur afin de l'aider dans son processus de prise de décision. La combinaison des règles et motifs extraits de la base de données avec les bases de données de gestion et de production permet d'obtenir une base de données dite inductive facilitant l'aide à la décision.

Dans le domaine de la fouille de données, la recherche de motifs a débuté par l'étude d'éléments (items) associés fréquemment ensemble dans des transactions d'achats. La recherche de règles d'association est une méthode dont le but est de découvrir, dans un ensemble de données, des motifs (pattern) qui se produisent fréquemment, puis de déduire de ces motifs fréquents des règles d'association [Agrawal *et al.* 1993b]. L'objectif de cette recherche est de trouver certaines formes de régularité dans les données. Par exemple, si on considère les achats de produits réalisés par des clients dans un supermarché, on souhaite connaître les produits qui sont souvent achetés en même temps.

Inspiré des techniques d'extraction de règles d'association, certaines études se sont intéressées à l'analyse de motifs séquentiels. Introduits par [Agrawal & Srikant 1995] et plus largement étudiés dans [Masseglia 2002], les motifs séquentiels peuvent être vus comme une extension de la notion de règle d'association, intégrant diverses contraintes temporelles. La recherche de motifs séquentiels consiste à extraire des ensembles d'éléments fréquemment associés sur une période de temps spécifique [Masseglia *et al.* 2004]. Cette recherche se démarque de celle des règles d'association par une mise en évidence de liaisons entre les différentes transactions contrairement aux règles d'associations qui extraient des combinaisons intra-transactions. Une autre différence importante consiste à identifier les individus ayant réalisé les transactions afin de pouvoir analyser leur comportement au cours du temps. Un exemple de règle obtenue par l'analyse d'une séquence temporelle pourrait être : "60% des gens qui achètent une télévision, achètent un magnétoscope dans les deux ans qui suivent".

La présence conjointe des dimensions spatiales et temporelles dans les données de mouvements d'objets mobiles rend ces données complexes à modéliser et à analyser. De nombreuses études s'intéressent plus particulièrement aux mouve-

ments des objets mobiles et plus particulièrement à l'étude de leurs trajectoires. Le suivi d'objets mobiles est utilisé couramment dans de nombreux domaines tels que la migration des animaux [Buard *et al.* 2011], le suivi de phénomènes naturels [Lee *et al.* 2008, Moisuc *et al.* 2005], les mouvements de foules, de piétons [Knorr *et al.* 2000, Viana *et al.* 2011], les déplacements de véhicules (automobiles, aéronefs, navires...) [Wan *et al.* 2007, Kharrat *et al.* 2008]. Toutes ces études ont pour objectif d'extraire des connaissances concernant le mouvement des objets mobiles.

Les mouvements des objets mobiles sont le résultat d'un processus complexe de prise de décision en fonction d'un ensemble d'objectifs et de contraintes liées à l'environnement dans lequel l'objet mobile évolue. [Nathan *et al.* 2008] indique que "le mouvement est une caractéristique fondamentale de la vie, induit par des processus agissant à des échelles spatiales et temporelles multiples". Les données spatio-temporelles obtenues par le biais de capteurs sont le résultat de ces processus. Une problématique intéressante consiste à découvrir et synthétiser ces processus à partir des données spatio-temporelles de mouvements.

Ces mouvements peuvent être impactés par des prises de décisions en fonction de l'environnement de l'objet mobile et du comportement des autres objets situés dans son entourage [Le Pors *et al.* 2009, Le Pors 2010]. De même, la nature de l'environnement dans lequel l'objet mobile évolue peut avoir un impact sur ses mouvements (courants d'air chaud pour les oiseaux, courants marins pour les navires, soleil ou pluie pour les déplacements des piétons).

Considérant que les déplacements des objets mobiles sont optimisés en fonction de différents objectifs à atteindre et de contraintes, ceux-ci peuvent par conséquent être déduits de l'étude du mouvement des objets. Ainsi, on peut citer différentes situations où les mouvements des objets mobiles reflètent le comportement de l'objet. C'est le cas par exemple d'un animal cherchant à fuir un prédateur [Nathan *et al.* 2008], d'un navire réalisant un évitement de collision, d'un piéton faisant ses courses dans un magasin.

Contexte applicatif maritime

Le cadre applicatif de cette thèse s'intéresse plus particulièrement au domaine maritime et à l'analyse de trajectoires de navires dont le trafic ne cesse d'augmenter. L'océan est un espace ouvert couvrant 70% de la surface de la terre. Le transport maritime représente 90% des échanges commerciaux mondiaux soit 7 billions de tonnes de marchandises transportées annuellement. C'est aussi un espace dangereux, difficilement contrôlable où les règles établies sont régulièrement transgressées (abordage, pêche illégale, trafic, pollution, immigration, piraterie, terrorisme).

L'institut de recherche de l'Ecole Navale ($IRENAV$) a établi de nombreux partenariats avec des acteurs du monde maritime. A ce titre, l'institut de recherche dispose d'une volumineuse base de données de positions de navires obtenue à partir de sources variées pendant plus de trois années. De plus, il bénéficie également de la présence d'experts du domaine maritime ainsi que de problématiques spécifiques liées à la surveillance du trafic et du renseignement maritime. Parmi ces partenaires, certains sont chargés de missions particulières liées à la navigation maritime.

Suivi et surveillance du trafic maritime

En France, le suivi, la surveillance et la coordination du trafic maritime est réalisé par les centres régionaux opérationnels de surveillance et de sauvetage ($CROSS$). Au niveau international, il font partie du réseau des centres de coordination de sauvetage maritime institués par la convention "Search And Rescue" (SAR) de "l'International Maritime Organization" (IMO). Les $CROSS$ sont chargés de la régulation et de la sécurité du trafic maritime ainsi que du service d'assistance maritime. Ils doivent également tenir à jour la situation maritime sur leur zone de couverture et transmettre aux navires des renseignements sur la sécurité maritime.

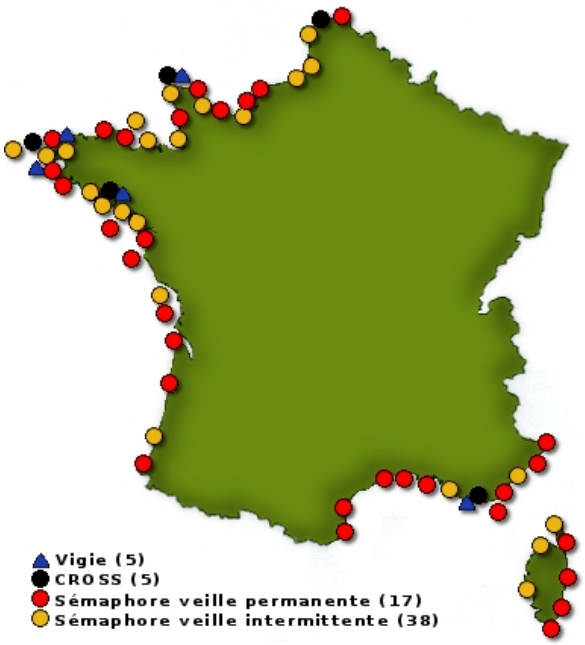


FIGURE 3 – Carte du réseau de surveillance du littoral maritime Français

Les $CROSS$ sont dotés de nombreux moyens de surveillance et de communication au sein d'un réseau de sémaphores et de vigies de la marine nationale complété par des aéronefs et des navires de surveillance et d'intervention. Le réseau français

de surveillance compte 5 *CROSS*, 5 vigies, 17 sémaphores en veille permanente et 38 en veille intermittente représentés sur la carte de la figure 3. L'ensemble est armé par 425 personnes.

Ils renseignent et exploitent différentes sources de données (*RADAR*, optronique, goniométrie, *AIS*, interrogations des navires). Ce réseau est interconnecté au système *SPATIONAV* [Michoux 2007]. Le système *SPATIONAV* permet de centraliser l'ensemble de ces informations pour ensuite les diffuser, de façon totalement automatique à l'ensemble des abonnés au système. Ainsi une image unique de la surveillance des approches maritimes est constituée. Elle doit être la plus exhaustive possible et tenue à jour en permanence.



FIGURE 4 – Opérateur de surveillance du trafic

Cette situation maritime est ensuite visualisée en temps réel sous forme de cartes affichant la dernière position des navires ainsi que les informations associées au navire. Dans certaines zones de navigation, le trafic peut être très dense. Actuellement, le système se contente d'afficher les positions des navires et d'indiquer de potentielle situations de risques de collisions en fonction de la cinématique des navires. Cependant, ce système n'est pas encore doté de module plus complexe d'analyse du comportement des navires.

De plus les opérateurs de surveillance du trafic (figure 4) peuvent avoir jusqu'à 250 navires à surveiller en même temps sur leur carte. Cette situation entraîne une surcharge cognitive leur rendant la tâche de surveillance complexe et fatigante comme l'illustre la figure 5. Cette surcharge cognitive peut entraîner une baisse de la vigilance des opérateurs [IALA 2009] dont les conséquences peuvent être désastreuses (collisions en mer, pollutions, retard dans le déclenchement des opérations de secours, mise en danger de la vie des marins). C'est pourquoi des outils d'aide à la décision basés sur l'analyse du comportement des navires sont intéressants à appliquer à cette problématique complexe.



FIGURE 5 – Carte de trafic dense dans la Manche

L'E-navigation

L'océan est un espace ouvert mais il n'est pas pour autant dénué de dangers. Dans certaines zones côtières, le relief marin comporte des irrégularités pouvant être dangereuses pour la navigation maritime. Ces dangers sont signalées par des cartes et des balises de navigation.

Cet espace est partagé par différents usagers ayant des activités et objectifs distincts (transports de marchandises, transports de passagers, pêche, plaisance, surveillance et protection). Dans certaines zones, le trafic est très dense et les risques pour la navigation sont nombreux (abordages, collisions). C'est pourquoi des règles de navigations internationales sont établies. La convention *SOLAS* (Safety Of Life At Sea) est un traité international visant à définir différentes règles relatives à la sécurité, la sûreté et l'exploitation des navires. Le règlement international pour prévenir les abordages en mer (*RIPAM*) fixe des "règles de priorité" entre les navires. Ces règles se basent sur la direction relative des navires, sur leur capacité de manœuvre et sur leur moyens de propulsion. Le règlement normalise aussi les feux et différentes marques des navires. Dans certaines zones particulières, des règles spécifiques de navigation peuvent être imposées (dispositifs de séparation du trafic, zones d'accès réservé à une certaine catégorie de navire).

Les navires ont besoin d'informations précises concernant l'environnement dans lequel ils naviguent afin d'éviter les accidents. Ils disposent de différents outils. Le marin effectue une veille permanente de son environnement de navigation en surveillant visuellement l'espace aux alentours de son navire. Cette veille est complétée par différents outils d'aide à la navigation. Le marin dispose de cartes marines détaillées qui répertorient les côtes, les dangers, les obstacles immuables ainsi que les profondeurs (cartes bathymétriques).

Ces cartes sont enrichies par des informations qui évoluent au cours du temps provenant d'une multitude de capteurs tels que :

- les sondeurs : pour connaître la hauteur d'eau en fonction de la marée,
- le *RADAR* (RAdio Detection And Ranging), afin de détecter les obstacles en surfaces (les côtes, les balises, les autres navires, les gros objets à la dérive),
- le système *AIS* complète les informations du *RADAR* par un système actif (les autres navires signalent leur positions et caractéristiques).



FIGURE 6 – Console *ECDIS*

Toutes ces informations sont ensuite fusionnées pour être visualisées sur un système de visualisation de cartes électroniques et d'information nommé *ECDIS* (Electronic Chart Display and Information System). Les systèmes *ECDIS* [IMO 1995] ont pour principale fonction de contribuer à renforcer la sécurité de la navigation en faisant apparaître sur un écran tous les renseignements cartographiques nécessaires à la sécurité et à l'efficacité de la navigation. La figure 6 présente une illustration de console *ECDIS* installée à bord d'un navire.

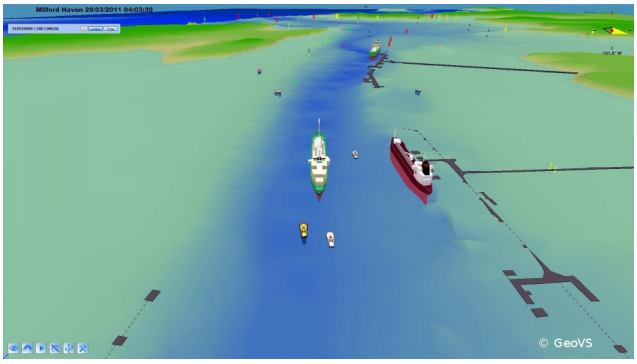


FIGURE 7 – Exemple de visualisation 3D du port de Milford Haven

La visualisation d'informations sur les *ECDIS* est pour le moment limitée à une représentation en deux dimensions. De récents travaux de recherche visent à représenter en 3D temps réel la situation maritime [Goralski & Gold 2008, Ray *et al.* 2011] comme présenté sur la figure 7.

Le système AIS (Automatic Identification System)

Le système AIS (Automatic Identification System) est un transpondeur permettant d'identifier et de localiser les navires. L'AIS utilise un signal radio digital pour diffuser des informations aux autres navires et aux stations côtières sur des fréquences VHF (Very High Frequency) maritimes dédiées au niveau mondial détaillé en annexe A.3. L'AIS ne remplace pas le radar car il est incapable de détecter les côtes et les marques de navigation, mais c'est un complément significatif aux technologies actuelles. Le radar fournit des informations basiques sur les obstacles et l'AIS fournit des informations supplémentaires sur la navigation des bateaux environnants (s'ils sont équipés de transpondeurs AIS).

La figure 8 extraite de [Pettersson 2004] illustre cette complémentarité. L'image $RADAR$ située en bas à gauche renvoie les échos de la côte mais ne signale pas le navire arrivant à contre sens qui est par contre identifié par l'AIS (en bas à droite). Le rôle du système AIS est de transmettre et recevoir des informations nautiques de façon automatisée afin d'éviter les incompréhensions entre navigateurs, d'éviter de surcharger les canaux VHF et améliorer la sécurité en mer. Ces équipements sont de plus en plus répandus. Ils sont employés essentiellement dans le cadre de la sécurité maritime à des fins d'identification, de visualisation temps réel et de détection des risques de collision.

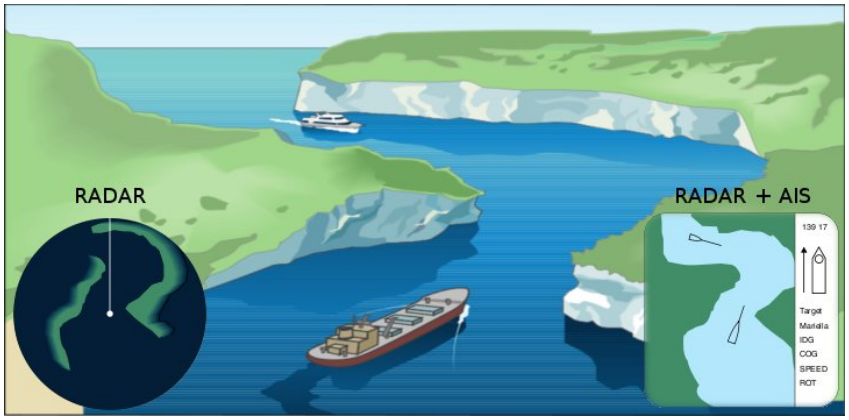


FIGURE 8 – Illustration du fonctionnement complémentaire d'un RADAR et d'un AIS

Le système AIS est composé des différents éléments lui permettant de mesurer et de transmettre des informations concernant la cinématique du navire. Ce système est couplé à un récepteur GPS pour obtenir la position du navire, un émetteur VHF pour transmettre ces données ainsi que d'autres capteurs (gyro-compas, indicateur de taux de giration). L'AIS communique différents types de données concernant le navire (données statiques) et sa cinématique (données dynamiques). Les transpondeurs (ou émetteurs) AIS sont des équipements permettant d'émettre des messages

AIS. Les transpondeurs *AIS* de classe A équipent les navires soumis aux conventions *SOLAS* (Safety Of Life At Sea). Les navires soumis à cette convention sont les cargos de plus de 300 tonnes, les navires effectuant du transport à l'international et les navires affectés au transport de passagers. Pour ce type de navire, un transpondeur *AIS* est obligatoire. Pour les autres types de navires (non soumis aux conventions *SOLAS*), il existe le système *AIS* de classe B qui est moins coûteux. Il n'est pas obligatoire pour ces navires mais participe de façon importante à la sécurité en mer.

L'*AIS* fonctionne de manière continue et autonome par transmission de messages contenus dans 27 types différents de trames *AIS*. Ces 27 trames sont présentées en annexe dans le tableau A.2. Parmi les informations contenues dans ces différentes trames *AIS*, on peut distinguer diverses catégories.

Des données statiques :
- Nom du bateau,
- Type,
- Call Sign (indicatif),
- Numéro unique *IMO* (International Maritime Organization),
- Longueur et largeur,
- Position de l'antenne *GPS* sur le navire.

Des données relatives au trajet du navire :
- Tirant d'eau (partie immergée du bateau),
- Cargaison,
- Destination,
- Temps estimé d'arrivée (*ETA*).

Des données dynamiques :
- Heure de la trame,
- Position du navire,
- Route vraie (Course Over Ground) *COG*,
- Vitesse vraie (Speed Over Ground) *SOG*,
- Cap compas (Heading),
- Taux de rotation (Rate Of Turn) *ROT*,
- Statut de navigation (au mouillage, en cours de navigation).

Des données informatives :
- Alarme,
- Sécurité,
- Informations textuelles.

La fréquence de transmission des trames varie en fonction du type d'information transmises et du statut de navigation du navire comme décrit dans le tableau en annexe. Les informations statiques sont diffusées toutes les 6 minutes ou lorsque les

données sont modifiées. Les informations dynamiques sont diffusées en fonction de la vitesse et des changements de cap du navire (navire au mouillage, en déplacement à petite ou vive allure). Les variations de fréquences de transmission des messages AIS de position influent directement sur la précision de représentation des trajectoires des navires et leur échantillonnage.

Les stations à terre sont également équipées de systèmes AIS disposant de fonctionnalités complémentaires (contrôle et gestion du système, interrogation individuelle des transpondeurs pour obtenir les statuts ...). Ces stations peuvent transmettre des aides à la navigation concernant l'état de la mer, la météo ou bien servir de relai de transmission pour étendre la couverture du système AIS. Ces équipements se trouvent en général dans les phares, les bouées, les centres de surveillances du trafic ($CROSS$), les capitaineries de ports.

En parallèle des transpondeurs, il existe des récepteurs AIS qui captent les messages émis par les transpondeurs de classe A ou B. Ces récepteurs AIS commencent à se développer de plus en plus y compris dans la navigation de plaisance. Ils sont souvent couplés à des systèmes de visualisation fusionnant les informations reçues par le système AIS, le radar et des cartes numériques de la zone de navigation. Ces systèmes de visualisation disposent parfois de capacités limitées d'analyse de la cinématique des navires. Ils peuvent calculer des routes de collision entre les navires et fournir des informations sur les points de passage au plus proche (Closest Point of Approach : CPA). En cas de risque de collision, la distance entre les points de passage au plus proche (Distance at Closest Point of Approach : $DCPA$) est très faible voire nulle. Cette situation est potentiellement dangereuse en fonction du temps restant avant le passage au plus proche (Time to Closest Point of Approach : $TCPA$). Plus le $TCPA$ est faible, plus le risque est imminent. Cela laisse peu de temps pour prendre une décision et corriger le cap du navire. En cas de $DCPA$ nul et de $TCPA$ faible, une alarme averti le navigateur du danger de collision.

Les navires hors couverture AIS peuvent néanmoins transmettre leur position grâce au système $LRIT$ (Long Range Identification and Tracking). Ce système effectue des transmissions AIS de façon moins fréquente via une couverture satellite. Afin d'étendre la portée des systèmes AIS basés à terre, la station spatiale internationale a été récemment équipée de systèmes permettant de récupérer et traiter des signaux AIS. Cette évolution permettra de disposer d'une nouvelle source de données couvrant l'intégralité des océans. La carte de la figure 9 est composée des 90 000 signaux reçus en l'espace de 14 heures entre le 2 et le 3 juin 2010.

En ce qui concerne l'entrepôt de données spatio-temporelles utilisé dans le cadre de l'exemple applicatif maritime de cette thèse, il contient 4 821 447 positions de 1005 navires différents recueillies sur une période de 30 mois dans la rade de Brest.

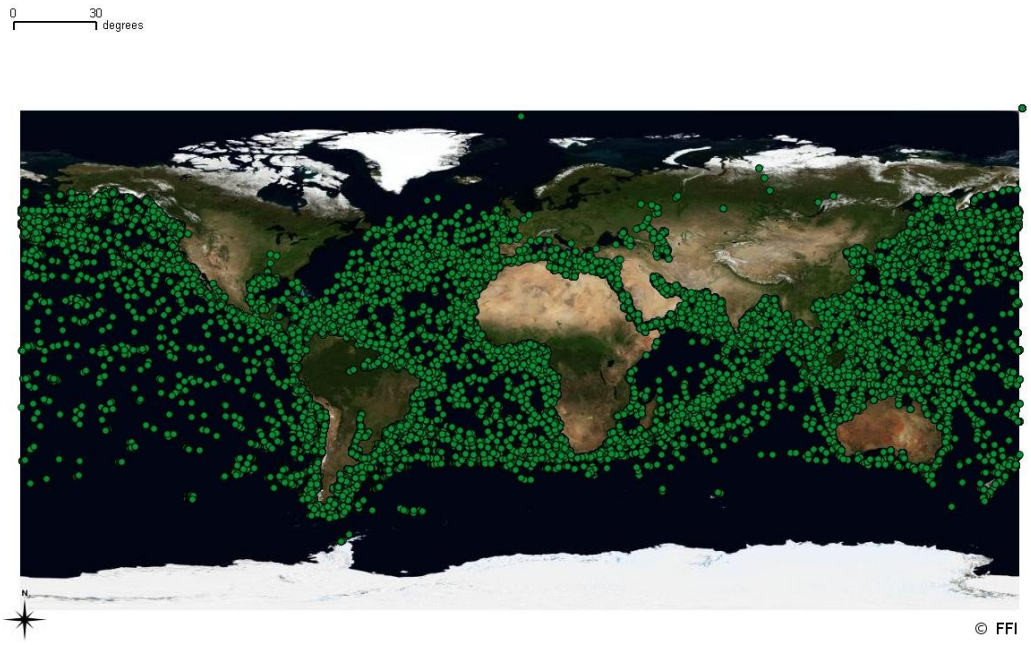


FIGURE 9 – Signaux AIS reçus par la station spatiale internationale

Ces données sont issues de trames AIS et comportent par conséquent toutes les caractéristiques des données statiques, dynamiques et de trajets énoncées dans cette section.

Problématique et objectifs de recherche

La problématique adressée dans cette thèse est multiple. Elle porte principalement sur l'étude du mouvement d'objets mobiles évoluant dans un espace ouvert et la définition de nouveaux outils permettant de qualifier le comportement de ces objets mobiles.

Cette thèse se propose de définir un processus complet d'aide à la décision qualifiant le comportement d'un objet mobile en fonction de son mouvement. Ce processus d'aide à la décision est basé sur une extraction de connaissances dans des bases de données historique de positions. L'analyse de données spatio-temporelles et la définition de processus d'informatique décisionnelle sont au cœur des problématiques adressées dans cette thèse.

En fonction des objectifs des objets mobiles, ceux-ci peuvent avoir des comportements similaires se traduisant par des trajectoires identiques. C'est le cas pour des objets dont l'objectif se résume à se déplacer d'un point A à un point B (migration des oiseaux, trajet domicile-travail, transport de marchandises entre deux ports). En revanche, dans certains cas, des comportements similaires peuvent engendrer des trajectoires différentes (animal qui s'enfuie ou qui chasse, piéton qui flâne, navire de plaisance).

Nous partons du postulat suivant qui servira de base à notre réflexion concernant l'étude du mouvement d'objets mobiles évoluant dans un espace ouvert.

Postulat
Des objets mobiles d'un même type évoluant dans un même espace ont des comportements similaires et suivent une route optimisée en fonction d'un ensemble d'objectifs.

En ce qui concerne les données maritimes, ce postulat peut être illustré par la figure 10 représentant la carte de densité des trajectoires de navires à l'échelle de l'Europe avec un taux de transparence de 1%. Sur cette figure, on peut observer des zones de forte densité en bleu foncé de navigation ou de nombreuses trajectoires similaires se superposent.

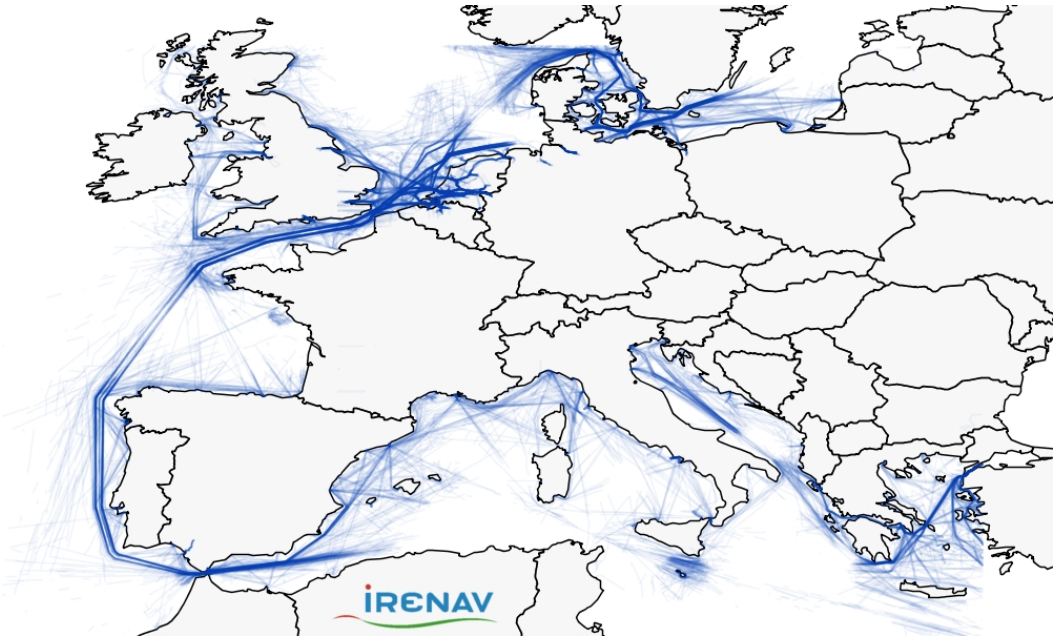


FIGURE 10 – Carte de densité des trajectoires de navires circulant en Europe sur une durée d'un mois (décembre 2010)

Comme nous l'avons indiqué dans la section introduisant l'étude du mouvement, les flux de données de positions sont désormais de plus en plus nombreux. Les progrès de l'informatique permettent le stockage et l'analyse de ces gros volumes de données. Cependant, la première problématique réside dans l'intégration et la modélisation des données de mouvement des objets mobiles dans des entrepôts de données.

Actuellement, les systèmes de gestion sont capables d'afficher sur des cartes les positions des objets mobiles et d'effectuer certains calculs sommaires concernant leur cinématique comme nous l'avons introduit dans la section traitant du contexte maritime. Cependant, pour analyser le mouvement des objets mobiles, les données de positions à elles seules ne sont pas suffisantes. Il est nécessaire de modéliser ce mouvement afin de définir des concepts plus riche tels que les notions de trajectoires, d'itinéraires au sein d'un modèle conceptuel. En fonction du niveau de détail et du modèle choisi pour représenter ce concept, différents types de connaissances peuvent être extraits.

Une fois ce modèle conceptuel défini et les données intégrées au sein d'un entrepôt de données, une seconde problématique apparaît. Cette problématique consiste à extraire de l'entrepôt de données des informations intéressantes concernant les comportements habituels des objets mobiles. Cela implique d'être capable d'analyser le comportement d'un groupe d'objets mobile disposant de mouvements similaires.

Le concept de similarité entre des mouvements est une notion complexe à définir car elle doit prendre en compte à la fois la dimension spatiale et temporelle. De plus, elle peut porter sur la trajectoire de l'objet mobile prise dans son ensemble ou sur une sous-partie de cette trajectoire. En fonction de cette notion de similarité, il est ensuite possible de regrouper les trajectoires des objets ayant des mouvements similaires.

Finalement, à partir de ces groupes de trajectoires similaires, il est intéressant d'extraire des informations concernant le comportement habituel des objets du groupe sous la forme de motifs. Ces motifs, correspondant à la synthèse des mouvements des groupes d'objets mobiles, sont intégrés dans une base de connaissances.

Enfin, une troisième problématique consiste à qualifier le comportement d'un objet mobile en fonction de son mouvement. A partir des connaissances extraites de la base de données (motifs), la première difficulté consiste à comparer le mouvement de l'objet aux motifs de la base de connaissance. Cette comparaison doit permettre de détecter des comportements inhabituels et de restituer cette information à un utilisateur dans un souci d'aide à la décision au sein d'une base de données inductive.

Toutes ces problématiques peuvent être synthétisées sous la forme de questions simples :

Q1 : Comment définir la notion de similarité entre des trajectoires afin de les comparer ?

Q2 : Comment sélectionner et regrouper des trajectoires similaires ?

Q3 : Comment décrire le comportement d'un groupe de trajectoires similaires sous la forme de motifs ?

Q4 : Comment qualifier le comportement d'un objet mobile à l'aide de ces motifs ?

Cette thèse traite les différentes problématiques exposées ci-avant suivant le plan proposé dans la section suivante.

Plan de thèse

Le premier chapitre présente le modèle conceptuel proposé pour caractériser les déplacements d'objets mobiles dans un espace ouvert ainsi que le schéma fonctionnel détaillant les différentes étapes nécessaires à la qualification de trajectoires. Le second chapitre détaille les différentes techniques liées à l'analyse de similarité appliquée à l'étude du mouvement. Le troisième chapitre décrit le processus de génération du patron spatio-temporel synthétisant le comportement d'un groupe homogène de trajectoires. Le quatrième chapitre traite des outils de mesure de similarité entre un patron et une trajectoire. Enfin, la conclusion rappelle les apports de cette thèse et présente quelques perspectives associées à ces travaux de recherche. La figure 11 reprend les différentes étapes de ce plan de thèse.

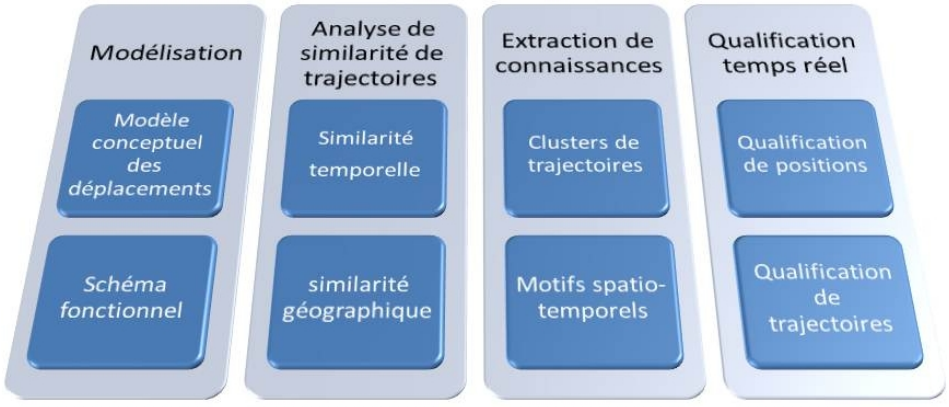


FIGURE 11 – Plan de thèse

Chapitre 1

Modélisation des déplacements d'objets mobiles

Sommaire

1.1 Modèle conceptuel 20
1.1.1 Représentation d'un objet mobile 21
1.1.2 Modélisation de positions dans l'espace 22
1.1.3 Représentation du mouvement 22
1.1.4 Modélisation des trajectoires 23
1.1.5 Définition d'un itinéraire 24
1.1.6 Spécification du concept de graphe de zones 24
1.1.7 Regroupement de trajectoires similaires 26
1.1.8 Modélisation d'une route-type 27
1.1.9 Synthèse du modèle conceptuel 28
1.2 Présentation du schéma fonctionnel 29
1.2.1 Acquisition de données de positions 31
1.2.2 Fouille de données spatio-temporelles 31
1.2.3 Analyse spatio-temporelle 34
1.2.4 Visualisation 35

Les capteurs de positions, en pleine expansion actuellement, fournissent de nombreuses informations concernant les déplacements des objets mobiles. Afin d'analyser ces mouvements et d'en extraire des comportements types, il est nécessaire de modéliser ces informations pour pouvoir les sauvegarder et les exploiter. Les bases de données spatio-temporelles ($BDST$) actuelles permettent de stocker et d'effectuer des requêtes sur des données de positions brutes [Spaccapietra *et al.* 2008]. Ces outils ne permettent cependant pas de représenter et d'analyser un comportement d'objets mobiles. Dans ce chapitre, nous présentons les différents éléments composant le modèle conceptuel permettant de modéliser les objets mobiles, leurs déplacements ainsi que l'espace dans lequel ils évoluent. Nous introduisons également le modèle de patron spatio-temporel synthétisant le comportement de groupes d'ob-

jets mobiles. Finalement, le schéma fonctionnel permettant de qualifier en temps réel un flux de données de positions est présenté.

1.1 Modèle conceptuel

Dans un premier temps, un modèle conceptuel permettant de décrire la façon dont un objet mobile évolue dans l'espace est défini. Le mouvement d'un objet peut être modélisé comme une suite de différents trajets ordonnée temporellement. Chacun de ces trajets occupe un intervalle temporel plus ou moins important dans la durée de vie de l'objet entre sa naissance et sa mort [Turchin 1998].

En fonction de l'application et du niveau d'analyse recherché, différentes granularités peuvent être choisies pour définir le mouvement tant au niveau spatial que temporel. Ces trajets réalisés tout au long de la vie de l'objet peuvent être découpés en différentes séquences temporelles telles que le trajet journalier d'un employé se rendant au travail, le parcours hebdomadaire d'un camion réalisant des livraisons à des clients au sein d'une région ou la migration annuelle des oiseaux cherchant des lieux propices à la reproduction. Chacun de ces trajets est réalisé en vu d'atteindre un objectif et peut être considéré comme un déplacement distinct que nous appellerons ci-après *trajectoire*.

En plus de la modélisation du déplacement des objets mobiles, s'ajoutent potentiellement certaines contraintes spécifiques au domaine étudié. Par exemple, certains oiseaux ne voyagent jamais la nuit. Ou bien, un navire dispose de moteurs ne lui permettant pas d'aller plus vite qu'une certaine vitesse. Le tirant d'eau d'un navire ne lui permet d'emprunter un canal qu'à marée haute.

Un modèle conceptuel de trajectoires est nécessaire afin de pouvoir sauvegarder, manipuler et interroger des bases de données d'objets mobiles [Wolfson 2002, Trajcevski *et al.* 2002a, Parent *et al.* 2006, Macedo *et al.* 2008, Frentzos *et al.* 2008]. Le modèle conceptuel est une représentation d'un ensemble de concepts liés sémantiquement. Ce modèle représente les objets principaux d'un domaine, leurs caractéristiques ainsi que les relations qui s'établissent entre ces différents objets.

Le sujet de recherche lié aux bases de données d'objet mobiles (Moving Objects Databases *MOD*) est un axe relativement vaste dans lequel on retrouve des travaux relatifs à la modélisation et à la sélection de positions d'objets mobiles [Sistla *et al.* 1997, Wolfson *et al.* 1999, Vazirgiannis & Wolfson 2001], des schémas d'indexation visant à optimiser les requêtes et les mises à jour [Tayeb *et al.* 1998, Kollios *et al.* 1999, Agarwal *et al.* 2000, Šaltenis *et al.* 2000], la gestion de l'incertitude dans la localisation des ob-

jets mobiles [Sistla *et al.* 1998, Pfoser & Jensen 1999, Trajcevski *et al.* 2002a, Trajcevski *et al.* 2002b, Cheng *et al.* 2004]. Certaines approches permettent également de prendre en compte le mouvement d'objets mobiles dont la forme évolue au cours du temps tels que des nuages de gaz ou des inondations qui se déplacent et évoluent en même temps [Güting *et al.* 2000]. Dans le reste de cette thèse, nous nous focalisons sur le mouvement d'objets mobiles dont la géométrie n'évolue pas au cours du temps (objet rigide).

1.1.1 Représentation d'un objet mobile

Les *objets mobiles* (OM) sont des entités qui se déplacent dans un environnement et dont la *position* p évolue au cours du temps. Ces objets mobiles peuvent être indifféremment des véhicules, des êtres humains, des animaux, des phénomènes naturels. Chaque objet mobile dispose de caractéristiques qui lui sont propres telles que sa forme géométrique, ses capacités de déplacement ainsi qu'un *type* définissant la catégorie sémantique à laquelle l'objet appartient. En fonction du contexte d'étude, le choix de la représentation géométrique de l'objet peut avoir un impact direct sur l'étude.

Ainsi, la taille de l'objet mobile aura un impact différent sur l'étude en fonction de l'espace dans lequel celui ci évolue ainsi que des autres objets qui s'y trouvent. Si une étude s'intéresse au déplacement d'un objet au sein d'un espace réduit, la taille de l'objet aura un impact direct sur ses capacité de déplacement ainsi que ses interactions possibles avec les autres objets (croisements, évitements, collisions). C'est le cas par exemple pour l'étude de flux de véhicules sur une autoroute ; les camions, les voitures et les motos ont une représentation géométrique différente directement liée à leur taille. Dans d'autres études, l'objet mobile évolue dans un espace ouvert, dans lequel la taille de l'objet mobile peut être considéré comme négligeable. L'étude de la migration des oiseaux par exemple ne nécessite pas de représenter de façon détaillée la géométrie de l'oiseau au vu de son espace d'étude (échelle d'un continent).

En ce qui concerne l'exemple des navires, l'analyse de leur déplacement peut porter sur différents aspects. Un premier niveau d'étude macroscopique s'intéresse au mouvement des navires à l'échelle d'un océan. Un second niveau d'étude microscopique se focalise sur une partie précise du mouvement du navire (par exemple lors d'une phase d'évitement de collision entre deux navires).

Dans cette thèse, nous nous intéressons à l'étude du mouvement d'objets mobiles évoluant dans un espace ouvert, peu contraint, et dont la taille de la géométrie est négligeable comparée à l'espace dans lequel l'objet évolue. C'est pourquoi, dans un objectif de simplification, l'objet mobile est représenté sous la forme d'un *point*.

1.1.2 Modélisation de positions dans l'espace

Un objet mobile évolue dans un *espace*. En fonction du cas d'étude, l'*espace* peut être ouvert et sans contraintes. Il peut aussi être restreint par des obstacles ou différents réseaux tels que des routes, des couloirs.

La *position* p d'un objet mobile est représentée en fonction des repères liés à cet espace et au temps. L'espace dans lequel évolue l'objet mobile peut être géoréférencé (déplacement de l'objet mobile dans un espace géographique tel que les animaux, véhicules, êtres humains, phénomènes météorologiques) ou non-géoréférencés (déplacement d'objet mobile dans un espace non géographique tel que le déplacement d'un crayon sur une feuille de papier, le suivi des mouvements des yeux). La localisation géographique dans l'espace peut être représentée sous la forme d'un *point* $\mathfrak{p}$ dont les coordonnées sont formulées en fonction du repère choisi. Ces coordonnées peuvent être exprimées au format géographique (World Geodetic System 1984 $WGS84$) [Hooijberg *et al.* 1997] sous la forme d'une latitude et d'une longitude $\mathfrak{p} = (\varphi, \lambda)$, dans le repère Cartésien $\mathfrak{p} = (x, y, (z))$ ou dans n'importe quel autre système de coordonnées de référence (SRC) permettant de localiser la position de l'objet dans l'espace.

L'évolution de la position de l'objet mobile est à la fois spatiale et temporelle. De même que pour la représentation des coordonnées spatiales dans l'espace, le temps est également défini par un système de référence. Le *temps absolu* t_A est représenté au format du "temps universel coordonné" (UTC). A partir du *temps absolu*, il est possible de définir la notion de *temps relatif*. Le *temps relatif* t_R correspond au temps écoulé entre deux *temps absolus*. La *position* p d'un objet mobile est représentée par un n-uplet associant un *point* $\mathfrak{p}$ à un *temps absolu* t_A ($p = (\mathfrak{p}, t_A)$).

1.1.3 Représentation du mouvement

Le mouvement d'un *objet mobile* est défini de différentes façons. Comme indiqué dans [Nathan *et al.* 2008], "le mouvement est une caractéristique fondamentale de la vie, induit par des processus agissant à des échelles spatiales et temporelles multiples". L'intérêt de cette thèse se porte sur l'extraction et la synthèse de ces différents processus liés au mouvement défini par [Nathan *et al.* 2008] comme "un changement dans la localisation spatiale d'un individu au cours du temps".

Différents types de mouvement sont définis par [Frank 2001], tels que les évolutions concernant l'objet lui-même (apparition, disparition, fusion et division) ou des modifications relatives à sa position dans l'espace. En ce qui concerne l'aspect temporel, [Turchin 1998] classe ces changements dans 3 catégories différentes (la naissance, la mort et le mouvement).

Certains paramètres représentant le mouvement d'un individu peuvent être mesurés. Ces paramètres sont classés en deux catégories, les paramètres instantanés observables à un instant t (la *position*, la vitesse, l'accélération, la direction...) et les paramètres relatifs mesurables sur des intervalles de temps (vitesse relative, angle de rotation, sinuosité) [Dykes & Mountain 2003, Laube *et al.* 2007, Giannotti & Pedreschi 2008]. L'évolution de ces paramètres au cours du temps permet de définir le mouvement de l'objet mobile (modèle paramétrique).

1.1.4 Modélisation des trajectoires

Plusieurs modèles se focalisent sur l'évolution spatiale et temporelle d'objets mobiles en définissant le concept de *trajectoires*. [Hägerstrand 1970] propose un modèle dans lequel les *trajectoires* sont décrites sous la forme de lignes associées à une activité afin de représenter le déplacement d'une entité. Ce type de représentation a été étendu par [Hornsby & Egenhofer 2002, Thériault *et al.* 2002] au travers des *geospatial lifelines* dont le modèle varie en fonction du niveau de détail recherché. Cette approche définit des périodes de temps pendant lesquelles une entité occupe un *espace*. L'évolution de cette entité pendant cette période est modélisée par une *geospatial lifeline* [Hägerstrand 1970] représentée en bleu sur la figure 1.1.

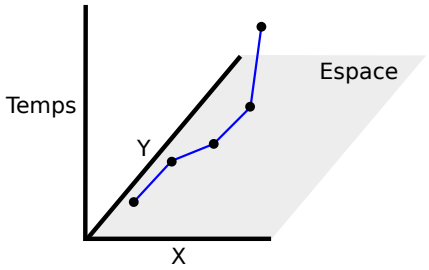


FIGURE 1.1 – Exemple de geospatial lifeline

La *trajectoire* d'un objet mobile peut être considérée comme une *geospatial lifeline* composée d'une suite de *positions* (points noirs sur la figure 1.1) ordonnée temporellement [Laube *et al.* 2007, Spaccapietra *et al.* 2008].

L'acquisition des données de *positions* est généralement réalisée grâce à l'utilisation de capteurs réalisant des mesures de manière discrète à différentes fréquences. Les données obtenues sont alors représentées sous la forme d'une séquence temporelle de *positions*. Cette représentation facilite les tâches d'acquisition, de traitement et de sauvegarde des données de *positions*. Ainsi, dans la suite de cette thèse, une trajectoire T est constituée d'une séquence ordonnée temporellement de positions discrètes formellement définie par $T = (p_1, ..., p_i, ..., p_n)$. On note $p_d = p_1$ la position de départ de la trajectoire et $p_a = p_n$ sa position d'arrivée.

1.1.5 Définition d'un itinéraire

Dans le postulat introduit page 16, nous avons émis l'hypothèse que les objets mobiles de même type et suivant le même *itinéraire* empruntent des *trajectoires* similaires. Afin de pouvoir étudier des *trajectoires* suivant un même *itinéraire*, il est nécessaire de définir plus en détail ce concept. Un *itinéraire* entre deux *zones* spatiales A et B noté ci-après $I_{(A,B)}$ représente le chemin à réaliser pour se rendre de la *zone* A à la *zone* B. C'est pourquoi, il est nécessaire de définir préalablement les différentes *zones* permettant de composer les *itinéraires* que nous souhaitons étudier.

1.1.6 Spécification du concept de graphe de zones

A la différence des objets mobiles se déplaçant sur un réseau, les navires, les avions ou les piétons, peuvent quant à eux se déplacer dans un espace ouvert. Cependant, en ce qui concerne leurs déplacements, un réseau implicite existe. Ce réseau est également lié à l'activité de l'objet mobile et son ensemble d'objectifs comme nous l'avons énoncé dans notre postulat (page 16). De plus, cet espace ouvert peut comporter des obstacles (physiques ou liés à la réglementation) ou des zones d'intérêt formées des destinations des objets mobiles et des lieux de passage obligatoires ou réglementés.

Cette notion de point d'intérêt (Point Of Interest *POI*) est déjà couramment utilisée dans les outils d'aide à la navigation tels que les *GPS* intégrés dans les systèmes de guidage des véhicules. L'objectif à atteindre est une adresse ou un lieu d'intérêt pour l'utilisateur (son logement, son travail). Généralement, les adresses sont représentées sous la forme d'une représentation ponctuelle simplifiée basée sur des coordonnées dans le repère géographique. Cependant, en fonction du contexte, une représentation ponctuelle peut s'avérer insuffisante en fonction de la géométrie de la zone spatiale d'intérêt et de l'interaction entre l'objet mobile et cette zone.

En fonction du contexte étudié (piétons, animaux, véhicules) les zones d'intérêt peuvent être définies manuellement par un opérateur expert ou bien automatiquement par un processus de fouille de données basé sur une analyse de densité [D'Auria *et al.* 2006, Pelekis *et al.* 2007, Giannotti *et al.* 2007, Giannotti & Pedreschi 2008, Giannotti *et al.* 2009, Pelekis *et al.* 2009]. Une zone d'intérêt peut être une zone de passage, d'attente, d'arrêt, de forte densité. Par exemple, pour des piétons, une zone d'intérêt pourrait être l'entrée d'un magasin, un escalier mécanique ou ascenseur, une bouche de métro. Alors que pour des animaux, il s'agirait plutôt de zone de chasse, tanière, point d'eau. Enfin pour notre exemple maritime, nous avons choisi les ports, les zones de mouillage, les zones d'at-

tente, les goulets, les détroits et les dispositifs de séparation du trafic (DST) tels que le rail d'Ouessant comme zones d'intérêt pour former les sommets d'un graphe de zones maritimes.

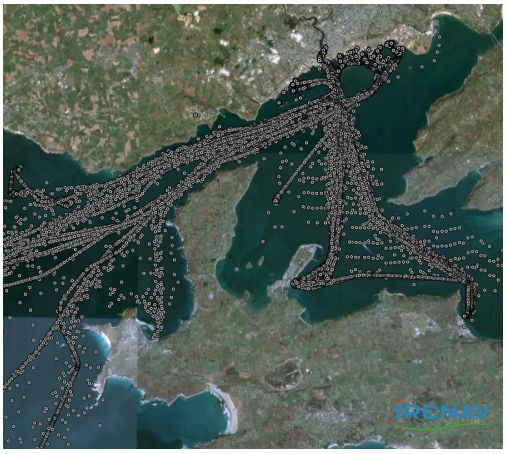


FIGURE 1.2 – Positions de navires dans la région de Brest

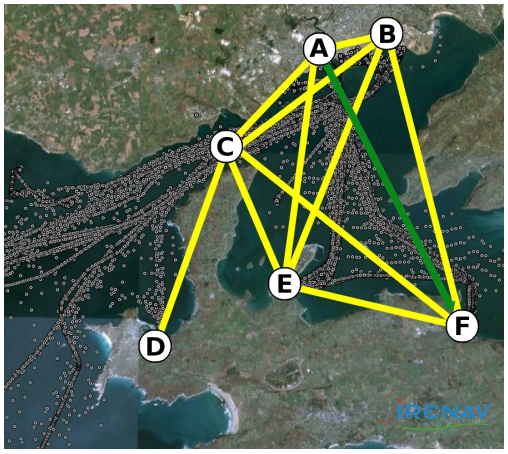


FIGURE 1.3 – Graphe de zones de la région de Brest

Ces zones d'intérêt disposent de limites géographiques permettant de définir leur emprise ainsi que les relations spatiales avec les autres objets mobiles (inclusion, intersection, exclusion...). Ces relations spatiales sont détaillées dans [Randell *et al.* 1992]. L'emprise géographique d'un objet représente la surface spatiale occupée par l'objet ainsi que sa position au sein de l'espace de référence.

Ces zones sont ensuite reliées par des arcs orientés pour former un graphe de zones (G_Z). La figure 1.3 présente le graphe de zones de la région de Brest. Un itinéraire est défini comme un arc reliant deux zones entres elles. Un itinéraire dispose d'une zone de départ Z_D et d'arrivée Z_A.

Une trajectoire T est considérée comme suivant un itinéraire I passant par des sommets du graphe G_Z si elle respecte les conditions suivantes :
Soit un itinéraire $I = (Z_D, Z_A)$
Soit une trajectoire $T = (p_d, ..., p_j, ..., p_a)$
La trajectoire T suit l'itinéraire I si :

$$\forall Z_i \in I, \exists p_j \in T, p_j \subset Z_i \tag{1.1}$$

$$\forall p_j \in T \wedge p_j \subset Z_l, \forall p_k \in T \wedge p_k \subset Z_m, Z_l \underset{I}{<} Z_m \rightarrow p_j < p_k \tag{1.2}$$

$$\forall p_j \in T \wedge p_j \subset Z_i \rightarrow Z_i \in I \tag{1.3}$$

$$p_j \subset Z_D \rightarrow p_j = p_d \tag{1.4}$$

$$p_j \subset Z_A \rightarrow p_j = p_a \tag{1.5}$$

En d'autres termes, pour chaque zone de l'itinéraire I, il existe au moins une position p de la trajectoire T dans cette zone [1.1] respectant la relation d'ordre temporel définie précédemment [1.2]. Aucune position p de la trajectoire T n'appartient à une zone du graphe autre que celle de l'itinéraire [1.3]. Seule la première position p_d de la trajectoire appartient à la zone de départ de l'itinéraire Z_D [1.4]. De même, seule la dernière position p_a de l'itinéraire appartient à la dernière zone (zone d'arrivée) de l'itinéraire Z_A [1.5]. Au vu de ces définitions, il est important de noter que pour toute trajectoire intersectant une zone, une position de cette trajectoire doit être incluse dans la zone. Ainsi, en fonction de l'emprise géographique de la zone, de la vitesse de déplacement de l'objet mobile et de sa fréquence d'émission, il est potentiellement nécessaire de sur-échantillonner les trajectoires ce qui peut être problématique.

La figure 1.3 présente en vert un exemple d'itinéraire entre les sommets A et F du graphe de zones (Brest Arsenal → Lanvéoc Ecole Navale). Le graphe étant incomplet, tous ses sommets ne sont pas reliés directement entre eux par une arête. De plus, certains paramètres extérieurs peuvent impacter le comportement des objets mobiles lors de leur déplacement. Par exemple, des règles de navigation peuvent être imposées afin d'éviter les risques de collision (dispositifs de séparation de trafic).

Certains phénomènes environnementaux peuvent également impacter ce comportement (tels que par exemple, les courants d'air chaud lors de la migration des oiseaux, les marées et courant marins). C'est pourquoi, les itinéraires aller et retour doivent être étudiés de façon distincte. Par conséquent, le graphe de zone proposé est un graphe orienté.

Les multiples points représentés sur la figure 1.2 correspondent aux *positions* enregistrées de navires. Des zones de navigation plus denses sont observables. Elles correspondent aux itinéraires empruntés par les navires. Ce sont ces zones denses que nous souhaitons étudier afin d'extraire des connaissances concernant les mouvements des navires évoluant dans ces zones.

1.1.7 Regroupement de trajectoires similaires

Un processus de fouille de données (data mining) peut être appliqué sur les mouvements des objets mobiles. Ce processus a pour objectif d'extraire des connaissances à partir d'un ensemble de *trajectoires* d'objets mobiles suivant un *itinéraire* commun. Afin de réaliser ce processus de fouille de données, la première étape consiste à extraire de la $BDST$ les trajectoires d'un même type S d'objets mobiles suivant le même itinéraire I.

Cet ensemble correspond à une partie de notre entrepôt de données appelé groupe homogène de trajectoires (GHT) d'objets mobiles de même type suivant le même itinéraire. Les notions de *trajectoires* et d'*itinéraires* ayant été formalisées, différents critères d'extraction de trajectoires sont définis. Les techniques d'extraction du GHT sont présentées au chapitre 3.1. Une fois ces GHT extraits de la base de données, ceux ci peuvent être employés pour étudier le comportement des objets mobiles composant ce GHT. Ce GHT est alors fournis à une seconde étape de fouille de données afin d'en extraire des connaissances liées au comportement habituel des objets de ce groupe suivant le même itinéraire.

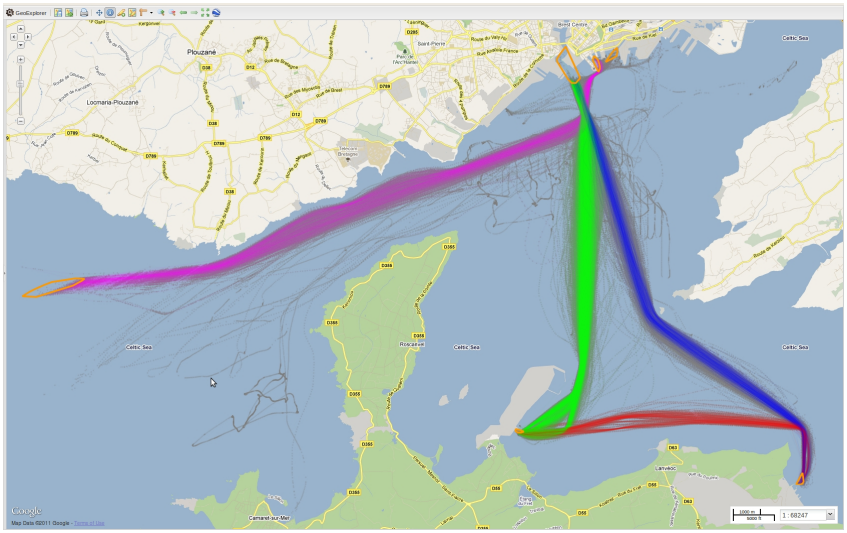


FIGURE 1.4 – Exemple de différents groupes homogènes de trajectoires dans la région de Brest

1.1.8 Modélisation d'une route-type

Disposant de groupes homogènes de trajectoires, nous souhaitons extraire de ces groupes une représentation synthétique du mouvement réalisé par tous les objets mobiles de ce groupe. Cette représentation nommée *route-type* synthétise le comportement d'un *groupe homogène de trajectoires* d'objets mobiles de même *type* suivant le même *itinéraire*. Cette *route-type* est constituée de plusieurs éléments. Comme nous avons pu l'observer sur l'exemple de la figure 1.4, les groupes homogènes de trajectoires disposent d'une emprise spatiale plus ou moins dense formant une sorte de couloir dans lequel les trajectoires du GHT évoluent.

L'objectif de l'étape de fouille de données sur les GHT consiste premièrement à extraire la *trajectoire type* suivie par l'ensemble des trajectoires du GHT ainsi que des informations concernant la dispersion spatiale et temporelle par rapport à

cette trajectoire type. Cette dispersion spatiale et temporelle peut être vue comme un *couloir spatial* et un *couloir temporel* dans lesquels la majorité des trajectoires du *GHT* évoluent. La *route-type* est constituée d'une *trajectoire type* complétée par un *couloir spatial* et un *couloir temporel*. Le processus de génération de cette route type est détaillé au chapitre 3.3.

1.1.9 Synthèse du modèle conceptuel

Les relations du modèle conceptuel conceptuel présenté dans les sections précédentes sont synthétisées dans l'extrait de diagramme *UML* ("Unified Modeling Language") de la figure 1.5. Sur cet extrait de diagramme UML, on retrouve les différentes entités présentées en italique dans les sections précédentes ainsi que les relations entre ces entités et certains de leurs attributs.

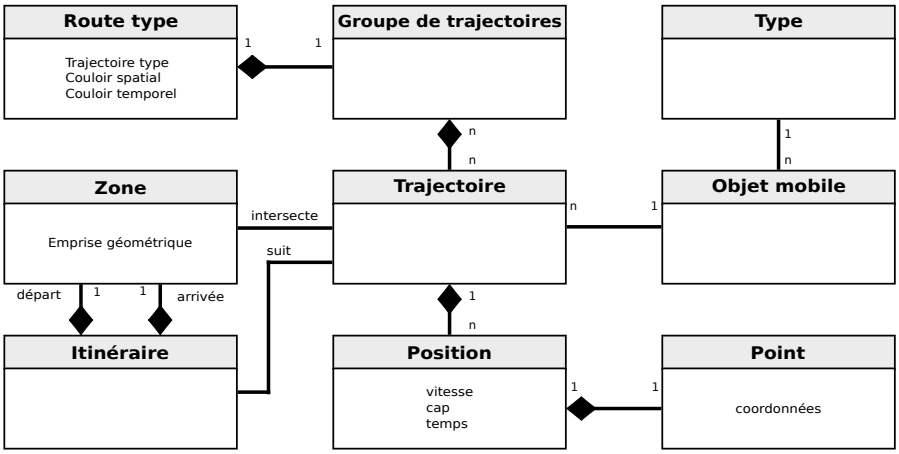


FIGURE 1.5 – Extrait du schéma UML du modèle conceptuel

On peut ainsi noter sur ce diagramme que :
- un *point* dispose d'attributs permettant de définir ses coordonnées ;
- une *position* est composée d'un *point* et d'autres attributs (vitesse, cap, temps relatif, temps absolu au format *UTC*) ;
- une *trajectoire* est composée de *positions* ;
- une *trajectoire* est liée à un *objet mobile* d'un certain *type* ;
- un *itinéraire* est constitué de deux *zones* (départ et arrivée) ;
- une *trajectoire* peut suivre un *itinéraire* ;
- un *groupe de trajectoires* est constitué de *trajectoires* suivant un même *itinéraire* ;
- une *route-type* est constituée à partir d'un *groupe de trajectoires* ;
- une *route-type* est composée d'une *trajectoire type*, d'un *couloir spatial* et d'un *couloir temporel*.

1.2 Présentation du schéma fonctionnel

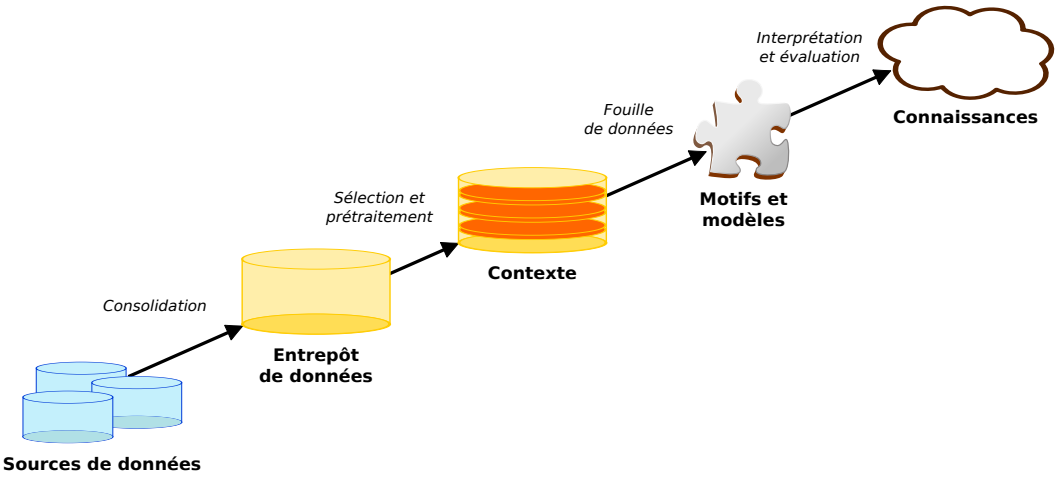


FIGURE 1.6 – Processus d'extraction de connaissances dans des données

A partir du modèle conceptuel proposé dans la section 1.1, les données peuvent être formatées dans un entrepôt de données dans l'objectif d'y découvrir des connaissances cachées à l'aide d'un processus d'extraction de connaissances. Le processus d'extraction de connaissances dans des données (ECD) est un ensemble d'opérations englobant l'extraction de motifs ainsi que des étapes de pré-traitement et de post-traitement.

L'objectif principal de la fouille de données consiste à tirer partie d'une grande quantité de données pour extraire de l'information implicite, pertinente et non triviale. Finalement, cette connaissance doit être extraite et présentée sous une forme compréhensible par l'utilisateur et utilisable pour raisonner. Quatre grandes étapes composent le processus d'extraction de connaissances dans des bases de données. Ces étapes sont présentées sur la figure 1.6.

Premièrement, l'étape de consolidation permet de rassembler et d'unifier dans un cadre conceptuel commun différentes sources de données au sein d'un entrepôt de données. Certains traitement et des mises en formes sont nécessaires pour intégrer les données dans l'entrepôt de données. Ces traitements sont réalisés par l'ETL dont le rôle est d'extraire les données provenant de sources diverses, les transformer pour les représenter dans un modèle unique et les charger dans un entrepôt de données.

La seconde étape consiste à sélectionner dans l'entrepôt de données celles qui concernent plus particulièrement notre problème d'étude. Dans notre cas, il s'agit de l'étude du mouvement d'objets mobiles. Dans certains cas, les données doivent être préparées afin de faciliter l'analyse en réalisant un pré-traitement des données. C'est également lors de cette étape de pré-traitement que les données erronées,

aberrantes ou manquantes sont filtrées.

La troisième étape constitue véritablement le cœur du processus d'extraction de connaissances dans les données. L'objectif de cette étape consiste à extraire des motifs ou des modèles à partir des données. L'espace de recherche étant très vaste, les contraintes spécifiées lors de la seconde étape d'extraction sont particulièrement importantes pour cette phase d'extraction de motifs.

Enfin, la dernière étape s'intéresse à l'interprétation et à l'évaluation des motifs et des modèles retournés par la phase de fouille de données.

Cette section introduit les différentes étapes du schéma fonctionnel présentant la chaîne de traitement des trajectoires en partant de l'acquisition du flux de données de positions jusqu'à la visualisation de trajectoires qualifiées. Ce schéma fonctionnel est composé de plusieurs étapes s'inspirant du modèle de processus d'extraction de connaissances dans des données. Les différentes étapes du schéma fonctionnel sont synthétisées sur la figure 1.7 et sont détaillées dans les sections suivantes.

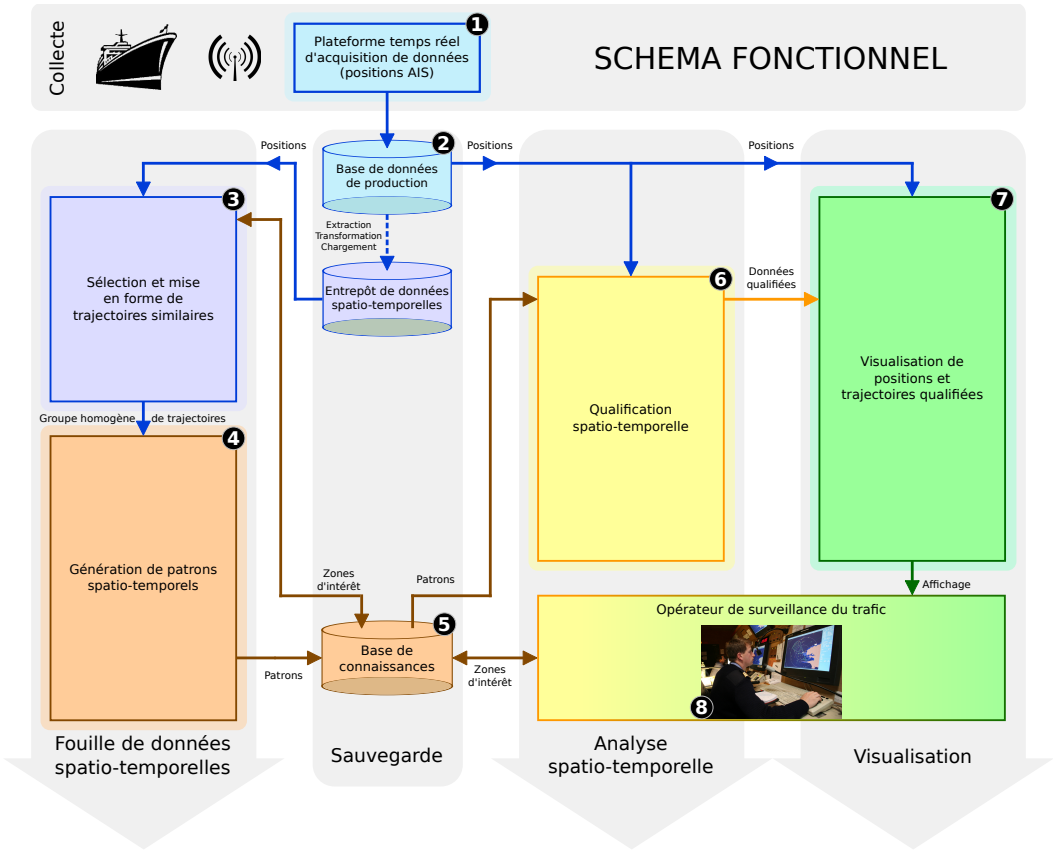


FIGURE 1.7 – Schéma fonctionnel simplifié

1.2.1 Acquisition de données de positions

La première étape du schéma fonctionnel consiste à acquérir un flux de données de positions (figure 1.7 étape 1). Les capteurs placés sur des objets mobiles génèrent une quantité importante de données de position. Celles-ci permettent pour chaque objet identifié de disposer à différent instants de leur localisation et retracer leur trajectoire comme indiqué dans la section 1.1.4.

Le processus d'acquisition des données permet de recevoir en temps réel et de sauvegarder dans des bases de données de production les positions des objets mobiles. Dans notre cas d'étude maritime, le système *AIS* équipant les navires est destiné à la sécurité du trafic maritime, il transmet la position du navire à intervalle variable en fonction de la vitesse et du déplacement du navire [Etienne *et al.* 2009]. Cette phase d'acquisition de données est présentée à l'étape 1 de la figure 1.7.

Ce flux de données est sauvegardé dans une base de données spatio-temporelles (figure 1.7 étape 2). De plus, ces données provenant des bases de données de production sont ensuite formatées en reprenant le modèle conceptuel défini à la section 1.1 afin d'alimenter un entrepôt de données de façon périodique.

Ce processus d'alimentation de l'entrepôt de données est réalisé par le biais d'un *ETL* (Extract Tranform and Load) chargé d'extraire les données de plusieurs sources différentes et de les transformer pour pouvoir les charger dans l'entrepôt de données en respectant le modèle conceptuel défini.

Cet entrepôt de données est spécialement conçu pour faciliter l'analyse de l'historique des données de production. Sur la figure 1.7, l'entrepôt de données est représenté en bleu par la base de données spatio-temporelles.

1.2.2 Fouille de données spatio-temporelles

La fouille de données peut être définie par [Piatetsky-Shapiro & Frawley 1991] comme un "processus non trivial d'extraction d'informations potentiellement utiles, implicites, et inconnues auparavant à partir d'un ensemble de données".

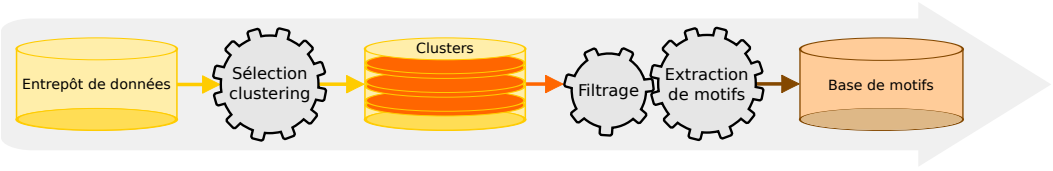


FIGURE 1.8 – Processus d'extraction de motifs à partir de clusters

La figure 1.8 présente en détail l'étape de fouille de données. La fouille de données spatio-temporelles est une étape qui est réalisée hors ligne et de façon périodique.

Cette étape est décomposée en deux phases distinctes. La première phase consiste à sélectionner au sein de l'entrepôt de données des groupes de données similaires (clusters) à analyser. La seconde phase consiste à extraire des connaissances à partir de ces clusters. Ces connaissances sont finalement sauvegardées dans une base de motifs.

1.2.2.1 Etape de sélection d'un groupe homogène de trajectoires

Comme indiqué dans la section 1.2.2, la première étape du processus de fouille de données consiste à sélectionner des clusters dans l'entrepôt de données. Dans le cadre de cette thèse, nous sommes intéressés par l'étude du mouvement des objets mobiles. C'est pourquoi, la sélection de ces clusters est réalisée de manière à obtenir des groupes homogènes de trajectoires. Différentes techniques d'extractions de ces clusters sont proposées au chapitre 3.1.

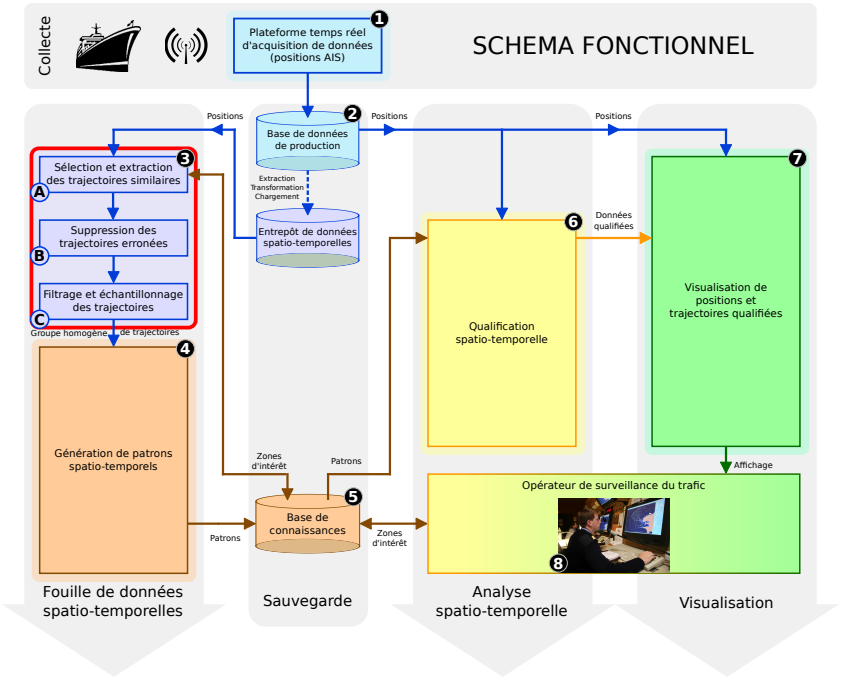


FIGURE 1.9 – Schéma fonctionnel détaillé de la phase d'extraction du groupe homogène de trajectoires

Une première technique est basée sur une analyse de la densité des trajectoires afin de classer les trajectoires dans différents clusters. Cette étape de classification est non-supervisée, le nombre de clusters n'est pas connu. Une autre solution consiste à classer les trajectoires dans des clusters en fonction de méta-connaissances telles que les caractéristiques des objets mobiles ou le graphe de zones et les itinéraires définis dans le modèle conceptuel aux sections 1.1.5 et 1.1.6. A contrario de la

première technique, cette dernière nécessite des connaissances supplémentaires : le graphe de zones. Le nombre de clusters est directement lié aux arcs de ce graphe de zones.

Ainsi, l'extraction des clusters au sein de l'entrepôt de données est déjà en sois une première étape de fouille de données (figure 1.9 étape 3.A). Ce groupe de trajectoires est ensuite épuré des trajectoires erronées comportant des erreurs de transmissions ou des données aberrantes (figure 1.9 étape 3.B). Enfin, chaque trajectoire du groupe est filtrée et ré-échantillonnée afin d'obtenir le groupe homogène de trajectoires d'objets mobiles de même type suivant le même itinéraire (GHT).

1.2.2.2 Etape de génération de patrons spatio-temporels

Une fois des clusters de données extrait de l'entrepôt de données, l'étape suivante consiste à extraire des connaissances synthétisées sous la forme de motifs. Cette étape s'inscrit dans la continuité du processus de fouille de données introduit à la section 1.2.2. A partir de GHT, un processus de génération de patrons spatio-temporels (figure 1.10 étape 4) est réalisé.

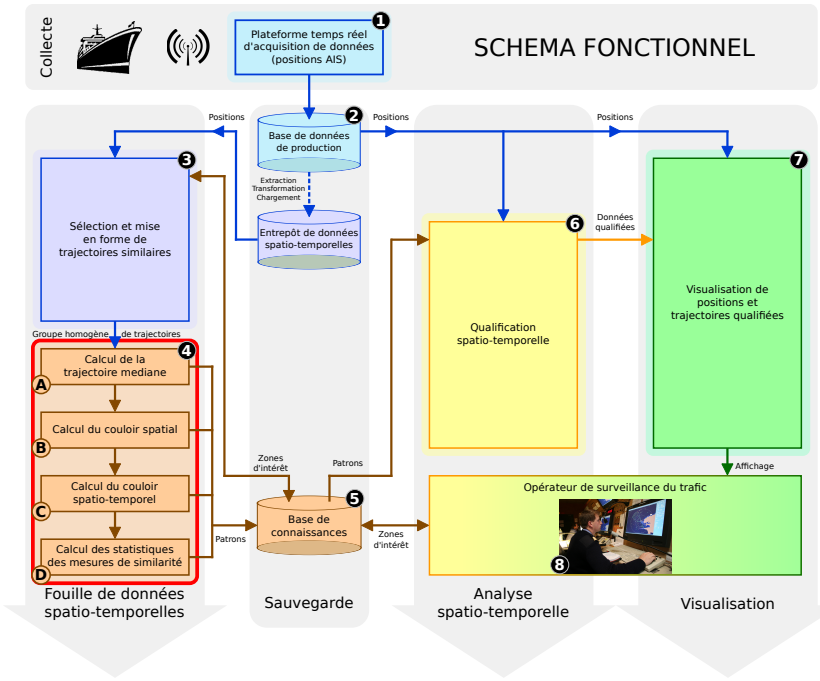


FIGURE 1.10 – Schéma fonctionnel détaillé de la phase de génération des patrons spatio-temporels

Ce processus est décomposé en 4 phases successives permettant d'obtenir un motif, appelé route-type, synthétisant le comportement des objets mobiles du GHT.

En premier lieu, la trajectoire-type représentant le comportement habituel des objets mobiles est calculée (figure 1.10 étape 4.A). Cette trajectoire-type est ensuite complétée par un couloir spatial (figure 1.10 étape 4.B) et un couloir spatio-temporel (figure 1.10 étape 4.C) dans lesquels la majorité des objets mobiles du GHT évoluent.

De plus, afin d'évaluer les motifs obtenus par cette étape de fouille de données, des statistiques sont calculées afin de comparer les trajectoires du GHT avec le patron obtenu. Ces statistiques permettent de connaitre la précision avec laquelle le motif synthétise le comportement des objets mobiles composant le GHT. Les résultats de ce processus sont sauvegardés dans une base de connaissances (figure 1.7 étape 5) contenant entre autre les motifs extraits (figure 1.8). Cette étape de fouille de données est réalisée hors ligne.

1.2.3 Analyse spatio-temporelle

A partir des motifs extrait de l'entrepôt de données, il est désormais intéressant de raisonner à l'aide de motifs afin de pouvoir qualifier le mouvement d'un objet mobile. Le cadre des bases de données inductives (BDI), introduit par [Imielinski & Mannila 1996], propose d'intégrer les motifs directement au sein des bases de données de production.

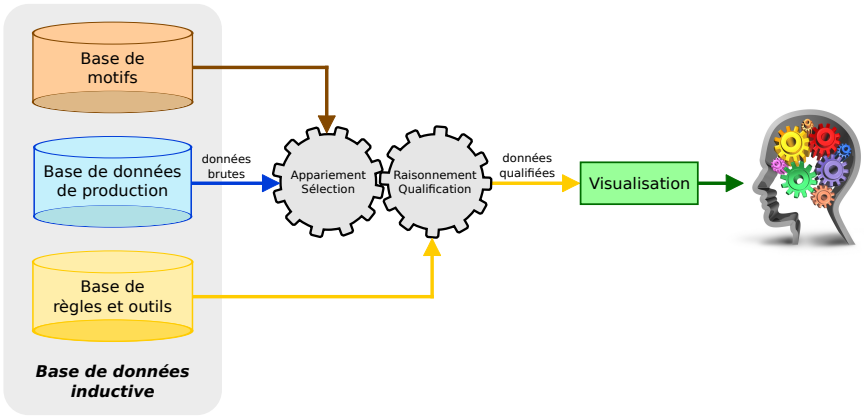


FIGURE 1.11 – Base de données inductive

Les bases de données inductives proposent d'intégrer dans un cadre commun des données brutes associées à des connaissance extraite à partir de ces données (figure 1.11). Une fois la base de connaissances renseignée à partir de données historisées, des outils d'analyse spatio-temporelle s'appuient sur les motifs extrait de l'entrepôt de données. Ces outils d'analyse sont basée sur une approche combinant des statistiques et la logique floue pour qualifier en temps réel des positions obtenues.

Cette étape de qualification de données (figure 1.12 étape 6) est décomposée en quatre phases successives. La première phase consiste à réaliser un appariement partiel d'une position et de sa trajectoire partielle à un patron spatio-temporel (figure 1.12 étape 6.A). Cette phase permet de sélectionner le patron spatio-temporel le plus proche (similaire) de la trajectoire à qualifier.

Une fois le patron spatio-temporel sélectionné, la position obtenue en temps réel peut être qualifiée par comparaison directe avec le couloir spatio-temporel (figure 1.12 étape 6.A). En ce qui concerne la trajectoire de l'objet mobile, elle peut être qualifiée grâce à la phase d'inférence floue (figure 1.12 étape 6.C). Cette phase d'inférence floue consiste à raisonner à l'aide d'un ensemble de règles floues afin d'obtenir un indice de similarité spatio-temporelle entre une trajectoire et un patron spatio-temporel. Cet indice est ensuite utilisé pour qualifier en temps réel la trajectoire (figure 1.12 étape 6.D). Les données de positions, trajectoires qualifiées et patrons spatio-temporels associés sont finalement transmis à l'étape de visualisation (figure 1.12 étape 7).

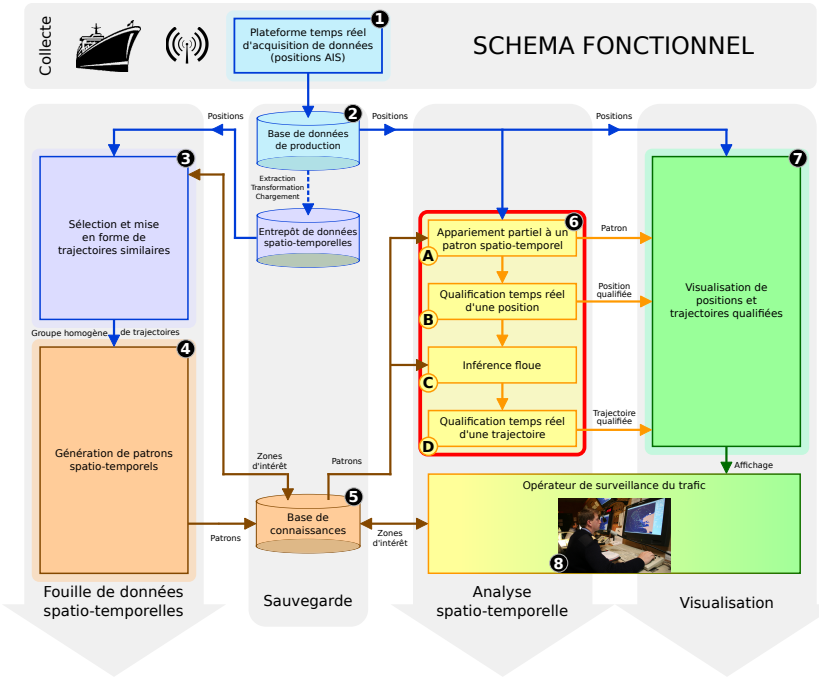


FIGURE 1.12 – Schéma fonctionnel détaillé de la phase de qualification spatio-temporelle

1.2.4 Visualisation

La dernière étape du schéma fonctionnel consiste à visualiser les positions et les trajectoires des objets mobiles (figure 1.13 étape 7.A). Cette visualisation est enrichie par l'affichage des patrons spatio-temporels sélectionnés (figure 1.13 étape 7.B),

des positions qualifiées (figure 1.13 étape 7.C) ainsi que les trajectoires qualifiées (figure 1.13 étape 7.D). L'opérateur de surveillance du trafic bénéficie donc de données de positions et de trajectoires enrichies par le processus de qualification des données (figure 1.13 étape 8) lui permettant de faciliter son analyse du trafic et la détection de situations anormales.

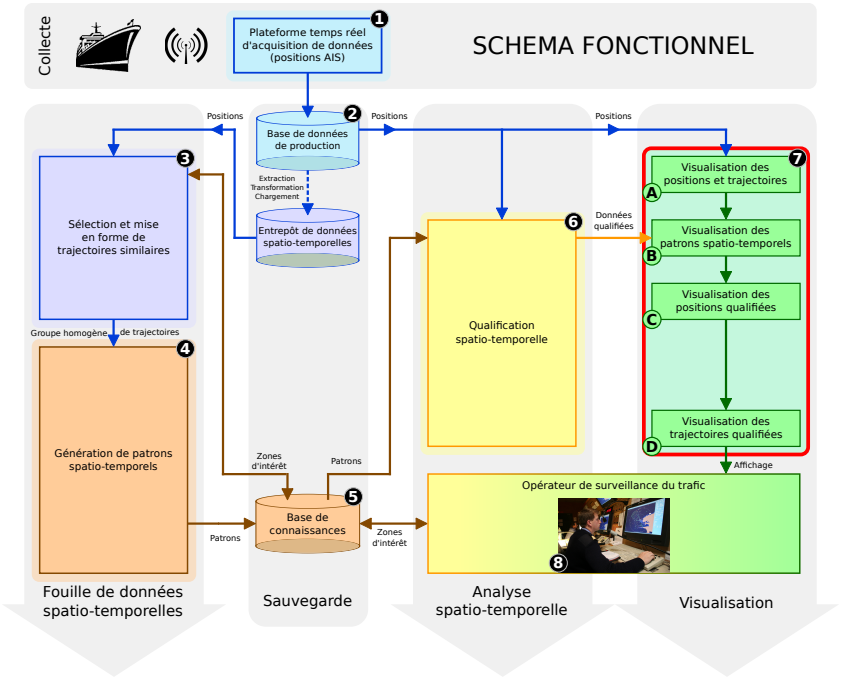


FIGURE 1.13 – Schéma fonctionnel détaillé de la phase de visualisation

Conclusion du chapitre

Dans ce chapitre nous avons dans un premier temps présenté le modèle conceptuel utilisé pour modéliser les déplacements d'objets mobiles. Différents concepts ont été formellement définis tels que la notion de position, de trajectoire, d'itinéraire, de route-type. De plus, nous avons réalisé une présentation d'un schéma fonctionnel mettant en œuvre les différentes étapes nécessaires de l'acquisition des données de positions à la qualification de positions et de trajectoires en temps réel. Nous avons proposé une approche d'extraction de connaissances dans des données (ECD) menant à la conception d'une base de données inductive comprenant à la fois des données, des motifs et des règles permettant de qualifier le mouvement d'objets mobiles en temps réel. Les différentes étapes de ce schéma fonctionnel sont détaillées dans les chapitres suivants.

Chapitre 2

Concepts de distances et similarité entre trajectoires

Sommaire

2.1 Mesures de similarité basées sur l'analyse de séries temporelles 39
2.1.1 Distance de Minkowski ($L_p - norm$) 39
2.1.2 Déformation temporelle dynamique ("Dynamic Time Warping") 41
2.1.3 Plus longue sous-séquence commune ("Longest common subsequence") 45
2.1.4 Distance d'édition 46
2.1.5 Comparatif des mesures de similarité basées sur l'analyse de séries temporelles 47
2.2 Mesures de similarité basées sur l'analyse de la géométrie . 47
2.2.1 Distance moyenne 48
2.2.2 Distance de Hausdorff 49
2.2.3 Distance de Fréchet 50
2.2.4 Distance de Fréchet discrète 51
2.2.5 Distance de Fréchet discrète moyenne 53
2.2.6 Distance de Fréchet discrète partielle 55
2.3 Mesures de similarité appliquées aux trajectoires 58
2.4 Synthèse 61

L'analyse du mouvement d'objets mobiles est une thématique de recherche active fortement liée à la fouille de données et à l'extraction de connaissances. Ces dernières années, les recherches liées à l'analyse de similarité de mouvements se sont multipliées.

Les bases de données de mouvements sont de plus en plus conséquentes grâce en particulier au développement des technologies de positionnement. Ces bases de données de mouvements sont étudiées dans de nombreuses disciplines telles que les systèmes d'information géographique (SIG), les sciences de l'environnement, sociales et cognitives.

C'est pourquoi, la notion même de similarité varie en fonction du domaine d'application adressé [Faloutsos *et al.* 1997]. L'analyse de similarité telle que définie par [Alt & Guibas 1996] peut être résumée aux questions suivantes :

- Soient deux objets A et B, à quel niveau ces deux objets se *ressemblent*-ils ?
- Quelles *transformations* géométriques (rotation, échelle) sont nécessaires pour faire ressembler au mieux l'objet B à l'objet A ?
- L'objet A peut-il être *simplifié* tout en restant ressemblant à l'objet original selon une tolérance fixée ?
- Quelles sont les *sous-parties* de l'objet B qui ressemblent le plus à l'objet A ?

La réponse à ces questions peut être utile dans différentes tâches telles que le regroupement de données "clustering", le classement de données "indexing" et la comparaison. L'analyse de similarité entre deux objets est répartie en deux grande catégories [Agrawal *et al.* 1993a, Alt & Guibas 1996] :

- L'analyse de similarité complète : les deux objets sont comparés l'un à l'autre dans leur ensemble.
- L'analyse de similarité partielle : certaines parties des deux objets sont similaires, il faut alors identifier les parties similaires et les comparer.

Dans la littérature, le concept de similarité est fortement lié à celui de la différence entre deux objets. Cette différence peut être mesurée en fonction du coût nécessaire pour transformer un objet en un autre ou bien basée sur une notion de distance entre ces objets [Faloutsos *et al.* 1997].

La notion de distance entre deux objets permet de quantifier la différence entre ces objets en fournissant l'outil de base des techniques d'analyse de similarité. En fonction des cas applicatifs et des caractéristiques des objets comparés, différentes mesures de distance peuvent être utilisées. L'objectif de ce chapitre est de présenter ces différentes mesures de distance ainsi que les techniques d'analyse de similarité s'appuyant sur ces mesures. Cette thèse porte sur l'analyse de similarité de trajectoires d'objets mobiles, c'est pourquoi nous avons focalisé notre attention sur les techniques d'analyse de similarité utilisées pour l'étude des mouvements.

La trajectoire d'un objet mobile pouvant être représentée comme indiqué dans la section 2.3 par une séquence temporelle ordonnée de positions constituant une forme géométrique, les techniques d'analyse de similarité de séries temporelles et de formes géométriques sont donc appropriées à l'étude de similarité de trajectoires. Ainsi, la première section de ce chapitre traite des techniques d'analyse de similarité de séries temporelles. Ensuite, les techniques de mesure de similarité basées sur une analyse des évolutions géométriques des objets mobiles sont présentées à la section 2.2. Enfin, un comparatif de synthèse des différentes techniques appliquées à l'analyse de similarité de trajectoires conclut ce chapitre.

2.1 Mesures de similarité basées sur l'analyse de séries temporelles

L'analyse de similarité de séries temporelles est un thème de recherche actif lié à de nombreux domaines d'applications tels que l'analyse de données provenant de l'industrie (marchés financiers, productions, ventes) ou l'analyse de données scientifiques (évolution de données météo ou biologiques...).

Ces analyses visant principalement à prédire l'évolution future des données ou valider certaines hypothèses [Agrawal *et al.* 1993a, Faloutsos *et al.* 1994, Ding *et al.* 2008b], la plupart des techniques employées peuvent être classées parmi différents groupes :

- distance de Minkowski ("Lp-norm"),
- déformation temporelle dynamique ("Dynamic Time Warping"),
- distance d'édition ("Edit distance"),
- la plus longue sous-séquence commune ("Longest common subsequence").

Une présentation de ces mesures de similarité basée sur l'analyse de séries temporelles est détaillée ci-après.

2.1.1 Distance de Minkowski ($L_p - norm$)

La similarité entre deux objets peut être basée sur l'analyse de différentes caractéristiques variables de ces objets (position dans l'espace, temps, vitesse, orientation, taille, etc...). Ces différentes mesures peuvent être réalisées à différentes niveaux, ce qui pose problème lorsque l'on souhaite les comparer.

Les processus de calcul de distances entre variables ayant des représentations et des unités différentes, ces représentations peuvent être quantitatives, binaires, ensemblistes, floues. Les paramètres caractéristiques d'un objet peuvent être représentés sous la forme d'un vecteur X constitué de n variables distinctes. Dans les exemples suivants, nous comparerons deux vecteurs X_j et X_k d'une même taille $|X_j| = |X_k| = n$, définis selon le formalisme suivant :

$$X_j = [x_{j_1}\ x_{j_2}\ ...\ x_{j_i}\ ...\ x_{j_n}]$$
$$X_k = [x_{k_1}\ x_{k_2}\ ...\ x_{k_i}\ ...\ x_{k_n}]$$

Pour calculer la différence entre X_j et X_k, il existe plusieurs méthodes basées sur un calcul de distances. La distance de Minkowski ($L_p - norm$) entre les vecteurs X_j et X_k est la distance la plus souvent utilisée pour des variables quantitatives. Cette distance est définie selon l'équation 2.1.

$$d_m(j,k) = \left(\sum_{i=1}^{n} |x_{j_i} - x_{k_i}|^p \right)^{\frac{1}{p}} \text{ avec } p \geq 1 \tag{2.1}$$

Différentes appellations de cette distance existent en fonction de la valeur du paramètre p. Cette distance est appelée distance de Manhattan (équation 2.2) pour $p = 1$ ou distance Euclidienne (équation 2.3) pour $p = 2$.

$$d_M(j,k) = \left(\sum_{i=1}^{n} |x_{j_i} - x_{k_i}| \right) \tag{2.2}$$

La distance de Manhattan se résume à calculer la somme des valeurs absolues entre les différents attributs des deux vecteurs.

$$d_E(j,k) = \sqrt{\sum_{i=1}^{n} (x_{j_i} - x_{k_i})^2} \tag{2.3}$$

La distance Euclidienne est très largement employée comme fonction de distance dans les techniques de mesures de similarité de séries temporelles [Agrawal *et al.* 1993a, Faloutsos *et al.* 1994, Faloutsos *et al.* 1997]. Enfin lorsque $p \rightarrow +\infty$, cela revient à calculer la distance maximale (distance de Chebychev) définie comme suit :

$$d_C(j,k) = \underset{i}{Max} |x_{j_i} - x_{k_i}| \tag{2.4}$$

Dans la plupart des techniques d'analyse de séries temporelles, la distance de Minkowski utilisée est principalement celle de Manhattan ($p = 1$) ou la distance Euclidienne ($p = 2$) [Chen & Ng 2004].

Dans le cas de notre exemple maritime, les positions des objets mobiles sont représentées dans le système géodésique $WGS84$ (World Geodetic System 1984) par des coordonnées géographiques au format latitude et longitude exprimées sous la forme d'angles [Hooijberg *et al.* 1997]. Le calcul de la distance en mètres entre deux positions géographiques $p1(\varphi 1, \lambda 1)$ et $p2(\varphi 2, \lambda 2)$ peut être effectué en utilisant la formule de Haversine (équation 2.5). Les latitudes et longitudes des positions $p1$ et $p2$ sont préalablement exprimées en radians et le paramètre R correspond au rayon de la terre exprimé en mètres.

$$
\begin{aligned}
a = sin^2\left(\frac{(\varphi 2 - \varphi 1)}{2} + cos(\varphi 1)\ cos(\varphi 2)\ sin^2\left(\frac{(\lambda 2 - \lambda 1)}{2}\right)\right) \\
b = 2\ atan2\left(\sqrt{a}, \sqrt{(1-a)}\right) \\
d_h(p1, p2) = R\ b
\end{aligned}
\tag{2.5}
$$

Cependant, si les variables disposent d'unités de mesures très différentes, les variables dotées des valeurs fortes et d'une variance importante auront un poids plus fort que les autres dans le calcul. Pour limiter cet effet, il est possible de normaliser les données afin de s'assurer que chaque variable contribue de manière identique au calcul de la distance. Une technique de normalisation [Hogg & Tanis 2009] consiste à s'assurer que chaque variable x_i est transformée en une nouvelle variable x_l de manière à ce que sa moyenne soit nulle ($\overline{x}_l = 0$) et son écart-type vaille un ($\sigma_l = 1$).

$$
\begin{aligned}
\overline{x}_l &= \frac{1}{N}\sum_{i=1}^{N} x_{l_i} \\
\sigma_l &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{l_i} - \overline{x}_l)^2} \\
x_{S_{l_i}} &= \frac{(x_{l_i} - \overline{x}_l)}{\sigma_l}
\end{aligned}
\tag{2.6}
$$

Une étude plus approfondie des différentes distances et mesures de similarité entre vecteurs est présentée dans [Cha 2007, Dahabiah 2010]. Une fois la définition des outils de calcul de distance établie, il est possible de les utiliser comme base pour l'analyse de similarité de séries temporelles.

2.1.2 Déformation temporelle dynamique ("Dynamic Time Warping")

La déformation temporelle dynamique (Dynamic Time Warping (DTW)) est un algorithme permettant d'aligner deux suites d'objets évoluant au cours du temps. Introduit par [Sakoe & Chiba 1978], cet algorithme permet de mesurer la similarité de l'évolution d'objets mobiles y compris en cas de différence de vitesse d'évolution pour peu que cette évolution soit représentable sous forme linéaire.

Cet algorithme est très souvent utilisé dans les applications de reconnaissance de la parole pour lesquelles les vitesses d'élocution peuvent varier de façon significative alors que la phrase prononcée reste la même. Le DTW est une méthode visant à optimiser l'appariement entre deux séries de données temporelles en les déformant

en vue de minimiser leur distance, ces séries étant déformées par une transformation non linéaire de la variable temporelle afin de mesurer leur similarité de façon indépendante du temps.

Formellement, la fonction de déformation temporelle $\varphi(k)$ permettant d'aligner les séries temporelles peut être définie de la façon suivante :
N correspond au nombre d'éléments de L_a, M au nombre d'éléments de L_b et T au nombre de couples de positions.
$\varphi(k) = (\varphi_a(k), \varphi_b(k))$, un couple de positions à aligner tel que : $k \in \{1...T\}$
$\varphi_a(k) \in \{1...N\}$ tel que $\varphi_a(k) \leq \varphi_a(k+1) \leq \varphi_a(k) + 1$
$\varphi_b(k) \in \{1...M\}$ tel que $\varphi_b(k) \leq \varphi_b(k+1) \leq \varphi_b(k) + 1$
Où $\varphi_a(k)$ est une reprojection de l'indice temporel de L_a, $\varphi_b(k)$ est une reprojection de l'indice temporel de L_b et $p(k)$ une pondération en fonction du patron de parcours choisi entre k et $k-1$ (incrémentation de φ_a, de φ_b ou φ_a et φ_b en même temps). Cette pondération permet de pénaliser ou d'autoriser la duplication ou l'omission d'éléments lors du parcours. La complexité de l'algorithme du DTW est de $O(N * M)$.

$$d_{DTW}(L_a, L_b) = min \left\{ \sum_{k=1}^{T} d(\varphi_a(k), \varphi_b(k)) p(k) \right. \tag{2.7}$$

$$MD_{[i,j]} = d_E(L_{a_i}, L_{b_j}) \tag{2.8}$$

L'algorithme 1 présente une méthode de calcul par programmation dynamique de la distance de déformation temporelle dynamique présenté dans l'équation 2.7. Considérant $d(i,j)$ comme étant la distance entre les positions respectives L_{a_i} et L_{b_j} des lignes L_a et L_b, le choix de la fonction de distance (cf section 2.1.1) est dépendant du type de données manipulées et de l'application.

La matrice de distance MD entre chaque élément des lignes L_a et L_b est calculée en utilisant l'équation 2.8. Ensuite, la matrice $MDTW$ est remplie par programmation dynamique en utilisant l'algorithme 1. Une fois la matrice $MDTW$ remplie, on recherche parmi cette matrice le chemin qui minimise la distance totale entre les couples d'éléments. Cette recherche peut être optimisée en introduisant certaines contraintes concernant le parcours du chemin.

Tout d'abord, le chemin commence forcément par le couple de positions correspondant aux deux éléments de départ des séquences L_a et L_b et se termine par le couple de positions correspondant aux deux éléments de fin des séquences L_a et L_b. Le parcours de la matrice s'effectue par incrémentations successives, le chemin avance d'une cellule adjacente seulement à chaque fois et ne reviens jamais en arrière.

Algorithme 1 Distance de déformation temporelle dynamique

Algorithme d_{DTW}=Distance_DTW(L_a,L_b)
Soit N le nombre de points de L_a
Soit M le nombre de points de L_b
Soit W la taille de la fenêtre de recherche
pour $i = 0$ à N **faire**
  pour $j = 0$ à M **faire**
    $MDTW_{[i,j]} = +\infty$
pour $i = 1$ à N **faire**
  pour $j = max(1, i - W)$ à $min(M, i + W)$ **faire**
    $MDTW_{[i,j]} = MD_{[i,j]} + min(MDTW_{[i-1,j]}, MDTW_{[i,j-1]}, MDTW_{[i-1,j-1]})$
$d_{DTW} = MDTW_{[N,M]}$
retour d_{DTW}

La figure 2.1 illustre l'application de l'algorithme DTW sur deux séries temporelles L_a (en vert) et L_b (en bleu). Les couples de positions non encore alignées et appariées au même indice temporel sont représentées sur la figure du milieu.

L'algorithme DTW aligne les deux séries temporelles en appariant les éléments en fonction du chemin minimal représenté par des points rouges dans la matrice de la figure 2.1. Le i^{eme} élément de la ligne L_a est alors apparié avec le $i+2^{eme}$ élément de la ligne L_b comme illustré en rose sur la figure 2.1.

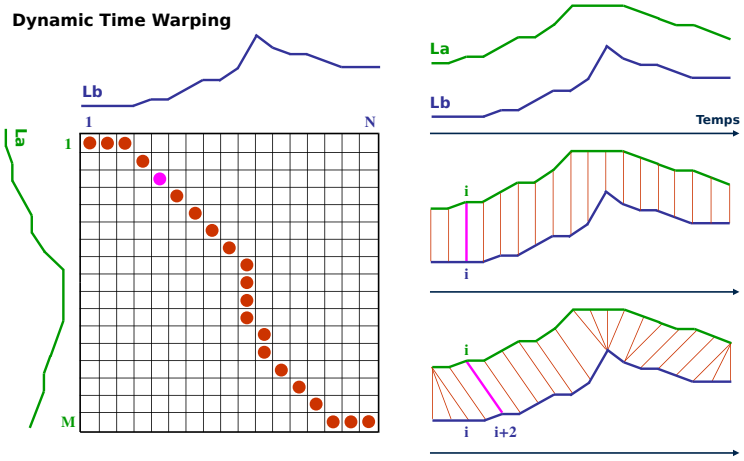


FIGURE 2.1 – Illustration de la déformation temporelle dynamique entre deux lignes L_a et L_b

Il est également possible d'ajouter une contrainte à cet algorithme en fixant la largeur maximale de la fenêtre de recherche W permettant d'éviter de calculer l'intégralité de la matrice [Itakura 1975, Sakoe & Chiba 1978,

Keogh & Ratanamahatana 2005]. Cependant, l'optimisation du paramètre spécifiant la taille de la fenêtre à utiliser reste difficile. De plus, la comparaison de séquences de tailles différentes est problématique. En effet, l'algorithme apparie tous les éléments des deux séquences, ce qui implique un bruit dû à la différence de taille, les éléments supplémentaires de la chaîne la plus longue étant appariés au dernier élément de la chaîne la plus courte.

Une étude comparative concernant le *DTW* appliqué à des séquences de tailles différentes est présentée dans [Kim *et al.* 2004, Ratanamahatana & Keogh 2004]. Certaines solutions telles que la réinterpolation ou la normalisation des séquences sont suggérées. De plus, le *DTW* permettant des accélérations et décélérations locales de fréquence de la série temporelle, celui-ci ne préserve donc pas les caractéristiques naturelles de la séquence [Yi *et al.* 1998]. Le *DTW* ne respecte pas l'inégalité triangulaire [Kim *et al.* 2004].

Ainsi, différents patrons de parcours peuvent être utilisés afin de déterminer la pondération à affecter au calcul de distance pour définir son coût. Ces patrons de parcours peuvent être symétriques ou asymétriques et disposer de poids différents. Une étude comparative de ces différentes fonctions de pondération de parcours est présentée dans [Sakoe & Chiba 1978, Myers *et al.* 1980].

Il est intéressant de noter que contrairement à l'algorithme de Levenshtein [Levenshtein 1966], le *DTW* ne pénalise pas les parcours qui engendrent une duplication d'éléments ($\varphi_a(k) = \varphi_a(k+1)$ ou $\varphi_b(k) = \varphi_b(k+1)$) ou bien leur omission ($\varphi_a(k+1) - \varphi_a(k) > 1$ ou $\varphi_b(k+1) - \varphi_b(k) > 1$).

Les figures 2.2, 2.3, 2.4 et 2.5 présentent de façon non exhaustive quelques exemples de pondérations calculées en fonction du parcours effectué entre deux éléments (valeur de pondération indiquée sur les arcs) [Rabiner & Juang 1993].

$$p(k) = min((\varphi_a(k) - \varphi_a(k-1)), (\varphi_b(k) - \varphi_b(k-1))) \tag{2.9}$$

$$p(k) = max((\varphi_a(k) - \varphi_a(k-1)), (\varphi_b(k) - \varphi_b(k-1))) \tag{2.10}$$

$$p(k) = (\varphi_a(k) - \varphi_a(k-1)) \tag{2.11}$$

$$p(k) = (\varphi_a(k) - \varphi_a(k-1)) + (\varphi_b(k) - \varphi_b(k-1)) \tag{2.12}$$

La pondération $p(k)$ des arcs du parcours étant obtenue par combinaison des évolutions des indices entre $\varphi_a(k)$,$\varphi_a(k-1)$ et $\varphi_b(k)$,$\varphi_b(k-1)$, les équations 2.9 (figure 2.2), 2.10 et 2.12 (figure 2.3) permettent de calculer les pondérations de parcours symétriques. L'équation 2.11 est un exemple de calcul de pondération de parcours asymétrique (figure 2.4).

De récentes recherches pour améliorer la rapidité de cet algorithme ont été réalisées [Sakurai *et al.* 2005, Ding *et al.* 2008b] en introduisant une limite de distance minimale permettant de filtrer un grand nombre de calculs de distances. Elles permettent de s'extraire des contraintes liées aux tailles différentes des séquences tout en diminuant la complexité de l'algorithme.

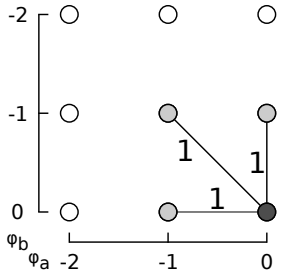


FIGURE 2.2 – Pondération de parcours symétrique (min)

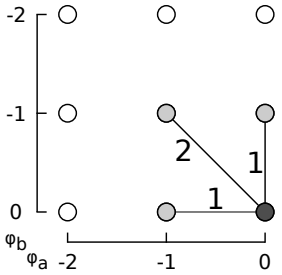


FIGURE 2.3 – Pondération de parcours symétrique (max)

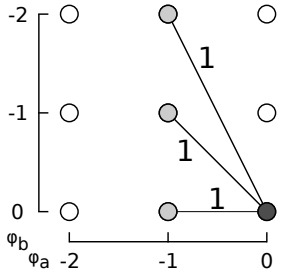


FIGURE 2.4 – Pondération de parcours asymétrique

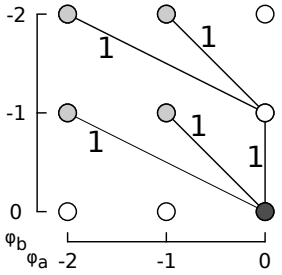


FIGURE 2.5 – Pondération de parcours de Rabiner et Juang

2.1.3 Plus longue sous-séquence commune ("Longest common subsequence")

Une autre mesure de distance est nommée "plus longue sous-séquence commune" (Longuest Common SubSequence $LCSS$). Cette mesure basée sur la distance d'édition présentée section 2.1.4 compare deux suites d'éléments en comptant le nombre maximal d'éléments composant la plus grande suite d'éléments communs [Agrawal *et al.* 1995].

Cependant la technique de la distance d'édition est initialement prévue pour être appliquée sur des séquences de caractères. Par conséquent, pour l'appliquer sur des séquences de valeurs numériques, il est nécessaire de déterminer un seuil ε permettant de définir si deux valeurs proches doivent être considérées comme égales par l'algorithme ($L_{a_i} \simeq L_{b_j}$). Les résultats de cet algorithme sont donc fortement dépendants du choix de ce paramètre.

L'idée principale de cette technique consiste à apparier deux séquences en leur permettant d'être étirées sans néanmoins réordonner les séquences d'éléments. L'algorithme $LCSS$ permet de ne pas prendre en compte certains éléments de la série à comparer, ce qui est intéressant lorsque les séquences comportent des données aberrantes [Agrawal *et al.* 1995, Das *et al.* 1997, Vlachos *et al.* 2002b].

Tout comme la distance d'édition et le DTW, la complexité de cet algorithme est de l'ordre de $O(N*M)$. La distance $LCSS$ entre $L_{a_{1..i...N}}$ et $L_{b_{1..j...M}}$ est définie par l'équation 2.13 [Vlachos *et al.* 2002b]. Une comparaison des techniques du DTW et $LCSS$ ainsi qu'une optimisation de l'algorithme basée sur la définition de bornes minimales est présentée dans [Vlachos *et al.* 2002b, Vlachos *et al.* 2003].

$$LCSS(L_{a_{1..i}}, L_{b_{1..j}}) = \begin{cases} 0 & Si\ (i = 0 \vee j = 0) \\ LCSS(L_{a_{1..i-1}}, L_{b_{1..j-1}}) & Si\ (L_{a_i} \simeq L_{b_j}) \\ max(\ LCSS(L_{a_{1..i-1}}, L_{b_{1..j}}), & Si\ (L_{a_i} \neq L_{b_j}) \\ \quad LCSS(L_{a_{1..i}}, L_{b_{1..j-1}})) & \end{cases} \tag{2.13}$$

2.1.4 Distance d'édition

Cette distance est principalement utilisée dans le cadre de la comparaison de chaînes de caractères. Cette distance calcule le nombre d'opérations nécessaires pour transformer une chaîne de caractères en une autre [Levenshtein 1966].

$$D_L(L_{a_{1..i}}, L_{b_{1..j}}) = \begin{cases} j & Si\ (i = 0) \\ i & Si\ (j = 0) \\ D_L(L_{a_{1..i-1}}, L_{b_{1..j-1}}) & Si\ i,j > 0 \wedge (L_{a_i} = L_{b_j}) \\ 1 + min \begin{cases} (D_L(L_{a_{1..i-1}}, L_{b_{1..j-1}}) & \\ (D_L(L_{a_{1..i-1}}, L_{b_{1..j}}) & autrement \\ (D_L(L_{a_{1..i}}, L_{b_{1..j-1}}) & \end{cases} & \end{cases} \tag{2.14}$$

La distance de Levenshtein est l'une des plus connues et très souvent utilisée dans l'analyse de séquences de caractères. Les trois opérations de type insertion, suppression ou substitution sont utilisées dans le processus de conversion d'une chaîne de caractères en une autre, le coût de chaque opération étant égal à 1 [Crochemore & Rytter 1994, Bozkaya *et al.* 1997, Cohen *et al.* 2003]. La distance d'édition entre deux séquences $L_{a_{1..N}}$ et $L_{b_{1..M}}$ est calculée par programmation dynamique en utilisant l'équation 2.14. Sa complexité est de l'ordre de $O(N*M)$.

Une version modifiée de cette distance d'édition où l'opérateur de substitution est supprimé a été proposée par [Bozkaya *et al.* 1997]. Cette version associée à une technique d'indexation permet d'extraire des séquences similaires de tailles différentes provenant d'une base de données volumineuse.

La technique de l'ERP ("Edit distance with Real Penalty") introduite par [Chen *et al.* 2004, Chen & Ng 2004, Chen *et al.* 2005] propose une combinaison de la distance de Manhattan ($L_1 - norm$) et de la distance d'édition permettant d'affecter une pénalité aux sous-parties non similaires. Cette distance d'édition est métrique et respecte l'inégalité triangulaire [Chen *et al.* 2005].

2.1.5 Comparatif des mesures de similarité basées sur l'analyse de séries temporelles

Une étude comparative des différentes mesures de similarité basées sur l'analyse de séries temporelles a été réalisée par [Ding *et al.* 2008b]. Dans cette étude, de nombreuses techniques présentées ci-avant sont comparées (distance Euclidienne, DTW, LCSS, ERP). Cette étude conclut que les techniques DTW, LCSS et ERP sont beaucoup plus précises que la distance Euclidienne pour des ensembles de données restreints alors que cette différence s'estompe pour des ensembles de données plus volumineux.

La distance Euclidienne entre deux séries temporelles appariées à un même pas de temps est une distance métrique ayant l'avantage d'être simple à calculer dans un temps linéaire $O(n)$ (pas de phase d'alignement). Cependant, cette mesure ne permet pas la comparaison de séries temporelles comportant des variations temporelles locales (changement de vitesse entre deux trajectoires similaires) et doit s'appliquer sur des séries temporelles de même nombre d'échantillons. En outre, cette mesure est fortement impactée par la présence de données aberrantes et de bruits.

Les techniques ERP, DTW et LCSS peuvent supporter des variations temporelles locales [Chen *et al.* 2005]. L'ERP est également une distance métrique alors que le DTW et LCSS ne le sont pas. LCSS n'est pas sensible aux bruits et données aberrantes.

2.2 Mesures de similarité basées sur l'analyse de la géométrie

Outre les mesures de similarité basées sur l'analyse de séries de données temporelles, les trajectoires peuvent également être comparées sur leur formes géométriques. Les techniques basées sur l'analyse de la géométrie et la forme des objets sont utilisées dans de nombreuses disciplines telles que la vision numérique, la robotique, la reconnaissance de formes, la cartographie, la biologie moléculaire.

Cette technique, tout comme celle de l'analyse de séries temporelles, est utilisée dans la fouille de données relatives au mouvement, particulièrement pour la détection de patrons de mouvements et de similarité entre trajectoires.

Ces mesures cherchent à mesurer la similarité (ou la distance) entre deux formes géométriques. Certaines transformations géométriques sont parfois utilisées pour maximiser la similarité entre les deux formes telles que la translation, la mise à l'échelle ou la rotation [Goodrich *et al.* 1999, Alt & Guibas 1996].

Cette section présente les différentes techniques liées au calcul de distances entre formes géométriques, ainsi que les mesures de similarité basées sur la géométrie algorithmique. Les deux principales distances présentées (la distance de Hausdorff et la distance de Fréchet) sont à la base d'un nombre important de recherches visant principalement à optimiser ces algorithmes pour les rendre plus rapides à calculer [Alt & Godau 1995, Eiter & Mannila 1994, Goodrich *et al.* 1999, Buchin *et al.* 2008, Alt 2009].

2.2.1 Distance moyenne

La distance moyenne (d_{moy}) entre deux lignes peut être représentée de différentes façons. [McMaster 1986] propose de calculer la surface entre deux lignes homologues divisée par la longueur de la ligne de référence (L_a). Par conséquent, cette distance n'est pas symétrique. La surface entre les deux lignes étant obtenue en reliant entre eux les points de départ des deux lignes et en faisant de même pour les points d'arrivée. Cette surface est représentée en gris sur l'exemple de la figure 2.6.

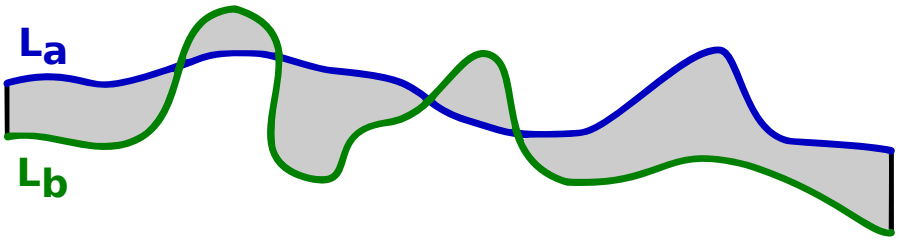


FIGURE 2.6 – Illustration du calcul de la surface entre deux lignes L_a et L_b

Cette distance moyenne permet d'obtenir une information concernant l'écart spatial moyen entre deux lignes. Cependant, à elle seule, cette information ne permet pas de savoir si les deux formes sont vraiment similaires ; elle doit être complétée par d'autres calculs, par exemple par un calcul de distance maximale entre les deux formes.

2.2.2 Distance de Hausdorff

Cette distance correspond à l'écart maximal existant entre deux lignes L_a et L_b [Hausdorff 1918]. Elle est définie formellement par l'équation 2.15.

$$d_H(L_a, L_b) = Max\left(\underset{p_a \in L_a}{Max}\left(\underset{p_b \in L_b}{Min}\left(d(p_a, p_b)\right)\right), \underset{p_b \in L_b}{Max}\left(\underset{p_a \in L_a}{Min}\left(d(p_a, p_b)\right)\right)\right) \quad (2.15)$$

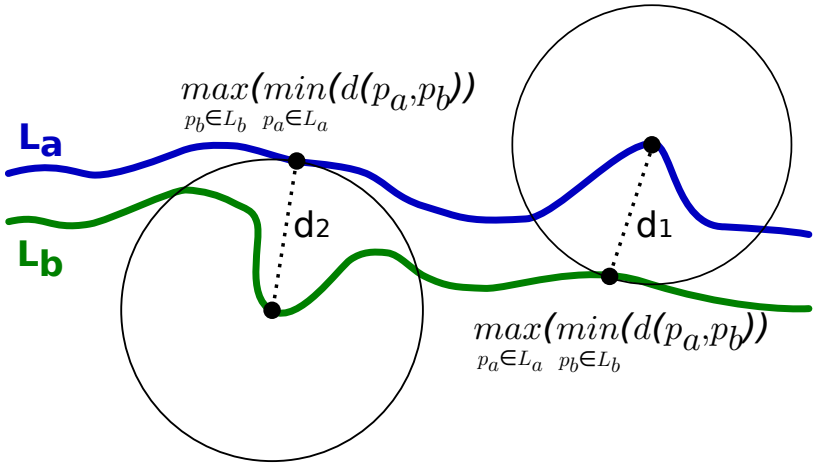


FIGURE 2.7 – Illustration de la distance de Hausdorff entre deux lignes L_a et L_b

La fonction $d(p_a, p_b)$ est une fonction de calcul de distance entre les points p_a et p_b choisie parmi celles introduites à la section 2.1.1. La distance de Hausdorff (d_H) est la plus grande des deux composantes 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 représentées sur l'illustration de la figure 2.7.

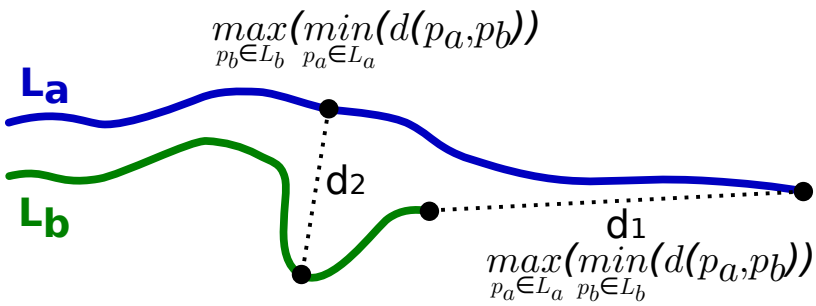


FIGURE 2.8 – Distance de Hausdorff entre deux lignes L_a et L_b d'emprise différente

Cependant, cette distance n'est pas adaptée lorsque les lignes comparées ont une emprise différente comme l'illustre la figure 2.8. L'emprise est définie comme la surface de l'espace occupée par l'objet. Dans ce cas particulier, la distance maximale entre la dernière position de la première ligne et celle de la seconde peut être fortement éloignée. La complexité de calcul de la distance de Hausdorff est de l'ordre de $O(N_a\ N_b)$ [Alt 2009].

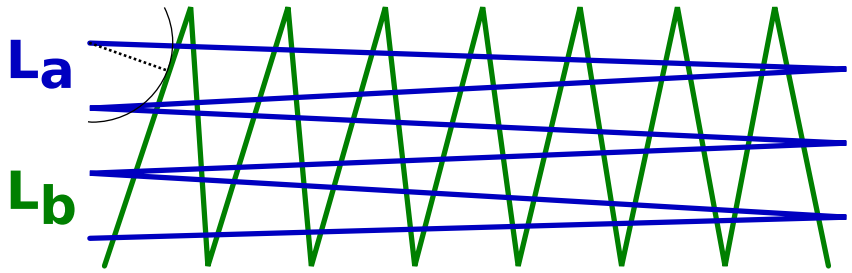


FIGURE 2.9 – Distance de Hausdorff faible entre deux lignes L_a et L_b sinueuses

La distance de Hausdorff n'est pas calculée sur des points homologues mais sur les points les plus proches spatialement, aussi, cette distance ne semble pas appropriée pour comparer des lignes sinueuses proches spatialement comme exposé en détail dans [Mascret 2010, Alt & Guibas 1996, Alt & Guibas 2000, Aronov *et al.* 2006] et sur l'exemple de la figure 2.9.

2.2.3 Distance de Fréchet

La distance de Fréchet [Fréchet 1905] est une distance permettant de calculer la distance maximale entre deux lignes [Alt & Godau 1992, Aronov *et al.* 2006]. La définition mathématique formelle de la distance de Fréchet peut être exprimée de la façon suivante :
Soit S un espace métrique ; considérant une ligne comme une suite de points orientés équivalente à une fonction continue de S, la ligne L_a représentée par la fonction continue $f : [a, a'] \rightarrow S$ et la ligne L_b représentée par la fonction continue $g : [b, b'] \rightarrow S$. Avec les conditions suivantes : $(a, a', b, b' \in \mathcal{R})$, $(a < a')$, $(b < b')$.

$$d_F(f,g) = \underset{\substack{\alpha:[0,1]\rightarrow[a,a'] \\ \beta:[0,1]\rightarrow[b,b']}}{Inf} \ \underset{t\in[0,1]}{Max} \left(d(f(\alpha(t)), g(\beta(t)))\right) \tag{2.16}$$

Cette distance peut être 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 représentation de cette distance sous forme de fonction continue rend sa complexité importante, $O(N_a\ N_b\ log^2(N_a\ N_b))$ avec N_a et N_b représentant le nombre de segments des lignes L_a et L_b [Alt & Godau 1995, Aronov *et al.* 2006].

2.2.4 Distance de Fréchet discrète

Une approximation de la distance de Fréchet dont la complexité est réduite à $O(N_a\ N_b)$ est proposée dans [Eiter & Mannila 1994]. Cette distance est nommée distance de Fréchet discrète (d_{Fd}). Les lignes L_a et L_b sont discrétisées sous la forme d'une suite de points.

$L_a = \{L_{a_1}...L_{a_N}\}$

$L_b = \{L_{b_1}...L_{b_M}\}$

La distance de Fréchet discrète d_{Fd} représente la longueur minimale 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}). Ce chemin permet d'obtenir une suite ordonnée de couples de points (L_{a_i}, L_{b_j}). L'évolution de ces couples de points est réalisée en fonction d'une des 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 2.17.

$$d_{Fd}(L_a, L_b) = max \left(\begin{array}{l} d_E(L_{a_n}, L_{b_m}) \\ \quad min \left(\begin{array}{ll} d_{Fd}(\{L_{a_1}...L_{a_{n-1}}\}, \{L_{b_1}...L_{b_m}\}) & \forall n > 1 \\ d_{Fd}(\{L_{a_1}...L_{a_n}\}, \{L_{b_1}...L_{b_{m-1}}\}) & \forall m > 1 \\ d_{Fd}(\{L_{a_1}...L_{a_{n-1}}\}, \{L_{b_1}...L_{b_{m-1}}\}) & \forall n > 1 \\ & \forall m > 1 \end{array} \right) \end{array} \right) \tag{2.17}$$

L'erreur due à l'approximation réalisée par la distance de Fréchet discrète est bornée par l'écart maximal ($LongMaxSeg$) entre deux points consécutifs d'un segment des lignes L_a et L_b [Eiter & Mannila 1994].
$d_F(L_a, L_b) \leq d_{Fd}(L_a, L_b) \leq d_F(L_a, L_b) + LongMaxSeg$

Le calcul de la distance de Fréchet est détaillé pour un exemple de lignes présenté sur la figure 2.10. La première ligne L_a est composée de 8 points ($N = 8$). La seconde

ligne L_b est composée de 7 points ($M = 7$). Les valeurs (X,Y) des points de L_a et L_b sont indiquées dans le tableau 2.15.

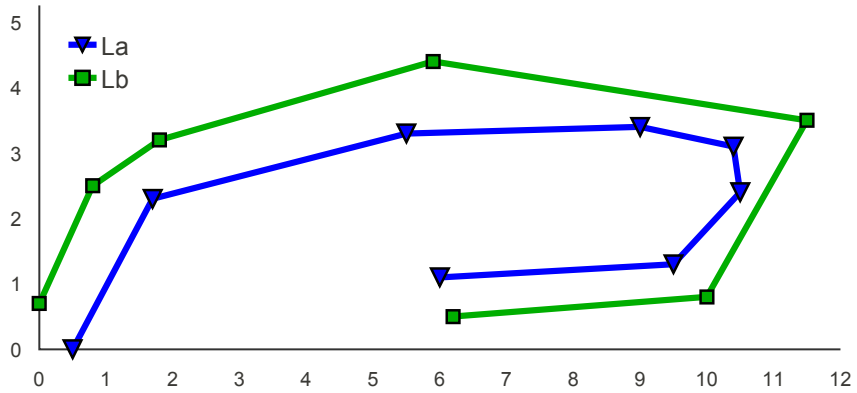


FIGURE 2.10 – Exemple de deux polylignes L_a et L_b

Pour chaque couple de points de L_a et L_b, la distance entre les points est calculée et sauvegardée dans la matrice de distances (MD) de taille ($N \times M$) (équation 2.8). En fonction du contexte et des objets manipulés, la fonction de calcul de distance doit être choisie en fonction des caractéristiques des points parmi ceux présentés à la section 2.1.1. Par exemple, dans le cadre de trajectoires représentées par une suite ordonnée de positions géographiques, la distance de Haversine (équation 2.5) est utilisée. Pour l'exemple de la figure 2.10, la fonction de distance retenue est la distance Euclidienne (d_E).

	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			Matrice de distances								
N°	Xb	Yb	MD	1	2	3	4	5	6	7	8
1	0	0,7	1	0,86	2,33	6,08	9,40	10,67	10,64	9,52	6,01
2	0,8	2,5	2	2,52	0,92	4,77	8,25	9,62	9,70	8,78	5,39
3	1,8	3,2	3	3,45	0,91	3,70	7,20	8,60	8,74	7,93	4,70
4	5,9	4,4	4	6,97	4,70	1,17	3,26	4,68	5,02	4,75	3,30
5	12	3,5	5	11,54	9,87	6,00	2,50	1,17	1,49	2,97	6,00
6	10	0,8	6	9,53	8,43	5,15	2,79	2,33	1,68	0,71	4,01
7	6,2	0,5	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 2.11 – Matrices de distances et de Fréchet de deux polylignes L_a et L_b

La matrice de Fréchet (MF) est ensuite calculée par programmation dynamique en utilisant la formule 2.17. Cette formule est très proche de celle utilisée pour le DTW. L'utilisation de la matrice de Fréchet de taille ($N \times M$) permet d'éviter le calcul récursif en mémorisant le résultat de l'étape de calcul précédente. Ainsi, la valeur d'une cellule de la matrice de Fréchet est directement dépendante des cellules adjacentes déjà calculées de la matrice de Fréchet et de la matrice de distance (MD), cette valeur est donnée par la formule 2.18.

$$\begin{aligned} MD_{[i,j]} &= d_E(La_i, Lb_j) \\ MF_{[i,j]} &= max(MD_{[i,j]}, min(MF_{[i-1,j]}, MF_{[i,j-1]}, MF_{[i-1,j-1]})) \end{aligned} \tag{2.18}$$

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 comparées dans leur totalité (pour notre exemple, les points L_{a_4} et L_{b_5} pour lesquels la distance maximale valant $2,50$ est atteinte).

2.2.5 Distance de Fréchet discrète moyenne

La distance de Fréchet discrète représente l'écartement maximal entre deux points homologues de deux lignes L_a et L_b. Deux points sont considérés comme homologues s'ils font partie de l'ensemble des couples de positions composant le chemin minimisant la distance de Fréchet discrète entre les deux lignes. Disposant de cette distance ainsi que de la matrice de Fréchet associée, il est intéressant de calculer la distance moyenne de Fréchet.

Cette distance moyenne est obtenue en utilisant les couples de points homologues dont l'appariement permet au maître et à son chien de parcourir le chemin avec la laisse la plus courte possible. Ces couples de points dont les distances d'écartement sont toutes inférieures ou égales à la distance de Fréchet composent le chemin minimum.

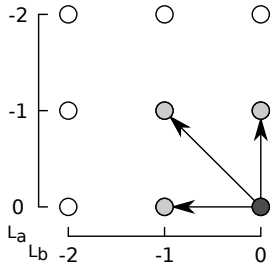


FIGURE 2.12 – Parcours des cases dans la matrice de Fréchet

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. Partant de la dernière cellule de la matrice de Fréchet contenant la distance de Fréchet ($MF_{[N,M]}$) il est possible de reconstituer le chemin minimal en sélectionnant les cellules précédentes de la matrice de Fréchet ayant la valeur la plus faible selon l'opérateur $\trianglelefteq$ défini par l'équation 2.19 et en comparant les trois cellules précédentes ($MF_{[N,M-1]}$,$MF_{[N-1,M]}$,$MF_{[N-1,M-1]}$) de la matrice de Fréchet comme indiqué dans l'équation 2.20 et illustré par la figure 2.12.

$$\begin{array}{l} MF_{[i,j]} \trianglelefteq MF_{[k,l]} \\ Si\ ((MF_{[i,j]} < MF_{[k,l]}) \vee ((MF_{[i,j]} = MF_{[k,l]}) \wedge (MD_{[i,j]} \leq MD_{[k,l]}))) \end{array} \tag{2.19}$$

$$P(i,j) = \begin{cases} Si\ (i=1) & MF_{[i,j-1]} \\ Si\ (j=1) & MF_{[i-1,j]} \\ Si\ (i>1 \wedge j>1) & \begin{cases} MF_{[i-1,j-1]} & Si((MF_{[i-1,j-1]} \trianglelefteq MF_{[i,j-1]}) \\ & \wedge (MF_{[i-1,j-1]} \trianglelefteq MF_{[i-1,j]})) \\ MF_{[i,j-1]} & Si((MF_{[i,j-1]} \trianglelefteq MF_{[i-1,j-1]}) \\ & \wedge (MF_{[i,j-1]} \trianglelefteq MF_{[i-1,j]})) \\ MF_{[i-1,j]} & Si((MF_{[i-1,j]} \trianglelefteq MF_{[i-1,j-1]}) \\ & \wedge (MF_{[i-1,j]} \trianglelefteq MF_{[i,j-1]})) \end{cases} \end{cases} \tag{2.20}$$

Les couples de points composant le chemin minimal sont obtenus en appliquant la fonction *Précédent* $P(N,M)$ en remontant dans la matrice (back tracking) jusqu'à obtenir le couple de points initial $C(1,1)$. Ce parcours de la matrice de Fréchet permet de constituer l'ensemble $\mathcal{C}$ de Z couples de points homogènes dont les cellules de la matrice de distance et de Fréchet sont surlignées en vert sur la figure 2.15.

$$\mathcal{C} = \{C_1(1,1), ..., C_k(i,j), ..., C_Z(N,M)\}$$

Disposant désormais de ces couples de points formant le chemin minimal, la moyenne des distances entre ces couples de points homogènes (cellules vertes) peut être calculée. Pour notre exemple, la distance de Fréchet moyenne ($d_{Fdm}(L_a, L_b)$) obtenue est de $((0,63+0,71+1,49+1,17+2,50+1,17+0,91+0,92+0,86)/9) = 1,15$.

$$d_{Fdm}(L_a, L_b) = \frac{1}{Z} \sum_{k=1}^{Z} \left(MD_{[\mathcal{C}_k]}\right) \tag{2.21}$$

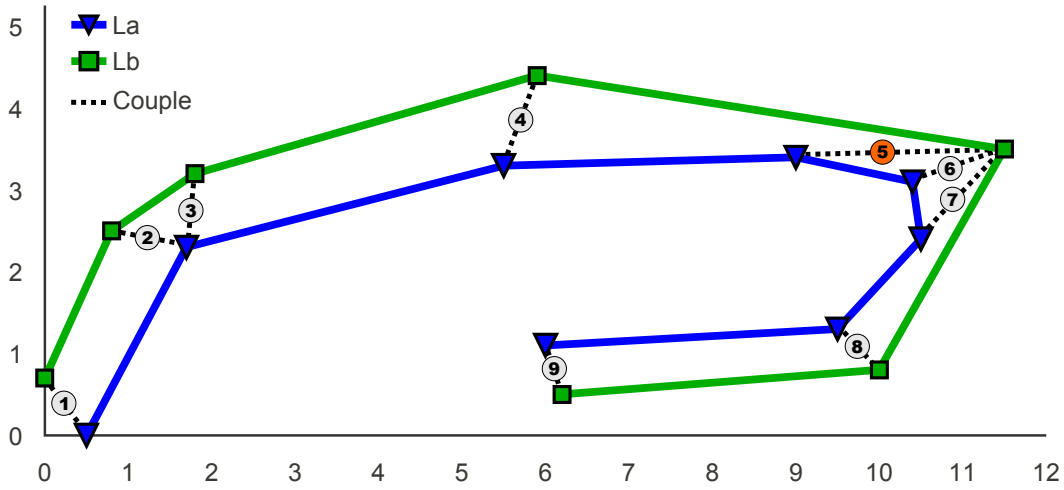


FIGURE 2.13 – Couples de points appariés de deux polylignes L_a et L_b

Les couples de points homologues de Fréchet correspondant aux cellules vertes de la matrice de distance sont présentés par des arcs en pointillés numérotés sur la figure 2.13. Le couple ayant la distance la plus éloignée correspondant à la valeur 2, 50 présenté en orange dans le tableau de la figure 2.15 (couple n°5) est représenté sur la figure 2.13 par l'arc dont le numéro est orange.

Cependant, cette technique est appliquée sur deux lignes ayant une emprise identique. Les points de départ et d'arrivée de chaque ligne sont automatiquement appariés ensemble, étant donné les conditions initiales et finales du processus d'appariement. Cette technique donne donc de bons résultats pourvu que les lignes soient d'emprise identique (*i.e.* que leurs surfaces géographiques soient similaires).

2.2.6 Distance de Fréchet discrète partielle

Pour deux lignes d'emprise différente, il est possible de calculer la distance de Fréchet discrète partielle introduite dans [Devogele 2000, Devogele 2002]. L'inconnue supplémentaire par rapport à la distance de Fréchet discrète est le couple des points de départ, d'arrivée ou bien les deux.

La technique de la distance de Fréchet partielle consiste à rechercher parmi la ligne de plus grande emprise (L_a), la sous-ligne L_c de L_a minimisant la distance de Fréchet discrète entre L_c et L_b. Cela signifie que l'écart de distance entre les points homologues de départ et d'arrivée recherché est inférieur à $d_{Fd}(L_a, L_b)$. L'algorithme 2 présente les différentes étapes de recherche de la sous-partie homologue de la ligne de plus grande emprise L_a avec la ligne L_b, dans le cas général d'appariement partiel où les points de départ et d'arrivée de L_a à apparier à L_b sont inconnus.

Une fois la distance de Fréchet discrète partielle connue, le chemin minimum permettant d'apparier les points homologues est obtenu de la même façon qu'à la section 2.2.5 pour la distance de Fréchet moyenne. Cet algorithme se simplifie aisément lorsqu'il n'y a plus qu'une seule inconnue à rechercher (point de départ ou d'arrivée à apparier).

Algorithme 2 Distance de Fréchet discrète partielle

Algorithme $[L_c, d_{Fdp}]$=Distance_de_Fréchet_discrète_partielle(L_a,L_b)
Soit L_a la ligne ayant l'emprise la plus grande (sinon inverser L_a et L_b)
Soit N le nombre de points de L_a
Soit M le nombre de points de L_b
Soit $d_{Fdp} = d_{Fd}(L_a, L_b)$
Soit $\mathcal{D}$ l'ensemble des points de départs candidats de L_a tels que :
$L_{a_i} \in \mathcal{D}$ Si $MD_{[i,1]} < d_{Fd}(L_a, L_b)$
Trier les points de $\mathcal{D}$ par ordre croissant de $MD_{[i,1]}$
Soit $\mathcal{A}$ l'ensemble des points de départs candidats de L_a tels que :
$L_{a_j} \in \mathcal{A}$ Si $MD_{[j,M]} < d_{Fd}(L_a, L_b)$
Trier les points de $\mathcal{A}$ par ordre croissant de $MD_{[j,M]}$
pour chaque point $L_{a_i} \in \mathcal{D}$ **faire**
  si $MD_{[i,1]} < d_{Fdp}$ **alors**
    pour chaque point $L_{a_j} \in \mathcal{A}$ **faire**
      si $i < j$ **alors**
        si $MD_{[j,M]} < d_{Fdp}$ **alors**
          $L_c =< L_{a_i}, ..., L_{a_j} >$
          si $d_{Fd}(L_c, L_b) < d_{Fdp}$ **alors**
            $d_{Fdp} = d_{Fd}(L_c, L_b)$
            $L_{a_{Dep}} = L_{a_i}$
            $L_{a_{Arr}} = L_{a_j}$
$L_c =< L_{a_{Dep}}, ..., L_{a_{Arr}}$
retour $[L_c, d_{Fdp}]$

Dans le cas du point de départ de L_a et L_b connu ($\mathcal{C}_1 = C(1,1)$), l'inconnue est le point d'arrivée L_{a_n} à apparier avec le dernier point de L_b (L_{b_M}). Ce point peut être trouvé en recherchant dans la matrice de Fréchet la cellule de la dernière ligne (correspondant au dernier point de L_b et surlignée en jaune sur la figure 2.15) ayant la valeur de d_{Fd} minimale (encadrée en bleu sur la figure 2.15) comme indiqué dans l'algorithme 3. Nous appellerons cette distance la distance de Fréchet discrète semi-partielle (d_{Fdsp}).

L'application de cet algorithme de distance de Fréchet discrète semi-partielle aux lignes L_a et L_b d'emprise différente représentées sur la figure 2.14 donne les

matrices de distances et de Fréchet de la figure 2.15. Les cellules colorées en vert correspondent aux couples de point homologues du chemin minimal déterminés de la même façon qu'expliqué dans la section 2.2.5 pour les cellules entourées en bleu soit les cellules des positions de départ et d'arrivée de la sous-ligne L_c de L_a homologue à L_b.

Algorithme 3 Distance de Fréchet discrète semi-partielle

1: **Algorithme** $[L_c, d_{Fdsp}]$=Distance_de_Fréchet_discrète_semi-partielle(L_a,L_b)
2: Soit L_a la ligne ayant l'emprise la plus grande (sinon inverser L_a et L_b)
3: Soit N le nombre de points de L_a
4: Soit M le nombre de points de L_b
5: Calculer la matrice de Fréchet entre L_a et L_b
6: Rechercher l'indice n compris entre 1 et N tel que $MD_{[n,M]} = min(MD_{[*,M]})$
7: $L_c =< L_{a_1}, ..., L_{a_n} >$
8: $d_{Fdsp} = MF_{[n,M]}$
9: **retour** $[L_c, d_{Fdp}]$

Afin de pouvoir apparier des trajectoires partielles, une des deux dernières méthodes doit être retenue. La technique de Fréchet [Fréchet 1905, Alt *et al.* 2001, Devogele 2002, Mascret *et al.* 2006] appliquée sur deux trajectoires discrètes permet d'apparier des trajectoires partielles en respectant la relation d'ordre temporel des positions des trajectoires tout en conservant l'intégralité des positions des trajectoires. Cependant, cette technique vise à minimiser la distance spatial maximale entre les deux trajectoires et ne tiens pas compte des différences de vitesses entre les positions. Cette technique a l'avantage de pouvoir s'appliquer à des trajectoires partielles et ne nécessite pas de calcul d'interpolation.

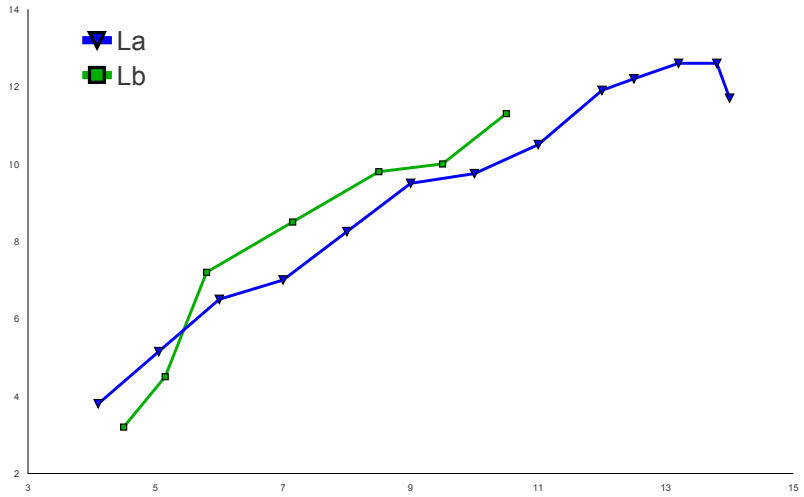


FIGURE 2.14 – Exemple de deux polylignes L_a et L_b d'emprise différente

Ligne a	N°	1	2	3	4	5	6	7	8	9	10	11	12	13
	Xa	4,1	5,05	6	7	8	9	10	11	12	12,5	13,2	13,8	14
	Ya	3,8	5,15	6,5	7	8,25	9,5	9,75	10,5	11,9	12,2	12,6	12,6	11,7

Ligne b		
N°	Xb	Yb
1	4,5	3,2
2	5,2	4,5
3	5,8	7,2
4	7,2	8,5
5	8,5	9,8
6	9,5	10
7	11	11

Matrice de distances

MD	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0,72	2,03	3,62	4,55	6,14	7,74	8,55	9,77	11,49	12,04	12,81	13,22	12,75
2	1,26	0,66	2,17	3,11	4,71	6,31	7,15	8,38	10,08	10,64	11,42	11,85	11,41
3	3,80	2,18	0,73	1,22	2,44	3,94	4,91	6,16	7,78	8,36	9,16	9,65	9,35
4	5,60	3,95	2,31	1,51	0,89	2,10	3,11	4,34	5,92	6,50	7,31	7,81	7,56
5	7,44	5,79	4,14	3,18	1,63	0,58	1,50	2,60	4,08	4,66	5,47	5,99	5,82
6	8,22	6,58	4,95	3,91	2,30	0,71	0,56	1,58	3,14	3,72	4,52	5,02	4,81
7	9,86	8,22	6,58	5,54	3,94	2,34	1,63	0,94	1,62	2,19	3,00	3,55	3,52

Matrice de Fréchet

MF	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0,72	2,03	3,62	4,55	6,14	7,74	8,55	9,77	11,49	12,04	12,81	13,22	13,22
2	1,26	0,72	2,17	3,11	4,71	6,31	7,15	8,38	10,08	10,64	11,42	11,85	11,85
3	3,80	2,18	0,73	1,22	2,44	3,94	4,91	6,16	7,78	8,36	9,16	9,65	9,65
4	5,60	3,95	2,31	1,51	1,22	2,10	3,11	4,34	5,92	6,50	7,31	7,81	7,81
5	7,44	5,79	4,14	3,18	1,63	1,22	1,50	2,60	4,08	4,66	5,47	5,99	5,99
6	8,22	6,58	4,95	3,91	2,30	1,22	1,22	1,58	3,14	3,72	4,52	5,02	5,02
7	9,86	8,22	6,58	5,54	3,94	2,34	1,63	1,22	1,62	2,19	3,00	3,55	3,55

FIGURE 2.15 – Matrices de distances et de Fréchet de deux polylignes L_a et L_b d'emprise différente

2.3 Mesures de similarité appliquées aux trajectoires

Basées sur une combinaison des techniques d'analyse de similarité géométrique et de séries temporelles présentées, de nombreuses recherches ont été proposées pour mesurer la similarité spatio-temporelle entre trajectoires.

Une technique basée principalement sur le $LCSS$ a été proposée par [Vlachos *et al.* 2002a, Vlachos *et al.* 2002b]. Cette technique nommée $LCSS_{\delta,\varepsilon}(L_a, L_b)$ permet de comparer la similarité de deux trajectoires L_a et L_b d'une durée et d'échantillonnage différents. Le paramètre δ permet l'étirement temporel des trajectoires de même que le paramètre ε qui définit la limite de translation spatiale autorisée. La complexité de cette technique calculée par programmation dynamique est de $O(\delta(N+M))$.

La distance Euclidienne est également très souvent utilisée pour mesurer la similarité entre deux trajectoires. [Yanagisawa *et al.* 2003] ont introduit une technique de recherche de similarité de forme de trajectoires appelée "k-Nearest Neighbor Queries". Basée sur une approximation linéaire par morceaux des trajectoires, cette technique calcule la distance Euclidienne moyenne entre les lignes.

Dans une approche similaire également basée sur une représentation des trajectoires par approximation linéaire par morceaux, [Lin & Su 2005] ont proposé la technique nommée "One Way Distance" (OWD). Cette technique calcule l'intégrale de la distance entre les points des trajectoires L_a et L_b divisée par la longueur de L_a.

La distance entre les deux trajectoires est la moyenne de leurs OWD. Ainsi cette technique peut être appliquée sur des trajectoires ayant des longueurs différentes. Elle peut être calculée grâce aux algorithmes proposés dans [Lin & Su 2005] d'une complexité de $O(n^2 log(n))$ et $O(n^2)$.

[Vlachos *et al.* 2004] a proposé une méthode qui a pour objectif de maximiser la similarité entre deux trajectoires en appliquant des transformations géométriques telles que des translations, mises à l'échelle ou des rotations. Les trajectoires sont représentées dans un espace basé sur les angles et longueurs des arcs, cet espace restant invariant aux translations, mises à l'échelle et rotations.

Afin de représenter des trajectoires dans cet espace, cette technique utilise l'angle de rotation comparé à un vecteur de référence ainsi que la longueur du vecteur de mouvement. Les coordonnées des positions des trajectoires sont alors transformées en une suite de paires d'angles et de longueurs d'arcs entre deux positions. Ensuite, le DTW est utilisé pour calculer la distance entre les deux trajectoires dans ce nouvel espace défini. De ce fait, les vitesses relatives entre les deux trajectoires ne sont pas préservées, ce qui fait que cette technique est une technique d'analyse de similarité spatiale.

D'autres techniques basées sur la distance d'édition ont également été proposées [Chen *et al.* 2004]. Pour optimiser le calcul de similarité avec cette méthode, une nouvelle représentation symbolique des trajectoires appelée "movement pattern strings" est utilisée. Elle consiste à représenter les distances et directions des segments de trajectoires par des caractères. La distance d'édition est ensuite employée pour calculer la similarité entre les deux chaînes de caractères.

Une autre approche nommée "Edit Distance on Real sequence" est également basée sur le concept de la distance d'édition [Chen *et al.* 2005]. Cette méthode recherche le nombre minimum d'opérations nécessaires pour transformer une trajectoire en une autre. Deux éléments des deux trajectoires L_a et L_b sont considérés comme similaires si ces éléments sont situés à une distance inférieure à un seuil ε. Cette technique a l'avantage de pouvoir s'appliquer sur des trajectoires d'une taille et d'un échantillonnage différents ou bruitées.

Les trajectoires sont principalement composées d'une suite de positions ordonnée dans le temps. A ce titre, elles disposent à la fois d'une dimension spatiale et temporelle. C'est pourquoi, certaines techniques d'analyse de similarité sont basées sur l'étude de la distance spatio-temporelle entre les trajectoires. [Sinha & Mark 2005] a utilisé la distance Euclidienne moyenne pour mesurer la similarité entre deux trajectoires régulièrement échantillonnées.

Prenant ainsi en compte la dimension spatiale et temporelle des trajectoires, ce travail a été étendu par [van Kreveld & Luo 2007] en ajoutant une fenêtre de recherche temporelle permettant de trouver les parties les plus similaires entre deux trajectoires. Dans cette méthode, le pas d'échantillonnage temporel entre les deux trajectoires doit être constant. Différents algorithmes ont été proposés par [Buchin *et al.* 2009] en fonction des paramètres initiaux de durée, de fenêtre temporelle, de durée minimale.

[Frentzos *et al.* 2007] propose une mesure nommée $DISSIM$ calculant la différence entre deux trajectoires d'une même longueur. La différence entre les deux trajectoires est définie par la somme des intégrales des distances Euclidiennes (équation 2.22).

$$DISSIM(L_a, L_b) = \sum_{k=1}^{n-1} \int_{t_k}^{t_{k+1}} d_{L_a,L_b}(t)dt \tag{2.22}$$

Cette approche nécessite que les trajectoires aient une même durée et un même nombre (k) de positions. Les trajectoires ayant un échantillonnage différent doivent donc être rééchantillonnées par interpolation linéaire. De plus, cette technique calculant la distance euclidienne cumulée entre les deux trajectoires, la distance obtenue est dépendante de la taille des trajectoires comparées.

[Pelekis *et al.* 2007] ont une approche similaire basée sur un opérateur de distance utilisant les coordonnées spatio-temporelles des positions combinées à la vitesse et à la direction. Cette mesure de distance est basée sur la surface créée entre les segments des points d'intersections ($1..i..I$) entre les deux trajectoires. Cet opérateur de distance nommé "Locality In-between Polylines" (LIP) est défini par la formule de l'équation 2.23.

$$LIP(L_a, L_b) = \sum_{\forall polygone_i} surface_i * \left(\frac{longueur(L_{a_i}, L_{a_{i+1}}) + longueur(L_{b_i}, L_{b_{i+1}})}{longueur(L_a) + longueur(L_b)} \right) \tag{2.23}$$

Cette équation, basée sur la surface géométrique entre les trajectoires ne prend pas la dimension temporelle en compte. Une extension nommée "Spatio Temporal Locality In-between Polylines" $STLIP$ permet de prendre en compte cet aspect temporel en lui affectant un poids dépendant d'un facteur multiplicatif fixé par l'utilisateur.

Une autre étude de [Trajcevski *et al.* 2007] s'est intéressée aux transformations géométriques de rotations et translations nécessaires pour calculer la similarité spatio-temporelle entre deux trajectoires. Cette étude nommée "Rigid Transformation Similarity Distance" ($RTSD$) est basée sur la distance de Fréchet.

La mesure proposée vise à minimiser la valeur maximale de la distance Euclidienne entre deux trajectoires en appliquant une combinaison de translations et de rotations à l'une des trajectoires. La distance Euclidienne est calculée entre les positions des trajectoires à un même temps t.

Pour comparer des trajectoires ayant des durées différentes, [Trajcevski *et al.* 2007] utilise une technique nommée "Temporal-Containment Similarity Distance" ($TCSD$) basée sur la mesure $RTSD$. La trajectoire la plus courte est déplacée le long de la trajectoire la plus longue. A chaque décalage, la $RSTD$ est calculée. La valeur minimale de $RSTD$ est conservée comme étant la $TCSD$. La complexité de l'algorithme $RTSD$ est de $O((n+m)^2)$ et de $O(nm(n+m)^2)$ pour la $TCSD$.

Toujours basée sur la distance de Fréchet, [Ding *et al.* 2008a] a proposé une mesure de similarité pseudo-métrique nommée "w-constrained discrete Fréchet distance" (wDF). Cette approche ajoute une contrainte de fenêtre temporelle à la distance de Fréchet. La taille de cette fenêtre temporelle w a un impact direct sur la précision et la complexité de l'algorithme qui est réduite à $O(w * n^2)$. Afin d'optimiser cet algorithme pour des trajectoires plus longues, [Ding *et al.* 2008b] a introduit dans son approche la notion de boîtes englobantes permettant d'approximer les trajectoires.

2.4 Synthèse

A la croisée entre les séries temporelles et les formes géométriques, les trajectoires peuvent se voir appliquer les techniques d'analyse de similarité de ces deux grandes catégories. L'analyse de similarité sur des critères temporels et/ou spatiaux est un domaine de recherche particulièrement actif dont les différentes techniques présentées au cours de ce chapitre sont résumées dans le tableau 2.1 inspiré des travaux de [Dodge 2011].

Les différentes techniques proposées dans ce tableau et appliquées sur des trajectoires sont toutes basées sur les méthodes de bases introduites dans ce chapitre et traitant d'analyse de similarité de séries temporelles ou de formes géométriques. Certaines de ces techniques combinent des approches des deux catégories.

En fonction du niveau d'analyse recherché, certaines techniques disposent d'avantages non négligeables tels que la robustesse aux bruits, aux transformations géométriques, la comparaison partielle, la combinaison des dimensions spatiales et temporelles. Ces différentes techniques sont la base que nous utiliserons dans les chapitres suivants, dans l'objectif de définir des patrons spatio-temporels et également pour comparer des trajectoires à ces patrons ainsi qu'à des trajec-

toires partielles. La technique employée devra donc pouvoir aligner deux trajectoires tout en respectant la relation d'ordre temporel entre les positions des trajectoires, sans effectuer de transformations géométriques sur les trajectoires et en conservant toutes les données de positions (pas d'omission de valeurs). De plus, cette technique devra également pouvoir être appliquée pour comparer des trajectoires partielles. Au vu des contraintes imposées, la distance de Fréchet discrète partielle [Devogele 2000, Devogele 2002] dispose de toutes les caractéristiques nécessaires pour réaliser cette tâche de comparaison.

Conclusion du chapitre

Dans ce chapitre nous avons introduit le concept de similarité entre des objets. Ce concept est fortement dépendant du domaine applicatif et des caractéristiques des objets manipulés ainsi que du résultat recherché. Différents niveaux de similarité peuvent être recherchés tels que la similarité totale ou partielle, ou bien la similarité obtenue après transformations. Considérant la formalisation du concept de trajectoire proposé au chapitre 1.1.4, deux grandes catégories de techniques d'analyse de similarité ont été présentées. La première technique consiste à effectuer une analyse de séries temporelles. La seconde est basée sur une analyse de la géométrie des objets à comparer. Les trajectoires manipulées sont à la fois composées d'une série temporelle de positions et disposent d'une représentation géométrique. C'est pourquoi, certains outils de mesure de similarité de trajectoires combinent les techniques d'analyse de séries temporelles avec celles basées sur la géométrie. Cette thèse portant sur l'analyse de similarité de trajectoires d'objets mobiles, une synthèse et un tableau récapitulatif présentant les techniques d'analyse de similarité utilisées pour l'étude des mouvements ont été proposés. Les caractéristiques essentielles à la comparaison de trajectoires partielles pour notre cas d'étude ont été citées. Au vu des contraintes imposées, la distance de Fréchet discrète partielle [Devogele 2000, Devogele 2002] est adaptée à cette tâche de comparaison.

Référence	Modèle de trajectoire	Appariement	Similarité	Concept	Caractéristiques	Contraintes
[Vlachos *et al.* 2002a, Vlachos *et al.* 2002b]	discret	partiel	spatial	LCSS	robuste au bruit, durée et échantillonnage différents	non métrique, complexité importante
[Yanagisawa *et al.* 2003]	discret	complet	spatial	Distance Euclidienne		sensible au bruit, trajectoires de même tailles
[Vlachos *et al.* 2004]	discret	complet	spatial	DTW	invariant aux translations, rotations, changements d'échelle	sensible au bruit
[Chen *et al.* 2004]	symbolique	complet	spatial	Distance d'édition	compare à la fois les distances et les directions, robuste aux bruits, alignement temporel	
[Chen *et al.* 2005]	discret	complet	spatial	Distance d'édition	robuste aux bruits, alignement temporel	
[Lin & Su 2005]	discret & continu	complet	spatial	Distance Euclidienne	échantillonnage spatial et temporel variable	
[Sinha & Mark 2005]	discret	complet	spatio-temporel	Distance Euclidienne		sensible aux bruits

[van Kreveld & Luo 2007]	discret	complet & partiel	spatio-temporel	Distance Euclidienne	similarité de sous-parties de trajectoires, alignement temporel	durée minimale à spécifier, sensible aux bruits
[Frentzos *et al.* 2007]	discret avec interpolation linéaire	complet	spatio-temporel	Distance Euclidienne		même durée, sensible aux bruits
[Pelekis *et al.* 2007]	discret	complet	spatio-temporel	basé sur la surface	prise en compte de la vitesse et la direction	sensible à la forme géométrique et aux points d'intersections
[Trajcevski *et al.* 2007]	discret avec interpolation linéaire	complet & partiel	spatio-temporel	Distance Euclidienne & Distance de Fréchet		mêmes durées, ignore la vitesse, complexité importante
[Ding *et al.* 2008b]	discret	complet	spatio-temporel	Distance de Fréchet	alignement temporel	ignore la vitesse, non métrique, sensible aux bruits
[Buchin *et al.* 2009]	continu	complet & partiel	spatio-temporel	Distance Euclidienne	similarité partielle, alignement temporel	durée minimale à spécifier, sensible aux bruits
[Devogele 2002]	discret	complet & partiel	spatio-temporel	Distance de Fréchet	similarité partielle, alignement temporel	sensible aux bruits

TABLE 2.1 – Comparatif des mesures de similarité

Chapitre 3

Clusters et motifs de trajectoires

Sommaire

3.1 Sélection et extraction d'un groupe homogène de trajectoires 66
3.2 Préparation des données 72
3.2.1 Filtrage des trajectoires erronées 72
3.2.2 Recalage spatial 73
3.2.3 Filtrage de Douglas et Peucker spatio-temporel 74
3.2.4 Échantillonnage d'un groupe homogène de trajectoires 76
3.2.5 Normalisation temporelle d'un groupe homogène de trajectoires 76
3.3 Génération d'un patron spatio-temporel 77
3.3.1 Le concept des boîtes à moustaches 78
3.3.2 Calcul d'une trajectoire-type 80
3.3.3 Calcul d'un couloir spatial 90
3.3.4 Calcul d'un couloir temporel 99
3.3.5 Synthèse d'un couloir spatio-temporel et d'une route-type . . . 100
3.4 Distribution des distances spatiales et temporelles des nuages de positions 103
3.5 Sauvegarde de la route-type 112

Disposant d'une volumineuse base de données de positions spatio-temporelles, nous souhaitons extraire de ces données des connaissances relatives aux déplacements des objets mobiles. Partant du postulat que des objets mobiles de même type suivant un itinéraire commun ont un comportement identique, l'objectif de ce chapitre est de définir un patron spatio-temporel synthétisant ce comportement habituel observé.

A partir des positions des objets mobiles, des trajectoires d'objets mobiles suivant un itinéraire doivent préalablement être constituées. Une première problématique consiste à extraire et regrouper ces trajectoires en un groupe homogène de trajectoires suivant un itinéraire commun. Ensuite, à partir de ce groupe de trajectoires, nous cherchons à définir une trajectoire-type synthétisant le déplacement du

groupe. En fonction du contexte dans lequel évoluent les objets mobiles, des écarts spatiaux et temporels variables avec cette trajectoire-type peuvent être observés. Il est donc important de prendre en compte ces écarts dans le processus de génération de notre patron spatio-temporel.

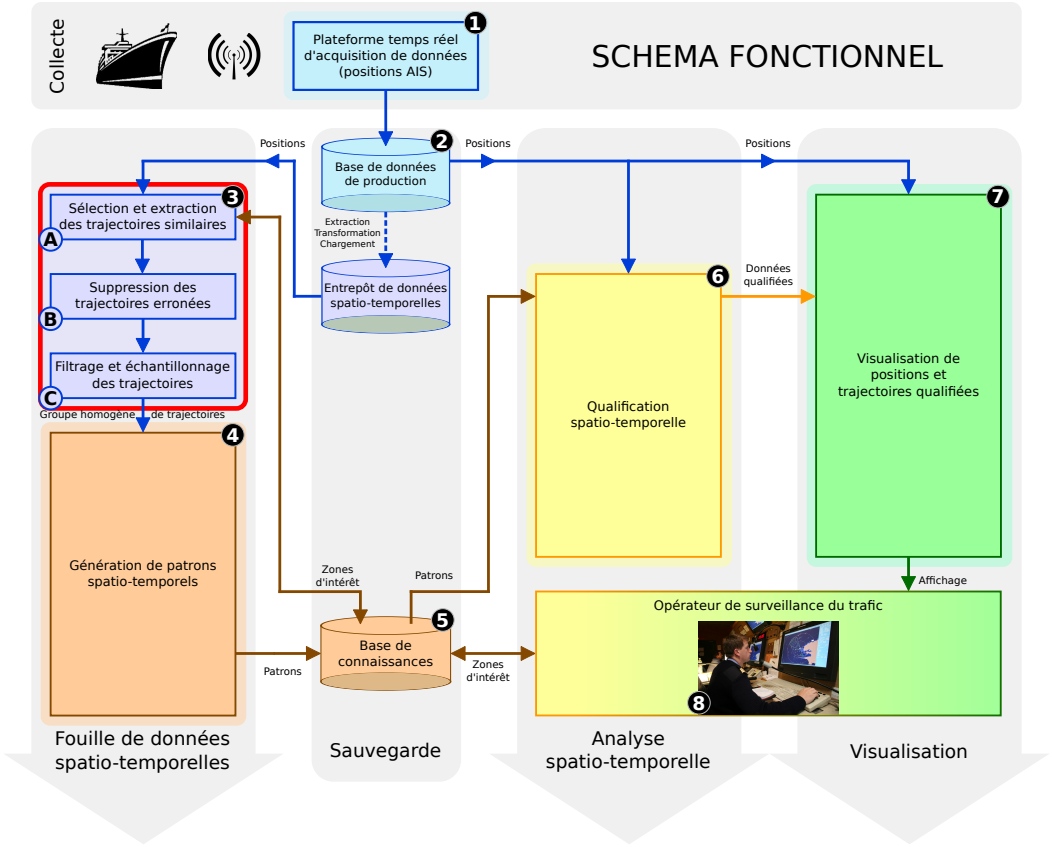


FIGURE 3.1 – Schéma fonctionnel de l'étape de préparation des données

Dans un premier temps, l'étape de préparation et d'extraction d'un groupe homogène de trajectoires sera présentée. Puis, un patron spatio-temporel composé d'une trajectoire-type et d'un couloir spatial et temporel sera défini. Enfin, une analyse statistique des écarts ainsi que certaines mesures seront présentées.

3.1 Sélection et extraction d'un groupe homogène de trajectoires

Les notions de *trajectoires* et d'*itinéraires* ayant été formalisées au chapitre 1.1.4 et 1.1.5, les critères d'extraction de trajectoires suivant un itinéraire vont être détaillés. L'objectif de cette section est de présenter les différentes méthodes

d'extraction de trajectoires d'un même type S d'objets mobiles suivant le même itinéraire I de la $BDST$.

Cet ensemble est appelé groupe homogène de trajectoires d'objets mobiles de même type suivant le même itinéraire (GHT). Différentes techniques peuvent être utilisées pour obtenir ce groupe de trajectoires.

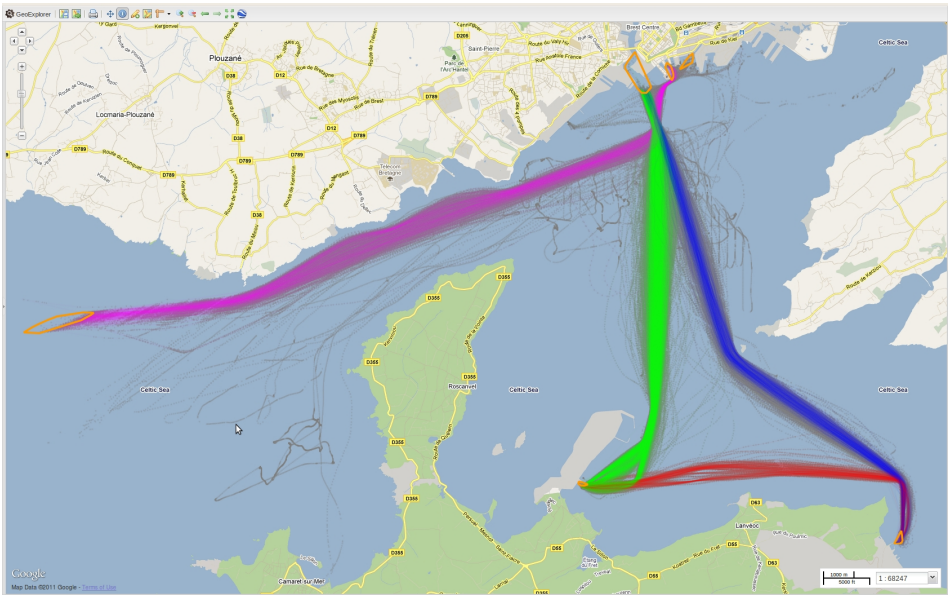


FIGURE 3.2 – Différents groupes homogènes de trajectoires obtenus par application de l'outil $T-Clustering$ du logiciel $M-Atlas$

Une première famille d'algorithmes s'appuient sur la technique du "Partionning Around Medoids" (PAM) [Kaufman & Rousseeuw 1990]. PAM est une technique de partitionnement de l'espace qui choisit aléatoirement comme référence certains éléments ($k-medoids$) et apparie de façon récursive le reste des données à ces éléments en fonction de leur similarité.

Fondé sur PAM, l'algorithme $CLARA$ (Clustering LARge Applications) [Kaufman & Rousseeuw 1990] optimise le calcul des $k-medoids$ en limitant le volume de données à un sous-échantillon sélectionné aléatoirement. Enfin, l'algorithme $CLARANS$ (Clustering Large Applications based on RANdomized Search) [Ng & Han 1994], limite quant à lui le nombre de données testées à chaque itération.

Ces algorithmes permettent de partitionner les données distribuées dans l'espace en k différents clusters dont les centres de gravité sont les $k-medoids$. Cependant, le nombre k d'objets choisis comme références à l'initialisation de l'algorithme est un paramètre qui doit être fixé par l'utilisateur en fonction du domaine d'application, ce qui pose problème lorsque ce nombre est inconnu a priori.

Une seconde famille d'algorithmes sont basés sur une décomposition hiérarchique de l'ensemble des données représenté par un arbre divisant cet ensemble en sous-ensembles plus petits jusqu'à obtenir les feuilles de l'arbre constituées d'une seule donnée. Deux approches existent, la première consiste à réaliser cet arbre (dendrogramme) en partant des feuilles et en remontant à la racine (approche agglomérative). La seconde approche consiste à partir de la racine en descendant jusqu'aux feuilles (approche divisive). Ces deux approches sont fondées sur des opérateurs permettant soit de fusionner, soit de diviser les ensembles à chaque étape [Jain & Dubes 1988, Guha *et al.* 1998, Hattori & Torii 1993].

L'avantage de ces algorithmes réside dans le fait qu'ils ne nécessitent pas de connaître le nombre k de partitions à obtenir. Néanmoins, il est nécessaire de définir une condition d'arrêt du processus de fusion ou de division. Cette condition repose couramment sur la distance minimale à obtenir entre deux clusters. $DBSCAN$ (Density Based Spatial Clustering of Applications with Noise) [Ester *et al.* 1996] repose sur l'analyse de la densité de points concentrés autour d'un point de référence. Cet algorithme s'appuie sur le nombre minimal de points $MinPts$ situés à la distance maximale Eps du point de référence. L'algorithme sélectionne un point p de référence aléatoirement puis recherche parmi les données le nombre de points situés à une distance inférieure à Eps.

Grâce à ces paramètres, les différents points de l'ensemble de données sont regroupés en clusters en fonction de la densité observée entre ces points. [Lee *et al.* 2007] propose une technique nommée $TRACLUS$ (Trajectory Clustering) construite également sur une analyse de la densité de segments de trajectoires représentés par des lignes.

Le logiciel $M-Atlas$ [Trasarti *et al.* 2010] regroupe différents outils d'analyse de mouvements et plus particulièrement l'outil $T-Clustering$ [Andrienko *et al.* 2009] permettant de regrouper les trajectoires similaires (la notion de similarité entre trajectoires est définie plus en détail au chapitre 2). Cet outil de regroupement de trajectoires est basé sur la technique nommée $OPTICS$ ("Ordering Points To Identify the Clustering Structure") [Ankerst *et al.* 1999]. L'outil $T-Clustering$ a été appliqué à notre base de données de mouvements afin d'en extraire des groupes homogène de trajectoires. Le résultat obtenu est présenté sur la figure 3.2.

Une autre famille d'algorithmes s'appuie sur l'ajout de connaissances a priori (horaires, catégories d'objets, POI, graphes de zones d'intérêts). La technique proposée consiste à extraire un groupe homogène de trajectoires en s'appuyant sur la notion d'itinéraire préalablement définie au chapitre 1.1.5. Les zones d'intérêt sont alors utilisées comme critère de sélection des trajectoires. D'autres critères, liés aux caractéristiques de l'objet mobile, sont également utilisés pour extraire le GHT.

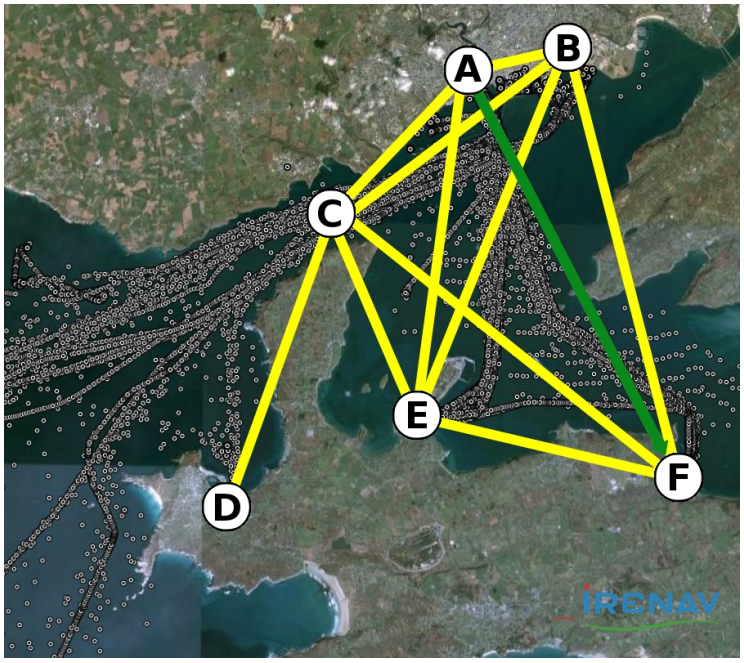


FIGURE 3.3 – Graphe de zones de la région de Brest

Ainsi, le premier critère de sélection est le type ou la catégorie des objets mobiles. Ce critère de sélection a pour objectif d'obtenir une classe homogène d'objets mobiles à analyser. Il dépend des caractéristiques des objets mobiles et du domaine applicatif étudié, il peut être défini sur les attributs de l'objet mobile, l'environnement spatial ou temporel (par exemple, une espèce d'animal sur une période précise de migration, un navire marchand avec chargement et prise en compte de la marée, déplacement d'un véhicule à une heure de pointe). Dans l'exemple applicatif, seuls les navires de type "navire à passagers" sont sélectionnés, soit 5 navires parmi 1 005. Les résultats d'autres classes sont présentés en annexes A.9 (Calais-Douvres) et A.10 (Manche-Mer du nord).

Cette première sélection d'une classe homogène d'objets mobiles nous permet d'obtenir des résultats statistiques satisfaisants au vu du nombre important de trajectoires analysées. Le second critère de sélection est basé sur des critères géographiques. La première position de la trajectoire doit nécessairement se trouver dans la première zone de l'itinéraire. De même, la dernière position de la trajectoire doit être incluse dans la dernière zone de l'itinéraire. De plus, pour chaque trajectoire, il ne doit exister qu'une seule position de départ appartenant à la zone de départ (de même pour la zone d'arrivée).

Finalement, le dernier critère employé est temporel, certains des objets mobiles pouvant effectuer ces itinéraires périodiquement. Il est donc nécessaire d'affecter ces différentes positions de départ et d'arrivée à des trajectoires distinctes.

L'algorithme 4 reprend de façon ordonnée l'ensemble des critères de sélection d'un GHT. Considérant la durée D_{ZiZj} approximative nécessaire à ce type d'objets mobiles pour réaliser le trajet entre chaque arc Z_iZ_j de l'itinéraire, il est possible de sélectionner des positions uniques de départ et d'arrivée de trajectoires pour chaque arc de l'itinéraire.

Algorithme 4 Extraction d'un groupe homogène de trajectoires

Algorithme GHT(Itinéraire I, Type d'objet mobile S)
Soit Z_D la zone de départ de l'itinéraire I
Soit Z_A la zone d'arrivée de l'itinéraire I
Soit D la durée estimée de l'itinéraire (Z_D, Z_A)
Rechercher l'ensemble $\mathcal{E}p_d$ des positions des objets de type S tel que $\mathcal{E}p_d \subset Z_D$
Rechercher l'ensemble $\mathcal{E}p_a$ des positions des objets de type S tel que $\mathcal{E}p_a \subset Z_A$
Sélectionner parmi $\mathcal{E}p_d$ le sous-ensemble $\mathcal{EV}p_d$ tel que :
$\mathcal{EV}p_d \subset \mathcal{E}p_d \wedge \forall p_{di} \in \mathcal{EV}p_d, \nexists p_{dj} \in \mathcal{EV}_d,\ t_{p_{di}} < t_{p_{dj}} \wedge t_{p_{dj}} < (t_{p_{di}} + 2 * D)$
Sélectionner parmi $\mathcal{E}p_a$ le sous-ensemble $\mathcal{EV}p_a$ tel que :
$\mathcal{EV}p_a \subset \mathcal{E}p_a \wedge \forall p_{ai} \in \mathcal{EV}p_a, \nexists p_{aj} \in \mathcal{EV}p_a,\ t_{p_{ai}} < t_{p_{aj}} \wedge t_{p_{aj}} < (t_{p_{ai}} - 2 * D)$
pour chaque $p_{di} \in \mathcal{EV}p_d$ **faire**
  si $\exists p_{ja}, dure(p_{id}, p_{ja}) < (2 * D)$ **alors**
    Extraire la trajectoire T_k entre p_{id} et p_{ia}
    si $\forall p_k \in T_k \wedge p_k \subset Z_i \Rightarrow Z_i = Z_D \vee Z_i = Z_A$ **alors**
      Ajouter T_k au GHT

En effet, si un objet même mobile met un temps D_{ZiZj} pour effectuer le trajet entre les deux zones Z_i et Z_j et qu'il réalise des allers-retours, chaque point de départ d'une nouvelle trajectoire de ce même objet mobile est donc séparé d'environ $2 * D_{ZiZj}$. Par conséquent, le point d'arrivée p_a de la trajectoire est forcément compris dans l'intervalle entre t_{p_d} et $t_{p_d} + 2 * D_{ZiZj}$. Ainsi, une position de départ p_d est considérée comme valide s'il n'existe aucune autre position p_j d'estampille temporelle t_{p_j} incluse dans la zone de départ Z_D dans l'intervalle de temps $t_{p_j} + D_{ZiZj}$.

De même, une position d'arrivée p_a est considérée comme valide s'il n'existe aucune autre position p_k d'estampille temporelle t_{Pk} incluse dans la zone d'arrivée Z_A dans l'intervalle de temps $t_{Pk} - D_{ZiZj}$. Finalement, une position de départ p_d peut être appariée à une position d'arrivée p_a si $t_{p_d} < t_{p_a} \wedge t_{p_a} < (t_{p_d} + 2 * D_{ZiZj})$. Ceci permet d'extraire de la base de données spatio-temporelles toutes les positions constituant la trajectoire de l'objet mobile sur l'intervalle compris entre t_{p_d} et t_{p_a}.

Pour chacune de ces positions composant la trajectoire, il est nécessaire de vérifier la propriété [1.3] en s'assurant qu'aucune d'elles n'est incluse dans une zone du graphe G_Z n'appartenant pas à l'itinéraire I. L'ensemble des trajectoires ainsi reconstitué compose le GHT à analyser.

Dans le cadre de notre exemple, cette méthode d'extraction de données a permis d'extraire 554 trajectoires de navires à passagers suivant l'itinéraire Brest Arsenal → Lanvéoc Ecole Navale représenté par l'arc (A,F) de la figure 3.3. Ce groupe homogène de trajectoire est représenté sur la figure 3.4. Le temps CPU nécessaire à l'exécution de cette étape appliqué à notre exemple comportant 4 821 447 positions est de 17 306 ms[1]. Cette technique d'extraction est beaucoup plus rapide que la première testée avec $M - Atlas$.

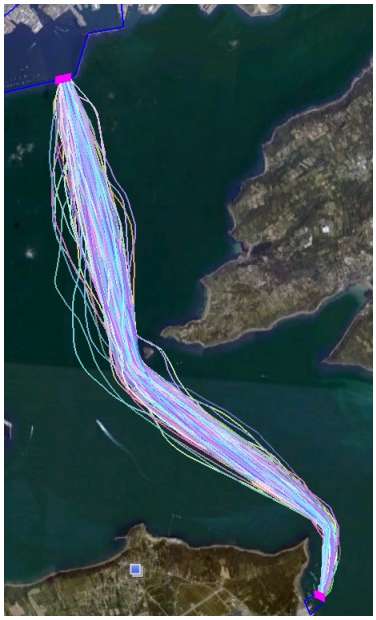


FIGURE 3.4 – Groupe homogène de trajectoires de navires de type "navire à passager" suivant l'itinéraire "Brest Arsenal → Lanvéoc Ecole Navale"

Le GHT de la figure 3.4 est très proche de celui obtenu en utilisant la technique du $T - Clustering$ et représenté en bleu sur la figure 3.2. Néanmoins, la réduction du nombre de données à traiter par l'introduction de critères de sélections combinés engendre une optimisation des temps de calculs nécessaires à l'extraction du groupe homogène de trajectoires. C'est pourquoi, dans un souci de simplicité et de rapidité de calculs, la technique d'extraction de groupes homogènes de trajectoires basée sur l'algorithme 4 a été retenue dans notre prototype maritime.

1. Mesure réalisée sur un PC disposant de deux CPU Intel Core2 6600 à 2,4 GHz et de 2 Go de mémoire

3.2 Préparation des données

Afin de pouvoir effectuer une analyse statistique optimale d'un *GHT* extrait à l'étape 3.A du schéma fonctionnel de la figure 3.1 comme présenté au chapitre 1.1.7, il est nécessaire de s'assurer au préalable que ce groupe de trajectoires est exempt d'erreurs. Aussi, la première étape du processus de génération de patron spatio-temporel repose sur la préparation et le filtrage des données erronées. Cette étape de prétraitement est présentée sur l'étape 3.B et 3.C de la figure 3.1.

3.2.1 Filtrage des trajectoires erronées

Disposant d'un *GHT* extrait de la base de données, il est nécessaire d'effectuer un premier filtrage des données pouvant amener un biais dans notre analyse. En effet, les données manipulées proviennent d'instruments de mesures ayant une précision relative plus ou moins importante et dépendant de différents facteurs.

En fonction de différents paramètres tels que les conditions météorologiques, la puissance de réception du signal, la fiabilité du capteur GPS utilisé ou de perturbations extérieures, les données collectées peuvent être imprécises voire erronées. La figure 3.5 montre un effet de crénelage sur une trajectoire disposant d'un capteur *GPS* peu précis.

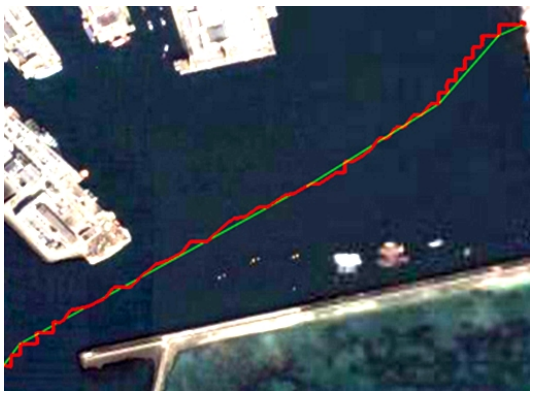


FIGURE 3.5 – Lissage des trajectoires par filtre de Douglas et Peucker spatio-temporel

Différents types d'imprécisions peuvent alors être observés comme présenté sur les figures 3.6 qui illustre le cas d'une position aberrante et d'une rupture de communication. C'est pourquoi il est nécessaire d'effectuer cette phase de préparation des données afin d'obtenir une représentation des trajectoires homogènes en termes de précision et de pas d'échantillonnage. Les trajectoires contenant des ruptures de communication comme présenté sur la figure 3.6 doivent être filtrées. Ces ruptures de communication sont détectées à l'aide d'un seuil temporel maximum (Δt_{max}) entre

deux transmissions. Si la durée entre deux positions consécutives d'une trajectoire est supérieure à ce seuil, la trajectoire est alors supprimée du *GHT* à analyser.

De même, certaines trajectoires peuvent comporter des positions aberrantes (comme indiqué sur la figure 3.6) dues à un dysfonctionnement du système de géolocalisation ou de transmission. Ces positions sont détectées lors du calcul des données de cap et de vitesse en comparant la vitesse calculée (v_{calc}) avec la vitesse maximale (Δv_{max}) pouvant être atteinte par un objet mobile du type étudié. Les trajectoires contenant des positions aberrantes sont également retirées du *GHT* étudié.

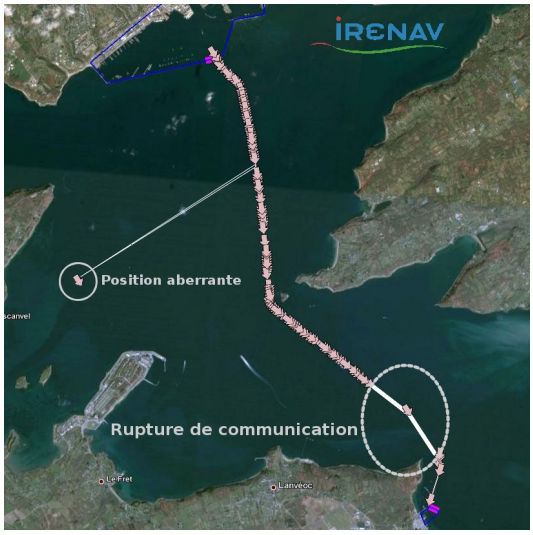


FIGURE 3.6 – Suppression des données erronées

Dans le cadre de notre exemple, parmi les 554 trajectoires, 500 trajectoires ont été conservées après filtrage des trajectoires erronées. Le nombre de trajectoires constituant le *GHT* après filtrage est par conséquent très largement suffisant pour la génération par analyse statistique de la route-type et du couloir spatio-temporel présentés dans la suite de cette thèse.

3.2.2 Recalage spatial

Afin de disposer de trajectoires dont les positions de départ et d'arrivée sont indépendantes du moment d'émission, les parties de la trajectoire à l'intérieur des zones de départ et d'arrivée doivent également être filtrées. Sans ce filtrage, des biais importants peuvent être mesurés sur le patron spatio-temporel défini dans les étapes suivantes. Dans notre exemple, le nuage des positions de départ d'un *GHT* est représenté sur la figure 3.7(a). Il induit un écart entre les positions les plus éloignées pouvant atteindre jusqu'à 200 m.

Les nouvelles positions de départ sont calculées en interpolant la position qui intersecte la ligne de départ virtuelle (représentant la frontière de la zone Z_D) et chaque trajectoire du GHT. Il en est de même pour la zone d'arrivée. L'estampille temporelle de cette nouvelle position de départ est interpolée en fonction de la vitesse initiale de la position de départ. Le résultat du recalage spatial appliqué à notre exemple pour la zone de départ de l'itinéraire Brest → Lanvéoc, est illustré par la figure 3.7(b).

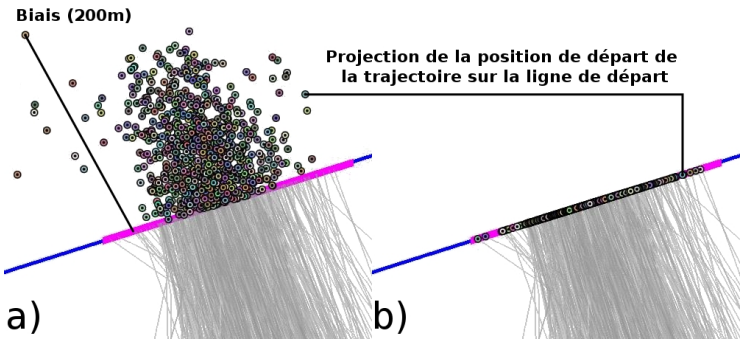


FIGURE 3.7 – Recalage spatial des trajectoires

3.2.3 Filtrage de Douglas et Peucker spatio-temporel

Une fois le recalage spatial effectué, afin d'optimiser les temps de calcul, les trajectoires sont simplifiées en utilisant un filtrage basé sur une technique initialement proposée par Douglas et Peucker [Douglas & Peucker 1973]. Différentes variantes de cet algorithme sont comparées dans les travaux de [Wu & Pelot 2007]. Dans le cadre de cette étude, l'algorithme de Douglas et Peucker spatio-temporel [Bertrand *et al.* 2007] est utilisé.

L'objectif de ce filtrage est de ne conserver que les positions significatives représentant la trajectoire, à savoir des changements de vitesse ou de cap. Pour effectuer cette simplification, la plus grande distance d_{max} entre les positions p_i de la trajectoire T et leurs projections spatio-temporelles p'_i sur la droite reliant les positions de départ p_d et d'arrivée p_a est calculée.

La figure 3.8 illustre l'algorithme 5 en représentant une trajectoire dont les positions de départ et d'arrivée sont p_d et p_a. Le point p'_i correspond à l'interpolation de la position p_i sur la droite $[p_d p_a]$ au temps t_{p_i} en utilisant la vitesse v_{p_d} de p_d. Si la distance d entre p_i et p'_i est supérieure à un seuil fixé (εd), la position la plus éloignée (p_{max}) est conservée. La trajectoire est alors divisée en deux à cette position et l'algorithme est appliqué de façon récursive aux deux sous-parties. Si la distance d est inférieure à εd, seules les positions p_d et p_a sont conservées.

Algorithme 5 Filtrage de Douglas et Peucker spatio-temporel

Algorithme Douglas_Peucker_ST(Trajectoire T)
Soit $[p_d p_a]$ la droite reliant les positions de départ p_d et d'arrivée p_a de T
Soit v_d la vitesse de l'objet mobile en p_d
pour chaque Position p_i de T **faire**
    Calculer la position p'_i interpolée sur la droite $[p_d p_a]$ à l'instant t_{p_i} à partir du point p_d à la vitesse v_d
    Calculer la distance d entre p_i et p'_i
    si $d > d_{max}$ **alors**
        $d_{max} = d$
        $p_{max} = p_i$
si $d_{max} > seuil$ **alors**
    Scinder T à la position p_{max} en deux sous-trajectoires T_1 et T_2
    Concaténer $Douglas_Peucker_ST(T_1)$ et $Douglas_Peucker_ST(T_2)$
sinon
    Ne conserver que les positions p_d et p_a de la trajectoire T

L'algorithme 5 présente en détail les différentes étapes de cette simplification de trajectoire. Celui-ci permet également de supprimer les erreurs locales dues par exemple à des imprécisions des appareils de mesure [Bertrand *et al.* 2007]. Dans le cadre de notre exemple, l'application du filtrage de Douglas et Peucker spatio-temporel au *GHT* a permis de réduire le nombre nécessaire de positions pour représenter efficacement les trajectoires de 104 201 à 16 110 positions soit un taux de compression de 84,54 % pour un seuil de 10 m (correspondant à la précision d'un GPS). Le temps CPU nécessaire à l'exécution des étapes de filtrage du *GHT* appliqué à notre exemple est de 934 ms.

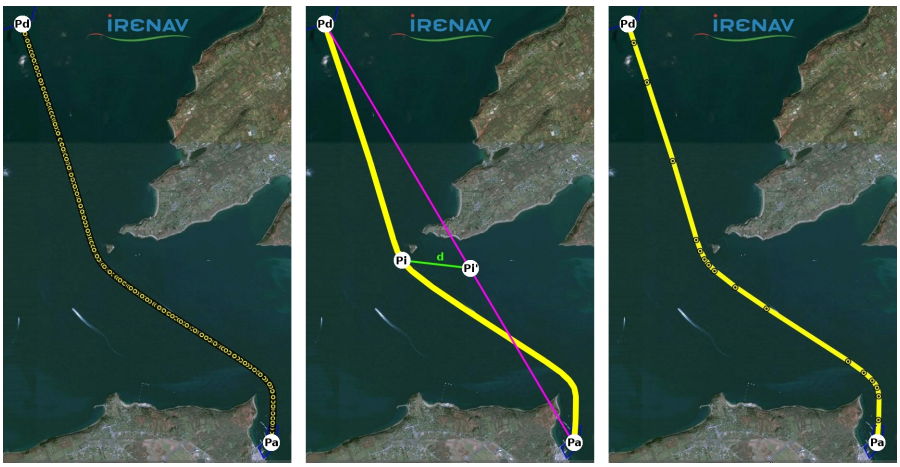


FIGURE 3.8 – Illustration de l'algorithme de Douglas et Peucker spatio-temporel

3.2.4 Échantillonnage d'un groupe homogène de trajectoires

Après avoir été filtrées afin de corriger les données, les trajectoires du GHT disposent d'un pas d'échantillonnage variable. C'est pourquoi, les trajectoires du GHT sont toutes rééchantillonnées au même pas d'échantillonnage spatial (une position tous les 100 mètres soit environ 120 positions par trajectoire pour le GHT choisi comme exemple applicatif), afin de faciliter le processus d'appariement décrit dans la section 2.2.6.

Le choix de ce pas d'échantillonnage spatial aura un impact direct sur les temps de calculs et la précision du patron spatio-temporel généré à partir du GHT. En effet, plus le pas d'échantillonnage spatial sera fin, plus le patron spatio-temporel sera précis mais plus le temps de calcul pour apparier les positions homologues de trajectoires sera important.

Une étude de l'impact de l'échantillonnage sur la précision du calcul de distance entre deux trajectoires est présenté dans le tableau 7 de [Mascret 2010]. Le groupe homogène de trajectoires filtrées et rééchantillonnées ainsi obtenu est désormais propice à une étude statistique spatiale et temporelle.

3.2.5 Normalisation temporelle d'un groupe homogène de trajectoires

Afin d'obtenir la trajectoire-type suivie par les objets mobiles, une analyse statistique spatiale des positions des trajectoires du GHT est effectuée. Il est nécessaire de s'assurer que la dimension temporelle n'impacte pas l'analyse spatiale, c'est pourquoi une normalisation temporelle est réalisée. Celle-ci permet d'éviter les distorsions spatiales introduites par des vitesses légèrement différentes des trajectoires composant le GHT.

Pour réaliser cette normalisation temporelle, trois estampilles temporelles sont affectées à chaque position d'une trajectoire. Les positions initiales des trajectoires disposent toutes d'une estampille temporelle représentant le temps absolu (t_A) au format UTC auquel l'objet mobile était à cette position.

Connaissant le temps absolu de départ de la trajectoire, il est possible d'affecter à chaque position un temps relatif (t_R) représentant la durée écoulée depuis la position de départ de la trajectoire. Enfin, un temps relatif normalisé (t_{RN}) est affecté à chaque position du GHT par le calcul suivant.

En premier lieu, la durée médiane $\tilde{D}$ du GHT est calculée. Puis, une fonction de normalisation temporelle est appliquée à toutes les trajectoires du GHT afin que chaque trajectoire commence à un temps initial relatif t_{R_0} nul et finisse au même temps relatif médian normalisé $\tilde{t}_{RN} = t_0 + \tilde{D}$.

Au vu des données exploitées, le choix de la durée médiane du GHT permet d'obtenir une route-type moins sujette aux perturbations dues aux trajectoires ayant un comportement inhabituel. Le temps CPU nécessaire à l'exécution de la normalisation temporelle du GHT appliqué à notre exemple est de 33 ms.

3.3 Génération d'un patron spatio-temporel

Une fois le GHT extrait de la $BDST$, filtré, rééchantillonné et normalisé, une analyse statistique de ce GHT peut être réalisée dans l'objectif de définir le type de motifs spatio-temporels le plus adapté aux clusters de trajectoires. La figure 3.9 présente la carte de densité des trajectoires du GHT affichées avec un taux de transparence de 1%. L'observation de la figure 3.9 permet de constater des variations importantes de la densité et de l'amplitude du groupe de trajectoires. Dans les parties droites, on remarque que l'amplitude est plus importante que dans les virages. L'analyse statistique du GHT a pour objectif de qualifier cette différence de densité et d'amplitude.

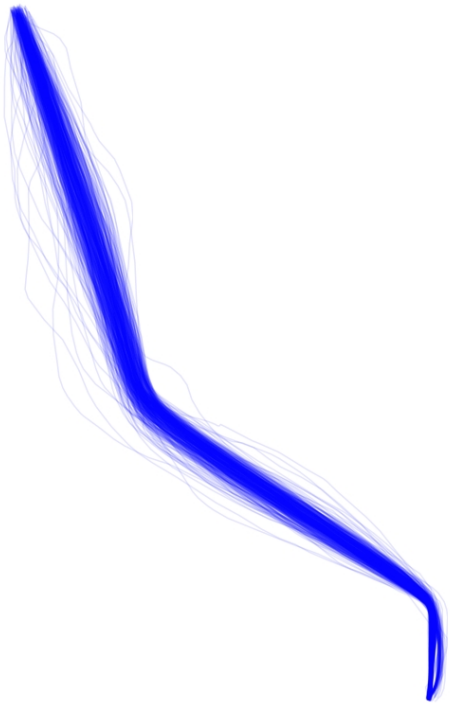


FIGURE 3.9 – Carte de densité des trajectoires du GHT (transparence 1%)

Le processus de génération du patron spatio-temporel est décomposé en plusieurs étapes qui sont détaillées sur l'étape 4 de la figure 3.10. Ces différentes étapes sont détaillées dans les sections suivantes.

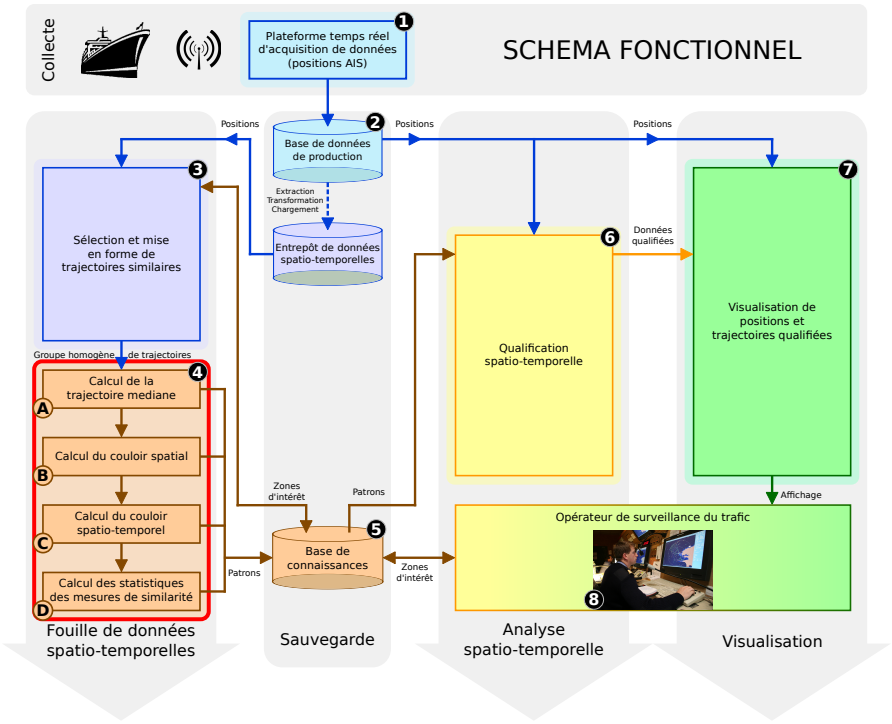


FIGURE 3.10 – Schéma fonctionnel de l'étape de génération du patron spatio-temporel

3.3.1 Le concept des boîtes à moustaches

Comme indiqué précédemment, nous souhaitons réaliser une analyse statistique du GHT afin d'étudier la densité, la symétrie et les variations d'amplitude d'un patron spatio-temporel représentant de manière synthétique le comportement des objets mobiles de ce GHT. En statistiques, une manière classique de résumer les caractéristiques de séries statistiques quantitatives qui varient en termes d'amplitude et de densité consiste à utiliser les boîtes à moustaches.

Ce concept des boîtes à moustaches (boxplot) a été introduit par Tukey dans les années 1970 [Tukey 1977, McGill *et al.* 1978]. Les boîtes à moustache permettent de représenter de façon synthétique la distribution d'un ensemble de données en fonction de certaines de ces caractéristiques (médiane, quartiles, écart interquartile, percentiles) [Le Guen 2001]. Différentes représentations ont été proposées dans la littérature [Potter 2006].

La figure 3.11 présente un exemple de boîte à moustache ainsi qu'une distribution gaussienne associée de moyenne μ et d'écart type σ. Généralement, une boite à moustache est constituée d'un rectangle (la boîte), bleu sur la figure 3.11, dont les extrémités correspondent respectivement au premier ($Q1$) et troisième quartile ($Q3$). 50% des données observées sont comprises dans cette boîte. La valeur de

la médiane est également représentée dans ce rectangle par un trait (vert). Enfin, les moustaches associées à la boîtes représentent les limites des valeurs au delà desquelles les données seront considérées comme exceptionnelles ou aberrantes (outlier).

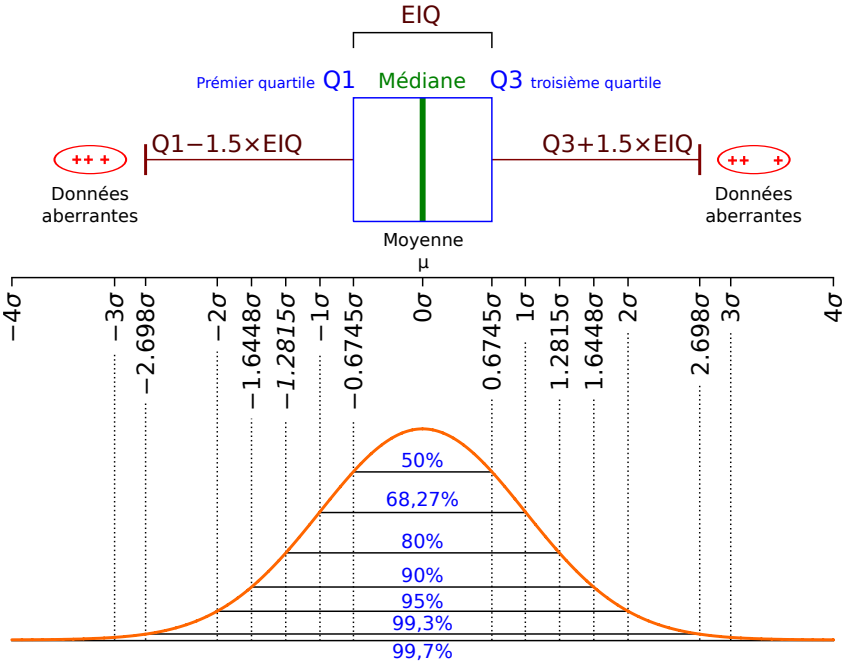


FIGURE 3.11 – Exemple de représentation d'une distribution gaussienne sous la forme d'une boite à moustache

Dans la littérature, différentes méthodes de calcul sont proposées pour fixer ces limites [Tukey 1977]. Dans certains cas, ces limites peuvent tout simplement représenter les valeurs extrêmes de la distribution (le minimum et le maximum). Une autre méthode consiste à utiliser l'écart interquartile (EIQ) pour spécifier les limites des moustaches. L'écart interquartile est donné par la formule 3.1 :

$$EIQ = |Q3 - Q1| \tag{3.1}$$

Les limite des moustaches au delà desquelles les données sont considérées comme aberrantes sont obtenues en ajoutant $\pm 1.5 * EIQ$ aux quartiles $Q1$ et $Q3$. La limite inférieure est donnée par la formule $L_{min} = Q1 - 1.5 * EIQ$. La limite supérieure est obtenue par la formule $L_{max} = Q3 + 1.5 * EIQ$.

En fonction du domaine applicatif et des besoins de visualisation, d'autres limites basées sur les percentiles peuvent être choisies. Ces limites sont définies en fonction de l'intervalle de confiance souhaité variant généralement entre 90 et 100%. Le tableau 3.1 présente la taille de l'intervalle de confiance exprimée en % en fonction de l'écart type σ.

Le motif spatio-temporel proposé dans les sections suivantes s'inspire de cette représentation statistique synthétique d'une variable à une dimension et propose une extension aux données spatio-temporelles.

$\mu \pm n * \sigma$	pourcentage de la population dans l'intervalle
$\mu \pm 0,6745\sigma$	50%
$\mu \pm 1\sigma$	68,27%
$\mu \pm 1,2815\sigma$	80%
$\mu \pm 1,6448\sigma$	90%
$\mu \pm 2\sigma$	95%
$\mu \pm 2,698\sigma$	99,3%
$\mu \pm 3\sigma$	99,7%

TABLE 3.1 – Pourcentage de la population contenue dans l'intervalle $\mu \pm n * \sigma$ pour une distribution gaussienne

3.3.2 Calcul d'une trajectoire-type

A partir du GHT obtenu et filtré, le calcul de la trajectoire-type suivie par la majorité des trajectoires de ce GHT peut être effectué. L'objectif de cette étape est d'obtenir, par analyse statistique spatiale, la trajectoire-type d'un itinéraire suivi par les objets mobiles d'un même type.

La première étape du processus de génération du patron spatio-temporel (figure 3.10 étape 4.A) consiste à calculer la trajectoire-type des objets mobiles suivie par le GHT (dans notre exemple il s'agit des "navires à passagers" effectuant l'itinéraire Brest → Lanvéoc). Cette trajectoire est composée d'une suite ordonnée des positions médianes des trajectoires du GHT.

Pour calculer ces positions médianes, il est nécessaire de regrouper les différentes positions homologues des trajectoires entre elles. Les différentes techniques d'appariement de positions homologues de trajectoires présentées au chapitre 2 peuvent être employées dans l'objectif d'obtenir ces groupes de positions homogènes.

3.3.2.1 Calcul d'une trajectoire-type par normalisation temporelle

Une première solution simple à mettre en œuvre consiste à regrouper toutes les positions des trajectoires à un même temps relatif normalisé calculé à la section 3.2.5. Ainsi, pour chaque temps normalisé, on dispose d'un nuage de positions homologues ($\mathcal{N}$) des trajectoires du GHT à ce temps normalisé.

Cette solution nécessite préalablement une normalisation temporelle du GHT ainsi qu'un calcul d'interpolation à chaque temps relatif normalisé pour chaque trajectoire du GHT. Afin d'optimiser les calculs, il est possible de ne sélectionner que les t_{RN} des positions des trajectoires filtrées par l'algorithme de Douglas et Peucker spatio-temporel présenté à la section 3.2.3.

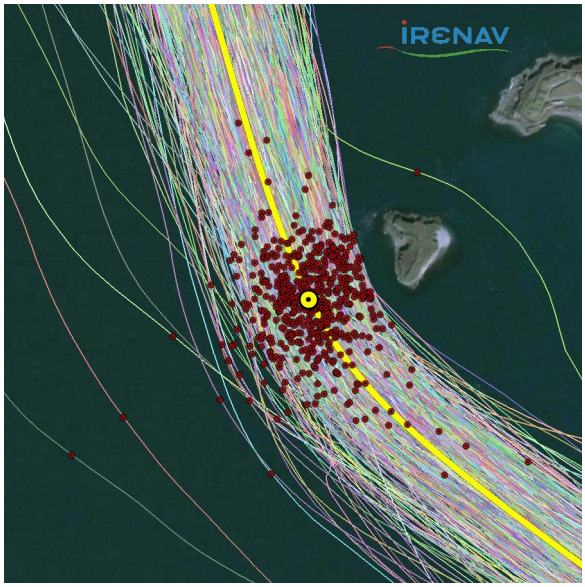


FIGURE 3.12 – Position médiane agrégée d'un nuage de positions homologues de GHT

Ainsi, les t_{RN} des positions clés conservées par le filtrage sont sélectionnés permettant d'intensifier les calculs sur les zones du GHT disposant de fortes variations de vitesse ou de direction et minimiser les calculs dans les zones. Considérant N le nombre de trajectoires du GHT et M le nombre de pas de temps normalisé utilisé pour générer les trajectoires types, la complexité du calcul de la trajectoire-type par normalisation temporelle est de $O(N \times M)$.

Les positions médianes ont pour coordonnées la latitude médiane et la longitude médiane des nuages de positions des trajectoires du GHT au même t_{RN}. Les positions médianes obtenues sont ordonnées temporellement afin de constituer une nouvelle trajectoire-type que nous nommerons ci-après trajectoire médiane ($\tilde{T}$). La figure 3.12 représente les trajectoires (traits clairs fins) du GHT, un nuage de positions à un même temps normalisé (points rouges) et la position médiane associée (cercle jaune).

L'algorithme 6 synthétise cette phase d'élaboration de la trajectoire médiane d'un GHT par normalisation temporelle. Le temps CPU nécessaire à l'étape de calcul de la trajectoire-type par normalisation temporelle du GHT appliqué à notre exemple est de 14 462 ms.

Algorithme 6 Calcul d'une trajectoire type par normalisation temporelle

Précondition :

pour chaque trajectoire T du GHT **faire**
  Suppression des trajectoires erronées
  Recalage spatial des trajectoires
  Douglas_Peucker_ST(Trajectoire T)
  Normalisation temporelle des trajectoires à une durée médiane $\tilde{D}$
Algorithme Calcul_Trajectoire_Médiane_Normalisé(GHT)
pour chaque trajectoire T_m du GHT **faire**
  pour chaque position p_{m_i} de T_m **faire**
    Soit $t_{RN_{m_i}}$ le temps relatif normalisé de la position p_{m_i}
    pour toutes les autres trajectoires T_n du GHT **faire**
      Interpoler la position p_{n_j} à $t_{RN_{m_i}}$
      Ajouter p_{n_j} au nuage $\mathcal{N}t_{RN_{m_i}}$
    Calculer la position médiane $\tilde{p_i}$ du nuage $\mathcal{N}t_{RN_{m_i}}$ à $t_{RN_{m_i}}$
    Affecter $t_{RN_{m_i}}$ comme temps relatif normalisé à $\tilde{p_i}$
    Ajouter $\tilde{p_i}$ à la trajectoire médiane $\tilde{T}$
retour $\tilde{T}$

La trajectoire-type obtenue par application de l'algorithme 6 sur l'exemple applicatif du GHT est représentée par la trajectoire jaune épaisse sur les figures 3.12 et 3.13.

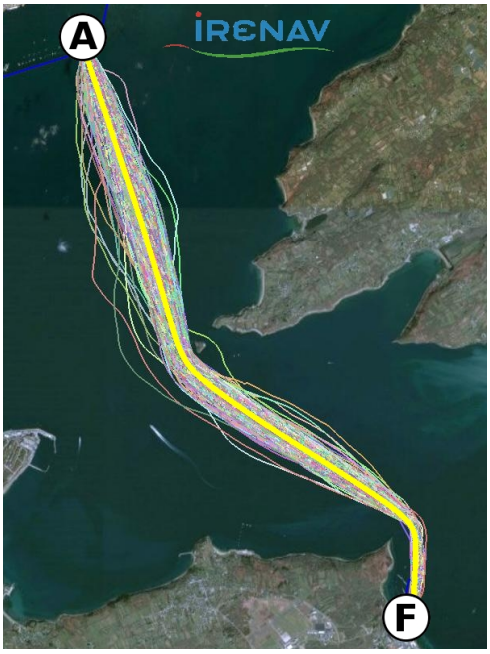


FIGURE 3.13 – Trajectoire médiane du GHT de l'itinéraire A-F (Brest → Lanvéoc)

3.3.2.2 Calcul d'une trajectoire-type par appariement de trajectoires

Une seconde approche consiste à choisir une trajectoire de référence (T_{ref}) afin d'y apparier toutes les autres trajectoires du GHT. Le choix d'une trajectoire de référence permet d'éviter d'effectuer un processus d'appariement de toutes les trajectoires du GHT deux à deux. Cette trajectoire de référence choisie parmi les trajectoires du GHT dispose d'une durée et d'une longueur la plus proche de la durée et de la longueur médiane du GHT.

Ainsi, pour choisir cette trajectoire de référence, il est nécessaire de calculer préalablement la durée médiane $\tilde{D}$ ainsi que sa longueur médiane $\tilde{L}$ du GHT. La sélection de cette trajectoire de référence permet de définir une suite de positions de références auxquelles seront rattachées les positions homologues des autres trajectoires.

Cette technique s'approche de la technique des K-mean [MacQueen 1967] pour laquelle le nombre de positions initiales (K) correspond au nombre de positions de la trajectoire de référence. Cependant, en cas de boucle dans la trajectoire, les positions d'entrée et de sortie de boucle (au niveau du croisement) sont toutes appariées au même cluster, cette technique ne prenant pas en compte l'aspect temporel des positions.

Les techniques d'appariement basées sur une approche utilisant la programmation dynamique (DTW, Fréchet) présentées au chapitre 3 permettent de respecter cet ordonnancement temporel en définissant des couples de positions appariées ($\mathcal{C}$) entre deux trajectoires.

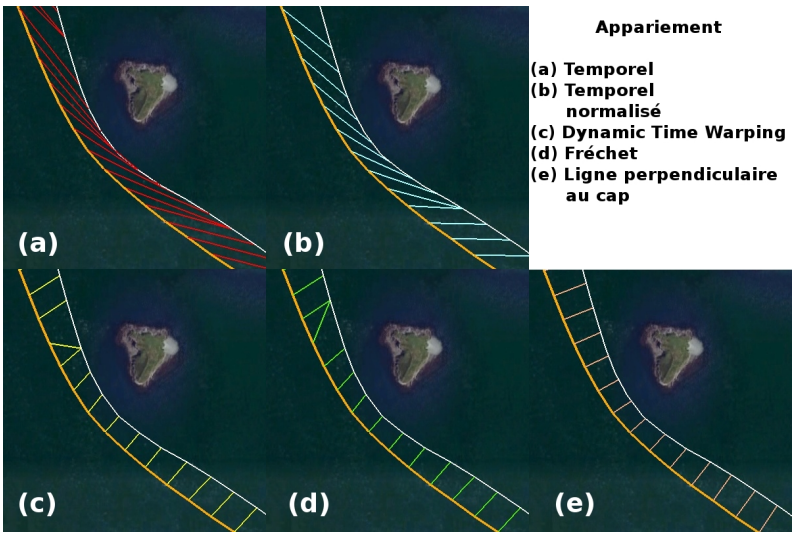


FIGURE 3.14 – Comparatif des techniques d'appariement de positions de trajectoires

Ces différentes techniques permettent d'apparier des positions de trajectoires en respectant l'ordonnancement temporel de ces positions. Les distances spatiales et temporelles entre les positions appariées de deux trajectoires peuvent être comparées afin d'en déduire leur similarité. Plus les distances spatiales et temporelles entre positions appariées sont grandes, moins la similarité est forte.
Plusieurs techniques sont proposées pour apparier les positions de deux trajectoires.

Différents critères permettent d'évaluer ces techniques d'appariement :
- la technique doit pouvoir apparier des trajectoires d'objets mobiles ayant des vitesses différentes (alignement temporel),
- la technique doit prendre en compte l'intégralité des positions des trajectoires tout en respectant leur ordonnancement temporel,
- l'appariement partiel doit être possible afin de pouvoir qualifier des trajectoires en temps réel (déplacements en cours).

Les techniques testées sont les suivantes :
- Appariement des positions au même temps relatif (durée depuis le départ).
  Ce choix donne de mauvais résultats si les vitesses sont dissemblables.
- Appariement des positions au même temps relatif normalisé (durée depuis le départ de la trajectoire exprimée en pourcentage de sa durée totale).
  Ce choix ne permet pas d'apparier des trajectoires partielles, non encore terminées.
- Appariement des positions selon la technique du "Dynamic Time Warping" (DTW) [Sakoe & Chiba 1978].
  Le DTW est une méthode visant à optimiser l'appariement entre deux séries de données temporelles. Ces séries sont déformées par une transformation non linéaire de la variable temporelle afin de mesurer leur similarité de façon indépendante du temps. Ce processus issu du traitement du signal donne de bons résultats, mais reste difficilement utilisable pour des trajectoires partielles.
- Appariement des positions selon la distance de Fréchet discrète [Devogele 2000, Devogele 2002, Mascret *et al.* 2006].
  Ce processus d'appariement fournit des couples proches de ceux définis par le DTW, il a l'avantage d'être utilisable pour des trajectoires partielles.
- Appariement des positions selon la droite perpendiculaire au cap (DPC) [Etienne *et al.* 2009, Etienne *et al.* 2010b].
  Cette technique consiste à relier une position d'une trajectoire avec la position de l'autre trajectoire qui intersecte la DPC de la première. Cette méthode est relativement rapide et utilisable pour des trajectoires partielles, cependant, elle n'est pas symétrique. Contrairement aux techniques précédentes, les positions de la trajectoire appariée doivent être interpolées.

La figure 3.14 illustre les 5 méthodes d'appariement proposées appliquées à deux trajectoires discrètes d'objets mobiles. Les couples de positions appariées sont reliés deux à deux par des trais fins. La technique de Fréchet [Fréchet 1905, Alt *et al.* 2001, Devogele 2000, Devogele 2002, Mascret *et al.* 2006] appliquée sur deux trajectoires discrètes permet d'apparier des trajectoires partielles en respectant la relation d'ordre temporel des positions des trajectoires. Cette technique a l'avantage de pouvoir s'appliquer à des trajectoires partielles et ne nécessite pas de calcul d'interpolation. C'est pourquoi, nous l'utiliserons pour les différentes tâches d'appariement de positions de trajectoires.

Ainsi, chaque trajectoire du GHT est appariée à la trajectoire de référence T_{ref} en utilisant la technique d'appariement de Fréchet. Les nuages de positions homologues ($\mathcal{N}$) associés à chaque position de T_{ref} sont constitués. De même que pour la première solution, la position médiane agrégée de chaque nuage est calculée, en ayant pour coordonnées la latitude médiane et la longitude médiane des nuages de positions.

En ce qui concerne le t_{RN} de cette position médiane, le temps relatif médian du nuage de positions homologues $\tilde{t}_{RN}$ est calculé et affecté à la position médiane. Les positions médianes obtenues sont ordonnées temporellement afin de constituer une nouvelle trajectoire-type ci après nommée trajectoire médiane ($\tilde{T}$).

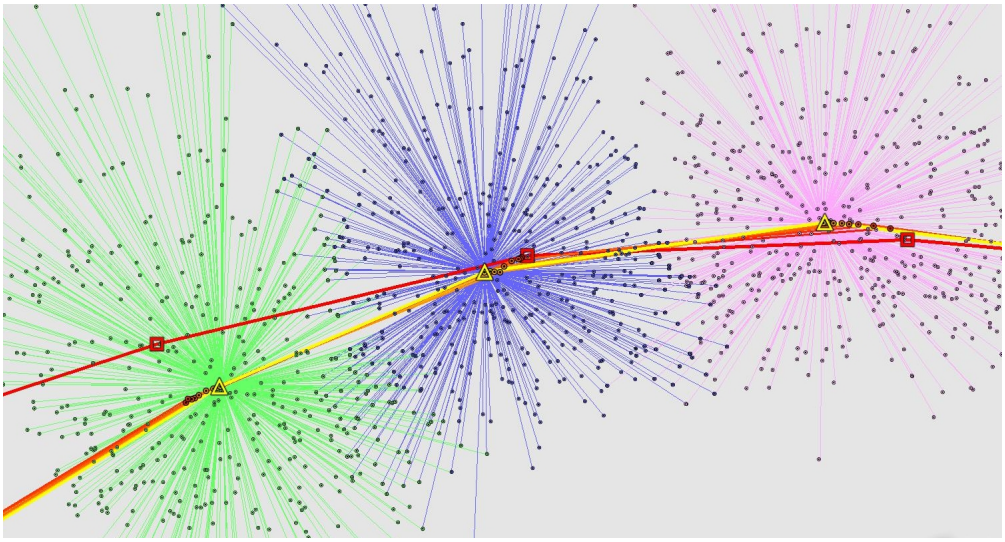


FIGURE 3.15 – Nuages de positions homologues et trajectoires de référence à différentes itérations de l'algorithme de génération de la trajectoire-type par appariement de Fréchet

De façon similaire à l'algorithme des K-mean, ce processus de génération de la trajectoire-type est appliqué récursivement en utilisant $\tilde{T}$ comme nouvelle trajectoire de référence jusqu'à ce que la différence entre T_{ref} et $\tilde{T}$ soit inférieure à un seuil de tolérance ε_{med} permettant d'accélérer cette étape du processus.

La figure 3.15 présente 3 nuages de positions homologues (petits points de couleur rouge, verte ou rose en fonction du nuage) appariées à chaque position médiane (triangle jaune) de la trajectoire médiane (ligne jaune).

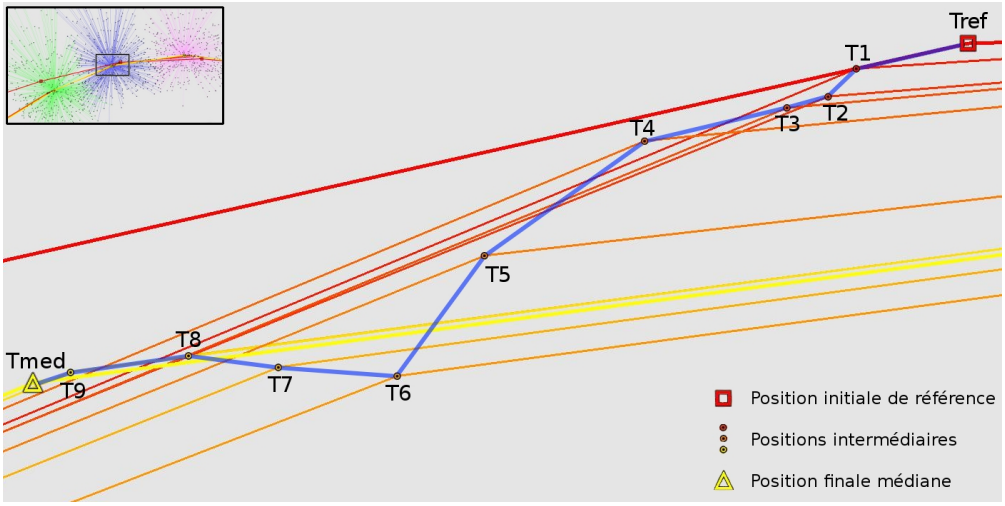


FIGURE 3.16 – Évolution de la position médiane d'un nuage de positions homologues en fonction du nombre d'itérations de l'algorithme de génération de la trajectoire-type par appariement de Fréchet

Le processus itératif de raffinement de la trajectoire-type est présenté plus précisément sur la figure 3.16 détaillant l'évolution de la position médiane d'un même indice de position de référence (correspondant à un zoom sur le nuage bleu de la figure 3.15). Sur cette figure, la trajectoire initiale T_{ref} est présentée en rouge, la position initiale servant de référence pour l'appariement des autres positions homologues des trajectoires du GHT est représentée par un carré rouge. Les positions intermédiaires calculées sont représentées par des points de couleurs rouge, orange et jaune et numérotés de $T1$ à $T9$. Les trajectoires intermédiaires et positions médianes du nuage associées aux différentes étapes du processus récursif de calcul de la trajectoire-type sont représentées par un dégradé de lignes de couleur orange. Les positions médianes du nuage sont représentées par des points orange numérotés indiquant le niveau de récursivité de l'algorithme. Finalement, la trajectoire jaune correspond à la trajectoire-type calculée et dont la position médiane du nuage est représentée par un triangle jaune. Dans le cadre de notre exemple, la trajectoire-type a été obtenue après 10 itérations.

L'algorithme 7 synthétise cette phase d'élaboration de la trajectoire-type d'un GHT par appariement. Notons que toutes les techniques d'appariement présentées au chapitre 2 peuvent être utilisées.

Algorithme 7 Calcul d'une trajectoire-type par appariement

Précondition :

pour chaque trajectoire T du GHT **faire**
  Suppression des trajectoires erronées
  Rééchantillonnage spatial du GHT
  Calcul de la durée médiane $\tilde{D}$ du GHT
  Calcul de la longueur médiane $\tilde{L}$ du GHT
  Choix de la trajectoire du GHT dont la longueur et la durée sont les plus proches de $\tilde{L}$ et $\tilde{D}$ comme T_{ref}
Algorithme Calcul_Trajectoire_Médiane_Appariement(GHT,T_{ref})
pour chaque trajectoire T_m du GHT **faire**
  Soit $\mathcal{C}T_m$ les couples de positions appariées entre T_m et T_{ref}
  pour chaque position p_{r_i} de T_{ref} **faire**
    Sélectionner le sous-ensemble $\mathcal{P}_i$ des positions de $\mathcal{C}T_m$ appariées à p_{r_i}
    si Cardinalité de $\mathcal{P}_i > 1$ **alors**
      Calculer la position moyenne $\overline{p}_i$ de $\mathcal{P}_i$
      Ajouter $\overline{p}_i$ au nuage $\mathcal{N}_i$ de positions homologues
    sinon
      Ajouter $\mathcal{P}_i$ au nuage $\mathcal{N}_i$ de positions homologues
pour chaque nuage $\mathcal{N}_i$ **faire**
  Calculer la position médiane $\tilde{p}_i$ du nuage $\mathcal{N}_i$
  Calculer le temps relatif médian $\tilde{t}_{R_i}$ du nuage $\mathcal{N}_i$
  Affecter $\tilde{t}_{R_i}$ comme temps relatif à $\tilde{p}_i$
  Ajouter $\tilde{p}_i$ à la trajectoire médiane $\tilde{T}$
si La distance de Fréchet entre $\tilde{T}$ et T_{ref} est inférieure à un seuil ε_{med} **alors**
  retour $\tilde{T}$
sinon
  retour Calcul_Trajectoire_Médiane_Appariement(GHT,$\tilde{T}$)

Pour l'exemple applicatif présenté, la technique de l'appariement de Fréchet a été utilisée, la trajectoire médiane a été obtenue par convergence après 10 itérations. Le temps CPU nécessaire à une étape d'appariement des 500 trajectoires du GHT à une trajectoire de référence pour un pas d'échantillonnage de 100 m (120 positions par trajectoire) est de 113 128 ms. Le temps CPU nécessaire à l'étape de calcul de la trajectoire-type par appariement de trajectoires du GHT appliqué à notre exemple est de 1 035 711 ms. Les deux techniques proposées pour calculer la trajectoire type sont très différentes. La première technique est plus rapide mais nécessite des calculs d'interpolation et pose problème dans le cas de trajectoires réalisant des boucles ou des virages serrés. La seconde technique est robuste à ce type de problèmes mais est 70 fois plus lente.

3.3.2.3 Impact du choix de la technique d'agrégation et de l'échantillonnage sur le calcul de la trajectoire-type

A partir des positions de la trajectoire-type, les vitesses et les caps de ces positions médianes sont déduits par calcul. L'utilisation d'une approche médiane permet de limiter le biais pouvant être lié à des trajectoires anormales ayant un écart très important. De plus, en fonction de l'échantillonnage spatial ou temporel choisi à l'étape 3.2.4 de rééchantillonnage des trajectoires, il est possible que plusieurs positions d'une même trajectoire soient appariées à la même position de T_{ref} comme présenté dans le cercle orange de la figure 3.17.

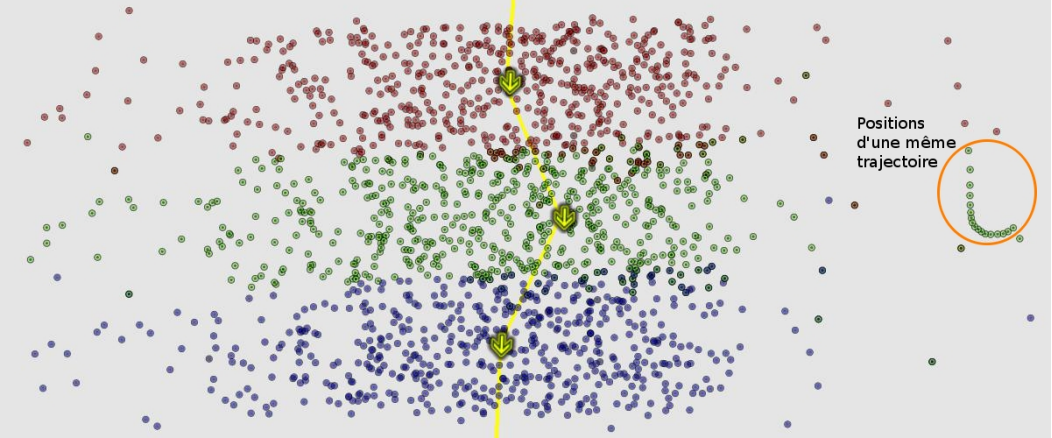


FIGURE 3.17 – Impact d'une trajectoire marginale disposant d'un nombre important de positions appariées à un même nuage

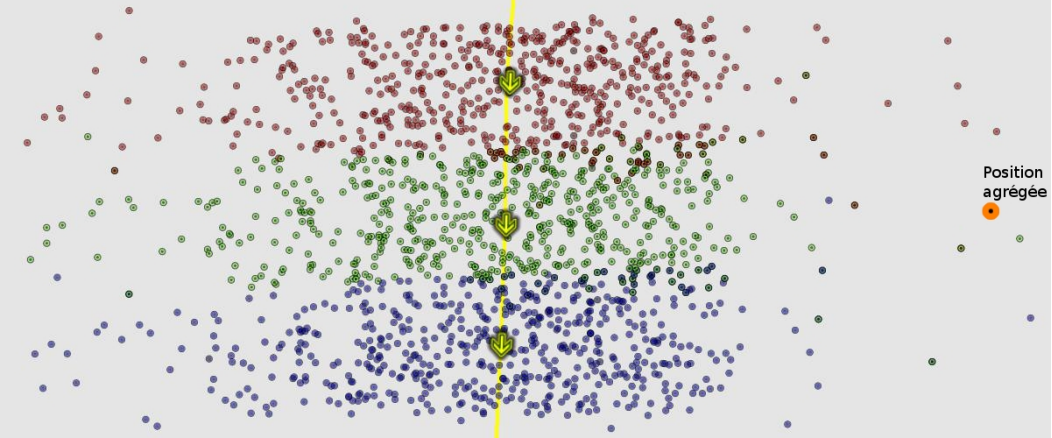


FIGURE 3.18 – Correction de l'impact d'une trajectoire disposant d'un nombre important de positions appariées à un même nuage par agrégation

Afin d'éviter de donner un poids plus important à ces trajectoires, une moyenne des positions d'une même trajectoire est calculée par nuage (représenté par le point

orange de la figure 3.18). Ce genre d'appariement multiple est problématique ; on souhaite éviter au maximum qu'un nombre important de positions d'une même trajectoire soit appariées à un même nuage.

Les techniques d'appariement telles que le Dynamic Time Warping ou l'appariement de Fréchet peuvent apparier plusieurs positions d'une trajectoire à une même position de la trajectoire de référence comme indiqué dans la section 2.2.4. Ce genre d'appariement multiple est cependant fortement diminué lorsque les trajectoires ont le même pas d'échantillonnage spatial.

L'exemple caractéristique de la figure 3.17 représente des nuages de positions appariées pour un GHT rééchantillonné à un pas temporel de 10 s. Une trajectoire anormale effectue un fort ralentissement dans la zone du nuage de positions vert.

De nombreuses positions de cette trajectoire (positions entourées par un cercle orange) sont donc appariées à une même position de référence ayant par conséquent un poids plus fort et un impact au niveau du calcul de la position médiane. Notons que cet impact aurait été encore plus important dans le cadre d'une approche moyenne pour le calcul de la position agrégée. C'est pourquoi le GHT est rééchantillonné au même pas d'échantillonnage spatial comme indiqué dans la section 3.2.4.

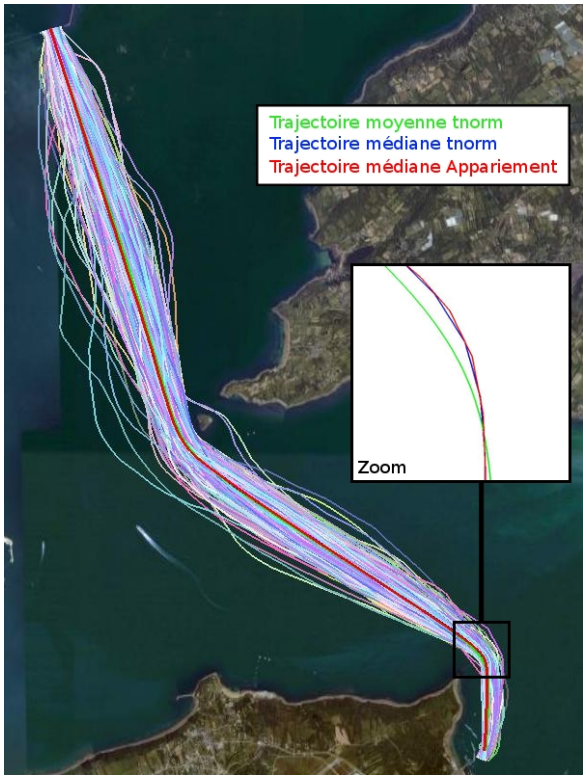


FIGURE 3.19 – Comparatif des trajectoires types obtenues par approche moyenne, médiane ou appariement

La figure 3.19 présente les trajectoires types obtenues en fonction des différentes techniques utilisées pour le GHT de l'exemple ainsi qu'un zoom sur les trois trajectoires types générées. La trajectoire verte représente la trajectoire-type générée en utilisant la technique de la normalisation temporelle et le calcul de la position moyenne de toutes les positions des trajectoires du GHT à un même temps normalisé. La trajectoire bleue représente la trajectoire-type obtenue en utilisant la même technique mais en prenant la position médiane. Enfin, la trajectoire rouge représente la trajectoire-type calculée à partir des positions médianes des nuages de positions homologues appariées comme présenté à la section 3.3.2.2.

Notons que les deux approches utilisant la médiane donnent des résultats très proches, alors que l'approche moyenne est plus fortement impactée par la présence de trajectoires ayant un comportement anormal et déviant de façon importante et asymétrique (principalement observable dans les zones de virages contraintes par un obstacle à terre comme présenté sur le zoom de la figure 3.19). Cette observation confirme l'intérêt de l'extension de l'approche statistique des boîtes à moustaches introduite à la section 3.3.1. C'est pourquoi dans la suite de cette étude, l'approche médiane sera utilisée et la trajectoire-type choisie sera la trajectoire médiane ($\tilde{T}$).

Une fois calculée, $\tilde{T}$ représente le comportement médian des objets mobiles d'un même type suivant le même itinéraire. Il est ainsi possible de comparer la trajectoire d'un objet mobile (de même type et itinéraire) avec $\tilde{T}$ et d'en déduire des similitudes ou différences de position, orientation, vitesse sur tout ou partie de cette trajectoire.

3.3.3 Calcul d'un couloir spatial

Une fois la trajectoire médiane obtenue, celle-ci doit être complétée afin de connaître l'écartement possible normal d'un point de vue spatial puis d'un point de vue temporel. Cette représentation peut être vue comme une extension 2D du concept des boîtes à moustaches pour la représentation de données statistiques spatiales et temporelles.

Dans cet objectif, premièrement, un couloir spatial est défini, contenant $P\%$ des trajectoires à droite et $P\%$ des trajectoires à gauche. La seconde étape du processus de génération du patron spatio-temporel (figure 3.10 étape 4.B) consiste à calculer ce couloir spatial dans lequel la majorité des trajectoires du GHT évoluent.

L'observation de la figure 3.20 permet de constater que les objets mobiles d'un même type effectuant un même itinéraire suivent globalement une même route. L'espace dans lequel ces objets évoluent étant ouvert, il est normal de constater de légers écarts spatiaux. Ceux-ci doivent être distingués d'écarts plus importants considérés comme inhabituels.



FIGURE 3.20 – Carte de densité et trajectoire médiane du Groupe homogène des trajectoires de navires à passagers effectuant l'itinéraire Brest Arsenal → Lanvéoc Ecole Navale

Par exemple, les deux trajectoires sortant du GHT pour passer à l'est de l'île sur la figure 3.20 doivent être considérées comme étant anormales. De plus, ont peut également observer sur cette figure que certaines trajectoires s'écartent plus ou moins du GHT puis y retournent. Ces écarts sont donc considérées comme anormaux mais seulement sur une portion de la trajectoire.

Ces écarts inhabituels représentent un certain pourcentage ($P\%$) du GHT. Afin de séparer les comportements normaux des situations inhabituelles, un couloir spatial est calculé à l'aide d'une analyse statistique des positions des trajectoires du GHT vis-à-vis de la trajectoire médiane calculée à la section 3.3.2. Ce paramètre P permet de fixer le seuil statistique au delà duquel un écart à la trajectoire médiane est considéré comme anormal.

Ainsi, les portions de trajectoires dépassant ce seuil ne seront pas intégrées dans le motif spatio-temporel synthétisant le comportement habituel des trajectoires. En fonction du paramètre P choisi, le nombre de trajectoires d'objets mobiles considérées comme anormales sera donc plus ou moins important. Afin de diminuer la charge cognitive de l'opérateur de surveillance du trafic comme souligné page 7, celui-ci peut se focaliser principalement sur ces trajectoires inhabituelles.

Ce pourcentage P, actuellement fixé empiriquement à 10%, sera utilisé dans le reste de cette étude. Il doit impérativement être fixé avant la génération du patron spatio-temporel en tenant compte du contexte applicatif et du niveau de discrimination recherché. Il serait intéressant d'effectuer une étude plus approfondie de l'impact de ce paramètre sur l'étape de génération du patron spatio-temporel dans de futurs travaux de recherche.

L'exemple de la figure 3.20 présente le GHT de la figure 3.4 ainsi que la carte de densité associée. La carte de densité est réalisée en affectant un taux de transparence de 1% aux trajectoires du GHT. Cet affichage permet de visualiser les zones de forte densité en bleu foncé ou de nombreuses trajectoires se superposent.

3.3.3.1 Calcul d'un couloir spatial par la technique de la droite perpendiculaire au cap

Disposant de la trajectoire médiane du GHT nous souhaitons désormais connaître pour chaque position médiane l'écartement spatial habituel des trajectoires du GHT ainsi qu'une limite au-delà de laquelle l'écartement est considéré comme anormal. Cependant, cet écartement spatial n'est pas constant ni symétrique comme on peut l'observer sur la figure 3.4.

En effet, on remarque visuellement sur cette figure que, sur les portions en ligne droite sans obstacles, les trajectoires ont tendance à s'écarter plus facilement de la trajectoire médiane et de façon symétrique. Alors qu'à l'approche de l'île, les trajectoires contraintes par l'obstacle ont tendance à se concentrer au plus proche de celui-ci entraînant une forte densité proche de l'île ainsi qu'un écart asymétrique. Une analyse statistique de la distribution des positions par rapport à la position médiane permettra de confirmer cette observation et sera présentée en détail à la section 3.4.

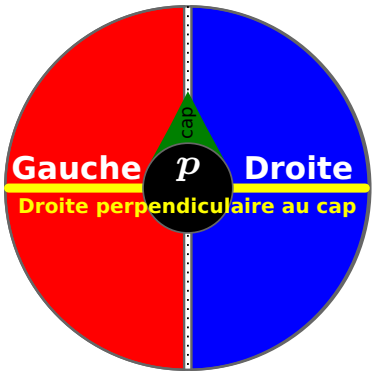


FIGURE 3.21 – Classement des positions à droite ou à gauche en fonction du cap de la position médiane

Afin de connaître cet écartement spatial pour chaque position médiane, les mesures des distances spatiales entre la trajectoire médiane et toutes les autres trajectoires du GHT doivent être calculées et ordonnées. Pour ordonner par distance les positions des trajectoires du GHT par rapport à la trajectoire médiane, la droite perpendiculaire au cap (DPC) de chaque position de la trajectoire médiane est utilisée. Cette droite est représentée en jaune sur la figure 3.21.

Les points d'intersection interpolés entre la DPC et chaque trajectoire du GHT sont classés de part et d'autre de $\tilde{T}$ par ordre de distance. Les points sont classés comme étant à gauche ou à droite en fonction de leur position relative au cap de la position de la trajectoire médiane comme l'illustre la figure 3.21 pour laquelle le cap de la position p est représenté par un triangle vert. Par convention, les distances spatiales des positions situées à gauche de la position médiane sont notées négativement et celles situées à droite positivement. Cette distinction de côté permet de calculer des limites statistiques différentes pour les côtés gauche et droit du couloir spatial autorisant ainsi la génération d'un couloir asymétrique.

Une fois les positions ordonnées par distance et par côté sur la DPC, une analyse statistique de la distribution des distances spatiales peut alors être réalisée. Cette analyse est basée sur le paramètre P préalablement introduit. Les positions dont les distances spatiales correspondent respectivement aux percentiles $P\%$ et $1-P\%$ sont sélectionnées comme étant les bornes de gauche (Δ_{G_i}) et de droite (Δ_{D_i}) du couloir spatial. Ces bornes calculées pour chaque position médiane $\tilde{p}_i$ sont sauvegardées. Les trajectoires composées des positions correspondant aux percentiles précédemment cités composent le couloir spatial de navigation.

Le choix de ces déciles statistiques permet d'obtenir les bornes spatiales du couloir au sein duquel la majorité des objets mobiles suivant cet itinéraire circulent. De plus, celui-ci conditionne le nombre de trajectoires considérées comme étant anormales. Ce paramètre P pouvant être adapté en fonction de la sensibilité et du volume de données disponibles est actuellement fixé de manière empirique à 10%.

De même que pour la trajectoire médiane comme exposé à la section 3.3.2.3, l'utilisation d'une approche médiane/décile plutôt que moyenne/variance permet d'éviter que la génération du couloir spatio-temporel soit perturbée par des trajectoires ayant un écart à la route-type très important. Le temps CPU nécessaire à la génération du couloir spatial appliqué à notre exemple est de 1 887 ms.

La figure 3.22 illustre le calcul de ce couloir pour notre exemple. La DPC est représentée sur la figure 3.22.b par une ligne jaune. Les frontières spatiales du couloir de droite et de gauche obtenues en fonction du choix de différents percentiles sont présentées par les frontières rouges (25%), vertes (50%) et bleues (90%). Les points d'intersection interpolés entre la DPC et les trajectoires du GHT sont représentés

par des points rouges pour ceux positionnés à droite de la position médiane et bleus pour ceux positionnés à gauche.

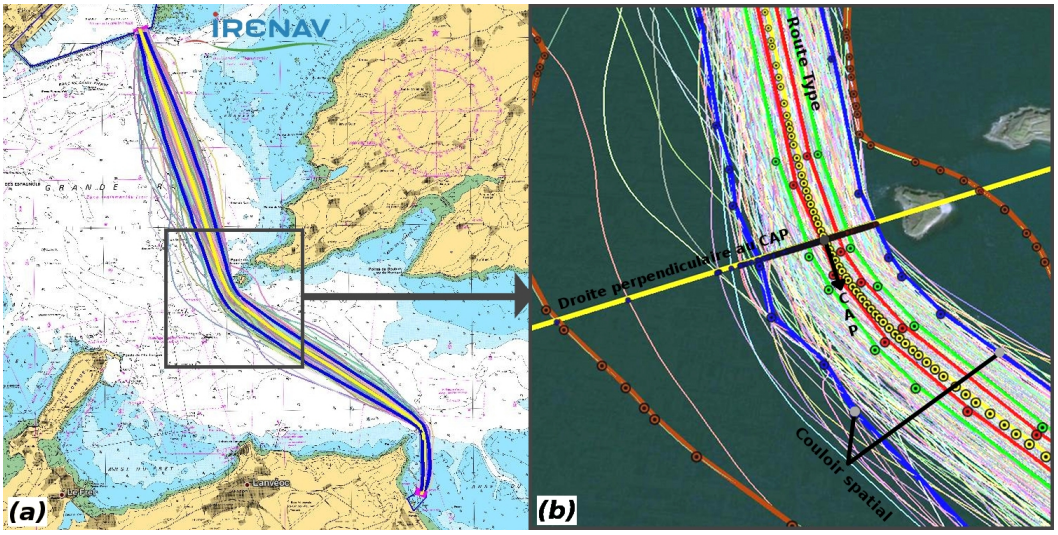


FIGURE 3.22 – Calcul du couloir spatial par la technique de la droite perpendiculaire au cap

Il est donc possible d'identifier pour chaque GHT les positions inhabituelles se trouvant en dehors du couloir spatial. Il faut noter que la distance entre la route-type et la frontière de droite et de gauche n'est pas identique. Effectivement, dans un espace ouvert, il est plus facile pour un objet mobile de dévier vers l'extérieur lorsqu'il tourne. De même, la largeur du couloir fournit une indication concernant l'écartement des trajectoires vis-à-vis de la route-type.

Dans notre exemple, comme le montre la figure 3.22.a, cet écartement est plus faible au départ, à l'arrivée et dans les virages. Par contre, dans les lignes droites, celle-ci a tendance à augmenter. Notons que le choix du paramètre P fixé empiriquement à 10% semble visuellement être un bon choix pour discriminer les trajectoires anormales (figure 3.22.b, frontière bleue). L'algorithme 8 synthétise les différentes étapes de calcul du couloir spatial. Ainsi, chaque position de la trajectoire médiane dispose d'une limite de distance spatiale à gauche Δ_G et à droite Δ_D orientée en fonction du cap de la position médiane. Cette technique peut néanmoins poser problème en cas de boucles dans le GHT. En effet, dans cette situation, plusieurs intersections seraient calculées pour chaque trajectoire du GHT avec la DPC. De plus, dans les situations où les différences de cap entre deux positions successives de la trajectoire médiane sont importantes, l'ordonnancement temporel des positions des trajectoires du GHT risque de ne pas être respecté.

Algorithme 8 Calcul d'un couloir spatial par calcul de DPC

Algorithme Calcul_Couloir_Spatial_DPC(GHT)
pour chaque position $\tilde{p}_i$ de la trajectoire médiane $\tilde{T}$ **faire**
  Calculer la droite DPC_i perpendiculaire au cap en $\tilde{p}_i$
  pour chaque trajectoire T_j du GHT **faire**
    Calculer le point d'intersection p'_i entre T_j et DPC_i
    si p'_i est à droite de $\tilde{p}_i$ **alors**
      Ajouter p'_i à l'ensemble $\mathcal{D}_i$
    sinon
      Ajouter p'_i à l'ensemble $\mathcal{G}_i$
  Classer $\mathcal{G}_i$ par ordre de distance spatiale avec $\tilde{p}_i$ décroissante
  Classer $\mathcal{D}_i$ par ordre de distance spatiale avec $\tilde{p}_i$ croissante
  p_{G_i} = position correspondant au percentile P de $\mathcal{G}_i$
  p_{D_i} = position correspondant au percentile $(1-P)$ de $\mathcal{D}_i$
  Δ_{G_i} = - distance spatiale entre p_i et p_{G_i}
  Δ_{D_i} = distance spatiale entre p_i et p_{D_i}
  Ajouter p_{D_i} à la trajectoire frontière T_D du couloir spatial
  Ajouter p_{G_i} à la trajectoire frontière T_G du couloir spatial

L'exemple de la figure 3.23 illustre ce problème. La DPC de la position p_a de $\tilde{T}$ est représentée par la droite rouge. La DPC de la position p_b de $\tilde{T}$ est représentée par la droite orange. Les points d'intersection entre les DPC et la trajectoire d'exemple T (bleu clair) située de l'autre côté de l'île sont représentés par des carrés de couleur orange et rouge (position p_1 et p_2).

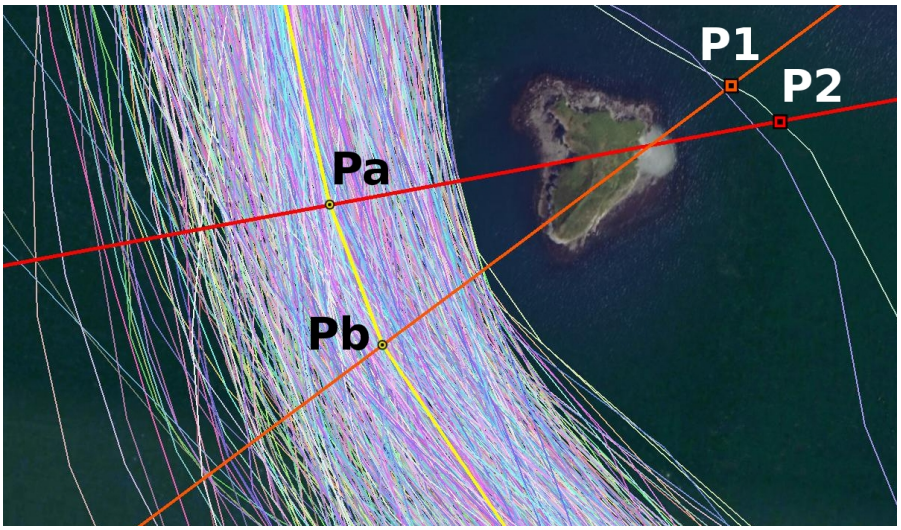


FIGURE 3.23 – Intersection de deux lignes perpendiculaires aux caps de positions de la trajectoire médiane

L'ordre des positions de $\tilde{T}$ est tel que $p_a < p_b$. De même, en ce qui concerne la trajectoire T, $p_1 < p_2$. On peut noter que ces deux DPC s'intersectent, ce qui a pour conséquence d'inverser la relation d'ordre temporelle entre les positions p_1 et p_2 lors de leur prise en compte dans le calcul du couloir.

Ainsi, c'est la position p_2 de la trajectoire T qui sera utilisée pour le calcul des statistiques du couloir en p_a alors que la position p_1 sera utilisée pour le calcul du couloir en p_b. La technique de la DPC est donc une technique basée sur un appariement de positions purement spatial regroupant les points les plus proches sur une même droite passant par la position médiane.

Cette technique ne tient pas compte de l'aspect temporel des positions des trajectoires du GHT. Ce non respect de l'ordonnancement des positions des trajectoires du GHT est problématique car il induit un bais dans les calculs statistiques. C'est pourquoi, au cours de cette thèse, nous avons cherché à améliorer la technique de génération du couloir spatial en proposant une autre technique basée sur les appariements de positions homologues présentée dans la section suivante.

3.3.3.2 Calcul d'un couloir spatial par la technique des appariements de positions homologues

Une autre technique peut également être employée pour obtenir les bornes du couloir spatial. Cette technique est basée sur l'analyse statistique des nuages de positions homogènes calculés lors de la génération de la trajectoire médiane par appariement de Fréchet comme présenté à la section 3.3.2.2.

Pour chaque nuage de positions homologues de la trajectoire médiane, les positions du nuage sont classées en deux sous-ensembles $\mathcal{G}$ et $\mathcal{D}$ en fonction du côté où se trouve la position par rapport au cap de la position médiane comme indiqué sur la figure 3.21. Les distances spatiales entre le sous-ensemble $\mathcal{G}$ des positions du nuage situées à gauche et la position médiane sont calculées et notées négativement. Les distances spatiales entre le sous-ensemble $\mathcal{D}$ des positions du nuage situées à droite et la position médiane sont calculées et notées positivement. Ces distances spatiales sont ensuite triées par ordre croissant.

De façon similaire à la méthode précédente, une analyse statistique de la distribution des distances spatiales peut alors être réalisée. Cette analyse est également dépendante du paramètre P introduit à la section 3.3.3. Les positions pour lesquelles les distances spatiales correspondent respectivement aux percentiles $P\%$ et $1-P\%$ sont sélectionnées comme étant les bornes de gauche (Δ_{G_i}) et de droite (Δ_{D_i}) du couloir spatial. Ces bornes calculées pour chaque position médiane $\tilde{p}_i$ sont sauvegardées. Les trajectoires composées des positions correspondant aux percentiles précédemment cités définissent le couloir spatial de navigation.

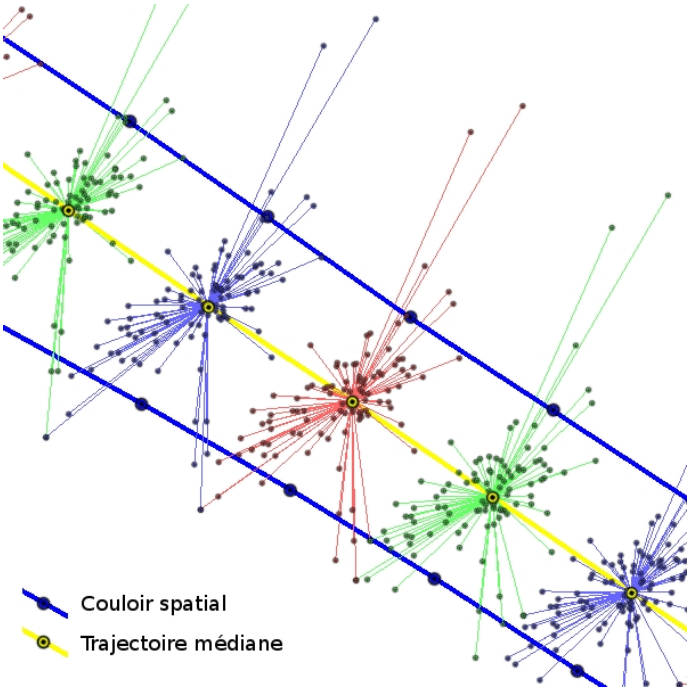


FIGURE 3.24 – Bornes du couloir spatial du GHT obtenu par analyse statistique des distances entre positions homogènes de nuages et position de la trajectoire médiane obtenue par appariement de Fréchet

De même qu'indiqué à la section 3.3.3, le choix du paramètre P permet d'obtenir les bornes spatiales du couloir au sein duquel la majorité des objets mobiles suivant cet itinéraire circulent et conditionne le nombre de trajectoires considérées comme étant anormales. L'algorithme 9 synthétise les différentes étapes de calcul du couloir spatial.

La figure 3.24 présente des nuages de positions homologues (petits points de couleur rouge, vert ou bleu) appariées par la technique de Fréchet aux positions de la trajectoire médiane représentées par des gros points jaunes. Pour chaque nuage, les positions dont la distance spatiale à la médiane correspondent au percentile P à gauche et $(1 - P)$ à droite sont représentées par de gros points bleus qui constituent les bornes du couloir spatial représenté en bleu épais sur la figure 3.24. La valeur P fixée empiriquement à 10% se situe visuellement dans la zone ou des variations significatives sont présentes dans l'histogramme. Dans un souci de lisibilité, le nombre de points composant chaque nuage a été restreint par rapport au GHT de l'exemple.

Cette technique reste cependant fortement dépendante de l'échantillonnage spatial du GHT. En effet, plus l'échantillonnage spatial des trajectoires du GHT est fin, plus les nuages obtenus ressemblent à une courbe s'approchant de la droite perpendiculaire au cap.

Algorithme 9 Calcul d'un couloir spatial par appariement de positions homologues

Algorithme Calcul_Couloir_Spatial_Appariement(GHT)
Appariement des trajectoires du GHT à la trajectoire médiane $\tilde{T}$
pour chaque position $\tilde{p}_i$ de la trajectoire médiane $\tilde{T}$ **faire**
    Soit $\mathcal{N}_i$, le nuage des positions homologues appariées à $\tilde{p}_i$
    pour chaque positions p_j du nuage $\mathcal{N}_i$ **faire**
        si p_j est à droite de $\tilde{p}_i$ **alors**
            Ajouter p_j à l'ensemble $\mathcal{D}_i$
        sinon
            Ajouter p_j à l'ensemble $\mathcal{G}_i$
    Classer $\mathcal{G}_i$ par ordre de distance spatiale avec $\tilde{p}_i$ décroissante
    Classer $\mathcal{D}_i$ par ordre de distance spatiale avec $\tilde{p}_i$ croissante
    p_{G_i} = position correspondant au percentile P de $\mathcal{G}_i$
    p_{D_i} = position correspondant au percentile $(1-P)$ de $\mathcal{D}_i$
    Δ_{G_i} = - distance spatiale entre p_i et p_{G_i}
    Δ_{D_i} = distance spatiale entre p_i et p_{D_i}
    Ajouter p_{D_i} à la frontière T_D du couloir spatial
    Ajouter p_{G_i} à la frontière T_G du couloir spatial
    Sauvegarder Δ_{G_i}
    Sauvegarder Δ_{D_i}

La figure 3.25 présente la trajectoire type (en jaune) ainsi que le couloir spatial (en bleu) du groupe homogène de trajectoires de type "navires à passagers" entre Brest et Lanvéoc. Les trajectoires types et couloirs spatiaux de trois différents GHT sont présentées en annexes A.8, A.9 et A.10.

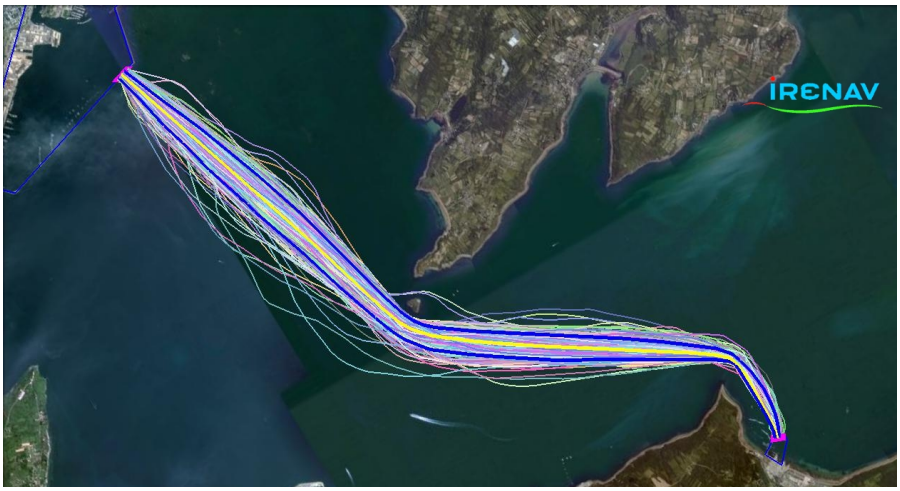


FIGURE 3.25 – Trajectoire type et couloir spatial du GHT "Brest-Lanvéoc"

3.3.4 Calcul d'un couloir temporel

Sachant désormais qu'un objet mobile se situe ou pas dans le couloir spatial de l'itinéraire qu'il est censé suivre, il est intéressant de calculer si celui-ci est dans les temps par rapport à la route-type. La troisième étape du processus de génération du patron spatio-temporel (figure 3.10 étape 4.C) consiste à calculer le couloir temporel dans lequel la majorité des trajectoires du GHT évolue.

De manière similaire au couloir spatial, il est possible de calculer un couloir temporel permettant d'analyser le retard ou l'avance d'un objet mobile. Il regroupe $P\%$ des trajectoires en retard et $P\%$ des trajectoires en avance à l'intérieur du couloir spatial. Seules les positions à l'intérieur du couloir spatial sont prises en considération pour ce calcul statistique. Pour cela, lors du calcul du couloir spatial, les temps relatifs (durée depuis le départ) des positions des nuages se trouvant dans le couloir spatial sont pris en compte.

Le choix des nuages de positions dépend de la technique utilisée. Ce sont les positions interpolées sur la droite perpendiculaire au cap pour la méthode de la section 3.3.3.1, ou les nuages de positions homologues pour la méthode de la section 3.3.3.2. Les temps relatifs de chacune de ces positions sont ensuite classés par ordre croissant pour chaque nuage.

Par convention, les distances temporelles des positions en retard par rapport à la position médiane sont notées négativement et celles en avance positivement. La sélection des percentiles P et $(1 - P)$ pour les positions par rapport au temps normalisé de la position de la trajectoire médiane permet d'obtenir des bornes temporelles d'avance (Δ_{A_i}) et de retard (Δ_{R_i}) de notre couloir pour chaque position $\tilde{p}_i$ de la trajectoire médiane. Le temps CPU nécessaire à la génération du couloir temporel appliqué à notre exemple est de 1 889 ms.

Il faut noter que les écarts temporels entre une position de la route-type et les bornes associées du couloir temporel ne sont pas identiques. Effectivement, il est plus facile pour un objet mobile de prendre du retard que de l'avance. De plus, seules les positions à l'intérieur du couloir spatial sont prises en compte car une position qui sort du couloir spatial peut soit prendre un raccourci, soit effectuer un détour, ce qui ne nous permet donc pas de qualifier la position comme étant en avance ou en retard. L'algorithme 10 résume les différentes étapes de calcul du couloir temporel.

Algorithme 10 Calcul d'un couloir temporel

Précondition :

Appariement des trajectoires du GHT à $\tilde{T}$ (Fréchet ou DTW)
Ou
Calcul des DPC des trajectoires du GHT aux positions médianes de $\tilde{T}$
Algorithme Calcul_Couloir_Temporel(GHT)
pour chaque position $\tilde{p}_i$ de la trajectoire médiane $\tilde{T}$ **faire**
    Soit $\mathcal{N}_i$, le nuage des positions homologues appariées à $\tilde{p}_i$
    pour chaque position p_j du nuage $\mathcal{N}_i$ **faire**
        si p_j se trouve dans le couloir spatial **alors**
            si p_j est en avance par rapport à $\tilde{p}_i$ **alors**
                Ajouter p_j à l'ensemble $\mathcal{A}_i$
            sinon
                Ajouter p_j à l'ensemble $\mathcal{R}_i$
    Classer $\mathcal{R}_i$ par ordre de distance temporelle avec $\tilde{p}_i$ décroissante
    Classer $\mathcal{A}_i$ par ordre de distance temporelle avec $\tilde{p}_i$ croissante
    p_{R_i} = position correspondant au percentile P de $\mathcal{R}_i$
    p_{A_i} = position correspondant au percentile $(1 - P)$ de $\mathcal{A}_i$
    Δ_{R_i} = - distance temporelle entre p_i et p_{R_i}
    Δ_{A_i} = distance temporelle entre p_i et p_{A_i}
    Sauvegarder Δ_{R_i}
    Sauvegarder Δ_{A_i}

3.3.5 Synthèse d'un couloir spatio-temporel et d'une route-type

Finalement, la route-type (RT) du GHT est composée de la trajectoire médiane enrichie avec les frontières du couloir spatial (Δ_G et Δ_D) et temporel (Δ_A et Δ_R). Ainsi, pour chaque $\tilde{p}_i$ de $\tilde{T}$, on dispose d'un écart spatial toléré à droite (Δ_{D_i}), d'un écart spatial toléré à gauche (Δ_{G_i}), d'un écart temporel toléré de retard (Δ_{R_i}) et d'un écart temporel toléré d'avance (Δ_{A_i}).

Entre deux positions médianes consécutives de la trajectoire médiane, ces écarts spatiaux et temporels sont interpolés linéairement en fonction du temps relatif. A chaque $\tilde{p}_i$, il est donc possible de définir cinq zones spatiales représentées sur la figure 3.26 :

- dans le couloir à l'heure (zone de normalité représentée en vert),
- dans le couloir en retard (zone de retard représentée en rouge),
- dans le couloir en avance (zone d'avance représentée en orange),
- à droite du couloir (zone droite représentée en gris),
- à gauche du couloir (zone gauche représentée en gris).

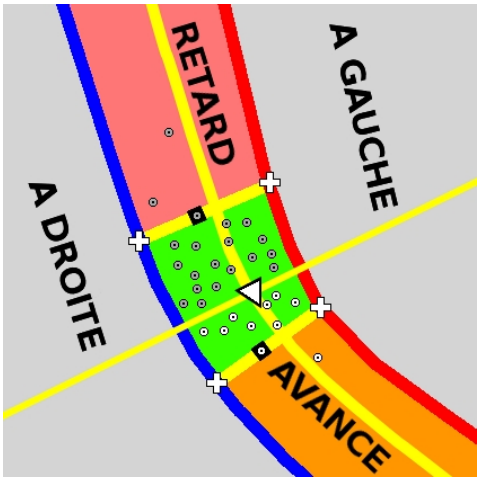


FIGURE 3.26 – Bornes du couloir spatial du GHT obtenu par analyse statistique des distances entre positions homogènes de nuages et position de la trajectoire médiane obtenue par appariement de Fréchet

Une présentation des trajectoires du GHT, de $\tilde{T}$ et du couloir spatial et temporel peut être réalisée en 3D en utilisant l'axe Z pour représenter le temps. Le temps relatif des positions depuis le départ de la trajectoire étant utilisé comme altitude de la position. Le GHT de notre exemple est représenté en 3D sur la figure 3.27. Cette représentation est très proche de la notion de *geospatial lifeline* introduite au chapitre 1.1.4. Elle peut être vue comme une extension des *geospatial lifelines* par l'ajout d'une emprise spatiale et temporelle en plus de la représentation ponctuelle de la position de l'objet.

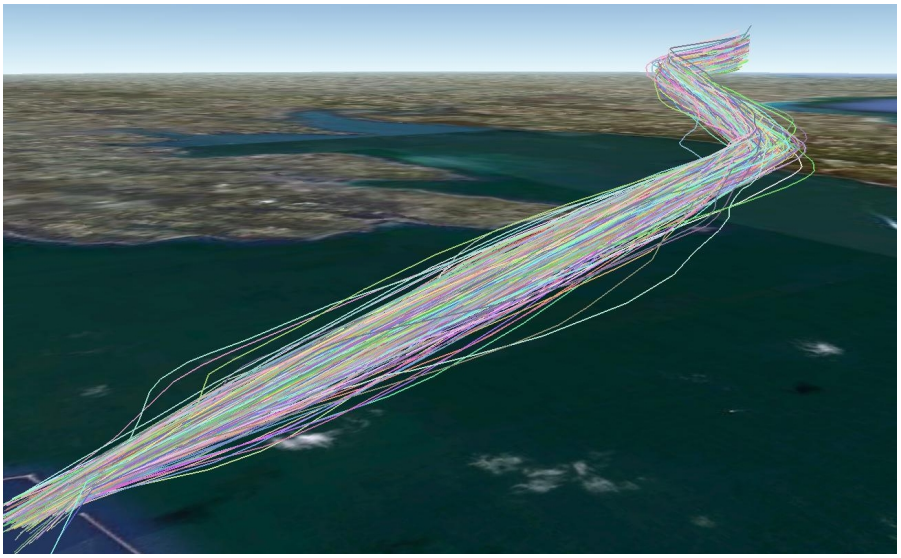


FIGURE 3.27 – Représentation 3D du GHT

Pour chaque position de la trajectoire médiane, les bornes spatiales et temporelles peuvent être combinées afin d'obtenir une zone de normalité spatio-temporelle 3D représentée par des quadrilatères verts sur la figure 3.28. La trajectoire médiane 3D du *GHT* est représentée en jaune, les limites interpolées du couloir spatio-temporel sont représentées par les frontières rouges (au-dessus se trouve la zone de retard) et vertes (au-dessous se trouve la zone d'avance) reliant les zones de normalité. Dans cette représentation 3D, la route type à la forme d'un polyèdre irrégulier à sections rectangulaires variables.

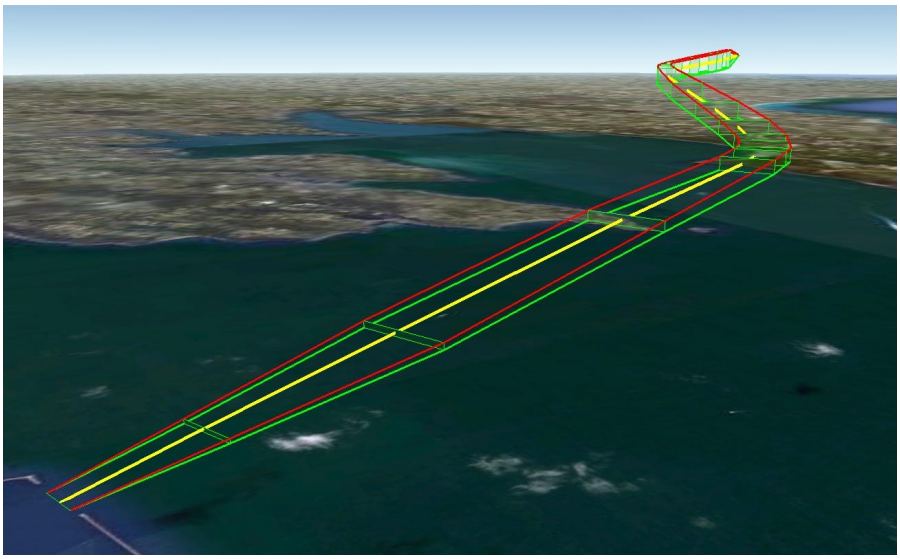


FIGURE 3.28 – Représentation 3D de la trajectoire médiane et du couloir spatio-temporel

Visuellement, ce couloir spatio-temporel est utilisable pour détecter des trajectoires inhabituelles d'objets mobiles n'étant pas contenues dans celui-ci. Cette phase de visualisation présentée à l'étape 7 de la figure 3.10 ne fait pas l'objet d'une étude détaillée dans cette thèse et demanderait une étude approfondie en terme de sémiologie graphique [Bertin 1967, Béguin & Pumain 2010].

Néanmoins, la visualisation 3D de certaines trajectoires inhabituelles permet d'observer le comportement anormal d'une trajectoire sortant du couloir spatio-temporel comme présenté sur la figure 3.29 illustrant une trajectoire déviant spatialement du couloir à droite et sur la figure 3.30 montrant une trajectoire prenant du retard.

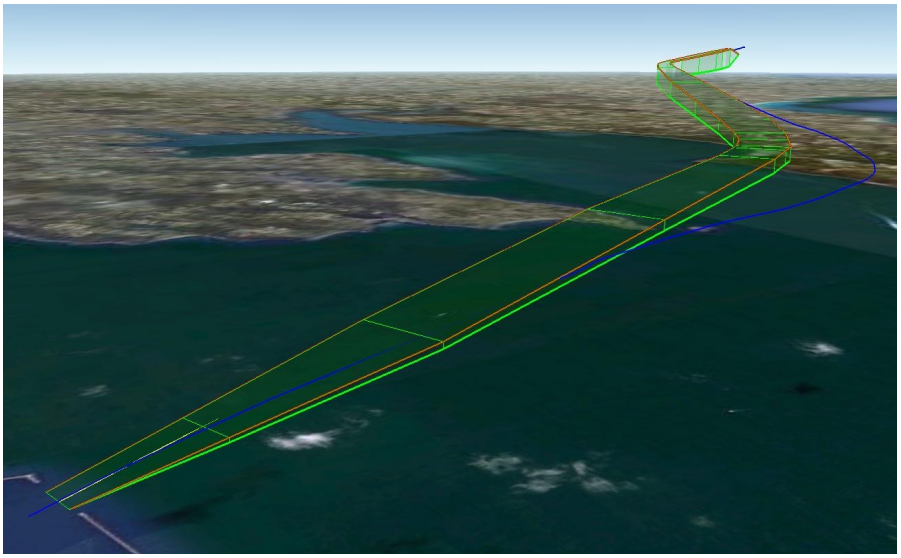


FIGURE 3.29 – Représentation 3D du couloir spatio-temporel et d'une trajectoire sortant du couloir à droite

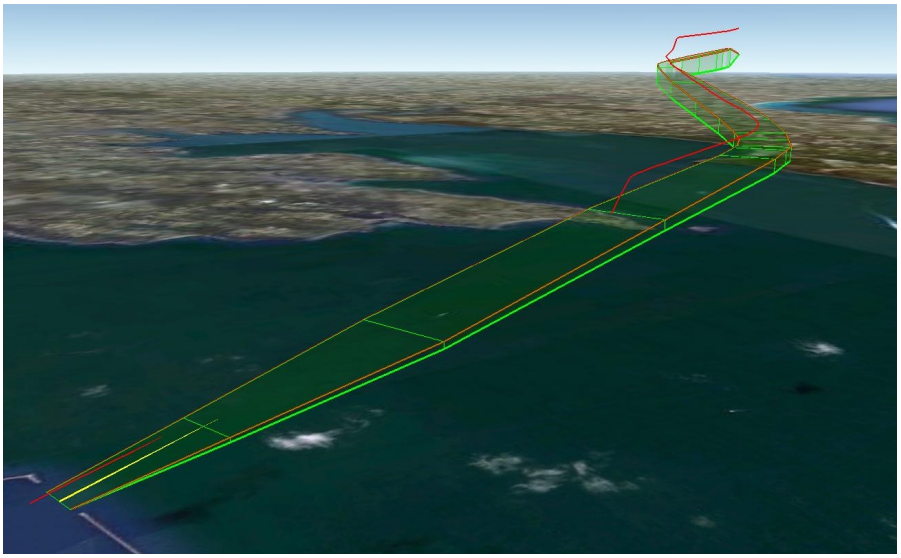


FIGURE 3.30 – Représentation 3D du couloir spatio-temporel et d'une trajectoire prenant du retard

3.4 Distribution des distances spatiales et temporelles des nuages de positions

Lors des étapes de calcul de la trajectoire médiane et du couloir spatio-temporel, des nuages de positions homogènes ont été utilisés. Ces nuages dépendent de la technique utilisée pour apparier les positions. Une approche médiane/percentiles a

été choisie comme exposé à la section 3.3.2.3 pour obtenir la position médiane du nuage ainsi que le choix des limites des frontières spatiales et temporelles du couloir.

Dans cette section, nous souhaitons valider l'approche statistique réalisée pour la génération de la route-type en effectuant une analyse des distributions statistiques des nuages de positions. De plus, nous souhaitons également connaître la façon dont les objets mobiles s'écartent.

Pour l'exemple applicatif, le GHT de 500 trajectoires de navires de type "navires à passagers" effectuant l'itinéraire Brest Arsenal → Lanvéoc Ecole Navale est échantillonné à un pas de 100m soit environ 120 positions par trajectoires pour une distance totale de 12 km. Les histogrammes de densité ont été calculés pour chaque nuage de positions homologues (soit 120 nuages pour notre exemple).

Dans un souci de lisibilité, 3 nuages de positions homologues répartis sur toute la longueur de l'itinéraire sont détaillés :

- le nuage n° 30 à 3 km du départ,
- le nuage n° 60 à 6 km du départ,
- le nuage n° 90 à 9 km du départ.

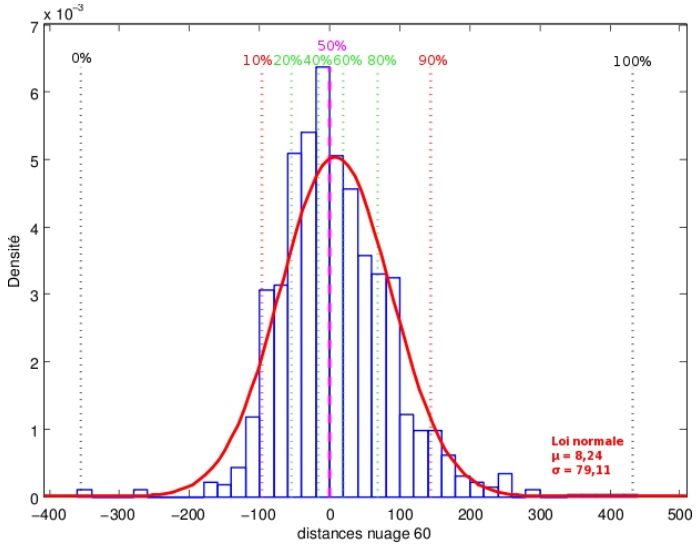


FIGURE 3.31 – Histogramme de densité des écarts de distances des positions du nuage 60/120

La figure 3.31 représente l'histogramme de densité des écarts de distances des positions comparées à la position médiane du nuage n° 60 et la figure 3.32 présente l'histogramme de densité des écarts de temps. Les valeurs des percentiles 20%, 40%, 60% et 80% sont représentées par des lignes pointillées vertes. Les valeurs extrêmes (minimum et maximum) sont représentées par des lignes pointillées noires. Les valeurs correspondant à la limite choisie pour les bornes du couloir (P% et $(1-P)$%)

sont représentées par des lignes pointillées rouges. Une courbe représentant une loi normale dont les paramètres s'approchent au plus près de ceux de la distribution observée est tracée en rouge sur ces figures.

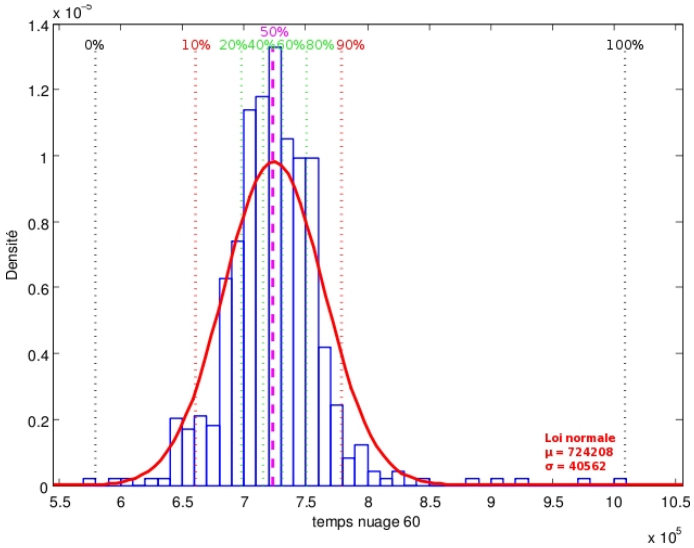


FIGURE 3.32 – Histogramme de densité des écarts de temps des positions du nuage 60/120

Les histogrammes de densité des autres nuages de positions (30, 60, 90) et d'autres caractéristiques des positions (distances, temps, vitesses, caps) sont présentés en annexe A.4, A.5, A.7, A.6.

Appariement :	Temporel	Temporel normalisé	DTW	Fréchet	DPC
Distance moy. (m) :	428,33	207,22	120,78	119,36	67,95

TABLE 3.2 – Comparaison des moyennes des distances spatiales entre le *GHT* et la trajectoire médiane en fonction des techniques d'appariement

Le tableau 3.2 présente les résultats des distances spatiales moyennes comparées pour les différentes techniques d'appariement appliquées au *GHT* de notre exemple. On note que les techniques simples basées sur un appariement au même temps ou temps normalisé donnent des résultats peu satisfaisants liés aux différences de vitesses de certains objets mobiles. Les techniques du Dynamic Time Warping ainsi que la méthode de Fréchet donnent des résultats très proches alors que la technique de la droite perpendiculaire au cap permet d'obtenir la distance spatiale moyenne la plus faible pour ce groupe homogène de trajectoires. Néanmoins, les distances, pour les techniques du *DTW* et de Fréchet, sont calculées entre des couples de positions appariées. Des résultats similaires à ceux de la *DPC* sont obtenus en projetant les

distances entre couples de positions sur la droite perpendiculaire au cap. De plus, ces résultats sont dépendant du pas d'échantillonnage des trajectoires du GHT. En effet, l'erreur d'approximation de la distance de Fréchet discrète est bornée par le pas d'échantillonnage spatial entre deux positions consécutive d'une trajectoire [Eiter & Mannila 1994].

La figure 3.33 illustre la comparaison de ces différentes techniques d'appariement en présentant les distances moyennes des 500 trajectoires du GHT en fonction de la technique d'appariement utilisée classées par ordre de distance moyenne croissante. On note également sur cette figure que les courbes des distances moyennes pour les techniques de Fréchet et du DTW sont pratiquement confondues.

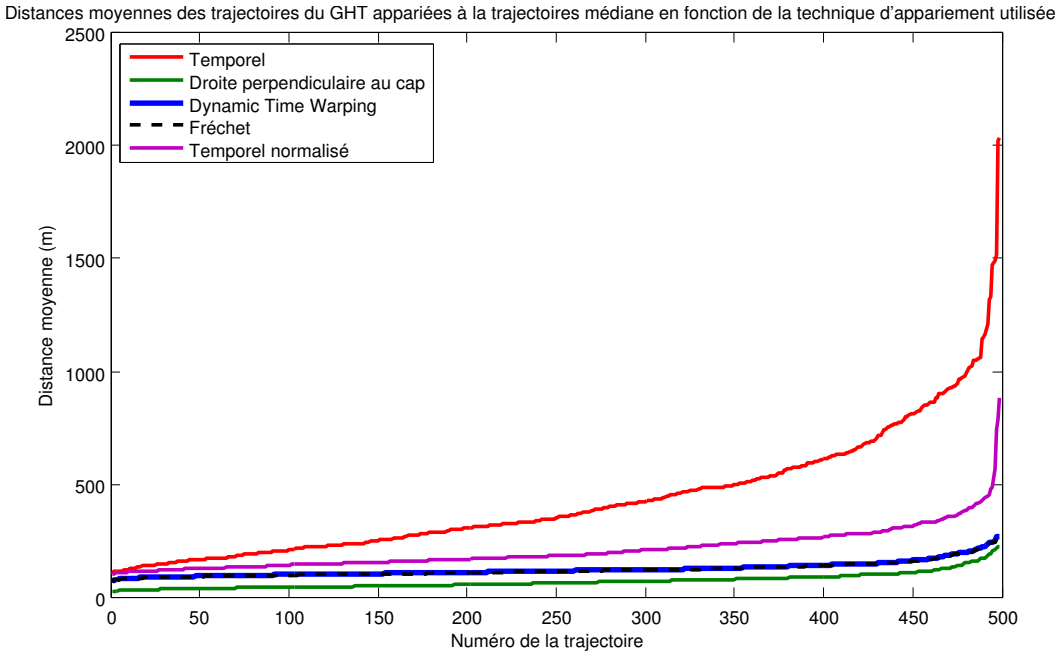


FIGURE 3.33 – Comparatif des distances moyennes entre les trajectoires du GHT et la trajectoire médiane en fonction de la technique d'appariement utilisée

Le calcul de ces valeurs statistiques est réalisé pour chaque nuage de positions. En complément de ces valeurs statistiques habituelles (percentiles, moyenne, écart-type), deux autres indicateurs sont calculés afin d'analyser la distribution des distances spatiales et temporelles des nuages et plus particulièrement la manière dont les masses de densité sont centrées ainsi que leurs étalement.

Le premier, le coefficient d'asymétrie (skewness) indique l'asymétrie de la densité de probabilité d'une variable aléatoire. Lorsque ce coefficient est négatif, cela signifie que la masse de probabilité est concentrée sur les valeurs fortes (à droite de la courbe sur la figure 3.34.a) et que la queue de distribution est étalée au niveau des valeurs faibles (à gauche de la courbe sur la figure 3.34.a).

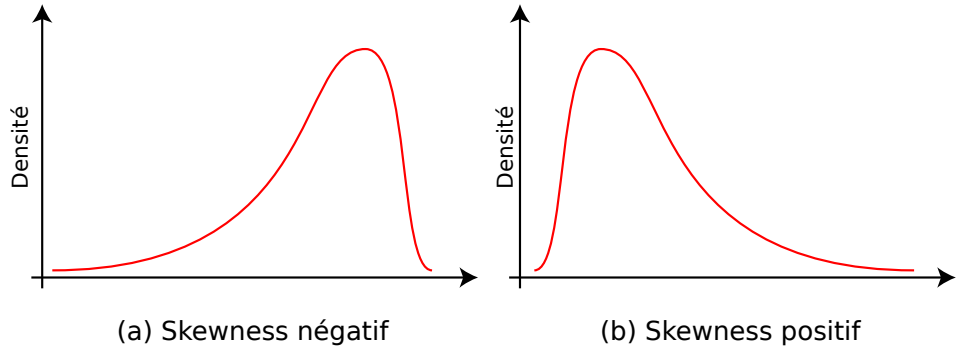


FIGURE 3.34 – Exemples de distributions ayant un coefficient d'asymétrie non nul

A contrario, lorsque ce coefficient est positif, cela signifie que la masse de probabilité est concentrée sur les valeurs faibles (à gauche de la courbe sur la figure 3.34.b) et que la queue de distribution est étalée au niveau des valeurs fortes (à droite de la courbe sur la figure 3.34.b). Ce coefficient d'asymétrie correspond au troisième moment standardisé [Groeneveld & Meeden 1984], il est calculé à partir du cube des écarts à la moyenne ($\overline{x}$) et de l'écart-type (σ) selon la formule 3.2 suivante :

$$S = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{\sigma} \right)^3 \tag{3.2}$$

Le second, le coefficient d'aplatissement (kurtosis) permet, outre l'écart-type de la distribution, de mesurer la disposition des masses de probabilité par rapport au centre de probabilité. Le kurtosis d'une distribution de probabilité suivant une loi normale centrée réduite vaut 3, cette distribution est alors mésokurtique.

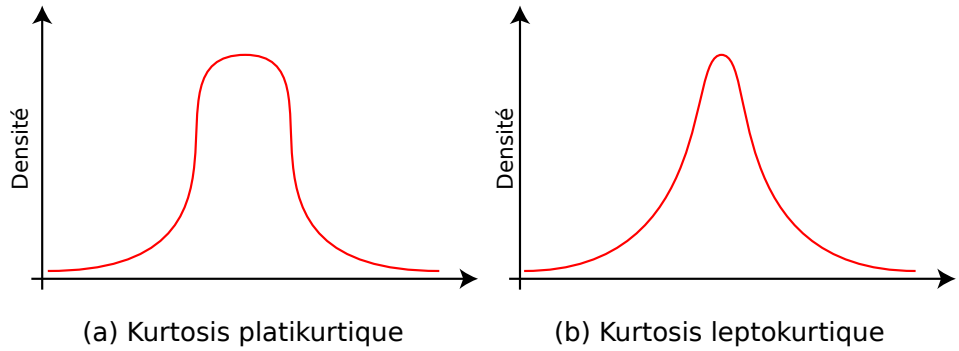


FIGURE 3.35 – Exemples de distributions ayant un coefficient d'aplatissement différent de 3

Lors que le kurtosis d'une distribution est supérieur à 3, cela signifie que la distribution a une forme pointue, cette distribution est alors leptokurtique comme présenté sur l'exemple de la figure 3.35.b). Enfin, une distribution pla-

tikurtique dont le kurtosis est inférieur à 3 aura une forme aplatie (figure 3.35.a). Ce coefficient d'aplatissement correspond au quatrième moment standardisé [Groeneveld & Meeden 1984], il peut être calculé à partir des écarts à la moyenne ($\overline{x}$) et de l'écart-type (σ) selon la formule 3.3 suivante :

$$K = \frac{(n+1)n}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \frac{(x_i - \overline{x})^4}{{\sigma^2}^2} - 3\frac{(n-1)^2}{(n-2)(n-3)} \tag{3.3}$$

La distribution des données de distances sous forme de boîtes à moustaches pour les trois nuages de positions homologues (n° 30,60 et 90) est représentée sur la figure 3.36. Cette représentation synthétique à l'avantage de combiner différentes indicateurs concernant la distribution des données afin de les comparer sur une même figure.

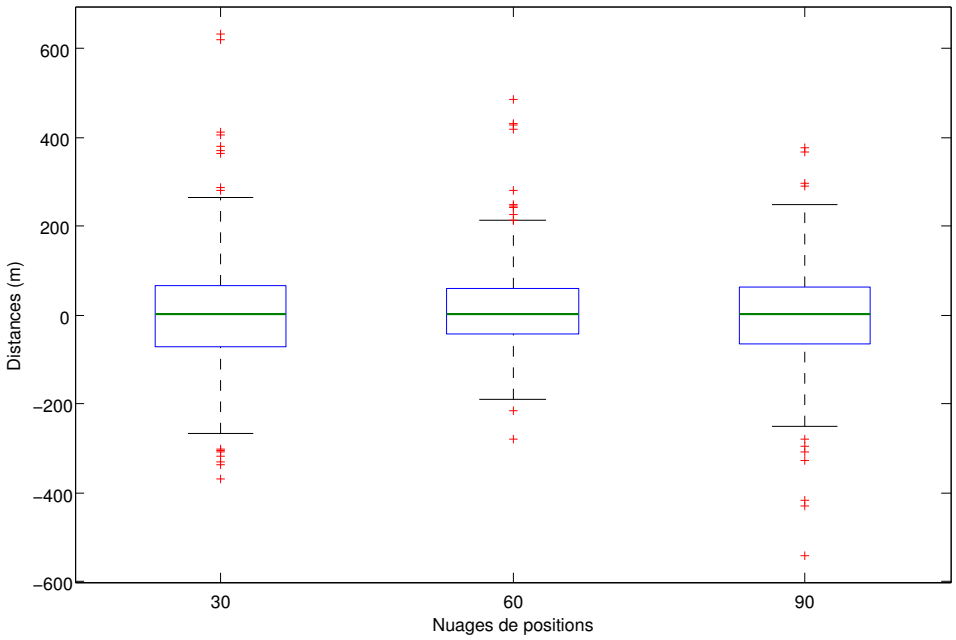


FIGURE 3.36 – Diagrammes en boîte de trois nuages de distances

On remarque sur la figure 3.36 que la taille de la boîte ainsi que les moustaches du nuage n° 60 sont plus petite que pour les autres nuages. De plus, la boîte n'est pas symétrique par rapport à la valeur médiane, celà permet de visualiser simplement le coefficient d'asymétrie de la distribution (skewness). De même, le coefficient d'aplatissement (kurtosis) peut être étudié en comparant la taille des boîtes. Cette représentation permet de voir l'évolution des limites du couloir spatial (ou temporel pour l'étude des distribution temporelles) ainsi que les données aberrantes affichées sous la forme de croix rouges. Le choix de notre extension spatio-temporelle des boîtes à moustaches s'avère par conséquent statistiquement pertinent.

Les courbes représentant l'évolution de différentes valeurs statistiques (moyenne, médiane, minimum, maximum et frontières du couloir pour $P = 10\%$) sont représentées sur la figure 3.37. On peut noter sur cet exemple une évolution significative de ces valeurs statistiques en fonction de l'action réalisée par l'objet mobile.

De plus, l'analyse de l'évolution des coefficients d'asymétrie (skewness) et d'aplatissement (kurtosis) (figure 3.38) permet de confirmer certaines de nos hypothèses concernant la distribution des distances spatiales et temporelles au sein des nuages de positions appariées. Dans les lignes droites, les distances des frontières spatiales ont tendance à s'écarter (la distribution spatiale s'étale et le kurtosis diminue légèrement). Dans les virages, les navires ont tendance à serrer le virage à la corde (le kurtosis augmente fortement et le skewness varie de façon importante).

Le kurtosis des distributions des nuages reste globalement supérieur à 3, ce qui indique que la distribution est globalement leptokurtique ; la courbe de densité ayant une forme "pointue", cela signifie que la masse de probabilité est regroupée à proximité du centre de probabilité. Il est intéressant de noter que la valeur du kurtosis augmente de façon très importante au niveau des virages, cela peut s'expliquer par la présence d'un obstacle au niveau du virage (île) forçant les trajectoires de navires à se regrouper au niveau de la trajectoire médiane.

A contrario, le kurtosis a tendance à être plus faible dans les phases où les navires se déplacent en ligne droite dans un espace ouvert, ce qui leur permet une plus grande expansion spatiale. De même, on note une variation importante du skweness au niveau du premier virage, qui peut être également liée à la présence de l'île regroupant une majorité de trajectoires proches de l'île (densité forte a proximité de la trajectoire médiane du côté de l'île) et sans obstacle de l'autre côté laissant les navires prendre plus de distance du côté opposé.

L'analyse des coefficients d'aplatissement et de symétrie des distributions spatiales et temporelles confirme certaines de nos hypothèses concernant la dissymétrie des frontières spatiales et temporelles de notre couloir ainsi que l'étalement du couloir en fonction du temps. Cette analyse se confirme sur l'observation de la figure 3.37 et est justifiée par l'analyse des coefficients d'aplatissement et de symétrie présentés sur la figure 3.38.

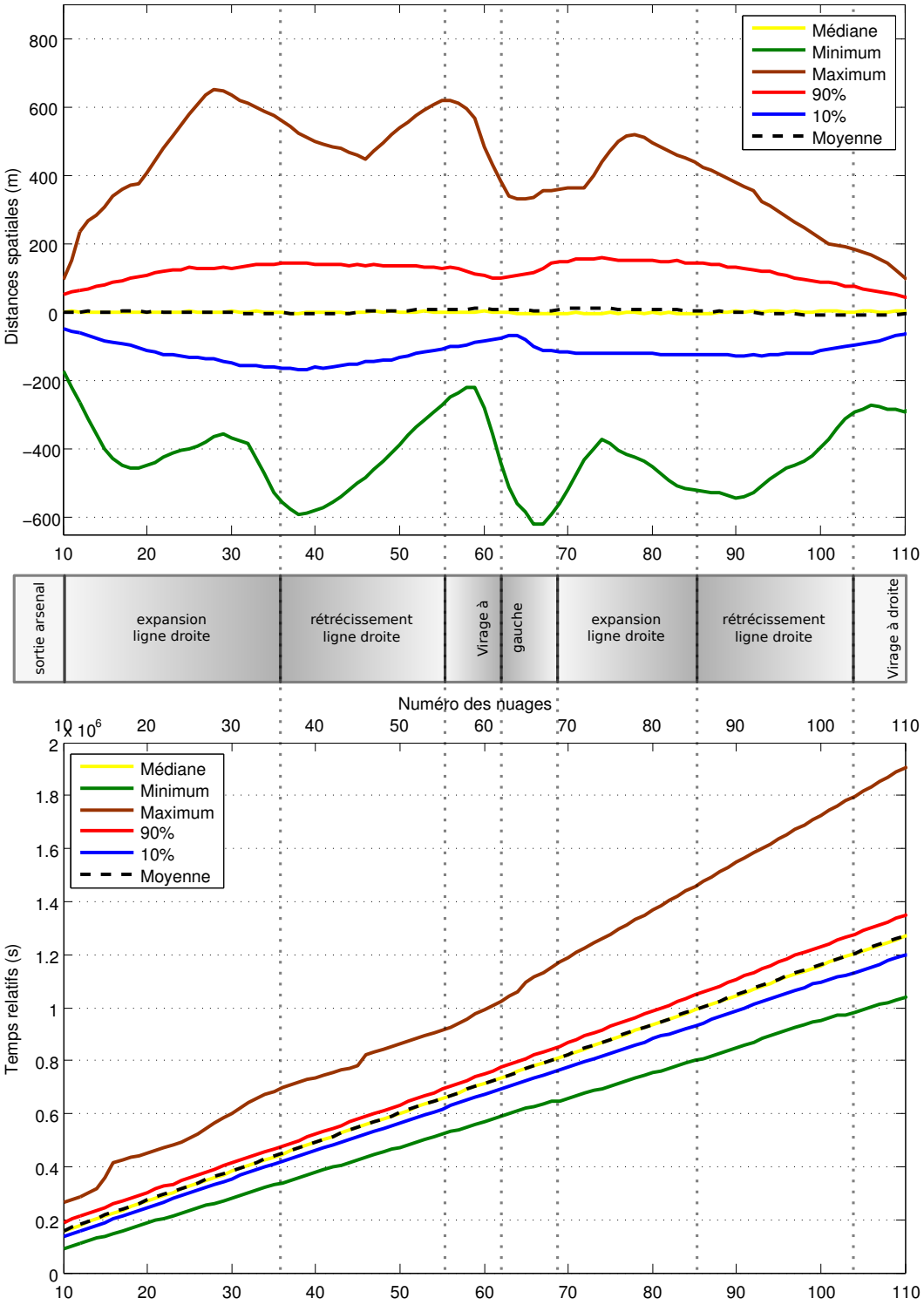


FIGURE 3.37 – Comparatif des courbes d'évolution des valeurs statistiques en fonction des nuages de positions appariées

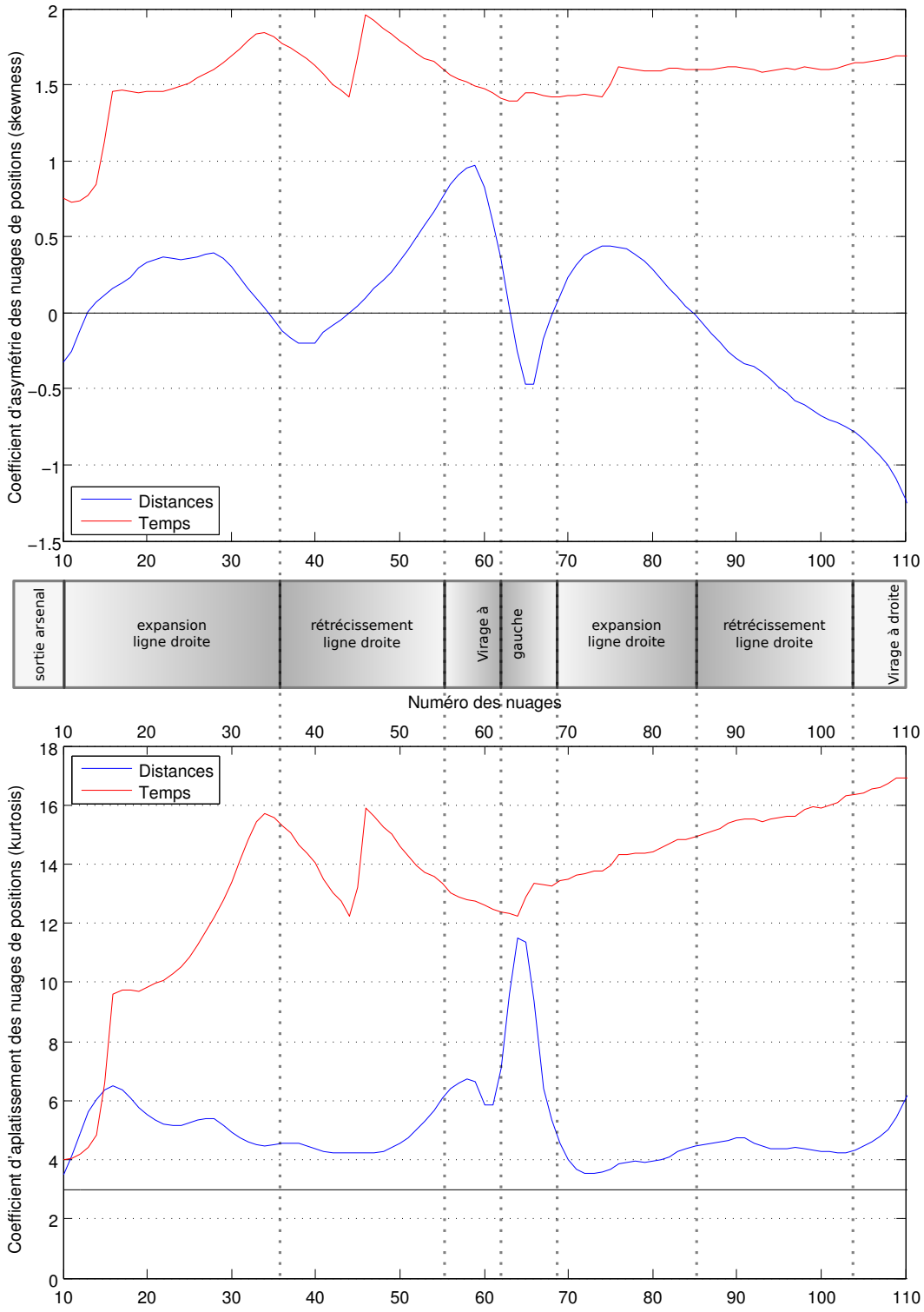


FIGURE 3.38 – Comparatif des courbes d'évolution des coefficients d'aplatissement et d'asymétrie en fonction des nuages de positions appariées

3.5 Sauvegarde de la route-type

La route-type composée de la trajectoire-type ainsi que des couloirs spatiaux et temporels associés est finalement sauvegardée dans une base de connaissance (figure 3.10 étape 5) afin d'être réutilisées pour une analyse temps réel. Le temps CPU global nécessaire à la génération du couloir spatio-temporel appliqué à notre exemple est de 36 511 ms. Ce temps est fortement dépendant du nombre de positions conservées pour chaque trajectoire du GHT ainsi que pour la route-type après filtrage de Douglas et Peucker spatio-temporel et rééchantillonnage.

Conclusion du chapitre

Dans ce chapitre, nous avons proposé un patron spatio-temporel constitué d'une trajectoire médiane et d'un couloir spatio-temporel associé. Ce patron spatio-temporel synthétise les déplacements d'un GHT d'objets mobiles d'un même type suivant un itinéraire commun. Une analyse statistique des densités spatiales et temporelles des nuages de positions utilisés pour générer notre patron spatio-temporel a été réalisée. Notre méthode propose un raffinement du concept des boîtes à moustaches en combinant des limites (moustaches) spatiales et temporelles au sein d'un couloir spatio-temporel. Ce couloir spatio-temporel peut être vu comme une extension du concept original des boîtes à moustaches 1D aux dimensions spatiales 2D(x,y) et temporelles (t). Cette analyse a montré que l'approche médiane/percentiles choisie est justifiée, les masses de densité étant parfois asymétriques et s'étalant de façon variable au cours du temps. Il serait intéressant d'appliquer cette technique de génération de patron spatio-temporel sur d'autre jeux de données tels que des déplacement de piétons ou d'animaux. De plus, certains phénomènes tels que la météo, les phénomènes climatiques, les saisons, peuvent impacter le comportement des objets mobiles. Ainsi, le patron spatio-temporel pourrait être raffiné en effectuant une analyse temporelle pour détecter des modifications cycliques temporelles des comportements des objets mobiles (liées aux cycles des marées, aux périodes de migrations des animaux, aux saisons). Différents patrons spatio-temporels plus ou moins précis pourraient alors être obtenus en fonction de critères plus ou moins stricts d'extraction du GHT (même type d'objet mobile, même saison temporelle...).

Chapitre 4

Qualification spatio-temporelle de trajectoires

Sommaire

4.1 Qualification d'une position 114
4.2 Mesures de similarité entre une trajectoire et un patron . . 116
4.2.1 Similarité spatiale d'une trajectoire comparée à un patron . . . 117
4.2.2 Similarité temporelle de trajectoires comparées à un patron . . 119
4.3 Indice de similarité spatio-temporelle entre une trajectoire et un patron 120
4.3.1 Définition des ensembles flous et des fonctions d'appartenance . 122
4.3.2 Fuzzification des variables linguistiques 126
4.3.3 Règles floues et indice de similarité spatio-temporelle 127
4.3.4 Calcul de l'indice de similarité spatio-temporelle 131

La base de connaissances ayant été renseignée, il est désormais possible de la combiner avec les bases de production afin d'obtenir une base de données inductive permettant de qualifier les positions sauvegardées dans la $BDST$ ainsi que celles obtenues en temps réel.

Il est cependant nécessaire d'apparier au préalable la trajectoire à un patron spatio-temporel. Une fois le patron spatio-temporel sélectionné, chaque position de la trajectoire peut être qualifiée. Des outils de comparaison entre une trajectoire et un patron doivent être définis afin de réaliser une analyse de la trajectoire dans sa globalité.

Ce chapitre présente en premier lieu les différentes techniques permettant d'apparier une trajectoire à un patron spatio-temporel. Des mesures de similarité sont ensuite introduites comme outil de comparaison entre une trajectoire et un patron. Ces mesures sont combinées à l'aide d'un module de logique floue. Les différentes phases de ce processus de qualification spatio-temporelle sont représentées à l'étape 6 de la figure 4.1.

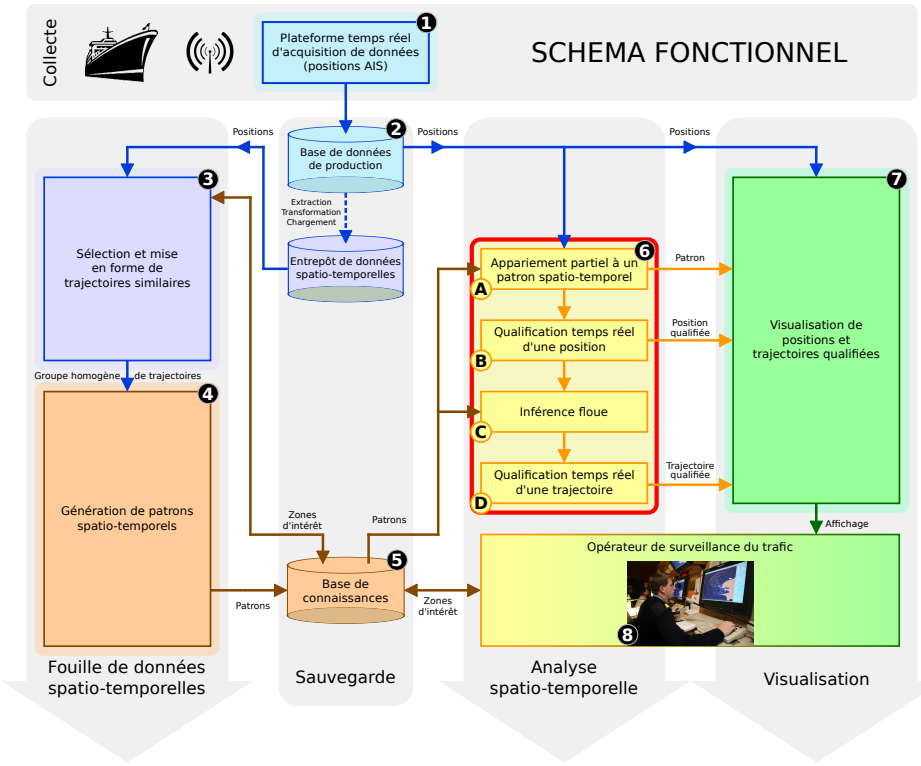


FIGURE 4.1 – Schéma fonctionnel détaillé de l'étape de qualification spatio-temporelle

4.1 Qualification d'une position

Pour chaque nouvelle position p reçue, le processus de qualification de position (figure 4.1 étape 6.A) suppose d'abord d'associer cette position à un des patrons spatio-temporels préalablement calculés (figure 4.2 étape 1). Dans cet objectif, sa trajectoire doit être appariée avec une des trajectoires médianes de la base de connaissances. Cet appariement s'effectue en fonction :

- de la position actuelle de l'objet mobile,
- du type de l'objet mobile,
- de l'historique de sa trajectoire,
- de la proximité de sa trajectoire avec des routes-types,
- de l'éventuelle information concernant sa destination prévue.

Cet appariement est variable. En effet, s'il existe plusieurs trajectoires-types $\tilde{T}$ proches de l'objet mobile, et si sa destination finale est inconnue, l'appariement peut être multiple. En revanche, si les destinations des objets mobiles sont connues (préalablement indiquées dans une feuille de route par exemple) ou bien si une unique $\tilde{T}$ est proche de l'objet mobile, alors l'appariement est plus fiable. Si la destination est connue, il est possible de déterminer plus facilement la $\tilde{T}$ à apparier.

Par exemple, lorsqu'un navire à passagers quitte la zone de Brest, sa trajectoire peut être appariée à différentes routes-types correspondant aux arcs du graphe de zones (A-E, A-C, A-B, A-F) représentées en gris clair sur la figure 4.2 étape 2 (les zones B et C du graphe sont situées hors figure). Plus le navire se dirige vers sa destination finale, plus il est facile de l'apparier de manière certaine à une trajectoire médiane. Ainsi, comme indiqué sur la figure 4.2 étape 3, la trajectoire médiane associée à la trajectoire du navire est sélectionnée en fonction de la position et de la trajectoire de l'objet mobile.

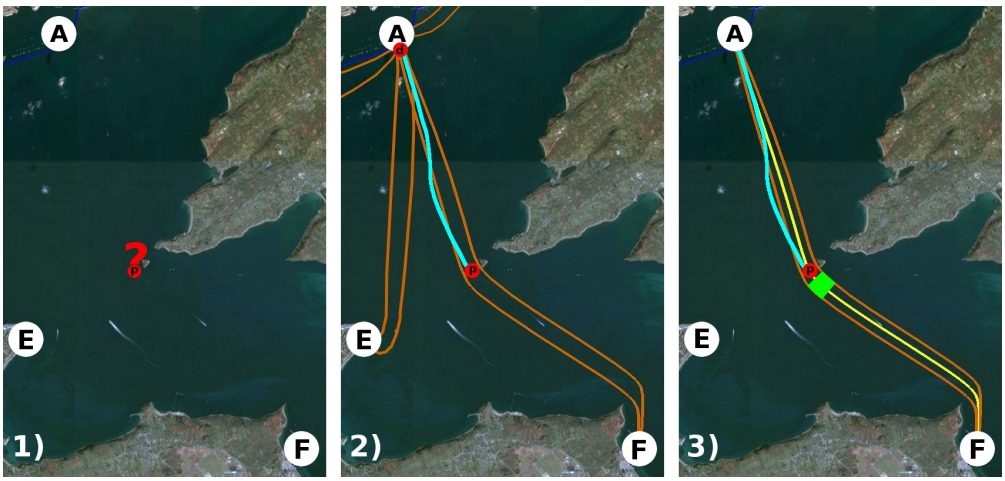


FIGURE 4.2 – Appariement d'une trajectoire à une route-type

Comme indiqué dans la section 2.2.6, une distance linéaire partielle (distance de Fréchet discrète) est appropriée à cette tâche d'appariement [Devogele 2002, Mascret *et al.* 2006]. La technique de Fréchet [Fréchet 1905, Alt *et al.* 2001] appliquée sur deux trajectoires discrètes permet d'apparier des trajectoires partielles en respectant la relation d'ordre temporel des positions des trajectoires. Cette technique a l'avantage de pouvoir s'appliquer à des trajectoires partielles et ne nécessite pas de calcul d'interpolation.

L'objectif est de mesurer les distances linéaires entre la trajectoire T partant de la zone de départ Z_D (zone A sur la figure 4.2) jusqu'à la position actuelle de l'objet mobile (représentée en bleu clair sur la figure 4.2 étape 3) et les trajectoires médianes ($\tilde{T}$) partant de la même zone Z_D (représentées en jaune sur la figure 4.2 étape 3). La trajectoire médiane ayant la distance linéaire partielle minimale est sélectionnée ainsi que son couloir spatio-temporel associé (couloir marron sur la figure 4.2).

Les écarts spatiaux et temporels entre la position de la trajectoire et la position de la trajectoire médiane appariée sont calculés. Ces écarts sont normalisés en utilisant les bornes ΔS_G, ΔS_D et ΔT_A, ΔT_R du couloir spatio-temporel de la position de la trajectoire médiane définies dans la section 3.3.5.

Ainsi, par exemple, une distance de 10 m entre une position et la position apparié de la trajectoire médiane est jugé importante si l'écart entre la trajectoire médiane et la frontière du couloir spatial vaut seulement 2 m. En revanche, cette distance sera considérée comme faible si l'écart entre la trajectoire médiane et la frontière du couloir spatial vaut 200 m. Le contexte dans lequel évolue l'objet mobile peut donc être pris en compte.

Dans l'exemple de la figure 4.2 étape 3, la zone de normalité (Z_N) du patron spatio-temporel apparié à la position p est représentée par la zone verte. Cette position est donc qualifiable comme étant dans le couloir, et en retard.

4.2 Mesures de similarité entre une trajectoire et un patron

La qualification des positions est un critère fondamental pour identifier les situations inhabituelles. Néanmoins, cette information, à elle seule, n'est pas suffisante.

Effectivement, une trajectoire peut être une suite ordonnée de positions qualifiées de "dans le couloir, à l'heure" et être inhabituelle. Par exemple, des suites de changements de direction peuvent entraîner des variations de la trajectoire, la rapprochant de la frontière droite puis de la frontière gauche comme l'illustre la figure 4.3. Cette trajectoire doit être qualifiée d'inhabituelle.

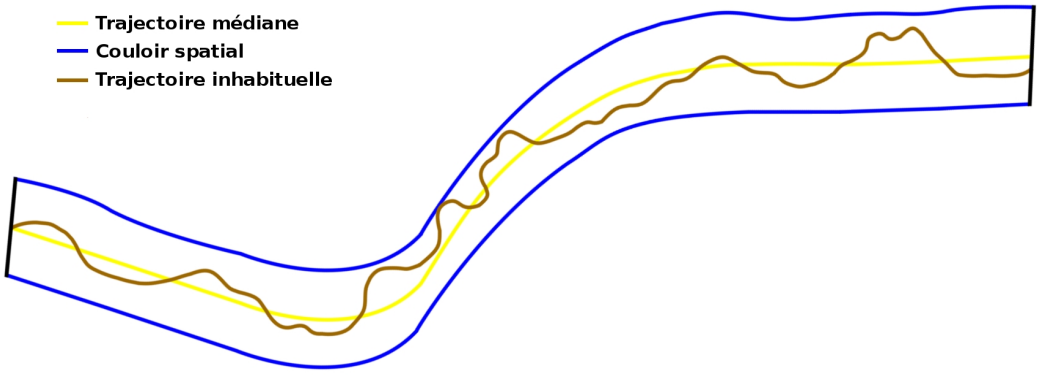


FIGURE 4.3 – Illustration d'une trajectoire inhabituelle contenue dans un couloir spatial

Il est donc indispensable de définir des mesures de similarité entre trajectoires. [Lee *et al.* 2008] et [Pelekis *et al.* 2007] ont défini des mesures de similarité tenant compte de l'aspect spatial et temporel. Cependant, ces dernières ne tiennent pas compte du contexte dans lequel se déplace l'objet mobile. En effet, une distance de 10 mètres comparée à un espace de 15 mètres de large peut être considérée comme

grande alors que la même distance dans un espace de 1000 mètres de large est considérée comme faible.

De même, il est souvent plus facile pour un objet mobile de prendre du retard que de l'avance, la vitesse maximale de l'objet étant limitée par ses propriétés physiques. Nous proposons donc de définir trois mesures de similarité géographique et trois mesures de similarité temporelle basées sur un appariement entre une trajectoire et un patron spatio-temporel, puis de les combiner.

4.2.1 Similarité spatiale d'une trajectoire comparée à un patron

A partir des couples de positions définis par le processus d'appariement présenté dans les sections 2.2.6 et 3.3.5, des mesures de similarité, tenant compte du contexte, peuvent être calculées. Le processus proposé se focalise sur la comparaison d'une trajectoire avec une route-type associée. Il est ainsi possible de normaliser les distances spatiales en fonction de l'écart spatial toléré (ΔS_j) de chaque position médiane $\tilde{p}_j$ de la route-type sélectionnée.

Le processus d'appariement entre la trajectoire T (suite ordonnée de positions p) et la trajectoire médiane $\tilde{T}$ (suite ordonnée de $\tilde{p}$) renvoie une suite de couples (p_i,$\tilde{p}_j$). La distance spatiale dS correspond à l'écart calculé entre les positions appariées p_i et $\tilde{p}_j$ exprimé en mètres. Les distances spatiales entre p_i et $\tilde{p}_j$ sont ensuite normalisées relativement à ΔS_j en $\tilde{p}_j$. Afin de distinguer la position relative de p_i par rapport à $\tilde{T}$, les distances normalisées sont notées négativement à gauche et positivement à droite.

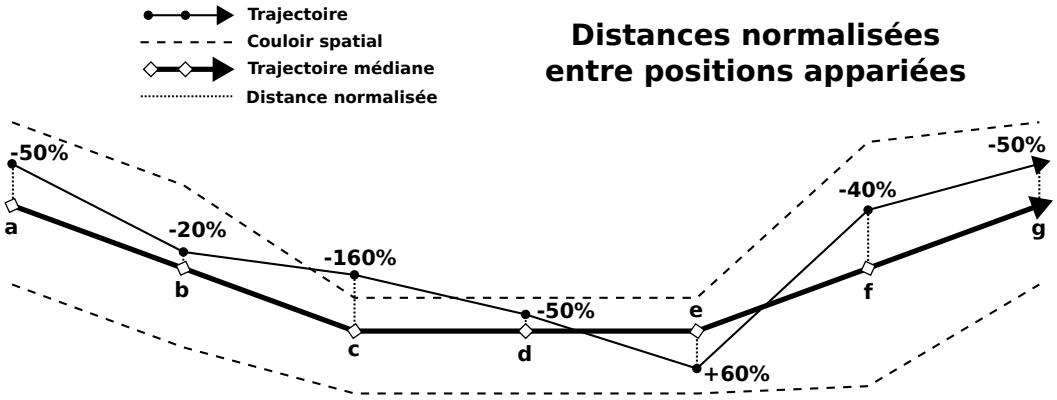


FIGURE 4.4 – Couples des positions appariées d'une trajectoire avec une trajectoire médiane, les distances spatiales normalisées sont exprimées en pourcentage de ΔS

La distance spatiale normalisée (dS_N) entre deux positions appariées est donc :

$$dS_N(p_i, \tilde{p}_j) = \begin{cases} \frac{-dS_{(p_i,\tilde{p}_j)}}{\Delta S_{G_j}} & \text{si } p_i \text{ est à gauche de } \tilde{p}_j \\ \frac{dS_{(p_i,\tilde{p}_j)}}{\Delta S_{D_j}} & \text{sinon} \end{cases} \tag{4.1}$$

Le couloir spatial pouvant être plus ou moins large et asymétrique, deux distances spatiales identiques peuvent avoir des distances spatiales normalisées différentes comme présenté sur la figure 4.4 pour les appariements c et f. De même, deux distances spatiales normalisées identiques peuvent avoir des distances spatiales complètement différentes comme représenté sur la figure 4.4 pour les appariements a et d.

Disposant de la distance spatiale normalisée dS_N, trois mesures sont proposées pour qualifier la similarité spatiale d'une trajectoire comparée à un patron :

– la distance spatiale maximale : DSM

$$DSM(T, \tilde{T}) = Max(|dS_N(p_i, \tilde{p}_j)|) \tag{4.2}$$

– la distance spatiale moyenne : DSm

$$DSm(T, \tilde{T}) = Moyenne(|dS_N(p_i, \tilde{p}_j)|) \tag{4.3}$$

– la moyenne des deltas spatiaux : δSm

$$\delta Sm(T, \tilde{T}) = \frac{\sum_{i=1}^{n-1} |dS_N(p_i, \tilde{p}_j) - dS_N(p_{i+1}, \tilde{p}_{j+1})|}{n-1} \tag{4.4}$$

La première mesure représente l'écart spatial maximal (DSM) entre des positions appariées, la deuxième (DSm) l'écart moyen. La troisième mesure (δSm) représente les variations entre les écarts successifs et renseigne sur un écart de forme entre la trajectoire et le patron. Pour l'exemple de la figure 4.4 :

– la DSM vaut -160%, ce qui signifie que la trajectoire est sortie de 60% du couloir au pire,
– la DSm vaut 61,43% ((50%+20%+160%+50%+60%+40%+50%)/7), ce qui signifie que la trajectoire est en moyenne à une distance de 61,43% de la trajectoire médiane par rapport à l'écart spatial toléré,
– la δSm vaut 83,33% (|-50%+20%| + |-20%+160%| + |-160%+50%| + |-50%-60%| + |+60%+40%| + |-40%+50%|)/6), ce qui signifie que l'écart entre la $\tilde{T}$ et la T varie de manière importante.

Il faut noter que ces mesures sont calculées de manière discrète, uniquement aux positions $\tilde{p}_j$. La trajectoire médiane étant filtrée et rééchantillonnée, elle surestime légèrement les mesures par rapport à une approche continue.

Afin d'illustrer l'intérêt de ces trois mesures, la figure 4.5 donne trois exemples de couples de T, $\tilde{T}$. Pour le couple (a), la DSM et la DSm sont très proches mais non nulles et la δSm est proche de zéro. Ces valeurs décrivent un couple de trajectoires décalées avec des formes très similaires.

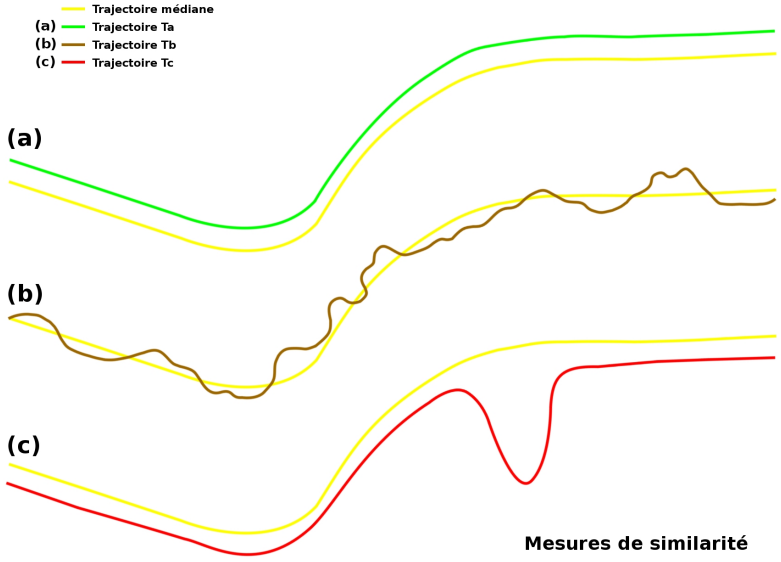


FIGURE 4.5 – Exemple de trois trajectoires associées à une même trajectoire médiane

Pour le couple (b), la DSM est assez faible et la DSm est faible. Par contre, la δSm est importante. Ces valeurs décrivent des trajectoires assez proches, mais ayant des formes dissimilaires.

Finalement, le couple (c) a une DSM importante, une DSm et δSm faible. Ces trajectoires sont donc la plupart du temps assez proches. Cependant, un écart important sur une courte période est présent. Ces mesures spatiales sont complétées par des mesures temporelles ayant des significations proches présentées dans la section suivante.

4.2.2 Similarité temporelle de trajectoires comparées à un patron

En ce qui concerne l'aspect temporel, la même méthode est employée pour définir trois mesures de similarité temporelle. La distance temporelle (dT) correspond à l'écart en secondes entre les estampilles temporelles relatives (durée depuis le départ) de positions appariées. Les distances temporelles entre p_i et $\tilde{p}_j$ sont ensuite normalisées relativement à ΔT_j en $\tilde{p}_j$. Les écarts de temps normalisés sont notés négativement lorsque p_i est en retard par rapport à $\tilde{p}_j$ et positivement lorsque p_i est en avance par rapport à $\tilde{p}_j$.

La distance temporelle normalisée (dT_N) entre deux positions appariées est donc :

$$dT_N(p_i, \tilde{p}_j) = \begin{cases} \frac{-|dT(p_i,\tilde{p}_j)|}{\Delta T_{R_j}} & \text{si } p_i \text{ est à en retard par rapport à } \tilde{p}_j \\ \frac{|dT(p_i,\tilde{p}_j)|}{\Delta T_{A_j}} & \text{sinon} \end{cases} \tag{4.5}$$

Disposant de la dT_N, trois mesures sont proposées pour qualifier la similarité temporelle d'une trajectoire comparée à un patron :

– la distance temporelle maximale : DTM

$$DTM(T, \tilde{T}) = Max(|dT_N(p_i, \tilde{p}_j)|) \tag{4.6}$$

– la distance temporelle moyenne : DSm

$$DTm(T, \tilde{T}) = Moyenne(|dT_N(p_i, \tilde{p}_j)|) \tag{4.7}$$

– la moyenne des deltas temporels : δTm

$$\delta Tm(T, \tilde{T}) = \frac{\sum_{i=1}^{n-1} |dT_N(p_i, \tilde{p}_j) - dT_N(p_{i+1}, \tilde{p}_{j+1})|}{n-1} \tag{4.8}$$

Nous disposons donc de six mesures de similarité. Ces mesures sont complémentaires, elles permettent de qualifier une trajectoire comparée à un patron dans sa globalité. Cependant, proposer six mesures aux opérateurs (figure 4.1 étape 8) n'est pas envisageable, trop d'informations seraient présentes pour chaque objet mobile et surchargeraient l'opérateur de surveillance du trafic. Ces différentes mesures doivent être fusionnées afin d'obtenir un indice de similarité global entre une trajectoire et un patron spatio-temporel.

4.3 Indice de similarité spatio-temporelle entre une trajectoire et un patron

Afin de définir un indice unique permettant de qualifier les trajectoires (figure 4.1 étape 6.D), deux approches sont possibles. La première consiste à agréger les six mesures à l'aide d'une somme pondérée. Cette solution n'est pas optimale ; elle a le défaut de renvoyer un résultat trop lisse et de ne pas correspondre exactement au jugement humain [Anouar & Wafa 2006].

La deuxième, plus pragmatique et proche du raisonnement humain, est basée sur la logique floue (fuzzy logic) [Zadeh 1978, Bouchon-Meunier 1995]. La logique floue s'appuie sur la théorie mathématique des ensembles flous introduit par [Zadeh 1965]. Ces ensembles sont une extension de la théorie des ensembles classiques permettant

une définition de ces ensembles de façon imprécise. La définition d'un sous-ensemble d'un univers donné est modélisée par des fonctions d'appartenance.

Contrairement à la logique booléenne, la logique floue permet de définir des degrés de vérité pour une condition autres que vrai ou faux. Ces conditions peuvent être combinées au sein de règles floues grâces à un ensemble d'opérateurs flous. La théorie des sous-ensembles flous se réduit à la théorie des sous-ensembles classiques dans le cas où les fonctions d'appartenance prennent des valeurs binaires.

La logique floue permet de s'appuyer sur des règles floues exprimées en langage naturel décrivant la similarité spatiale (SIM_S), temporelle (SIM_T) et spatio-temporelle (SIM_{ST}) telles que sur les exemples suivants :

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.
$((DSm = Faible) \wedge (DSM = Faible) \wedge (\delta Sm = Faible)) \Rightarrow (SIM_S = Tfort)$

Comme indiqué dans les exemples de règles floues (R1, R2 et R3), les mesures de similarités sont nommées dans les règles et peuvent potentiellement y être combinées. Ces mesures quantitatives de similarité doivent donc être transformées en variables linguistiques qui seront utilisées dans le module de logique floue présenté à l'étape 6.C de la figure 4.1.

La figure 4.6 présente les différentes étapes du module de logique floue qui seront détaillées dans les sections suivantes. A partir d'une trajectoire et d'un patron spatio-temporel apparié, les mesures de similarité entre la trajectoire et le patron sont calculées.

Ces mesures sont ensuite transformées en variables floues (étape de fuzzification) à l'aide d'ensembles flous préalablement définis à partir des statistiques des mesures de similarité du GHT du patron spatio-temporel.

Un raisonnement flou est réalisé grâce à un ensemble de règles floues basées sur ces variables floues. L'étape de défuzzification des conclusions de ces règles permet d'obtenir deux indices numériques de similarité spatiale et temporelle. Finalement, ces indices sont agrégés en un indice de similarité spatio-temporel caractérisant la similarité de la trajectoire comparée au patron spatio-temporel.

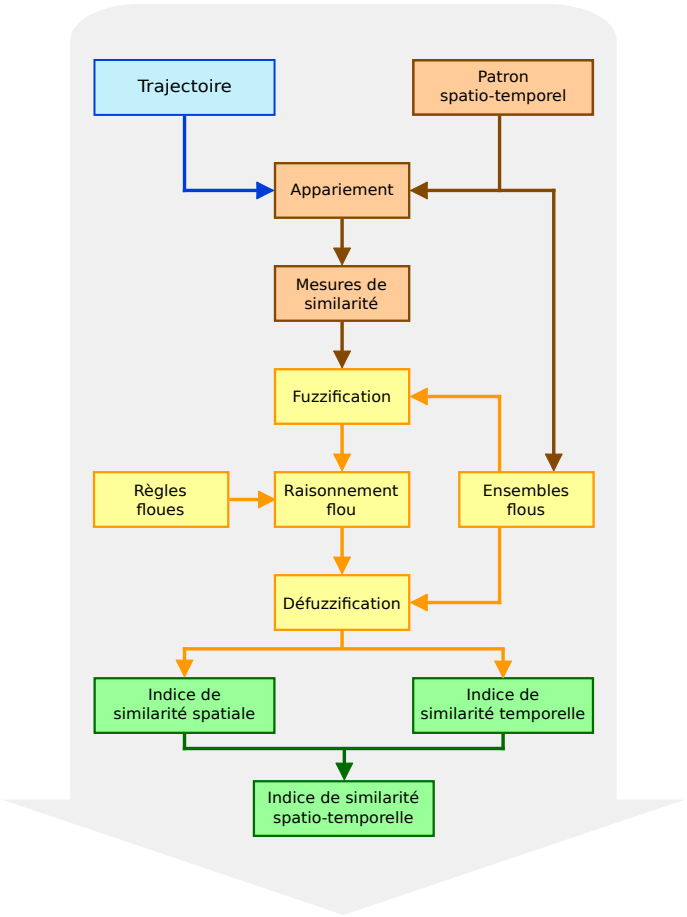


FIGURE 4.6 – Processus détaillé de l'étape d'inférence floue

4.3.1 Définition des ensembles flous et des fonctions d'appartenance

D'autres règles doivent être ajoutées pour prendre en compte l'ensemble des critères de similarité spatiale et temporelle. Pour pouvoir raisonner avec ce type de règles floues, la première étape consiste à transformer les valeurs des six mesures numériques de similarité en termes linguistiques. Lors de cette étape dite de "Fuzzification" (figure 4.6), des ensembles flous et des fonctions d'appartenance sont introduits.

Trois termes linguistiques sont créés ($Faible$, $Moyen$ et $Fort$) permettant de caractériser les variables linguistiques associées aux mesures de similarité. Pour toute valeur numérique d'entrée x, la fonction d'appartenance $\mu^{DSM}_{Faible}(x)$ définit le degré d'appartenance de x à l'ensemble flou $Faible$ compris entre 0 et 1 pour la mesure DSM. Il en est de même pour les fonctions d'appartenance $\mu^{DSM}_{Moyen}(x)$ et $\mu^{DSM}_{Fort}(x)$ ainsi que pour les autres mesures DSm et δSm.

Les fonctions d'appartenance liées aux mesures de similarité temporelle (DTM, DTm et δTm) sont également définies de façon identique. Différentes fonctions d'appartenance peuvent être utilisées, les plus communes étant les fonctions d'appartenance présentées sur la figure 4.7 et dont les expressions sont présentées dans le tableau 4.1.

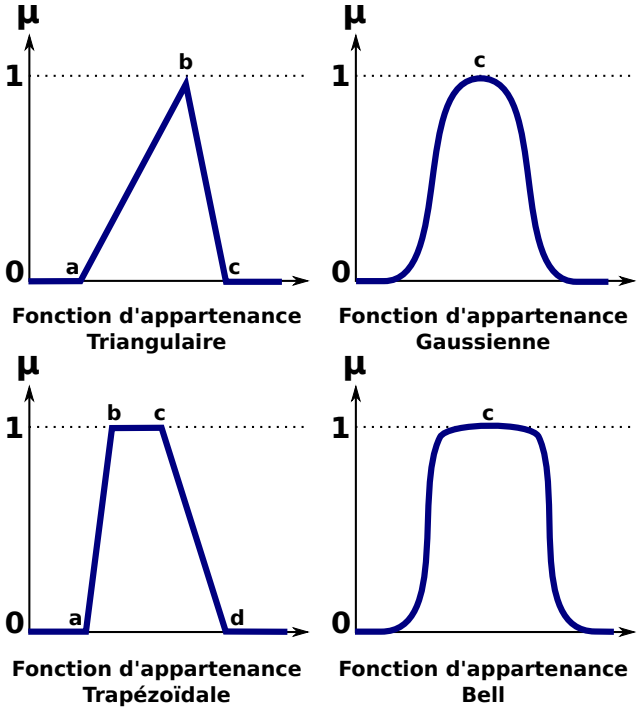


FIGURE 4.7 – Différents types de fonctions d'appartenance

Une comparaison de ces différentes fonctions d'appartenance est présentée dans les travaux de [Zhao & Bose 2002, Gayathri Monicka & Ramash Kumar 2011]. Cette étude montre que les fonctions d'appartenance du type triangulaire et trapézoïdale donnent les meilleurs résultats et sont les plus simples à intégrer dans un système de contrôle flou. Ces différentes fonctions d'appartenance dont les expressions sont indiquées dans le tableau 4.1 sont présentées sur la figure 4.7.

Type	Expression
Triangulaire	$f(x;a,b,c) = Max\{Min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\}$
Trapézoïdale	$f(x;a,b,c,d) = Max\{Min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\}$
Gaussienne	$f(x;\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}}$
Bell	$f(x;a,b,c) = \frac{1}{1+\left\lvert\frac{x-c}{a}\right\rvert^{2b}}$

TABLE 4.1 – Expressions de différents types de fonctions d'appartenance

Les fonctions d'appartenance utilisées dans la suite de cette étude sont de type linéaire par morceaux (trapézoïdales). Elles permettent de définir aisément les zones où la fonction est vraie et celles où la fonction est fausse ce qui facilite le recueil d'expertise. Pour chaque mesure de similarité, les limites des différentes zones ($Faible$, $Moyen$ et $Fort$) doivent être spécifiées. Ces limites de zones sont définies à partir de l'étude statistique de l'appariement de toutes les trajectoires du GHT avec la trajectoire médiane et du calcul des percentiles statistiques de chaque mesure de similarité. Ces statistiques sont obtenues en appariant chaque trajectoire du GHT à la route-type avec la même technique d'appariement que celle utilisée lors de la génération du patron spatio-temporel (Fréchet). Pour chaque couple de positions appariées, la distance spatiale normalisée exprimée en pourcentage du couloir spatial ainsi que la distance temporelle normalisée exprimée en pourcentage du couloir temporel est mesurée. Les percentiles de chaque mesure de similarité sont ainsi obtenus.

Mesure	0%	20%	40%	50%	60%	80%	100%
DSM	43,48	93,93	110,24	119,79	131,4	168,73	723,28
DSm	12,33	31,83	39,63	43,63	48,96	65,07	140,63
δSm	2,87	4,31	4,82	5,08	5,42	6,48	18,84
DTM	27,97	73	98,91	107,97	118,42	158,26	43535
DTm	6,88	33,08	47,38	55,24	62,42	94,3	592,98
δTm	1,5	2,74	3,22	3,5	3,84	5,19	681,45

Table 4.2 – Statistiques des différentes mesures de similarité du GHT

Mesure	Moyenne	Ecart type	coeff. d'asymétrie	coeff. d'aplatissement
DSM	141,59	84,91	3,45	19,37
DSm	49,31	22,48	1,28	4,80
δSm	5,53	1,77	2,37	13,37
DTM	994,77	4229,05	5,31	35,06
DTm	74,43	67,53	3,13	17,67
δTm	17,69	65,76	5,32	35,21

Table 4.3 – Indices statistiques des différentes mesures de similarité du GHT

D'autres travaux se basent également sur une étude statistique pour générer les fonctions d'appartenance [Wang & Mendel 1992, Idri *et al.* 2007]. L'histogramme de densité des valeurs de DSM du GHT et la courbe de probabilité cumulée associée sont présentés sur les figures 4.8. Les tableaux 4.2 et 4.3 listent les statistiques des différentes mesures de similarité du GHT exprimées en pourcentages.

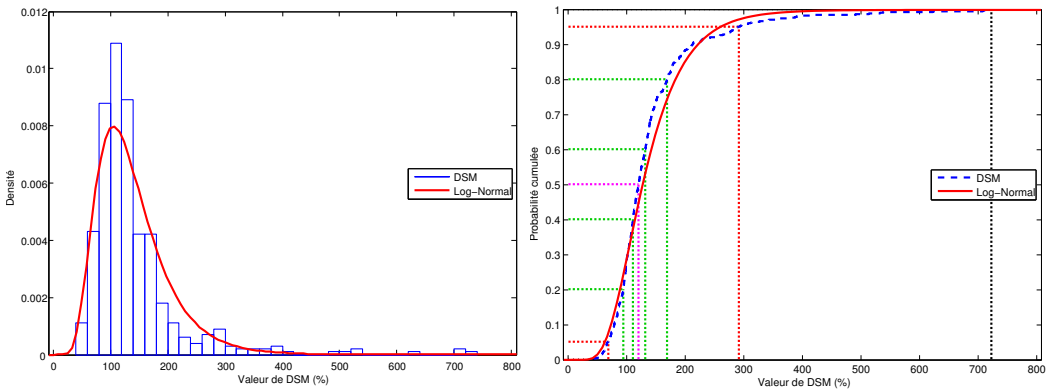


FIGURE 4.8 – Histogramme et courbe de probabilité cumulée de la DSM du GHT

Ainsi, pour chaque mesure de similarité, les fonctions d'appartenance proposées sont basées sur les percentiles des statistiques des mesures de similarité du GHT :

- $Faible$ en dessous de 20 %,
- entre $Faible$ et $Moyen$ de 20 à 40 %,
- $Moyen$ de 40 à 60 %,
- entre $Moyen$ et $Fort$ de 60 à 80 %,
- $Fort$ au delà de 80 %.

Les fonctions d'appartenance des ensembles flous de la mesure de similarité DSM reprenant les valeurs de la table 4.2 sont présentées sur l'exemple de la figure 4.9.

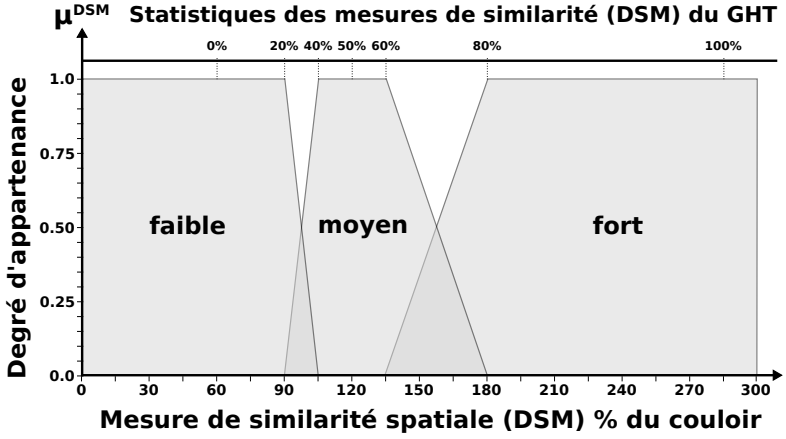


FIGURE 4.9 – Fonctions d'appartenance associées à la mesure de similarité DSM

Les limites des zones ($Faible$, $Moyen$ et $Fort$) sont représentées par des traits pointillés verts (20%,40%,50%,60%,80%) sur la courbe de probabilité cumulée de la DSM du GHT de la figure 4.8. Ces mêmes limites sont reportées en pointillés

en haut de la figure 4.9 présentant les fonctions d'appartenance de la mesure de similarité DSM. Les histogrammes et courbes de probabilité cumulées des autres mesures de similarité sont présentées en annexes A.12 et A.13. Les différentes fonctions d'appartenance définies pour chaque mesure de similarité d'un GHT sont présentées dans le tableau 4.4. Ces 18 fonctions d'appartenance sont finalement liée à la route-type du GHT puis sauvegardées dans la base de données inductive incluant par ailleurs les données de positions ainsi que les motifs spatio-temporels.

Mesures spatiales			Mesures temporelles		
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 4.4 – Fonctions d'appartenance des différentes mesures de similarité

Disposant des fonctions d'appartenance, il est désormais possible d'obtenir le degré d'appartenance de chaque mesure de similarité d'une trajectoire comparée à une route-type aux ensembles flous ($Faible$, $Moyen$ et $Fort$).

4.3.2 Fuzzification des variables linguistiques

Une fois les ensembles flous et fonctions d'appartenance définis, il est possible de rendre floue une mesure de similarité calculée entre une trajectoire appariée à une route-type via l'étape de "Fuzzification" (figure 4.6).

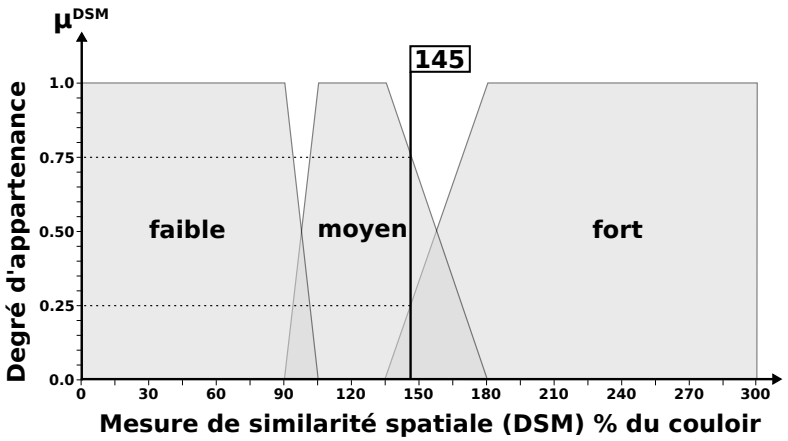


FIGURE 4.10 – Valeur floue de la DSM d'une trajectoire appariée à une route-type

Cette étape transforme une valeur numérique en degré d'appartenance flou par évaluation des fonctions d'appartenance présentées du tableau 4.4. Dans l'exemple de la figure 4.10, la trajectoire a une valeur numérique x de DSM valant 145% du couloir spatial représenté par un trait fin noir. Après évaluation des fonctions d'appartenance définies dans la section 4.3.1, les résultats suivants sont obtenus :
$\mu^{DSM}_{Faible}(x) = 0 \quad \mu^{DSM}_{Moyen}(x) = 0.75 \quad \mu^{DSM}_{Fort}(x) = 0.25$
Le degré d'appartenance de x à l'ensemble flou $Faible$ est donc de 0%, de 75% pour l'ensemble flou $Moyen$ et de 25% pour l'ensemble flou $Fort$.

Les valeurs des mesures de similarité ayant été associées à des variables linguistiques, l'étape suivante consiste à raisonner sur ces variables en les combinant avec des règles floues.

4.3.3 Règles floues et indice de similarité spatio-temporelle

Dans notre étude, nous souhaitons qualifier la similarité spatio-temporelle d'une trajectoire. Nous disposons de trois mesures de similarité spatiale et de trois mesures de similarité temporelle calculées pour une trajectoire appariée à une route-type ainsi que des fonctions d'appartenance associées aux ensembles flous définis dans la section 4.3.1.

Ainsi, il est possible de combiner ces mesures de similarité spatiale afin de définir un indice de similarité spatiale et faire de même pour l'indice de similarité temporelle. Pour obtenir cet indice de similarité, des règles floues sont définies (figure 4.6). Une règle floue est composée d'un prédicat associant une combinaison de variables linguistiques et d'une conclusion associant une variable linguistique à une proposition de terme linguistique comme présenté sur l'exemple de la figure 4.11.

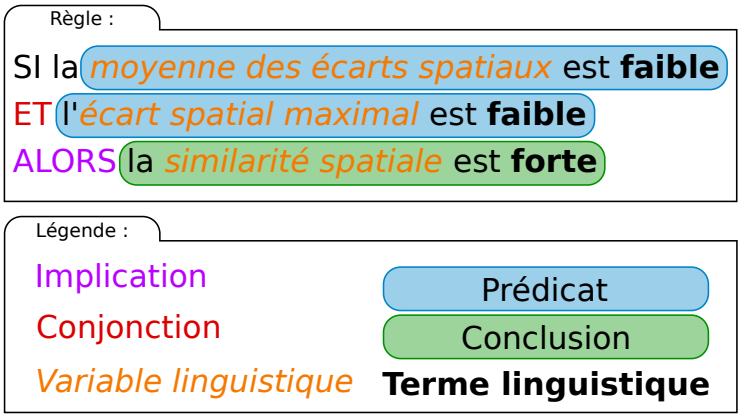


FIGURE 4.11 – Exemple de règle floue

L'utilisation de variables et termes linguistiques dans la rédaction des règles floues a l'avantage de faciliter l'écriture des règles floues ainsi que le recueil d'expertise des opérateurs. Ces règles sont également facilement compréhensibles par un être humain.

La combinaison des variables linguistiques composant les prédicats est réalisée grâce à des opérateurs de logique floue présentés en rouge sur l'exemple de la figure 4.11. Ces opérateurs permettent d'écrire des combinaisons logiques entre notions floues en réalisant des calculs basés sur leurs degrés de vérité.

Les opérations réalisées sur les ensembles flous sont réalisées grâce à différentes fonctions :

- t-norme, $\alpha \otimes \beta$, (conjonction, $ET : \wedge$) ;
- t-conorme, $\alpha \oplus \beta$, (disjonction, $OU : \vee$) ;
- négation, $\ominus\alpha$, ($NON : \neg$) ;
- implication, $\alpha \Rightarrow \beta$, ($IMPLIQUE :\rightarrow$).

Différents fonctions utilisées pour définir ces opérateurs sont présentés en détail dans [Klir & Yuan 1995, Bouchon-Meunier 2007, Kohout 2007]. De manière classique en logique floue [Zadeh 1965, Mamdani & Assilian 1975, Bobillo & Straccia 2008], ces opérateurs peuvent être définis en utilisant les fonctions Min et Max. Dans cette étude, nous utiliserons les opérateurs de Mamdani [Mamdani & Assilian 1975] qui sont résumés dans le tableau 4.5.

Règle	Degré de vérité
$(x \wedge y)$	$Min(\mu_x(x), \mu_y(y))$
$(x \vee y)$	$Max(\mu_x(x), \mu_y(y))$
$(\neg x)$	$(1 - \mu_x(x))$
$(x, y \rightarrow z)$	$Min(\mu_x(x), \mu_y(y))$

TABLE 4.5 – Opérateurs de Mamdani

Les bases des définitions de très nombreuses autres différentes fonctions ont été proposées par Dombi, Dummett, Fodor, Gaines, Gödel, Gougen, Kleene-Dienes, Łukasiewicz, Mamdani, Reichenbach, Rescher, Schweizer, Yager, Zadeh, Weber, Willmott [Gödel 1932, Dummett 1959, Hay 1963, Zadeh 1965, Mamdani & Assilian 1975, Willmott 1980, Weber 1983, Yager *et al.* 1995]. Les formules de ces fonctions sont détaillées dans les tables 3.2 et 3.3 de [Klir & Yuan 1995].

Cinq termes linguistiques sont créés pour les variables linguistiques correspondant à l'indice de similarité spatiale (SIM_S), l'indice de similarité temporelle (SIM_T) et l'indice de similarité spatio-temporel (SIM_{ST}). Ces cinq termes ($TFaible$, $Faible$, $Moyen$, $Fort$ et $TFort$) sont utilisés comme conclusion des règles floues dont les fonctions d'appartenance sont présentées sur la figure 4.12.

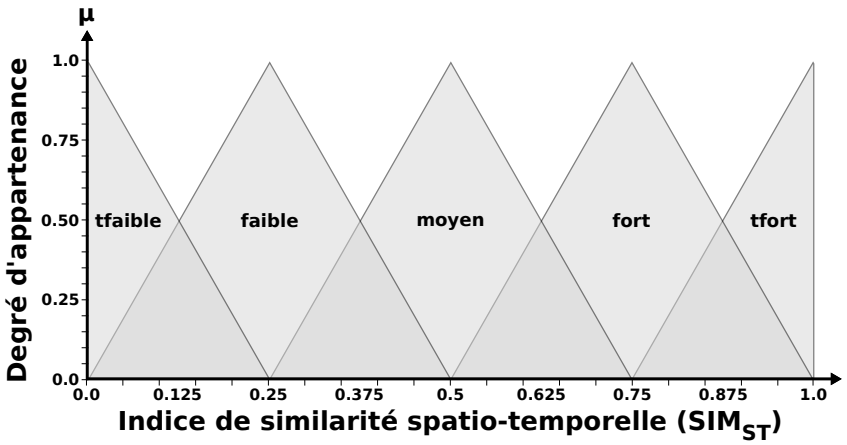


FIGURE 4.12 – Fonctions d'appartenance des indices de similarité

Ayant préalablement défini trois termes linguistiques pour chacune des trois variables linguistiques spatiales, il existe 3^3 combinaisons possibles de prédicats soit 27 règles pour le calcul de l'indice de similarité spatiale (de même pour l'indice de similarité temporelle). Ces 27 règles peuvent être représentées sous la forme d'une matrice associative floue et d'un arbre de décision associé.

Prédicats			**Conclusion**
Faible	*Moyen*	*Fort*	
3	0	0	*TFort*
2	1	0	*Fort*
2	0	1	*Fort*
1	2	0	*Fort*
0	3	0	*Moyen*
1	1	1	*Moyen*
0	1	2	*Faible*
1	0	2	*Faible*
0	2	1	*Faible*
0	0	3	*TFaible*

TABLE 4.6 – Conclusions des règles floues en fonction des prédicats

La conclusion de la règle dépendant des combinaisons des termes linguistiques associés aux trois différentes variables, il n'y a pas d'impact au niveau de l'ordre des variables linguistiques dans la représentation sous forme d'arbre de décision. Le choix de la conclusion de la règle est basé sur le nombre de prédicats associés aux termes *Faible*, *Moyen* ou *Fort* comme indiqué dans le tableau 4.6.

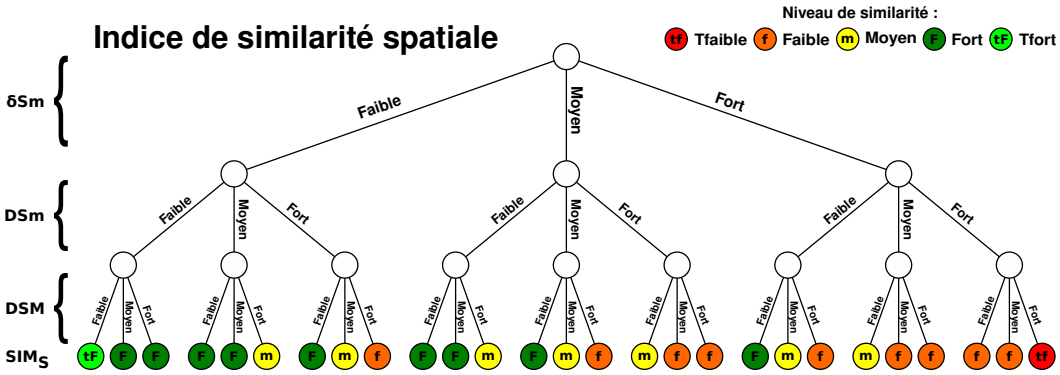


FIGURE 4.13 – Arbre de décision de la matrice associative floue de SIM_S

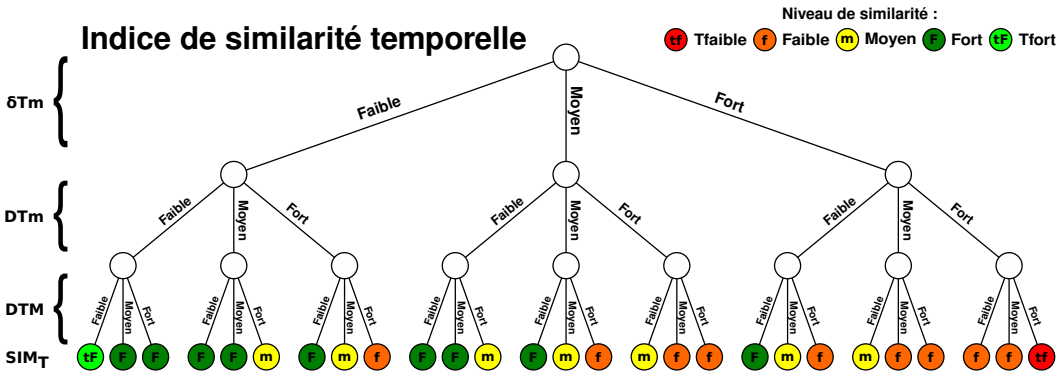


FIGURE 4.14 – Arbre de décision de la matrice associative floue de SIM_T

La figure 4.13 présente l'arbre de décision associé à la matrice associative floue de l'indice de similarité spatiale. Le premier niveau de l'arbre de décision correspond au δSm, le second à la DSm et le dernier à la DSM. Le terme linguistique associé à la variable linguistique des mesures de similarité est indiqué sur les arcs de l'arbre de décision. Le terme linguistique de l'indice de similarité proposé en conclusion de l'activation de la règle floue est représenté par une coloration des feuilles de l'arbre de décision.

Ainsi, une branche de l'arbre de décision correspond à la conjonction de trois propositions associant une variable linguistique à un terme linguistique. L'arbre de décision de l'indice de similarité temporelle représenté sur la figure 4.14 est réalisé de façon identique.

Par exemple, une trajectoire dont la δSm est $Forte$ à 75%, la DSm est $Faible$ à 25% et la DSM est $Forte$ à 50% active la règle présentée en gras sur la figure 4.15 dont la conclusion propose l'association de la variable linguistique SIM_S au terme linguistique $Faible$ à 25%.
$(\delta Sm = Fort \wedge DSm = Faible \wedge DSM = Fort) \Rightarrow (SIM_S = Faible)$

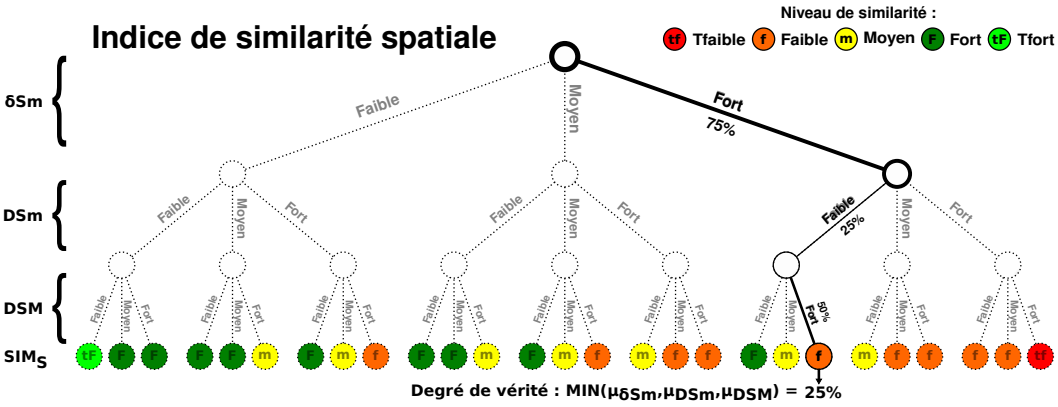


FIGURE 4.15 – Exemple d'activation d'une règle floue

Le degré d'activation de la conclusion de cette règle est calculé par évaluation de son prédicat en utilisant le minimum des degrés de vérité de ses propositions tel que proposé par les opérateurs de Mamdani [Mamdani & Assilian 1975] : $Min(\mu^{\delta Sm}_{Fort}, \mu^{DSm}_{Faible}, \mu^{DSM}_{Fort}) = Min(75, 25, 50) = 25\%$

Une fois les règles floues définies, l'inférence floue consiste à raisonner de manière approximative avec ces règles en partant du principe que "plus la condition sur les entrées est vraie plus la règle doit être appliquée". L'inférence floue est le cycle de calcul des degrés de vérité de toutes les règles ainsi que tous les ensembles flous des variables linguistiques se trouvant dans les prédicats de ces règles. Le degré de vérité de chacune des 27 règles de la matrice associative floue peut alors être calculé.

4.3.4 Calcul de l'indice de similarité spatio-temporelle

La dernière étape de "défuzzification" (figure 4.6) consiste à obtenir une valeur quantitative numérique à partir des degrés de vérité des règles floues et des conclusions associées. Cette étape transforme ces appartenances à des ensembles flous (valeurs qualitatives) en une valeur quantitative interprétable par l'utilisateur.

Dans notre cas différents indices de similarité sont introduits :
- l'indice de similarité spatiale (SIM_S),
- l'indice de similarité temporelle (SIM_T),
- l'indice de similarité spatio-temporelle (SIM_{ST}).

Pour obtenir ces valeurs quantitatives numériques, nous avons retenu la méthode du centre de gravité (COG) [Janikow 1998] utilisée de façon classique en logique floue. En fonction du degré de vérité des règles floues et de leurs conclusions, la variable linguistique SIM_S peut être associée à différents degrés aux termes linguistiques $TFaible$, $Faible$, $Moyen$, $Fort$ et $TFort$.

	$Valeur(\%)$	$Faible(\%)$	$Moyen(\%)$	$Fort(\%)$
DSM	145	0	75	25
DSm	31	90	10	0
δSm	4.5	100	0	0
DTM	157.5	0	50	50
DTm	50	0	100	0
δTm	12.75	75	25	0

TABLE 4.7 – Exemple de mesures de similarité et degrés d'appartenance associés

Chacun de ces termes disposant d'une fonction d'appartenance présentée sur la figure 4.12, les degrés de vérité associés aux termes permettent de définir les limites des surfaces actives des fonctions d'appartenance.

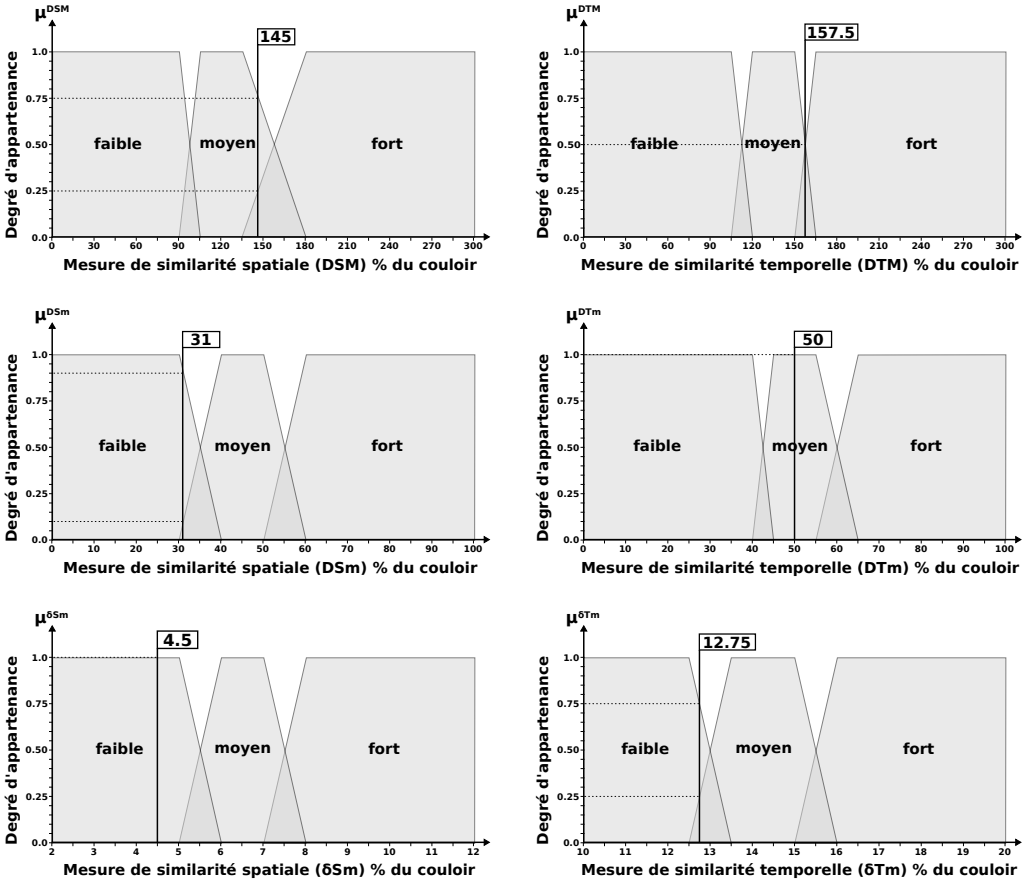


FIGURE 4.16 – Exemple de valeurs des mesures de similarité spatio-temporelle d'une trajectoire comparée à une route-type

Ainsi dans le cas de l'exemple de synthèse de la figure 4.16, les mesures de similarité entre T et $\tilde{T}$ et leurs degrés d'appartenance aux ensembles flous *Faible*, *Moyen* et *Fort* sont présentés dans le tableau 4.7.

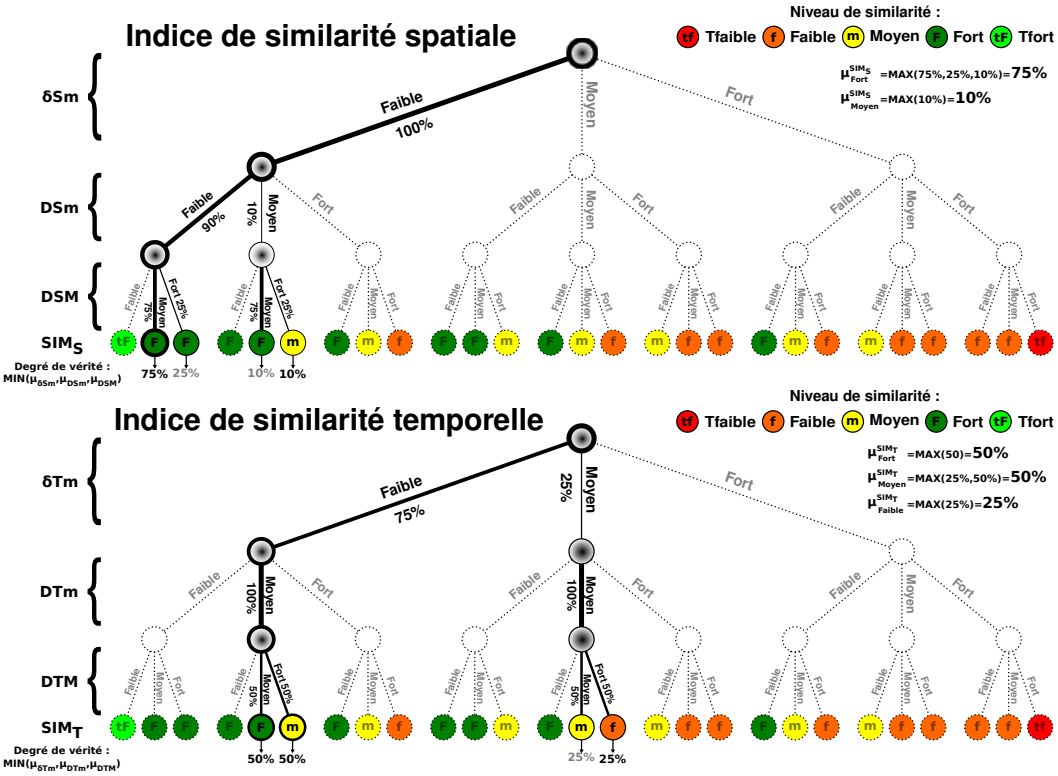


FIGURE 4.17 – Degrés de vérité des branches des arbres de décision de SIM_S et SIM_T

Les différents degrés de vérité des branches de l'arbre de décision sont calculés comme présenté dans la section 4.3.3 en fonction des degrés d'appartenance aux ensembles flous des mesures de similarité. Les 27 règles de l'arbre de décision étant évaluées en parallèle, plusieurs règles peuvent être activées avec des conclusions et des degrés d'activation différents voire, dans certains cas, contradictoires, comme indiqué sur la figure 4.17.

	$TFaible(\%)$	$Faible(\%)$	$Moyen(\%)$	$Fort(\%)$	$TFort(\%)$
SIM_S	0	0	10	75	0
SIM_T	0	25	50	50	0

TABLE 4.8 – Degrés de vérité des fonctions d'appartenance de SIM_S et SIM_T

Un ensemble flou global est construit par agrégation de chacun des ensembles flous résultant de l'évaluation des règles de l'arbre de décision. Une des méthodes

employable pour réaliser cette agrégation consiste à calculer la disjonction de toutes les règles de l'arbre (le maximum de chaque fonction d'appartenance est calculé). Les valeurs agrégées des différentes fonctions d'appartenance des indices de similarité spatiale et temporelle de l'exemple sont présentées dans le tableau 4.8.

A partir des degrés de vérité du tableau 4.8, la surface cumulée des fonctions d'appartenance est calculée. Cette surface est représentée en gris foncé sur l'exemple de la figure 4.18.

Le centre de gravité de cette surface est ensuite projeté sur l'axe des abscisses afin d'obtenir la valeur numérique de l'indice. Le centre de gravité de cette surface projeté sur l'axe des abscisses renvoie un indice SIM_S de 72% et un indice SIM_T de 55% pour l'exemple de la figure 4.18.

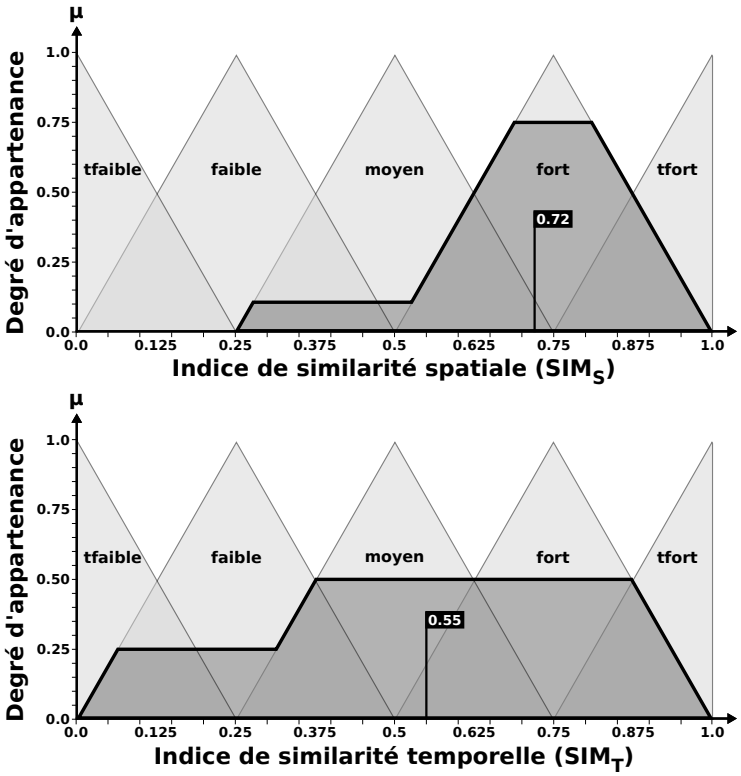


FIGURE 4.18 – Défuzzification de SIM_S et SIM_T par la méthode du centre de gravité sur les surfaces

SIM_S et SIM_T sont ensuite agrégés (figure 4.6) par calcul de leur minimum de manière à obtenir l'indice SIM_{ST} (valant 55% pour l'exemple choisi). Il est donc possible à l'aide de ces six mesures de similarité, de ces règles d'inférence floue et des fonctions d'appartenance d'obtenir un indice de similarité spatio-temporelle. Plus l'indice est fort, plus les trajectoires sont similaires.

Les courbes des différents indices de similarité spatiale, temporelle et spatio-temporelle des trajectoires du GHT de l'exemple maritime sont présentées sur les figures 4.19, 4.20 et 4.21.

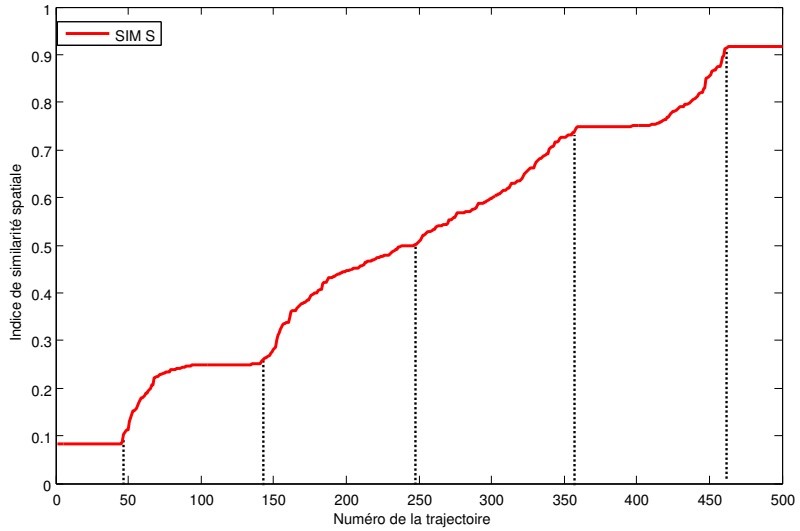


FIGURE 4.19 – Courbe de l'indice de similarité spatiale du GHT

Sur les courbes 4.20 et 4.21, on observe clairement un premier palier correspondant aux trajectoires dont l'indice de similarité spatiale et temporelle est minimal (inférieur à 0,1), ce qui correspond à une similarité avec la route-type *Très Faible*.

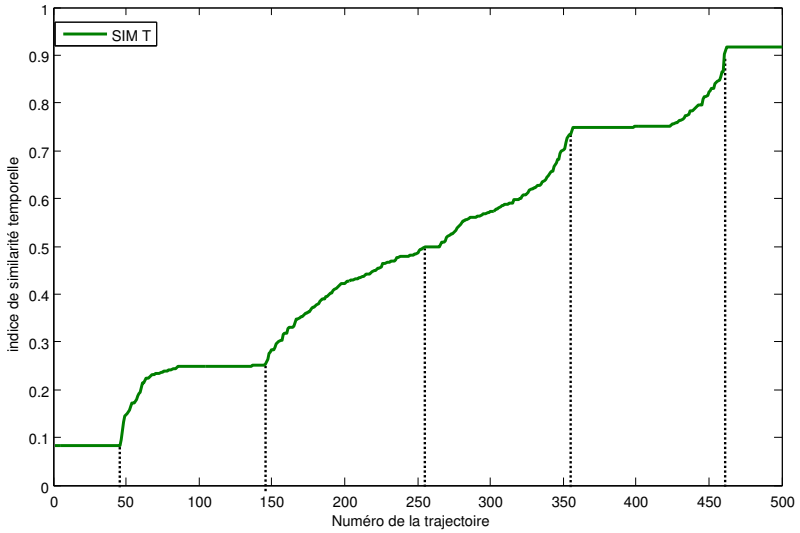


FIGURE 4.20 – Courbe de l'indice de similarité temporelle du GHT

Ce palier contient environ 50 trajectoires sur les 500 composant le GHT ce qui correspond à la valeur ($P = 10\%$) que nous avons spécifié comme seuil de détection des données aberrantes dans le processus de génération de la route-type.

De plus, on observe des variations de ces indices par paliers consécutifs d'environ 100 trajectoires (soit par tranches de 20% de notre GHT) ce qui correspondant aux percentiles statistiques des paliers choisis pour définir des limites des ensembles flous de chacune des mesures de similarité.

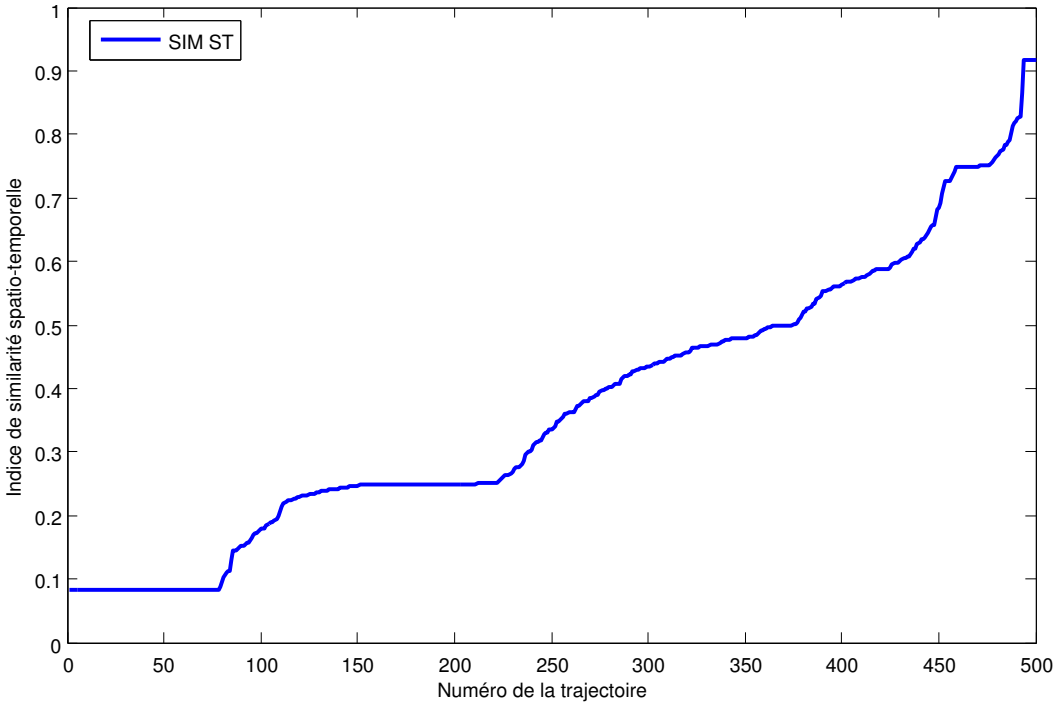


FIGURE 4.21 – Courbe de l'indice de similarité spatio-temporelle du GHT

L'analyse de la figure 4.21 permet de se rendre compte que l'indice de similarité spatio-temporelle proposé est relativement sévère. En effet, la sélection du minimum des deux indices de similarité spatiale et temporelle implique que la moindre imperfection dans une trajectoire induit une diminution de la valeur finale de l'indice de similarité spatio-temporelle.

Cet indice est cependant plus proche de la perception humaine qu'un indice renvoyé par une somme pondérée. Il est donc fort utile dans le contexte de l'analyse de trajectoires inhabituelles. Il doit être employé de manière complémentaire à l'analyse de positions.

Des réglages des différents paramètres permettant d'indiquer le seuil de tolérance de notre route-type ainsi que les limites des percentiles utilisés pour générer les limites des ensembles flous doivent désormais être réalisés en relation avec des experts du domaine applicatif afin entre autres d'obtenir des valeurs d'indices plus appropriées.

Conclusion du chapitre

Dans ce chapitre, les patrons spatio-temporels définis au chapitre 3 ont été mis à profit pour qualifier les positions et les trajectoires d'objets mobiles en temps réel. A l'aide des techniques d'analyse de similarité entre deux trajectoires présentées au chapitre 2.3, nous avons proposé un processus d'appariement entre une trajectoire partielle et une route-type. Chaque nouvelle position acquise peut ainsi être qualifiée.

De plus, différentes mesures de similarité spatiales et temporelles entre une trajectoire et une route-type ont été définies. Ces mesures sont synthétisées sous la forme d'un indice de similarité entre la trajectoire et le motif associé. Une approche logique floue a été employée dans cet objectif. Cette approche a la particularité de définir les limites des ensembles flous à l'aide d'une étude statistique portant sur les similarités spatiales et temporelles des trajectoires de GHT.

Ainsi nous obtenons un indice de similarité proche de celui perçu humainement et tenant compte du contexte via les distributions observées. De ce fait, il peut naturellement être employé pour faciliter l'analyse de trajectoires d'objets mobiles par un opérateur de surveillance du trafic.

Conclusion

Apport de la thèse

Dans cette thèse, nous nous sommes intéressés à l'étude du mouvement d'objets mobiles. Nous sommes partis du postulat que "des objets mobiles d'un même type évoluant dans un même espace ont des comportements similaires et suivent une route optimisée en fonction d'un ensemble d'objectifs" (page 16). Notre objectif est de définir un processus d'aide à la décision permettant de qualifier le comportement de ces objets mobiles. Cet objectif principal est décomposé en différentes sous-parties traitées dans les différents chapitres de cette thèse.

La première problématique est relative à la modélisation du mouvement. Cette thèse a proposé un processus d'intégration de données portant sur les objets mobiles et l'historique de leur déplacement. Un modèle conceptuel générique au déplacement d'objets mobiles évoluant dans un espace ouvert a été introduit dans le chapitre 1. Différents concepts ont été formellement définis tels que la notion de position, de trajectoire, d'itinéraire et de route-type.

Ce modèle a été instancié dans le cadre de déplacements d'objets maritimes. De plus, nous avons proposé un schéma fonctionnel mettant en œuvre les différentes étapes nécessaires au processus d'aide à la décision (de l'intégration des données dans un entrepôt à la restitution de données qualifiées en temps réel). Ainsi, un entrepôt de données volumineux de près de 5 millions de positions a été constitué sur la zone de Brest.

Une fois ce modèle conceptuel défini et les données intégrées au sein d'un entrepôt de données, l'objectif suivant consiste à extraire de l'entrepôt de données des informations intéressantes concernant les comportements habituels des objets mobiles. Pour celà, il est nécessaire au préalable de disposer d'outils d'analyse de similarité entre des trajectoires. Le concept de similarité entre trajectoires ainsi que de nombreuses techniques d'analyse de similarité ont été présentées dans le chapitre 2.

Ce concept est fortement dépendant du domaine applicatif et des caractéristiques des objets manipulés ainsi que du résultat recherché. Différents niveaux peuvent être recherchés tels que la similarité totale ou partielle, ou celle obtenue après transformations. Considérant la formalisation du concept de trajectoire proposée au chapitre 1.1.4, deux grandes catégories de techniques d'analyse de similarité ont été présentées.

La première technique consiste à effectuer une analyse de séries temporelles. La seconde est basée sur une analyse de la géométrie des objets à comparer. Les trajectoires manipulées sont à la fois composées d'une série temporelle de positions et disposent d'une représentation géométrique. C'est pourquoi, certains outils de mesure de similarité de trajectoires combinent les techniques d'analyse de séries temporelles avec celles basées sur la géométrie. Cette étude répond à la question Q1 posée en introduction (Comment définir la notion de similarité entre des trajectoires afin de les comparer ?).

Basé sur cette notion de similarité, un premier niveau de sélection de groupes de trajectoires disposant de mouvements similaires a été défini (chapitre 3.1) [Etienne *et al.* 2008, Etienne *et al.* 2009]. Nous avons proposé deux approches distinctes pour extraire ces groupes homogènes de trajectoires (GHT).

La première technique basée sur la notion de graphe de zones d'intérêt, considérant que les objets se déplacent avec l'objectif d'atteindre une zone particulière. Ce graphe de zone est défini grâce aux connaissances d'experts. La seconde technique est basée sur la notion de similarité entre trajectoires pour définir des clusters.

Dans le cadre de notre exemple applicatif maritime, nous avons obtenu comme résultat des groupes homogènes de trajectoires relativement proches. Cette partie de la thèse propose des solutions à la question Q2 (Comment sélectionner et regrouper des trajectoires similaires ?).

Au chapitre 3, nous avons proposé une technique d'extraction de motifs spatio-temporels de trajectoires à partir de ces groupes de trajectoires similaires. Ces motifs correspondent à la synthèse des mouvements des groupes d'objets mobiles, ils sont obtenu à partir d'une analyse statistique des distributions des différents paramètres de nuages de positions homologues du groupe de trajectoires.

Cette représentation sous la forme d'une trajectoire type et d'un couloir spatio-temporel associé a été présentée comme une extension du concept des boites à moustaches aux données spatio-temporelles. Une analyse des propriétés statistiques de ces motifs a également été réalisée. Cette analyse à mis en évidence la présence de variations dans la symétrie et l'étalement des masses de probabilité, confortant le choix de notre approche basée sur la médiane et les percentiles.

Ainsi, l'étude statistique d'un GHT permet d'obtenir des informations concernant les comportements de ces objets. Les résultats de cette étude sont synthétisés au sein d'une route-type dont le modèle a été formellement défini au chapitre 1 et intégrés dans une base de connaissances [Etienne *et al.* 2010a, Etienne *et al.* 2010b, Devogele *et al.* 2010]. Ce chapitre apporte une réponse à la question Q3 (Comment décrire le comportement d'un groupe de trajectoires similaires sous la forme de motifs ?).

La dernière problématique adressée dans cette thèse consiste à qualifier le comportement d'un objet mobile en fonction de son mouvement. Dans le chapitre 4, nous nous sommes intéressés à l'utilisation du patron spatio-temporel défini au chapitre 3 pour qualifier des trajectoires d'objets mobiles.

Les données temps réel de positions et de trajectoires partielles d'objets mobiles sont comparées aux motifs contenus dans la base de connaissances. Inspiré des techniques d'analyse de similarité entre deux trajectoires présentées au chapitre 2, nous avons proposé un processus d'appariement entre une trajectoire partielle et une route-type.

Ce processus d'appariement sélectionne au sein d'une base de connaissances la route-type permettant de qualifier la trajectoire d'un objet mobile suivant le même itinéraire. Nous avons également spécifié au chapitre 4 un processus de qualification spatio-temporelle des trajectoires d'objets mobiles comparées à des motifs basé sur un raisonnement flou [Etienne & Devogele 2012].

Un ensemble de mesures de similarité spatiales et temporelles a été proposé. Différentes variables floues ont été définies à partir de l'étude statistique des mesures spatiales et temporelles des trajectoires du GHT. La comparaison d'une trajectoire avec une route-type permet de calculer les différentes mesures de similarité spatiales et temporelles. Les valeurs des mesures de similarité peuvent ensuite être représentées sous la forme de variables floues ("fuzzifiées") afin d'être utilisées dans un ensemble de règles floues que nous avons déclinées à partir d'un arbre de décision.

Ces règles ont l'avantage de pouvoir être exprimées simplement à partir de variables floues ce qui facilite le recueil de connaissance auprès des experts du domaine. Le processus de raisonnement flou sur ces mesures permet d'obtenir un indice de similarité spatio-temporelle entre la trajectoire de l'objet mobile et le motif. Ce processus offre une solution à la question Q4 (Comment qualifier le comportement d'un objet mobile à l'aide de ces motifs ?).

L'indice de similarité spatio-temporel obtenu est plus proche de la perception humaine et peut être employé pour faciliter l'analyse de trajectoires d'objet mobiles pour un opérateur de surveillance de trafic. En fonction de cet indice de similarité spatio-temporelle des comportements inhabituels peuvent être détectés. Cette information est ensuite restituée à l'opérateur dans un souci d'aide à la décision [Etienne *et al.* 2011].

Grâce aux motifs spatio-temporels, un premier niveau d'analyse qualifie simplement des trajectoires sur des critères spatiaux et temporels. Ainsi, à partir d'une simple base de données gérant les postions d'objets mobiles et permettant uniquement de répondre à des questions simples du type "Où est tel objet ?", nous passons progressivement à une base de données contenant ces informations de positionne-

ment, mais aussi des connaissances de plus haut niveau liées au comportement des objets mobiles.

La démarche que nous avons présenté a permis de regrouper des connaissances (motifs), des outils de qualification (règles floues) et des données de production temps réel (les positions) au sein d'une même base de données [Devogele *et al.* 2012]. Ce type de bases appelées bases de données inductives permettent en manipulant simultanément les données et les connaissances d'analyser le comportement des objets mobiles et de poser des requêtes de plus haut niveau. Ce processus complet d'informatique décisionnelle nous permet de répondre à l'intégralité des questions posées dans l'introduction.

Dans notre exemple de surveillance du trafic maritime, les opérateurs peuvent avoir un grand nombre de navires à suivre. Ils disposent ainsi d'un outil d'aide à la surveillance réduisant leur charge cognitive en leur permettant de se focaliser sur les objets ayant des trajectoires inhabituelles. La transmission des informations qualitatives (concernant la trajectoire du navire) leur permet de concentrer leur attention sur un petit nombre de navires. De plus, ce module d'analyse est capable, une fois la trajectoire appariée, d'indiquer la destination du navire ou du prochain nœud du graphe ainsi que le temps nécessaire pour arriver à destination par comparaison avec les données de la route-type.

Un prototype fonctionnel basé sur les concepts proposés dans cette thèse a été développé. La base de données spatio-temporelles a été modélisée sous PostgreSQL en utilisant la cartouche spatiale PostGIS. Les processus d'extraction de groupes homogènes de trajectoires, d'appariement de trajectoires (par normalisation temporelle, Dynamic Time Warping, Fréchet), de génération de routes types et de qualification de positions et de trajectoires par raisonnement flou ont été implémentés sous l'environnement de développement Eclipse en Java (10000 lignes de code). L'analyse statistique des résultats des phases de génération de la route type a été réalisée à l'aide du logiciel Matlab.

Ce prototype est capable de réaliser une extraction d'un groupe homogène de trajectoires à partir d'un graphe de zones. L'étape de calcul et de sauvegarde dans une base de connaissances de la route-type de trois différents GHT a également été réalisée ; les résultats sont présentés en annexes A.8, A.9 et A.10.

Le module de qualification de trajectoires a été implémenté et testé sur différents GHT. Les résultats obtenus ont permis de classer un ensemble de trajectoires en fonction de leur similarité avec différentes routes-types. Cependant, cette phase d'analyse n'a pas encore été testée sur des données provenant de flux temps réels et n'a pas encore été présentée à des experts du domaine.

Perspectives

Dans cette section, nous présentons les perspectives à court terme permettant d'affiner les concepts et la méthodologie proposés dans cette thèse. Puis des perspectives concernant le développement de notre application sont détaillées. Enfin des perspectives à plus long terme, ouvrant sur d'autres domaines de recherche, sont décrites.

Perspectives à court terme

Cette analyse statistique nous a également permis d'obtenir des informations intéressantes concernant la répartition des masses de densité des données (skewness, kurtosis). Nous avons observé que ces paramètres évoluent en fonction du mouvement réalisé par l'objet mobile et de son environnement. Une étude plus détaillée de ces paramètres, en collaboration avec des statisticiens, serait intéressante à mettre en œuvre afin de confirmer notre observation et d'utiliser ces paramètres comme un outil de qualification du mouvement.

De plus, une étude approfondie nous permettrait de détecter les situations où les densités de probabilités ont plusieurs modes (temporel ou spatial) afin de raffiner les critères de sélection des GHT ou modifier automatiquement le graphe de zones pour éviter ce genre de situation. Ce problème a été rencontré sur l'étude du GHT de l'itinéraire du pas de calais comme l'illustre l'annexe A.10.

Dans cette thèse, nous avons utilisé une approche statistique pour construire la route-type ainsi que les fonctions d'appartenance utilisées dans l'étape de qualification des données par raisonnement flou. Le paramètre $P\%$, fixant les limites du couloir spatio-temporel est actuellement fixé empiriquement. Ce paramètre a un impact direct sur la génération du patron spatio-temporel et, par conséquent, sur les capacités d'analyse et de détection de comportements anormaux du processus de qualification. Aussi, une étude de l'impact de ce paramètre sur les performances du système de qualification serait intéressante à mettre en place.

De plus, nous avons observé que les techniques employées pour générer les patrons spatio-temporels sont fortement dépendantes de l'échantillonnage choisi pour représenter les trajectoires. Plus le pas d'échantillonnage est petit, plus le patron spatio-temporel est précis mais plus les calculs nécessaires à sa génération sont importants. Dans le chapitre 3.2.3, nous avons présenté un algorithme de filtrage des trajectoires permettant de réduire de façon importante le nombre de positions nécessaires pour représenter une trajectoire. Seules les positions clés (changement de vitesse ou d'orientation) de la trajectoire sont conservées. Cependant, les techniques employées pour apparier des trajectoires donnent de meilleurs résultats sur

des trajectoires disposant d'un même pas d'échantillonnage. Une étude de l'impact de l'échantillonnage sur la rapidité et la précision du calcul des patrons spatio-temporels serait à envisager.

Dans le chapitre 4, le processus de qualification des données est fondé sur une approche à base de logique floue. Les fonctions d'appartenance des variables floues sont directement définies à partir de statistiques de mesures de similarité. Le processus de génération de la route-type repose également sur une analyse de la distribution statistique des données (position, vitesse, cap etc...). Aussi, il serait envisageable de s'inspirer de l'approche du chapitre 4 pour spécifier les limites de la zone de normalité utilisée pour qualifier une position. Cette zone, actuellement définie par 4 bornes (avance, retard, gauche, droite) pourrait donc être présentée sous la forme d'une combinaison d'ensembles flous.

L'appariement en temps réel d'une position à une route-type reste problématique. Certaines parties d'un itinéraire peuvent être partagées avec d'autres (croisements, co-location). L'appariement en temps réel d'une position à une route-type est donc potentiellement multiple et variable au cours du temps. Le recoupement d'informations telles que l'historique de la trajectoire, l'identifiant de l'objet mobile et sa destination prévue facilitent cette tâche. Néanmoins, à chaque nouvelle position reçue de l'objet mobile, l'appariement est susceptible d'être modifié. C'est pourquoi, un degré de confiance pourrait être ajouté à l'appariement entre une trajectoire partielle et une route-type afin d'être intégré dans le processus de qualification des positions et des trajectoires.

Perspectives applicatives

Une autre versant de cette analyse serait de prendre en compte les spécificités dues à l'environnement dans lequel l'objet mobile évolue (les saisons et les marées dans le cadre de notre exemple). Certains phénomènes tels que la météo, les phénomènes climatiques, les saisons, peuvent impacter le comportement des objets mobiles. Ainsi, le patron spatio-temporel pourrait être raffiné en effectuant une analyse temporelle pour détecter des modifications cycliques temporelles des comportements des objets mobiles (liées aux cycles des marées, aux périodes de migrations des animaux, aux saisons). Différents patrons spatio-temporels plus ou moins précis pourraient alors être obtenus en fonction de critères plus ou moins stricts d'extraction du GHT (même type d'objet mobile, même saison temporelle...). Il serait intéressant d'appliquer d'autres critères de sélection du GHT ainsi que d'enrichir cette analyse en la couplant avec des travaux portant sur la prise en compte de l'environnement ([Richard 2007]), des statistiques d'accidents ([Degré 2007]) et des comportements illicites ([Morel *et al.* 2008]).

Il serait également intéressant d'effectuer un recueil d'expertise auprès de ces opérateurs de surveillance du trafic afin de compléter les mesures de similarité et optimiser les règles floues du système. En effet, le raisonnement d'un opérateur n'est pas basé sur une combinaison exhaustive des différentes mesures tel que nous l'avons proposé avec une matrice associative floue mais plutôt sur un sous-ensemble simplifié de ces combinaisons pour définir des règles floues. De plus, un protocole de validation, pourraient être réalisés afin de vérifier si les trajectoires soulignées comme ayant un comportement anormal par notre système sont également détectées comme telles par un opérateur de surveillance.

Ce protocole nécessite au préalable une étude de sémiologie graphique afin d'afficher les informations de similarité de manière optimale et fournir une interface adaptée aux opérateurs.

Les processus de définition des indices de similarité et d'analyse de trajectoires doivent être consolidées. Il conviendrait désormais de tester le modèle que nous avons proposé en le mettant à disposition d'opérateurs de surveillance du trafic maritime basés dans les $CROSS$ (Centres régionaux opérationnels de surveillance et de sauvetage). Ces opérateurs peuvent avoir à surveiller jusqu'à plus de 250 navires affichés simultanément sur leurs écrans de contrôle. Cet outil couplé à un système de visualisation adapté pourrait leur permettre de se focaliser sur les trajectoires de navires ayant un comportement considéré par le système comme anormal et ainsi diminuer leur charge cognitive. Cette phase de validation *in situ* devrait déboucher sur une industrialisation du prototype.

La méthodologie présentée bien qu'appliquée à un exemple maritime est générique. Elle pourrait par exemple être appliquée à l'étude du déplacement d'animaux, de phénomènes naturels ou de piétons. Cependant, la base de données maritime dont nous disposons est loin d'avoir livré toutes ses informations. Nous disposons également d'autres sources de données maritimes couvrant l'ensemble du globe. Une analyse à plus grande échelle est donc envisageable. Ce genre de bases de données peut servir de point de départ à une expérimentation sur la généralisation de routes-types ou la détection de nouveaux types de comportement.

Perspectives à plus long terme

Une fois une trajectoire appariée à un patron, il est également envisageable de prédire le comportement de l'objet mobile avec une précision connue. Cette précision est dépendante des bornes spatio-temporelles de la route-type sélectionnée. Nous disposons d'une base de données assez volumineuse pour être utilisée à la fois pour extraire des patrons spatio-temporels et ensuite tester les prédictions réalisées grâce à ces patrons sur des trajectoires observées.

De plus, la sélection ou la suppression manuelle par l'opérateur des trajectoires considérées comme anormales constituerait une extension possible du système qui permettrait une mise à jour des outils de qualification des trajectoires par apprentissage.

Les statistiques utilisées pour générer les fonctions d'appartenances des variables floues sont basées sur une étude des trajectoires du *GHT* prises dans leur ensemble. Elles sont donc intéressantes pour qualifier une trajectoire dans son ensemble mais peuvent se révéler insuffisantes pour qualifier une trajectoire partielle. Il serait judicieux d'effectuer une étude de l'évolution des statistiques des mesures de similarité sur une suite de trajectoires partielles du même *GHT* tronquées à différents pas d'échantillonnage.

Les trajectoires des objets mobiles sont extraites de la base de données en s'appuyant sur un graphe de zones. Par analogie avec ce graphe de zone, il serait intéressant de découper les trajectoires en une suite de tronçons afin de faciliter l'analyse du comportement d'un navire sur les sous-parties d'une trajectoire ayant des caractéristiques communes (vitesse, cap, taux de rotation...). Ce découpage permettrait de décrire, à l'aide de variables linguistiques, les trajectoires d'objets mobiles et dans un deuxième temps de comparer qualitativement des trajectoires entres elles (ou avec une route-type). Une description possible serait : *pour aller de la zone A à la zone B pour un objet de tel type, il faut aller tout droit avec une vitesse d'environ 20 km/h et un cap au sud-est, puis au bout de 10 minutes changer de cap pour aller à l'est et aller tout droit pendant 2 minutes.* Une fois encore, l'algorithme de Douglas et Peucker spatio-temporel conserve les positions remarquables d'une trajectoire. Celles-ci pourraient être utilisées afin de mettre en évidence les jonctions entre deux tronçons. Chaque position d'un tronçon ayant des propriétés communes, l'analyse et la détection de données anormales sur une période en seraient facilitées.

Nous avons par ailleurs défini des outils de comparaisons entre une trajectoire et une route-type. Néanmoins, nous ne disposons pas du même genre d'outil pour comparer deux routes types entre elles. Une étude comparative entre deux patrons spatio-temporels d'objets mobiles de type différent suivant un même itinéraire serait intéressante à réaliser dans l'objectif de définir des opérateurs de comparaison et de combinaison de patrons spatio-temporels. Ceci permettrait d'obtenir des outils de généralisation des patrons en fonction du niveau de détail recherché (critères de sélection plus ou moins restrictifs, gestion de la multi-modalité dans les statistiques).

Conclusion du chapitre

Cette thèse a proposé une méthode afin d'extraire des motifs spatio-temporels. Ces motifs sont définis à partir de clusters de trajectoires regroupant les déplacements d'objets mobiles de même type et ayant un itinéraire similaire. En intégrant ces motifs dans une base de données inductive, il a été montré qu'il était possible de qualifier les déplacements (positions et trajectoires) en temps réels et ainsi d'identifier des comportements inhabituels. Cette thèse est ainsi à la croisée de l'informatique décisionnelle et plus particulièrement de la fouille de données d'une part et de l'étude des déplacements d'objets mobiles d'autre part. Elle permet de répondre à des requêtes de hauts niveaux à l'aide des connaissances extraites. Le processus générique défini a été appliqué à un jeu de données réelles conséquent portant sur les déplacements des navires.

Cette thèse ouvre également de nombreuses perspectives dans différents axes allant de la prédiction des déplacements des objets mobiles à la segmentation de trajectoires en parties homogènes. Les résultats de cette thèse doivent pouvoir s'appliquer à des domaines variés tel l'aide à la surveillance du trafic, l'étude des déplacements journaliers, ou l'analyse de la migration d'animaux.

Bibliographie

[Agarwal *et al.* 2000] P.K. Agarwal, L. Arge et J. Erickson. *Indexing Moving Points*. Dans Proceedings of the nineteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pages 175–186. ACM, 2000. (Cité en page 20.)

[Agrawal & Srikant 1995] R. Agrawal et R. Srikant. *Mining Sequential Patterns*. Dans Proceedings of the Eleventh International Conference on Data Engineering, pages 3–14. IEEE Computer Society, 1995. (Cité en page 6.)

[Agrawal *et al.* 1993a] R. Agrawal, C. Faloutsos et A. Swami. *Efficient Similarity Search In Sequence Databases*. LNCS, Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms, vol. 730, pages 69–84, 1993. (Cité en pages 38, 39 et 40.)

[Agrawal *et al.* 1993b] R. Agrawal, T. Imieliński et A. Swami. *Mining Association Rules Between Sets of Items in Large Databases*. Dans ACM SIGMOD Record, volume 22, pages 207–216. ACM, 1993. (Cité en page 6.)

[Agrawal *et al.* 1995] R. Agrawal, K. Lin, H.S. Sawhney et K. Shim. *Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases*. Dans Proceedings of the 21th International Conference on Very Large Data Bases. Citeseer, 1995. (Cité en pages 45 et 46.)

[Alt & Godau 1992] H. Alt et M. Godau. *Measuring the Resemblance of Polygonal Curves*. Dans Proceedings of the eighth annual symposium on Computational geometry, pages 102–109. ACM, 1992. (Cité en page 50.)

[Alt & Godau 1995] H. Alt et M. Godau. *Computing the Fréchet Distance Between two Polygonal Curves*. International Journal of Computational Geometry and Applications, vol. 5, no. 1, pages 75–91, 1995. (Cité en pages 48 et 51.)

[Alt & Guibas 1996] H. Alt et L.J. Guibas. *Discrete Geometric Shapes : Matching, Interpolation, and Approximation. A survey.* Rapport technique, Handbook of Computational Geometry, 1996. (Cité en pages 38, 48 et 50.)

[Alt & Guibas 2000] H. Alt et L. J. Guibas. *Discrete Geometric Shapes : Matching, Interpolation, and Approximation*. Dans J.-R. Sack et J. Urrutia, editeurs, Handbook of Computational Geometry, pages 121 – 153. North-Holland, Amsterdam, 2000. (Cité en page 50.)

[Alt *et al.* 2001] H. Alt, C. Knauer et C. Wenk. *Matching Polygonal Curves with Respect to the Fréchet Distance*. page 74, 2001. (Cité en pages 57, 85 et 115.)

[Alt 2009] H. Alt. *The Computational Geometry of Comparing Shapes*. Dans Susanne Albers, Helmut Alt et Stefan Näher, editeurs, Efficient Algorithms,

volume 5760 of *Lecture Notes in Computer Science*, pages 235–248. Springer Berlin / Heidelberg, 2009. (Cité en pages 48 et 50.)

[Andrienko *et al.* 2009] G. Andrienko, N. Andrienko, S. Rinzivillo, M. Nanni et D. Pedreschi. *A Visual Analytics Toolkit for Cluster-Based Classification of Mobility Data*. Dans Advances in Spatial and Temporal Databases : 11th International Symposium, SSTD 2009 Aalborg, Denmark, July 8-10, 2009 Proceedings, volume 5644, page 432. Springer-Verlag New York Inc, 2009. (Cité en page 68.)

[Ankerst *et al.* 1999] M. Ankerst, M. M. Breunig, H. Kriegel et J. Sander. *OPTICS : Ordering Points to Identify the Clustering Structure*. Dans ACM SIGMOD International Conference on Management of Data, pages 49–60. ACM Press, 1999. (Cité en page 68.)

[Anouar & Wafa 2006] B.H. Anouar et M. Wafa. *Raisonnement analogique et similarité*. 2006. (Cité en page 120.)

[Aronov *et al.* 2006] B. Aronov, S. Har-Peled, C. Knauer, Y. Wang et C. Wenk. *Fréchet Distance for Curves, Revisited*. Algorithms–ESA 2006, pages 52–63, 2006. (Cité en pages 50 et 51.)

[Bertin 1967] J. Bertin. Sémiologie graphique. les diagrammes, les réseaux, les cartes. EHESS. Ecole des hautes etudes en sciences sociales, 1967. (Cité en page 102.)

[Bertrand *et al.* 2007] F. Bertrand, A. Bouju, C. Claramunt, T. Devogele et C. Ray. Web and wireless geographical information systems, volume 4857 of *Lecture Notes in Computer Science*, chapitre Web Architecture for Monitoring and Visualizing Mobile Objects in Maritime Contexts, pages 94–105. Springer Berlin / Heidelberg, 2007. (Cité en pages 74 et 75.)

[Bimonte 2007] S. Bimonte. *Intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation*. Thèse de doctorat, L'Institut National des Sciences Appliquées de Lyon, 2007. (Cité en pages 4 et 5.)

[Bobillo & Straccia 2008] F. Bobillo et U. Straccia. *FuzzyDL : an Expressive Fuzzy Description Logic Reasoner*. Dans EEE International Conference on Fuzzy Systems, 2008. FUZZ-IEEE 2008.(IEEE World Congress on Computational Intelligence)., pages 923–930. IEEE, 2008. (Cité en page 128.)

[Bouchon-Meunier 1995] B. Bouchon-Meunier. La logique floue et ses applications. Addison-Wesley France, 1995. (Cité en page 120.)

[Bouchon-Meunier 2007] B. Bouchon-Meunier. La logique floue, que sais-je ? Presses Universitaires de France, 4 édition, 2007. (Cité en page 128.)

[Bozkaya *et al.* 1997] T. Bozkaya, N. Yazdani et M. Özsoyoglu. *Matching and Indexing Sequences of Different Lengths*. Dans Proceedings of the sixth international conference on Information and knowledge management, pages 128–135. ACM, 1997. (Cité en page 46.)

[Buard *et al.* 2011] E. Buard, M. Brasebin et S. IGN. *Visual exploration of large animal trajectories*. Dans 25th International Cartographic Conference (ICC'11). ICC, 2011. (Cité en page 7.)

[Buchin *et al.* 2008] K. Buchin, M. Buchin et C. Wenk. *Computing the Fréchet Distance Between Simple Polygons*. Computational Geometry, vol. 41, no. 1-2, pages 2–20, 2008. (Cité en page 48.)

[Buchin *et al.* 2009] K. Buchin, M. Buchin, M. van Kreveld et J. Luo. *Finding long and similar parts of trajectories*. Dans Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pages 296–305. ACM, 2009. (Cité en pages 60 et 64.)

[Béguin & Pumain 2010] M. Béguin et D. Pumain. La représentation des données géographiques. statistique et cartographie (3 éd.). Cursus, 2010. (Cité en page 102.)

[Cha 2007] S.H. Cha. *Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions*. Dans International Journal Of Mathematical Models And Methods In Applied Sciences. Citeseer, 2007. (Cité en page 41.)

[Chen & Ng 2004] L. Chen et R. Ng. *On the Marriage of Lp-norms and Edit Distance*. Dans Proceedings of the Thirtieth international conference on Very large data bases-Volume 30, pages 792–803. VLDB Endowment, 2004. (Cité en pages 40 et 47.)

[Chen *et al.* 2004] L. Chen, M.T. Özsu et V. Oria. *Symbolic Representation and Retrieval of Moving Object Trajectories*. Dans Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, pages 227–234. ACM, 2004. (Cité en pages 47, 59 et 63.)

[Chen *et al.* 2005] L. Chen, M. T. Ozsu et V. Oria. *Robust and Fast Similarity Search for Moving Object Trajectories*. Dans SIGMOD '05 : Proceedings of the 2005 ACM SIGMOD international conference on Management of data, pages 491–502, New York, NY, USA, 2005. ACM. (Cité en pages 47, 59 et 63.)

[Cheng *et al.* 2004] R. Cheng, D.V. Kalashnikov et S. Prabhakar. *Querying Imprecise Data in Moving Object Environments*. IEEE Transactions on Knowledge and Data Engineering, vol. 16, no. 9, pages 1112–1127, 2004. (Cité en page 21.)

[Cohen *et al.* 2003] W.W. Cohen, P. Ravikumar et S.E. Fienberg. *A Comparison of String Distance Metrics for Name-Matching Tasks.* Dans Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web (IIWeb-03), 2003. (Cité en page 46.)

[COST 2009] COST. *Memorandum of Understanding for the Implementation of a European Concerted Research Action Designated as COST Action Ic0903 : Knowledge Discovery From Moving Objects (MOVE).* European Cooperation in the field of Scientific and Technical Research, June 2009. (Cité en page 3.)

[Crochemore & Rytter 1994] M. Crochemore et W. Rytter. Text algorithms. Oxford University Press, USA, 1994. (Cité en page 46.)

[Dahabiah 2010] A. Dahabiah. *Extraction de connaissances et indexation de données multimédia pour la détection anticipée d'événements indésirables.* Thèse de doctorat, Dépt. Image et Traitement Information (Institut Télécom-Télécom Bretagne), Laboratoire de traitement de l'information médicale, Université de Rennes 1, Université Européenne de Bretagne, 2010. (Cité en page 41.)

[Das *et al.* 1997] G. Das, D. Gunopulos et H. Mannila. *Finding Similar Time Series.* Principles of Data Mining and Knowledge Discovery, pages 88–100, 1997. (Cité en page 46.)

[D'Auria *et al.* 2006] M. D'Auria, M. Nanni et D. Pedreschi. *Time-focused Density-Based Clustering of Trajectories of Moving Objects.* Journal of Intelligent Information Systems, vol. 27, no. 3, pages 267–289, novembre 2006. (Cité en page 24.)

[Degré 2007] T. Degré. *The Use of Risk Concept to Characterize and Select High Risk Vessels for Ship Inspections.* 12th International Congress of the International Maritime Association of the Mediterranean, vol. 6, pages 37–49, 2007. (Cité en page 144.)

[Devogele *et al.* 2010] T. Devogele, L. Etienne et A. Bouju. *Mesures de similarité de trajectoires suivant le même itinéraire.* Dans Conférence internationale de Géomatique et Analyse Spatiale SAGEO'10, 2010. (Cité en page 140.)

[Devogele *et al.* 2012] T. Devogele, L. Etienne, C. Ray et C. Claramunt. Mobility data : Modeling, management, and understanding, chapitre Part III – Mobility Applications, Maritime Applications. Cambridge press, ouvrage à paraître, 2012. (Cité en page 142.)

[Devogele 2000] T. Devogele. *Mesure d'exactitude et processus de fusion à l'aide de la distance de Fréchet discrète.* Revue internationale de Géomatique, vol. 10, pages 359–381, 2000. (Cité en pages 55, 62, 84 et 85.)

[Devogele 2002] T. Devogele. *A New Merging Process for Data Integration Based on the Discrete Fréchet Distance*. Dans Advances in Spatial Data Handling : 10th International Symposium on Spatial Data Handling, pages 167–181. Springer Verlag, 2002. (Cité en pages 55, 57, 62, 64, 84, 85 et 115.)

[Ding *et al.* 2008a] H. Ding, G. Trajcevski et P. Scheuermann. *Efficient Similarity Join of Large Sets of Moving Object Trajectories*. Dans Proceedings of the 2008 15th International Symposium on Temporal Representation and Reasoning, pages 79–87, Washington, DC, USA, 2008. IEEE Computer Society. (Cité en page 61.)

[Ding *et al.* 2008b] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang et E. Keogh. *Querying and Mining of Time Series Data : Experimental Comparison of Representations and Distance Measures*. Proceedings of the VLDB Endowment, vol. 1, no. 2, pages 1542–1552, 2008. (Cité en pages 39, 45, 47, 61 et 64.)

[Dodge 2011] S. Dodge. *Exploring Movement Using Similarity Analysis*. Thèse de doctorat, Universität Zürich, 2011. (Cité en page 61.)

[Douglas & Peucker 1973] D.H. Douglas et T.K. Peucker. *Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature*. Cartographica : The International Journal for Geographic Information and Geovisualization, vol. 10, pages 112–122, 1973. (Cité en page 74.)

[du Mouza & Rigaux 2005] C. du Mouza et P. Rigaux. *Mobility Patterns*. GeoInformatica, vol. 9, no. 4, pages 297–319, 2005. (Cité en page 2.)

[Dummett 1959] M. Dummett. *Propositional Calculus with Denumerable Matrix*. Journal of Symbolic Logic, vol. 27, pages 97–106, 1959. (Cité en page 128.)

[Dykes & Mountain 2003] J.A. Dykes et D.M. Mountain. *Seeking Structure in Records of Spatio-Temporal Behaviour : Visualization Issues, Efforts and Applications*. Computational Statistics & Data Analysis, vol. 43, no. 4, pages 581–603, 2003. (Cité en page 23.)

[Eiter & Mannila 1994] T. Eiter et H. Mannila. *Computing Discrete Fréchet Distance*. Rapport technique, Technische Universitat Wien, 1994. (Cité en pages 48, 51 et 106.)

[Ester *et al.* 1996] M. Ester, H.P. Kriegel, J. Sander et X. Xu. *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. Dans Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining, volume 1996, pages 226–231. Portland : AAAI Press, 1996. (Cité en page 68.)

[Etienne & Devogele 2012] L. Etienne et T. Devogele. *Mesures de similarité de trajectoires basées sur l'utilisation de patrons spatio-temporels.* Ingénierie des Systèmes d'Information, numéro à paraître, 2012. (Cité en page 141.)

[Etienne *et al.* 2008] L. Etienne, T. Devogele et A. Bouju. *Outils d'aide aux décideurs concernant le suivi de navires : suivi de trajectoires relatives entre navires et détection de trajectoires inhabituelles.* Dans 7ème journées scientifiques et techniques du CETMEF, 2008. (Cité en page 140.)

[Etienne *et al.* 2009] L. Etienne, T. Devogele et A. Bouju. *Analyse de similarité de trajectoires d'objets mobiles suivant le même itinéraire : Application aux trajectoires de navires.* Ingénierie des Systèmes d'Information, vol. 14, no. 5/2009, pages 85–106, 2009. (Cité en pages 31, 84 et 140.)

[Etienne *et al.* 2010a] L. Etienne, T. Devogele et A. Bouju. *Analyse temps réel du comportement d'objets mobiles évoluant dans un espace ouvert.* Dans Actes du 5ème atelier Représentation et raisonnement sur le temps et l'espace (RTE 2010), 2010. (Cité en page 140.)

[Etienne *et al.* 2010b] L. Etienne, T. Devogele et A. Bouju. *Spatio-Temporal Trajectory Analysis of Mobile Objects Following the Same Itinerary.* Dans Proceedings of the International Symposium on Spatial Data Handling (SDH), pages 86–91, 2010. (Cité en pages 84 et 140.)

[Etienne *et al.* 2011] L. Etienne, C Ray et G. Mcardle. *Spatio-Temporal Visualisation of Outliers.* Dans Proceedings of the international workshop on Maritime Anomaly Detection (MAD), page 119, 2011. (Cité en page 141.)

[Faloutsos *et al.* 1994] C. Faloutsos, M. Ranganathan et Y. Manolopoulos. *Fast Subsequence Matching in Time-Series Databases.* Dans SIGMOD '94 : Proceedings of the 1994 ACM SIGMOD international conference on Management of data, pages 419–429, New York, NY, USA, 1994. ACM. (Cité en pages 39 et 40.)

[Faloutsos *et al.* 1997] C. Faloutsos, HV Jagadish, A.O. Mendelzon et T. Milo. *A Signature Technique for Similarity-Based Queries.* Dans Compression and Complexity of Sequences 1997. Proceedings, pages 2–20. IEEE, 1997. (Cité en pages 38 et 40.)

[Frank 2001] A.U. Frank. *Socio-Economic Units : Their Life and Motion.* Life and motion of socio-economic units, vol. 8, page 12, 2001. (Cité en page 22.)

[Frentzos *et al.* 2007] E. Frentzos, K. Gratsias et Y. Theodoridis. *Index-based Most Similar Trajectory Search.* ICDE 2007, IEEE 23rd International Conference on Data Engineering, pages 816–825, 2007. (Cité en pages 60 et 64.)

[Frentzos *et al.* 2008] E. Frentzos, N. Pelekis, I. Ntoutsi et Y. Theodoridis. *Trajectory Database Systems.* Mobility, Data Mining and Privacy, pages 151–187, 2008. (Cité en page 20.)

[Fréchet 1905] M. Fréchet. *Sur l'écart de deux courbes et sur les courbes limites.* Transactions of the American Mathematical Society, vol. 6, no. 4, pages 435–449, 1905. (Cité en pages 50, 57, 85 et 115.)

[Gayathri Monicka & Ramash Kumar 2011] J. Gayathri Monicka et K. Ramash Kumar. *Performance Evaluation of Membership Functions on Fuzzy Logic Controlled AC Voltage Controller for Speed Control of Induction Motor Drive.* International Journal of Computer Applications, vol. 13, no. 5, pages 8–12, 2011. (Cité en page 123.)

[Giannotti & Pedreschi 2008] F. Giannotti et D. Pedreschi. Mobility, data mining and privacy : Geographic knowledge discovery. Springer Publishing Company, Incorporated, 2008. (Cité en pages 23 et 24.)

[Giannotti *et al.* 2007] F. Giannotti, M. Nanni, F. Pinelli et D. Pedreschi. *Trajectory Pattern Mining.* Dans KDD '07 : Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 330–339, New York, NY, USA, 2007. ACM. (Cité en page 24.)

[Giannotti *et al.* 2009] F. Giannotti, M. Nanni, D. Pedreschi, C. Renso et R. Trasarti. *Mining Mobility Behavior from Trajectory Data.* Dans CSE (4), pages 948–951, 2009. (Cité en page 24.)

[Gomez *et al.* 2011a] L. Gomez, B. Kuijpers, B. Moelans et A. Vaisman. *A State-of-the-Art in Spatio-Temporal Data Warehousing, OLAP and Mining.* Integrations of Data Warehousing, Data Mining and Database Technologies : Innovative Approaches, page 200, 2011. (Cité en page 4.)

[Gomez *et al.* 2011b] L. Gomez, B. Kuijpers et A. Vaisman. *A Data Model and Query Language for Spatio-temporal Decision Support.* Geoinformatica, pages 1–42, 2011. (Cité en page 4.)

[Gonzalez *et al.* 2008] M.C. Gonzalez, C.A. Hidalgo et A.L. Barabási. *Understanding Individual Human Mobility Patterns.* Nature, vol. 453, no. 7196, pages 779–782, 2008. (Cité en page 2.)

[Goodrich *et al.* 1999] M.T. Goodrich, J.S.B. Mitchell et M.W. Orletsky. *Approximate Geometric Pattern Matching Under Rigid Motions.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 4, pages 371–379, 1999. (Cité en page 48.)

[Goralski & Gold 2008] R. Goralski et C.M. Gold. *Marine GIS : Progress in 3D Visualization for Dynamic GIS.* Dans SDH, pages 401–416, 2008. (Cité en page 11.)

[Groeneveld & Meeden 1984] R.A. Groeneveld et G. Meeden. *Measuring Skewness and Kurtosis.* The Statistician, pages 391–399, 1984. (Cité en pages 107 et 108.)

[Guha *et al.* 1998] S. Guha, R. Rastogi et K. Shim. *CURE : an efficient clustering algorithm for large databases.* Dans ACM SIGMOD Record, volume 27, pages 73–84. ACM, 1998. (Cité en page 68.)

[Gödel 1932] K. Gödel. *Zum intuitionistischen Aussagenkalkül.* Anzieger Akademie der Wissenschaften Wien, vol. 69, pages 65–66, 1932. (Cité en page 128.)

[Güting *et al.* 2000] R.H. Güting, M.H. Böhlen, M. Erwig, C.S. Jensen, N.A. Lorentzos, M. Schneider et M. Vazirgiannis. *A Foundation for Representing and Querying Moving Objects.* ACM Transactions on Database Systems (TODS), vol. 25, no. 1, page 42, 2000. (Cité en page 21.)

[Hattori & Torii 1993] K. Hattori et Y. Torii. *Effective Algorithms for the Nearest Neighbor Method in the Clustering Problem.* Pattern Recognition, vol. 26, no. 5, pages 741–746, 1993. (Cité en page 68.)

[Hausdorff 1918] F. Hausdorff. *Dimension und äußeres Maß.* Mathematische Annalen, vol. 79, no. 1, pages 157–179, 1918. (Cité en page 49.)

[Hay 1963] L.S. Hay. *Axiomatization of the Infinite-Valued Predicate Calculus.* The Journal of Symbolic Logic, vol. 28, no. 1, pages 77–86, 1963. (Cité en page 128.)

[Hogg & Tanis 2009] R.V. Hogg et E.A. Tanis. Probability and statistical inference. lavoisier, 8 édition, 8 2009. (Cité en page 41.)

[Hooijberg *et al.* 1997] M. Hooijberg, M. Hooijberg et M. Hooijberg. Practical geodesy. Springer, 1997. (Cité en pages 22 et 40.)

[Hornsby & Egenhofer 2002] K. Hornsby et M.J. Egenhofer. *Modeling Moving Objects over Multiple Granularities.* Annals of Mathematics and Artificial Intelligence, vol. 36, no. 1, pages 177–194, 2002. (Cité en page 23.)

[Hägerstrand 1970] T. Hägerstrand. *What About People in Regional Science ?* Papers in Regional Science, vol. 24, no. 1, pages 6–21, 1970. (Cité en page 23.)

[IALA 2002] IALA. *Guidelines on the Universal Automatic Identification System (AIS), Volume 1, Part II – Technical Issues, Edition 1.1.* Rapport technique, International Association of Marine Aids to Navigation and Lighthouse Authorities, 2002. (Cité en page 175.)

[IALA 2004] IALA. *The Automatic Identification System (AIS), Volume 1, Part I, Operational Issues.* Rapport technique, International Association of Marine Aids to Navigation and Lighthouse Authorities, 2004. (Cité en page 175.)

[IALA 2009] IALA. *Guide for VTS Authorities on Fatigue Awareness and Human Reliability.* Rapport technique, International Association of Marine Aids to Navigation and Lighthouse Authorities, 2009. (Cité en page 9.)

[IALA 2010] IALA. *Technical Clarification on Recommendation ITU-R M.1371-3 Edition 2.4*. Rapport technique, International Association of Marine Aids to Navigation and Lighthouse Authorities, 2010. (Cité en page 175.)

[Idri *et al.* 2007] A. Idri, A. Zahi, M. ElKoutbi et A. Abran. *Impacts des techniques de construction des ensembles flous sur la précision d'un modèle d'estimation des coûts de logiciels par analogie floue*. 4th International Conference : Sciences of Electronic, Technologies of Information and Telecommunications, 2007. (Cité en page 124.)

[Imielinski & Mannila 1996] T. Imielinski et H. Mannila. *A Database Perspective on Knowledge Discovery*. Communications of the ACM, vol. 39, no. 11, pages 58–64, 1996. (Cité en page 34.)

[IMO 1995] IMO. *Resolution A.817(19) : Performance Standards for Electronic Chart Display and Information Systems (ECDIS)*. Rapport technique, International Maritime Organization, 1995. (Cité en page 11.)

[IMO 2007] IMO. *Development of an E-Navigation Strategy*. Rapport technique, International Maritime Organization, 2007. (Cité en page 175.)

[Itakura 1975] F. Itakura. *Minimum Prediction Residual Principle Applied to Speech Recognition*. IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 23, no. 1, pages 67–72, 1975. (Cité en page 44.)

[ITU 2001] ITU. *ITU-R Recommendation M.1371-1, Technical Characteristics for a Universal Shipborne Automatic Identification System Using Time Division Multiple Access in the Maritime Mobile Band*. Rapport technique, International Telecommunications Union, 2001. (Cité en pages 174 et 175.)

[Jain & Dubes 1988] A.K. Jain et R.C. Dubes. Algorithms for clustering data. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988. (Cité en page 68.)

[Janikow 1998] C.Z. Janikow. *Fuzzy Decision Trees : Issues and Methods*. IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics., vol. 28, no. 1, pages 1–14, 1998. (Cité en page 131.)

[Kaufman & Rousseeuw 1990] L. Kaufman et P.J. Rousseeuw. Finding groups in data : an introduction to cluster analysis, volume 39. Wiley Online Library, 1990. (Cité en page 67.)

[Keogh & Ratanamahatana 2005] E. Keogh et C.A. Ratanamahatana. *Exact Indexing of Dynamic Time Warping*. Knowledge and Information Systems, vol. 7, no. 3, pages 358–386, 2005. (Cité en page 44.)

[Kharrat *et al.* 2008] A. Kharrat, I. S. Popa, K. Zeitouni et S. Faiz. Clustering algorithm for network constraint trajectories, chapitre Clustering Algorithm for Network Constraint Trajectories, pages 631–647. Springer Berlin Heidelberg, 2008. (Cité en page 7.)

[Kim *et al.* 2004] S.W. Kim, S. Park et W.W. Chu. *Efficient Processing of Similarity Search Under Time Warping in Sequence Databases : An Index-based Approach*. Information Systems, vol. 29, no. 5, pages 405–420, 2004. (Cité en page 44.)

[Klir & Yuan 1995] G.J. Klir et B. Yuan. Fuzzy sets and fuzzy logic : Theory and applications. Prentice Hall PTR Upper Saddle River, NJ, USA, 1995. (Cité en page 128.)

[Knorr *et al.* 2000] E.M. Knorr, R.T. Ng et V. Tucakov. *Distance-based Outliers : Algorithms and Applications*. The VLDB Journal, vol. 8, no. 3-4, pages 237–253, 2000. (Cité en page 7.)

[Kohout 2007] L.J. Kohout. *Power Sets and Implication Operators Revisited : A Retrospective Look at the Foundational and Conceptual Issues in Bandler and Kohout's Paper After 29 Years*. Dans 5th Conference of the European Society for Fuzzy Logic and Technology, pages 429–436, 2007. (Cité en page 128.)

[Kollios *et al.* 1999] G. Kollios, D. Gunopulos et V.J. Tsotras. *On Indexing Mobile Objects*. Dans Proceedings of the eighteenth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pages 261–272. ACM, 1999. (Cité en page 20.)

[Laube *et al.* 2007] P. Laube, T. Dennis, P. Forer et M. Walker. *Movement Beyond the Snapshot-Dynamic Analysis of Geospatial Lifelines*. Computers, Environment and Urban Systems, vol. 31, no. 5, pages 481–501, 2007. (Cité en page 23.)

[Le Guen 2001] M. Le Guen. *La boîte à moustaches de TUKEY, un outil pour initier à la Statistique*. Statistiquement Vôtre, vol. 4, 2001. (Cité en page 78.)

[Le Pors *et al.* 2009] T. Le Pors, T. Devogele et C. Chauvin. *Multi Agent System Integrating Naturalistic Decision Roles : Application to Maritime Traffic*. Dans IADIS International Conference Intelligent Systems and Agents, 2009. (Cité en page 7.)

[Le Pors 2010] T. Le Pors. *Simulation cognitive de la prise de décision d'experts. Application au trafic maritime*. Thèse de doctorat, Université européenne de Bretagne, Université de Bretagne Sud., 2010. (Cité en page 7.)

[Lee *et al.* 2007] J.G. Lee, J. Han et K. Whang. *Trajectory Clustering : A Partition-and-group Framework*. Dans SIGMOD '07 : Proceedings of the 2007 ACM SIGMOD international conference on Management of data, pages 593–604, New York, NY, USA, 2007. ACM. (Cité en page 68.)

[Lee *et al.* 2008] J.G. Lee, J. Han et X. Li. *Trajectory Outlier Detection : A Partition-and-Detect Framework*. Dans Data Engineering, 2008. ICDE 2008.

IEEE 24th International Conference on Data Engineering, pages 140–149, 2008. (Cité en pages 7 et 116.)

[Leonardi *et al.* 2010] L. Leonardi, G. Marketos, E. Frentzos, N. Giatrakos, S. Orlando, N. Pelekis, A. Raffaeta, A. Roncato, C. Silvestri et Y. Theodoridis. *T-Warehouse : Visual OLAP Analysis on Trajectory Data.* Dans IEEE 26th International Conference on Data Engineering (ICDE), pages 1141–1144. IEEE, 2010. (Cité en page 4.)

[Levenshtein 1966] V.I. Levenshtein. *Binary Codes Capable of Correcting Deletions, Insertions, and Reversals.* Dans Soviet Physics Doklady, volume 10, page 707, 1966. (Cité en pages 44 et 46.)

[Lin & Su 2005] B. Lin et J. Su. *Shapes Based Trajectory Queries for Moving Objects.* Dans GIS '05 : Proceedings of the 13th annual ACM international workshop on Geographic information systems, pages 21–30, New York, NY, USA, 2005. ACM. (Cité en pages 58, 59 et 63.)

[Macedo *et al.* 2008] J. Macedo, C. Vangenot, W. Othman, N. Pelekis, E. Frentzos, B. Kuijpers, I. Ntoutsi, S. Spaccapietra et Y. Theodoridis. *Trajectory Data Models.* Mobility, Data Mining and Privacy, pages 123–150, 2008. (Cité en page 20.)

[MacQueen 1967] J. MacQueen. *Some Methods for Classification and Analysis of Multivariate Observations.* Dans Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, page 14, 1967. (Cité en page 83.)

[Mamdani & Assilian 1975] E.H. Mamdani et S. Assilian. *An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller.* International Journal of Man-Machine Studies, vol. 7, no. 1, pages 1 – 13, 1975. (Cité en pages 128 et 131.)

[Mascret *et al.* 2006] A. Mascret, T. Devogele, I. Berre et A. Hénaff. *Coastline Matching Process Based on the Discrete Fréchet Distance.* Progress in Spatial Data Handling, pages 383–400, 2006. (Cité en pages 57, 84, 85 et 115.)

[Mascret 2010] A. Mascret. *Développement d'une approche SIG pour l'intégration de données Terre/Mer.* Thèse de doctorat, École Nationale Supérieure d'Arts et Métiers, 2010. (Cité en pages 50 et 76.)

[Masseglia *et al.* 2004] F. Masseglia, M. Teisseire et P. Poncelet. *Extraction de motifs séquentiels : Problèmes et méthodes.* Ingénierie des systèmes d'information, vol. 9, no. 3-4, pages 183–210, 2004. (Cité en page 6.)

[Masseglia 2002] F. Masseglia. *Algorithmes et applications pour l'extraction de motifs sequentiels dans le domaine de la fouille de données : de l'incrémental au temps réel.* Thèse de doctorat, Université de Versailes St-Quentin en Yvelines, 2002. (Cité en page 6.)

[McGill *et al.* 1978] R. McGill, J.W. Tukey et W.A. Larsen. *Variations of Box Plots.* American Statistician, vol. 32, pages 12–16, 1978. (Cité en page 78.)

[McMaster 1986] R.B. McMaster. *A Statistical Analysis of Mathematical Measures for Linear Simplification.* Cartography and Geographic Information Science, vol. 13, no. 2, pages 103–116, 1986. (Cité en page 48.)

[Michoux 2007] L. Michoux. *SPATIONAV : Pierre angulaire de la protection et de la sauvegarde maritime.* REE. Revue de l'électricité et de l'électronique, no. 6-7, pages 81–84, 2007. (Cité en page 9.)

[Moisuc *et al.* 2005] B. Moisuc, P. Davoine, J. Gensel et H. Martin. *GenGHIS. Un outil de modelisation spatio-temporelle pour le suivi historique des risques naturels.* INGENIERIE DES SYSTEMS D INFORMATION, vol. 10, no. 4, page 35, 2005. (Cité en page 7.)

[Morel *et al.* 2008] M. Morel, A. Napoli, A. Littaye, J.P. George, F. Jangal, B. Alhadef, C. Scapel, J. Lebrevelec et D. Dejardin. *Surveillance et contrôle des activités des navires en mer ScanMaris.* Dans Workshop Interdisciplinaire sur la Sécurité Globale, 2008. (Cité en pages 2 et 144.)

[Mountain & Raper 2001] D. Mountain et J. Raper. *Modelling Human Spatio-temporal Behaviour : A Challenge for Location-based Services.* GeoComputation, vol. 178, pages 24–26, 2001. (Cité en page 2.)

[Myers *et al.* 1980] C. Myers, L. Rabiner et A. Rosenberg. *Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition.* IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 28, no. 6, pages 623–635, 1980. (Cité en page 44.)

[Nathan *et al.* 2008] R. Nathan, W.M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz et P.E. Smouse. *A Movement Ecology Paradigm for Unifying Organismal Movement Research.* Proceedings of the National Academy of Sciences, vol. 105, no. 49, pages 19052–19059, 2008. (Cité en pages 2, 7 et 22.)

[Ng & Han 1994] R.T. Ng et J. Han. *Efficient and Effective Clustering Methods for Spatial Data Mining.* Dans Proceedings of the International Conference on Very Large Data Bases, pages 144–144. Citeseer, 1994. (Cité en page 67.)

[Parent *et al.* 2006] C. Parent, S. Spaccapietra et E. Zimányi. Conceptual modeling for traditional and spatio-temporal applications : The mads approach. Springer-Verlag, 2006. (Cité en page 20.)

[Pelekis *et al.* 2007] N. Pelekis, I. Kopanakis, G. Marketos, I. Ntoutsi, G. Andrienko et Y. Theodoridis. *Similarity Search in Trajectory Databases.* Dans Proceedings of the 14th International Symposium on Temporal Representation and Reasoning, pages 129–140. IEEE Computer Society, 2007. (Cité en pages 24, 60, 64 et 116.)

[Pelekis *et al.* 2009] N. Pelekis, I. Kopanakis, E.E. Kotsifakos, E. Frentzos et Y. Theodoridis. *Clustering Trajectories of Moving Objects in an Uncertain World.* Dans ICDM'09. Ninth IEEE International Conference on Data Mining., pages 417–427. IEEE, 2009. (Cité en page 24.)

[Pettersson 2004] B. Pettersson. *Automatic Identification System.* Rapport technique, Swedish Maritime Administration, 2004. (Cité en pages 12 et 175.)

[Pfoser & Jensen 1999] D. Pfoser et C. Jensen. *Capturing the Uncertainty of Moving-Object Representations.* Dans Advances in Spatial Databases, pages 111–131. Springer, 1999. (Cité en page 21.)

[Piatetsky-Shapiro & Frawley 1991] G. Piatetsky-Shapiro et W.J. Frawley. Knowledge discovery in databases, volume 272. AAAI press, 1991. (Cité en pages 5 et 31.)

[Potter 2006] K. Potter. *Methods for Presenting Statistical Information : The Box Plot.* Visualization of Large and Unstructured Data Sets,(LNI), vol. 4, pages 97–106, 2006. (Cité en page 78.)

[Rabiner & Juang 1993] L. Rabiner et B. Juang. Fundamentals of speech recognition. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993. (Cité en page 44.)

[Randell *et al.* 1992] D.A. Randell, Z. Cui et A.G. Cohn. *A Spatial Logic Based on Regions and Connection.* KR, vol. 92, pages 165–176, 1992. (Cité en page 25.)

[Ratanamahatana & Keogh 2004] C.A. Ratanamahatana et E. Keogh. *Everything you Know About Dynamic Time Warping is Wrong.* Dans Third Workshop on Mining Temporal and Sequential Data. Citeseer, 2004. (Cité en page 44.)

[Ray *et al.* 2011] C. Ray, R. Goralski, C. Claramunt et C. Gold. *Real-Time 3D Monitoring of Marine Navigation.* Information Fusion and Geographic Information Systems, pages 161–175, 2011. (Cité en page 11.)

[Richard 2007] A. Richard. *L'optimisation des transits de navires marchands par l'utilisation des courants de marée.* Canadian Journal of Transportation, vol. 1, 2007. (Cité en page 144.)

[Sakoe & Chiba 1978] H. Sakoe et S. Chiba. *Dynamic Programming Algorithm Optimization for Spoken Word Recognition.* IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, pages 43–49, 1978. (Cité en pages 41, 44 et 84.)

[Sakurai *et al.* 2005] Y. Sakurai, M. Yoshikawa et C. Faloutsos. *FTW : Fast Similarity Search Under the Time Warping Distance.* Dans PODS '05 : Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on

Principles of database systems, pages 326–337, New York, NY, USA, 2005. ACM. (Cité en page 45.)

[Sinha & Mark 2005] G. Sinha et D.M. Mark. *Measuring Similarity Between Geospatial Lifelines in Studies of Environmental Health.* Journal of Geographical Systems, vol. 7, no. 1, pages 115–136, 2005. (Cité en pages 2, 59 et 63.)

[Sistla *et al.* 1997] A.P. Sistla, O. Wolfson, S. Chamberlain et S. Dao. *Modeling and Querying Moving Objects.* Dans The 1997 IEEE 13 th International Conference on Data Engineering, ICDE, pages 422–432, 1997. (Cité en page 20.)

[Sistla *et al.* 1998] A.P. Sistla, O. Wolfson, S. Chamberlain et S. Dao. *Querying the Uncertain Position of Moving Objects.* Temporal Databases : Research and Practice, pages 310–337, 1998. (Cité en page 21.)

[Spaccapietra *et al.* 2008] S. Spaccapietra, C. Parent, M.L. Damiani, J.A. de Macedo, F. Porto et C. Vangenot. *A Conceptual View on Trajectories.* Data & Knowledge Engineering, vol. 65, no. 1, pages 126–146, 2008. (Cité en pages 19 et 23.)

[Stefanovic *et al.* 2000] N. Stefanovic, J. Han et K. Koperski. *Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes.* IEEE Transactions on Knowledge and Data Engineering, vol. 12, no. 6, pages 938–958, 2000. (Cité en page 4.)

[Tao *et al.* 2004] Y. Tao, G. Kollios, J. Considine, F. Li et D. Papadias. *Spatio-Temporal Aggregation Using Sketches.* Dans 20th International Conference on Data Engineering, 2004. Proceedings., pages 214–225. IEEE, 2004. (Cité en page 4.)

[Tayeb *et al.* 1998] J. Tayeb, O. Ulusoy et O. Wolfson. *A Quadtree-Based Dynamic Attribute Indexing Method.* The Computer Journal, vol. 41, no. 3, page 185, 1998. (Cité en page 20.)

[Thériault *et al.* 2002] M. Thériault, C. Claramunt, A.M. Séguin et P. Villeneuve. *Temporal GIS and Statistical Modelling of Personal Lifelines.* Dans Advances in Spatial Data Handling : 10th International Symposium on Spatial Data Handling, page 433. Springer Verlag, 2002. (Cité en page 23.)

[Trajcevski *et al.* 2002a] G. Trajcevski, O.Wolfson, C. Hu, H. Lin, F. Zhang et N. Rishe. *Managing Uncertain Trajectories of Moving Objects with DOMINO.* Proceedings of the International Conference on Enterprise Information Systems, pages 218–225, 2002. (Cité en pages 20 et 21.)

[Trajcevski *et al.* 2002b] G. Trajcevski, O. Wolfson, F. Zhang et S. Chamberlain. *The Geometry of Uncertainty in Moving Objects Databases.* Advances in Database Technology - EDBT 2002, pages 145–161, 2002. (Cité en page 21.)

[Trajcevski *et al.* 2007] G. Trajcevski, H. Ding, P. Scheuermann, R. Tamassia et D. Vaccaro. *Dynamics-aware similarity of moving objects trajectories*. Dans Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, page 11. ACM, 2007. (Cité en pages 60, 61 et 64.)

[Trasarti *et al.* 2010] R. Trasarti, S. Rinzivillo, F. Pinelli, M. Nanni, A. Monreale, C. Renso, D. Pedreschi et F. Giannotti. *Exploring Real Mobility Data With M-Atlas*. Machine Learning and Knowledge Discovery in Databases, pages 624–627, 2010. (Cité en page 68.)

[Tukey 1977] J.W. Tukey. Exploratory data analysis. Addison-Wesley, 1977. (Cité en pages 78 et 79.)

[Turchin 1998] P. Turchin. Quantitative analysis of movement : Measuring and modeling population redistribution in animals and plants. Sinauer Associates, Sunderland, Massachusetts, USA, 1998. (Cité en pages 2, 20 et 22.)

[van Kreveld & Luo 2007] M. van Kreveld et J. Luo. *The Definition and Computation of Trajectory and Subtrajectory Similarity*. Dans Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems, page 44. ACM, 2007. (Cité en pages 60 et 64.)

[Vazirgiannis & Wolfson 2001] M. Vazirgiannis et O. Wolfson. *A Spatiotemporal Model and Language for Moving Objects on Road Networks*. Advances in Spatial and Temporal Databases, pages 20–35, 2001. (Cité en page 20.)

[Viana *et al.* 2011] Windson Viana, Alina Dia Miron, Bogdan Moisuc, Jérôme Gensel, Marlène Villanova-Oliver et Hervé Martin. *Towards the semantic and context-aware management of mobile multimedia*. Multimedia Tools Appl., vol. 53, pages 391–429, June 2011. (Cité en page 7.)

[Vlachos *et al.* 2002a] M. Vlachos, D. Gunopulos et G. Kollios. *Robust Similarity Measures for Mobile Object Trajectories*. Dans DEXA, 13th International Workshop on Database and Expert Systems Applications, pages 721–726, 2002. (Cité en pages 58 et 63.)

[Vlachos *et al.* 2002b] M. Vlachos, G. Kollios et D. Gunopulos. *Discovering Similar Multidimensional Trajectories*. Dans Proceedings of the International Conference on Data Engineering, pages 673–684. IEEE Computer Society Press ; 1998, 2002. (Cité en pages 46, 58 et 63.)

[Vlachos *et al.* 2003] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos et E. Keogh. *Indexing Multi-dimensional Time-series with Support for Multiple Distance Measures*. Dans Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 216–225. ACM, 2003. (Cité en page 46.)

[Vlachos *et al.* 2004] M. Vlachos, D. Gunopulos et G. Das. *Rotation Invariant Distance Measures for Trajectories.* Dans Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 707–712. ACM, 2004. (Cité en pages 59 et 63.)

[Wan & Zeitouni 2005] T. Wan et K. Zeitouni. *Modélisation d'objets mobiles dans un entrepôt de données.* EGC : Extraction et Gestion des Connaissances, pages 33–53, 2005. (Cité en page 4.)

[Wan *et al.* 2007] T. Wan, K. Zeitouni et X. Meng. *An OLAP System for Network-Constrained Moving Objects.* Dans SAC '07 : Proceedings of the 2007 ACM symposium on Applied computing, pages 13–18, New York, NY, USA, 2007. ACM. (Cité en page 7.)

[Wang & Mendel 1992] L.X. Wang et J.M. Mendel. *Generating Fuzzy Rules by Learning from Examples.* IEEE Transactions on Systems, Man and Cybernetics, vol. 22, no. 6, pages 1414–1427, 1992. (Cité en page 124.)

[Weber 1983] S. Weber. *A General Concept of Fuzzy Connectives, Negations and Implications Based on T-norms and T-conorms.* Fuzzy Sets and Systems, vol. 11, pages 103–113, January 1983. (Cité en page 128.)

[Willmott 1980] R. Willmott. *Two Fuzzier Implication Operators in the Theory of Fuzzy Power Sets.* Fuzzy sets and Systems, vol. 4, no. 1, pages 31–36, 1980. (Cité en page 128.)

[Wolfson *et al.* 1999] O. Wolfson, A.P. Sistla, S. Chamberlain et Y. Yesha. *Updating and Querying Databases that Track Mobile Units.* Distributed and parallel databases, vol. 7, no. 3, pages 257–387, 1999. (Cité en page 20.)

[Wolfson 2002] O. Wolfson. *Moving Objects Information Management : The Database Challenge.* Dans NGITS '02 : Proceedings of the 5th International Workshop on Next Generation Information Technologies and Systems, pages 75–89, London, UK, 2002. Springer-Verlag. (Cité en page 20.)

[Wu & Pelot 2007] Y. Wu et R. Pelot. Geomatics solutions for disaster management, chapitre Comparison of Simplifying Line Algorithms for Recreational Boating Trajectory Dedensification, pages 321–334. Jonathan Li and Sisi Zlatanova and Andrea G. Fabbri, 2007. (Cité en page 74.)

[Yager *et al.* 1995] R.R. Yager, H. Nguyen, M. Sugeno et R. Tong. Theoretical aspects of fuzzy control. John Wiley & Sons, 1995. (Cité en page 128.)

[Yanagisawa *et al.* 2003] Y. Yanagisawa, J. Akahani et T. Satoh. *Shape-Based Similarity Query for Trajectory of Mobile Objects.* Dans Proceedings of the 4th International Conference on Mobile Data Management, MDM '03, pages 63–77, London, UK, UK, 2003. Springer-Verlag. (Cité en pages 58 et 63.)

[Yi *et al.* 1998] B.K. Yi, HV Jagadish et C. Faloutsos. *Efficient Retrieval of Similar Time Sequences Under Time Warping.* Dans Proceedings of the International Conference on Data Engineering, pages 201–208, 1998. (Cité en page 44.)

[Zadeh 1965] L.A. Zadeh. *Fuzzy Sets.* Information and control, vol. 8, no. 3, pages 338–353, 1965. (Cité en pages 120 et 128.)

[Zadeh 1978] L.A. Zadeh. *Fuzzy Sets as a Basis for a Theory of Possibility.* Fuzzy Sets and Systems, vol. 1, no. 1, pages 3–28, 1978. (Cité en page 120.)

[Zhao & Bose 2002] J. Zhao et B.K. Bose. *Evaluation of Membership Functions for Fuzzy Logic Controlled Induction Motor Drive.* Dans IEEE 2002 28th Annual Conference of the Industrial Electronics Society IECON 02, volume 1, pages 229–234. IEEE, 2002. (Cité en page 123.)

[Šaltenis *et al.* 2000] S. Šaltenis, C. S. Jensen, S. T. Leutenegger et M. A. Lopez. *Indexing the Positions of Continuously Moving Objects.* Dans SIGMOD '00 : Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pages 331–342, New York, NY, USA, 2000. ACM. (Cité en page 20.)

Liste des symboles

$[p_d p_a]$ Droite reliant la position de départ et d'arrivée de la trajectoire, page 74

Δ_{A_i} Borne d'avance du couloir temporel de la i^{eme} position de la trajectoire médiane, page 99

Δ_{D_i} Borne du couloir spatial de droite de la i^{eme} position de la trajectoire médiane, page 94

Δ_{G_i} Borne du couloir spatial de gauche de la i^{eme} position de la trajectoire médiane, page 94

Δ_{R_i} Borne de retard du couloir temporel de la i^{eme} position de la trajectoire médiane, page 99

ΔS_D Borne de l'écart spatial de droite, page 115

ΔS_G Borne de l'écart spatial de gauche, page 115

ΔS_j Écart spatial toléré à la position médiane j, page 117

δSm Moyenne des écarts de distances spatiales, page 118

ΔT_A Borne de l'écart temporel d'avance, page 115

ΔT_j Écart temporel toléré à la position médiane j, page 119

ΔT_R Borne de l'écart temporel de retard, page 115

Δt_{max} Écart maximal temporel toléré entre 2 positions consécutives d'une trajectoire, page 72

δTm Moyenne des écarts de distances temporelles, page 120

Δv_{max} Vitesse maximale atteignable par l'objet mobile du type étudié, page 73

εd Seuil de distance toléré entre la position d'une trajectoire et sa projection spatio-temporelle, page 74

$\mathcal{C}$ Ensemble des couples de positions appariées entre deux trajectoires, page 83

$\mathcal{N}$ Nuage de positions homologues des trajectoires du GHT, page 80

$\mathfrak{p}$ point, page 22

$\mu^{\delta Sm}_{Faible}(x)$ Fonction d'appartenance de la mesure de similarité δSm à l'ensemble flou $Faible$, page 126

$\mu^{\delta Sm}_{Fort}(x)$ Fonction d'appartenance de la mesure de similarité δSm à l'ensemble flou $Fort$, page 126

$\mu_{Moyen}^{\delta Sm}(x)$ Fonction d'appartenance de la mesure de similarité δSm à l'ensemble flou $Moyen$, page 126

$\mu_{Faible}^{DSM}(x)$ Fonction d'appartenance de la mesure de similarité DSM à l'ensemble flou $Faible$, page 126

$\mu_{Faible}^{DSm}(x)$ Fonction d'appartenance de la mesure de similarité DSm à l'ensemble flou $Faible$, page 126

$\mu_{Fort}^{DSM}(x)$ Fonction d'appartenance de la mesure de similarité DSM à l'ensemble flou $Fort$, page 126

$\mu_{Fort}^{DSm}(x)$ Fonction d'appartenance de la mesure de similarité DSm à l'ensemble flou $Fort$, page 126

$\mu_{Moyen}^{DSM}(x)$ Fonction d'appartenance de la mesure de similarité DSM à l'ensemble flou $Moyen$, page 126

$\mu_{Moyen}^{DSm}(x)$ Fonction d'appartenance de la mesure de similarité DSm à l'ensemble flou $Moyen$, page 126

σ Ecart type, page 78

$\tilde{D}$ Durée médiane des trajectoires du GHT, page 76

$\tilde{L}$ Longueur médiane des trajectoires du GHT, page 83

$\tilde{p}$ Position médiane, page 117

$\tilde{p}_j$ j^{eme} position de la trajectoire médiane, page 117

$\tilde{T}$ Trajectoire médiane, page 114

$\tilde{t}_R$ Temps relatif médian d'un nuage de position homologues d'un GHT, page 85

AIS Automatic Identification System, page 12

$BDST$ Base de données spatio-temporelle, page 2

BI Business Intelligence, page 3

Boxplot Boîte à moustache, page 78

$CLARA$ Clustering LARge Applications, page 67

$CLARANS$ Clustering Large Applications based on RANdomized Search, page 67

cluster groupe d'objets, page 38

COG Centre de gravité des surfaces des ensembles flous, page 131

COG Course Over Ground, page 13

CPA Closest Point of Approach, page 14

$CROSS$ Centres Régionaux Opérationnels de Surveillance et de Sauvetage, page 8

$CSTDMA$ carrier-sense time-division multiple-access, page 175

d_H Distance de Hausdorff, page 49

d_{Fdsp} Distance de Fréchet discrète semi-partielle, page 56

d_{Fd} Distance de Fréchet discrète, page 51

d_{max} Distance maximale entre la position d'une trajectoire et sa projection spatio-temporelle, page 74

d_{moy} Distance moyenne, page 48

D_{ZiZj} Durée temporelle nécessaire au parcours de l'itinéraire entre les zone Z_i et Z_j, page 70

Datawarehouse Entrepôt de données, page 4

$DBSCAN$ Density Based Spatial Clustering of Applications with Noise, page 68

$DCPA$ Distance at Closest Point of Approach, page 14

DPC Droite perpendiculaire au cap d'une position, page 93

dS Distance spatiale (exprimée en mètres), page 117

dS_N Distance spatiale normalisée (exprimée en % du couloir), page 118

DSM Distance spatiale maximale, page 118

DSm Distance spatiale moyenne, page 118

DST Dispositif de séparation du trafic, page 25

dT Distance temporelle, page 119

dT_N Distance temporelle normalisée, page 120

DTM Distance temporelle maximale, page 120

DTm Distance temporelle moyenne, page 120

DTW Dynamic Time Warping, page 41

$ECDIS$ Electronic Chart Display and Information System, page 11

EIQ Ecart interquartile, page 79

ERP Edit distance with Real Penalty, page 47

ETL Extract-Transform-Load, page 4

$Faible$ Ensemble flou $Faible$, page 122

FATDMA fixed-access time-division multiple-access, page 175

Fort Ensemble flou *Fort*, page 122

Fuzzy logic Logique floue, page 120

G_Z Graphe de zones, page 25

GeoPKDD Geographic Privacy-aware Knowledge Discovery and Delivery, page 2

GHT Groupe homogène de trajectoires, page 67

GNSS Global Navigation Satellite System, page 1

GPS Global Positioning System, page 1

GSM Global System for Mobile Communications, page 1

$I_{(A,B)}$ Itinéraire entre la zone A et la zone B, page 24

IMO International Maritime Organization, page 8

indexing Classement des données, page 38

IRENAV Institut de Recherche de l'Ecole Navale, page 8

KDD Knowledge Discovery from Data, page 5

$L_p - norm$ Distance de Minkowski, page 39

LCSS Longuest Common SubSequence, page 45

LIP Locality In-between Polylines, page 60

MD Matrice de distances, page 42

MDTW Matrice de calcul du Dynamic Time Warping, page 42

MDX Multidimensional Expressions, page 4

MF Matrice de Fréchet, page 52

MOD Moving objects databases, page 20

MODAP Mobility, Data Mining and Privacy, page 2

Moyen Ensemble flou *Moyen*, page 122

OLAP Online Analytical Processing, page 4

OM Objet mobile, page 21

OPTICS Ordering Points To Identify the Clustering Structure, page 68

Outlier Donnée exceptionnelle ou aberrante, page 79

OWD One Way Distance, page 58

p position, page 21

p_a Position d'arrivée d'une trajectoire, page 23

p_a Position d'arrivée de la trajectoire, page 74

p_d Position de départ d'une trajectoire, page 23

p_d Position de départ de la trajectoire, page 74

p_i i^{eme} position de la trajectoire, page 74

p'_i Projection spatio-temporelle de la i^{eme} position de la trajectoire, page 74

p_{max} Position la plus éloignée de sa projection spatio-temporelle sur la droite $[p_d p_a]$, page 74

PAM Partionning Around Medoids, page 67

POI Point Of Interest, page 24

$Q1$ Premier quartile, page 78

$Q3$ Troisième quartile, page 78

$RADAR$ RAdio Detection And Ranging, page 11

Reporting Génération de rapports depuis un entrepôt de données, page 4

$RFID$ Radio Frequency IDentification, page 1

$RIPAM$ Règlement International pour Prévenir les Abordages en Mer, page 10

ROT Rate Of Turn, page 13

RT Route-type d'un GHT composée de sa trajectoire médiane enrichie des bornes du couloir spatio-temporel, page 100

$RTSD$ Rigid Transformation Similarity Distance, page 60

S Type d'objet mobile, page 26

SAR Search And Rescue, page 8

$SIAD$ Système d'Information d'Aide à la Décision, page 3

SIG Systèmes d'information géographique, page 37

SIM_S Indice de similarité spatiale, page 121

SIM_T Indice de similarité temporelle, page 121

SIM_{ST} Indice de similarité spatio-temporelle, page 121

SOG Speed Over Ground, page 13

$SOLAS$ Safety Of Life At Sea, page 14

$SOTDMA$ self-organised time-division multiple-access, page 175

$SPATIONAV$ Système de surveillance des approches maritimes, page 9

$STLIP$ Spatio Temporal Locality In-between Polylines, page 60

T Trajectoire, page 23

$T - Clustering$ Outil de regroupement de trajectoires similaires, page 68

t_A temps absolu au format UTC, page 22

t_R temps relatif exprimé en secondes, page 22

t_{p_i} Temps de la i^{eme} position de la trajectoire, page 74

T_{ref} Trajectoire de référence, page 83

t_{RN} Temps relatif normalisé d'une position exprimé en pourcentage de la durée médiane du GHT, page 76

$TCPA$ Time to Closest Point of Approach, page 14

$TCSD$ Temporal-Containment Similarity Distance, page 61

$TFaible$ Ensemble flou *Très Faible*, page 128

$TFort$ Ensemble flou *Très Fort*, page 128

$TRACLUS$ Trajectory Clustering, page 68

v_{calc} Vitesse calculée, page 73

v_{p_d} Vitesse de la position de départ de la trajectoire, page 74

wDF w-constrained Discrete Fréchet distance, page 61

$WGS84$ World Geodesic System 1984, page 22

$WGS84$ World Geodetic System 1984, page 40

$WLAN$ Wireless Local Area Network, page 1

x Valeur numérique d'entrée, page 122

Z_A Zone d'arrivée d'un itinéraire, page 25

Z_D Zone de départ d'un itinéraire, page 25

Z_N Zone de normalité, page 116

p_i i^{eme} position d'une trajectoire, page 117

μ Moyenne, page 78

Annexe A

Annexes

A.1 Tableau des différents fréquences de transmission AIS

Manœuvre réalisée	**Fréquence de transmission**
AIS Classe A	
A l'ancre	3 minutes
< à 14 nœuds	10 secondes
< à 14 nœuds et changeant de direction	3 1/3 secondes
14 à 23 nœuds	6 secondes
14 à 23 nœuds et changeant de direction	2 secondes
> à 23 nœuds	2 secondes
> à 23 nœuds et changeant de direction	2 secondes
AIS Classe B	
< à 2 nœuds	3 minutes
2 à 14 nœuds	30 Secondes
14 à 23 nœuds	15 secondes
> à 23 nœuds	5 secondes
Avion de recherche et de sauvetage	10 secondes
Stations de base	
Aide à la navigation	3 minutes
Station de base	10 seconds

Table A.1 – Fréquences de transmission des différentes informations AIS.

A.2 Tableau des différents types de trames AIS

La recommandation M.1371-1 de l'ITU (International Telecommunications Union) définit 27 trames AIS différentes contenant chacune des informations concernant le navire, son déplacement ou sa route ainsi que les modalités de transmission de ces trames [ITU 2001].

01	Position Report with SOTDMA
02	Position Report with SOTDMA
03	Position Report with CSTDMA
04	Base Station Report
05	Ship static and voyage related data
06	Addressed Binary Message
07	Binary Acknowledge
08	Binary Broadcast Message
09	Standard SAR Aircraft position report
10	UTC and Date inquiry
11	UTC and Date response
12	Addressed safety related message
13	Safety related Acknowledge
14	Safety related Broadcast Message
15	Interrogation
16	Assigned Mode Command
17	GNSS Binary Broadcast Message
18	Standard Class B CS Position Report
19	Extended Class B Equipment Position Report
20	Data Link Management
21	Aids-to-navigation Report
22	Channel Management
23	Group Assignment Command
24	Class B CS Static Data Report
25	Single slot binary message
26	Multiple slot binary message with Communications State
27	Position report for long range applications
28-63	Undefined ; Reserved for future use

TABLE A.2 – Liste des différents types de messages AIS.

A.3 Protocole de communication de l'AIS

Le système AIS utilise deux fréquences VHF marine (canal 87B : 161,975 MHz et canal 88B : 162,025 MHz). Le temps d'émission sur les fréquences dédiées est partagé en créneaux de transmissions de 26,6 ms. Ces créneaux sont partagés par les différents navires transmettant dans une même zone de couverture grâce aux protocoles de communication suivants [ITU 2001, IALA 2002, IALA 2004, IMO 2007, IALA 2010] :

- self-organised time-division multiple-access ($SOTDMA$),
- carrier-sense time-division multiple-access ($CSTDMA$),
- fixed-access time-division multiple-access ($FATDMA$).

Le fonctionnement du protocole de communication de l'AIS est illustré sur la figure A.1 adaptée de [Pettersson 2004]. Le système AIS du navire envoie sa position dans un créneau de transmission. En même temps, il réserve le créneau suivant qui sera utilisé pour transmettre sa future position. Lorsqu'un grand nombre de bateaux est présent dans une zone, l'AIS réduit automatiquement son rayon d'action afin de ne prendre en compte que les signaux les plus forts c'est à dire les signaux des bateaux les plus proches. La couverture du système AIS est liée aux capacités de la VHF et dépend essentiellement de la hauteur de l'antenne et de la puissance de l'émetteur. Cette couverture peut être fortement augmentée par l'utilisation de répéteurs.

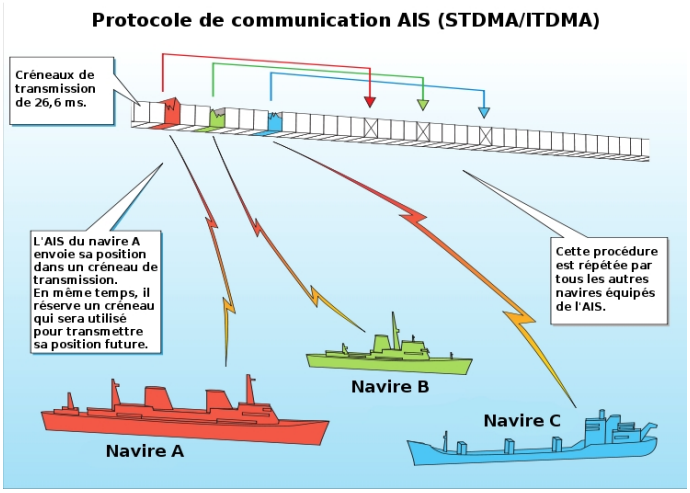


FIGURE A.1 – Illustration du fonctionnement du protocole de communication AIS.

A.4 Histogrammes de densité des distances des nuages de positions

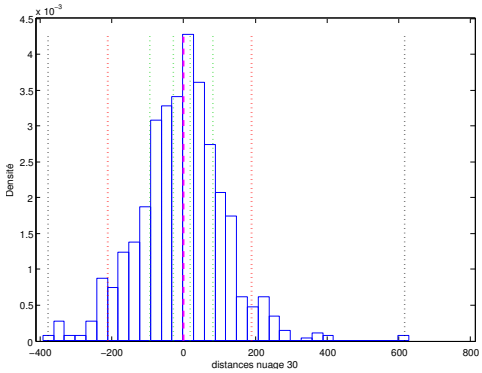


FIGURE A.2 – Histogrammes de densité des distances du nuage 30/120 de positions homologues du *GHT*

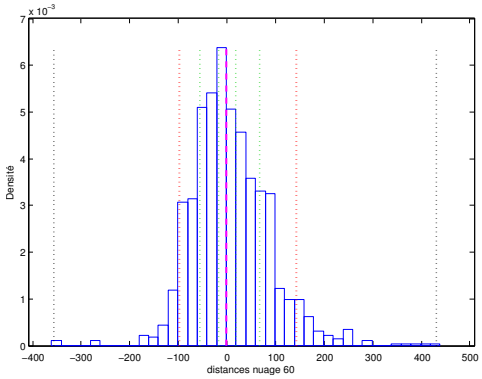


FIGURE A.3 – Histogrammes de densité des distances du nuage 60/120 de positions homologues du *GHT*

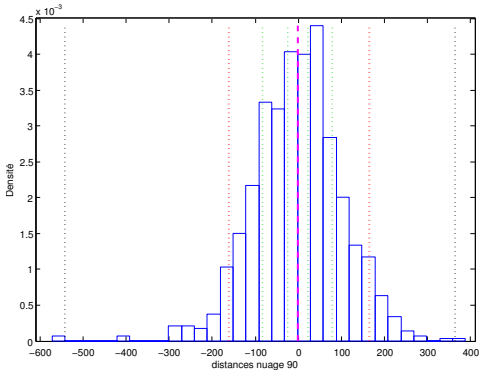


FIGURE A.4 – Histogrammes de densité des distances du nuage 90/120 de positions homologues du *GHT*

A.5 Histogrammes de densité des temps des nuages de positions

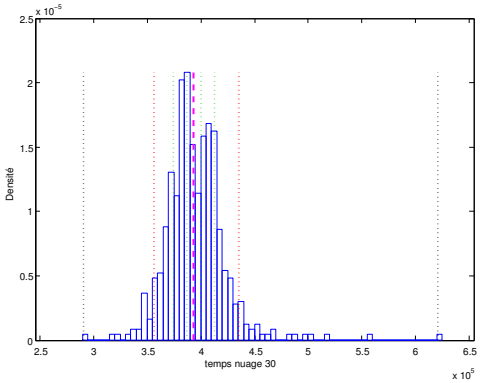


FIGURE A.5 – Histogrammes de densité des temps du nuage 30/120 de positions homologues du GHT

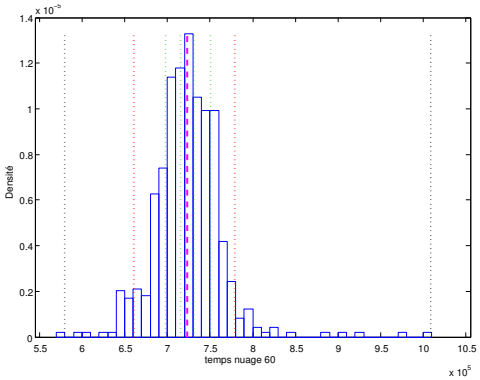


FIGURE A.6 – Histogrammes de densité des temps du nuage 60/120 de positions homologues du GHT

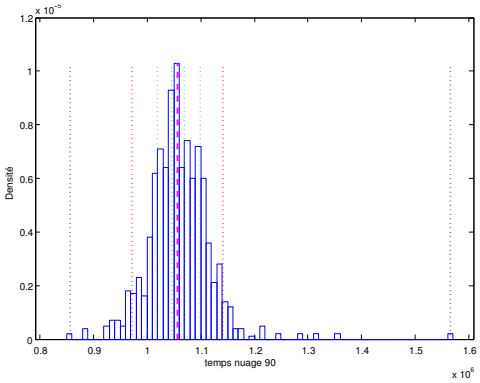


FIGURE A.7 – Histogrammes de densité des temps du nuage 90/120 de positions homologues du GHT

A.6 Histogrammes de densité des caps des nuages de positions

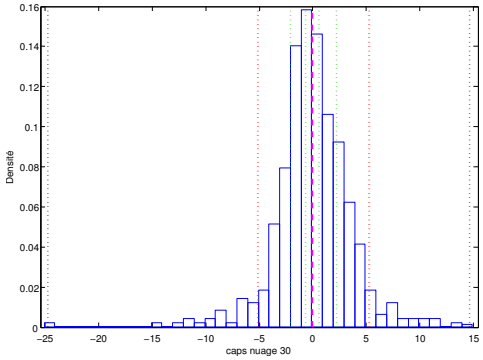


FIGURE A.8 – Histogrammes de densité des caps du nuage 30/120 de positions homologues du *GHT*

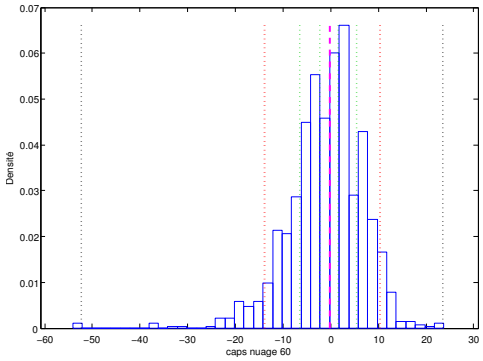


FIGURE A.9 – Histogrammes de densité des caps du nuage 60/120 de positions homologues du *GHT*

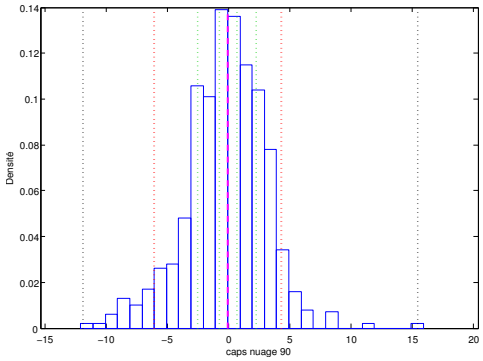


FIGURE A.10 – Histogrammes de densité des caps du nuage 90/120 de positions homologues du *GHT*

A.7 Histogrammes de densité des vitesses des nuages de positions

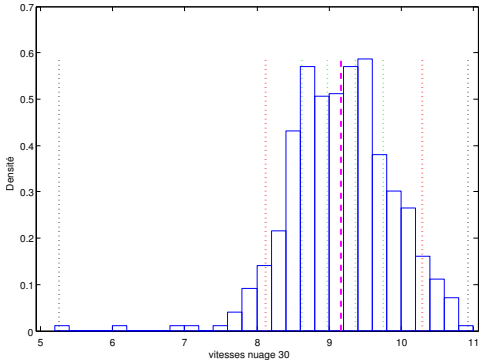


FIGURE A.11 – Histogrammes de densité des vitesses du nuage 30/120 de positions homologues du GHT

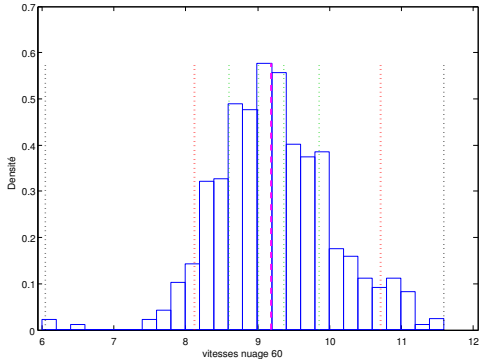


FIGURE A.12 – Histogrammes de densité des vitesses du nuage 60/120 de positions homologues du GHT

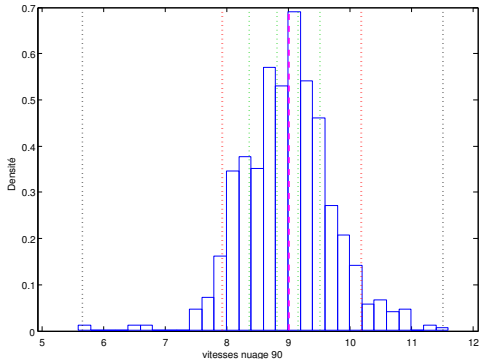


FIGURE A.13 – Histogrammes de densité des vitesses du nuage 90/120 de positions homologues du GHT

A.8 Route type de l'itinéraire "Brest-Lanvéoc"

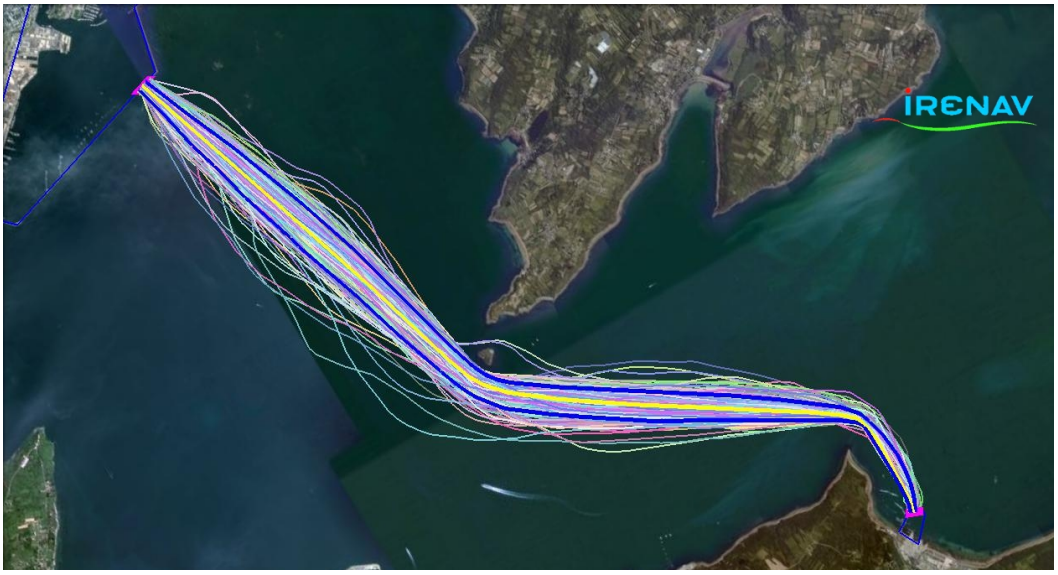


FIGURE A.14 – Route type de l'itinéraire "Brest-Lanvéoc" (navires à passagers)

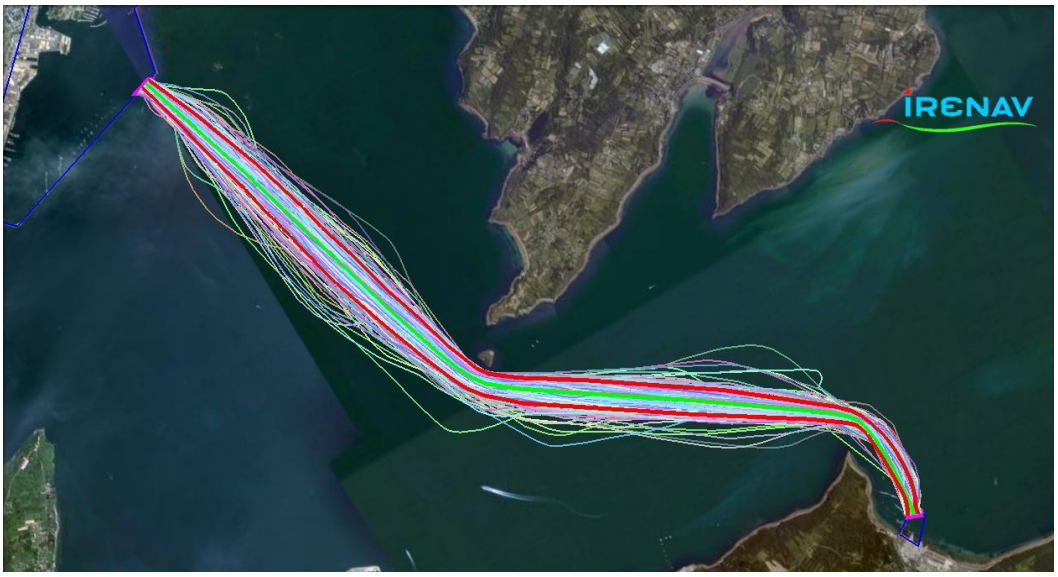


FIGURE A.15 – Route type de l'itinéraire "Lanvéoc-Brest" (navires à passagers)

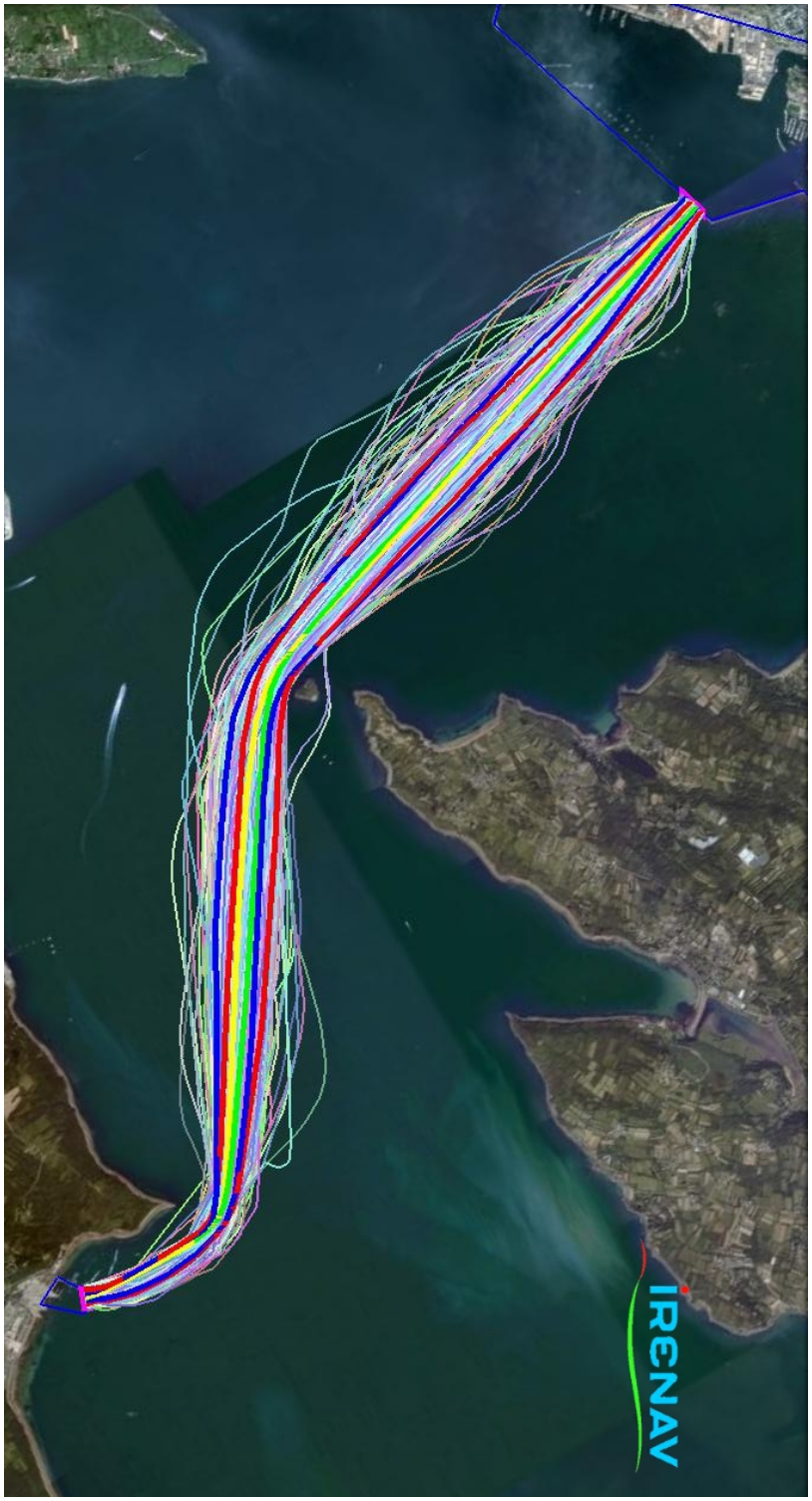


FIGURE A.16 – Routes types de l'itinéraire "Brest-Lanvéoc" et "Lanvéoc-Brest" (navires à passagers)

A.9 Route type de l'itinéraire "Calais-Douvres"

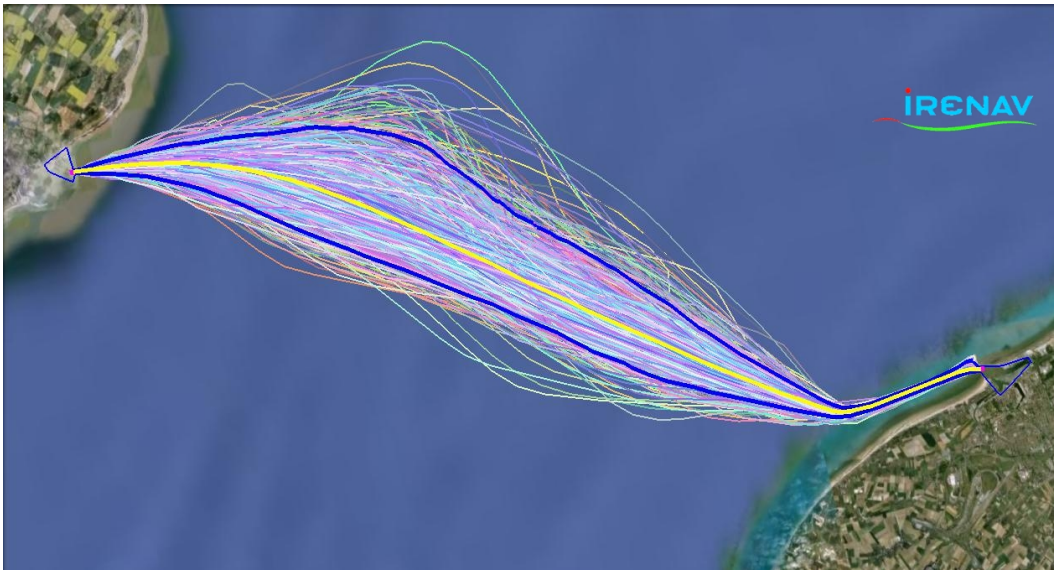


FIGURE A.17 – Route type de l'itinéraire "Calais-Douvres" (navires à passagers)

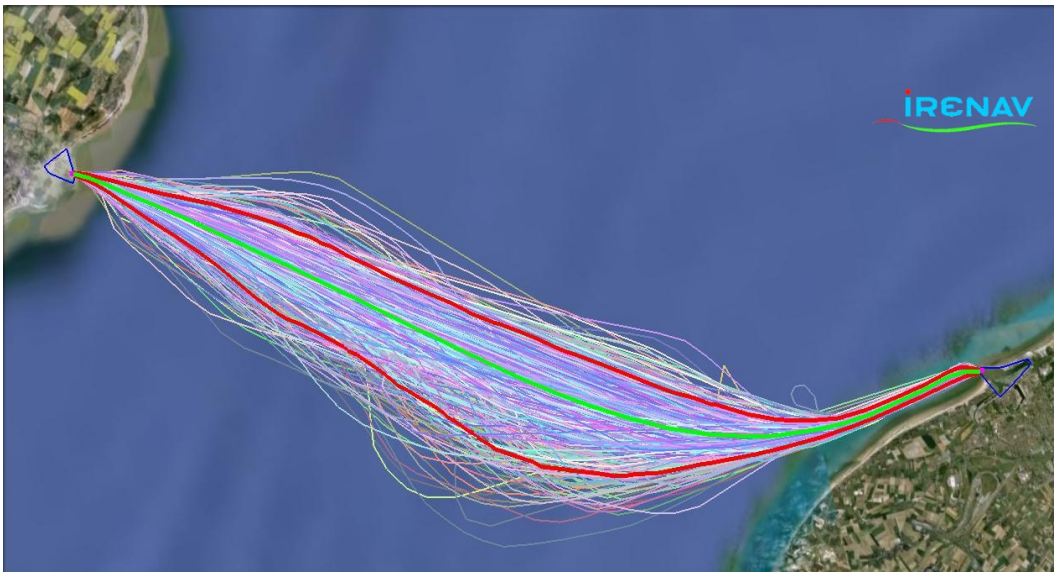


FIGURE A.18 – Route type de l'itinéraire "Douvres-Calais" (navires à passagers)

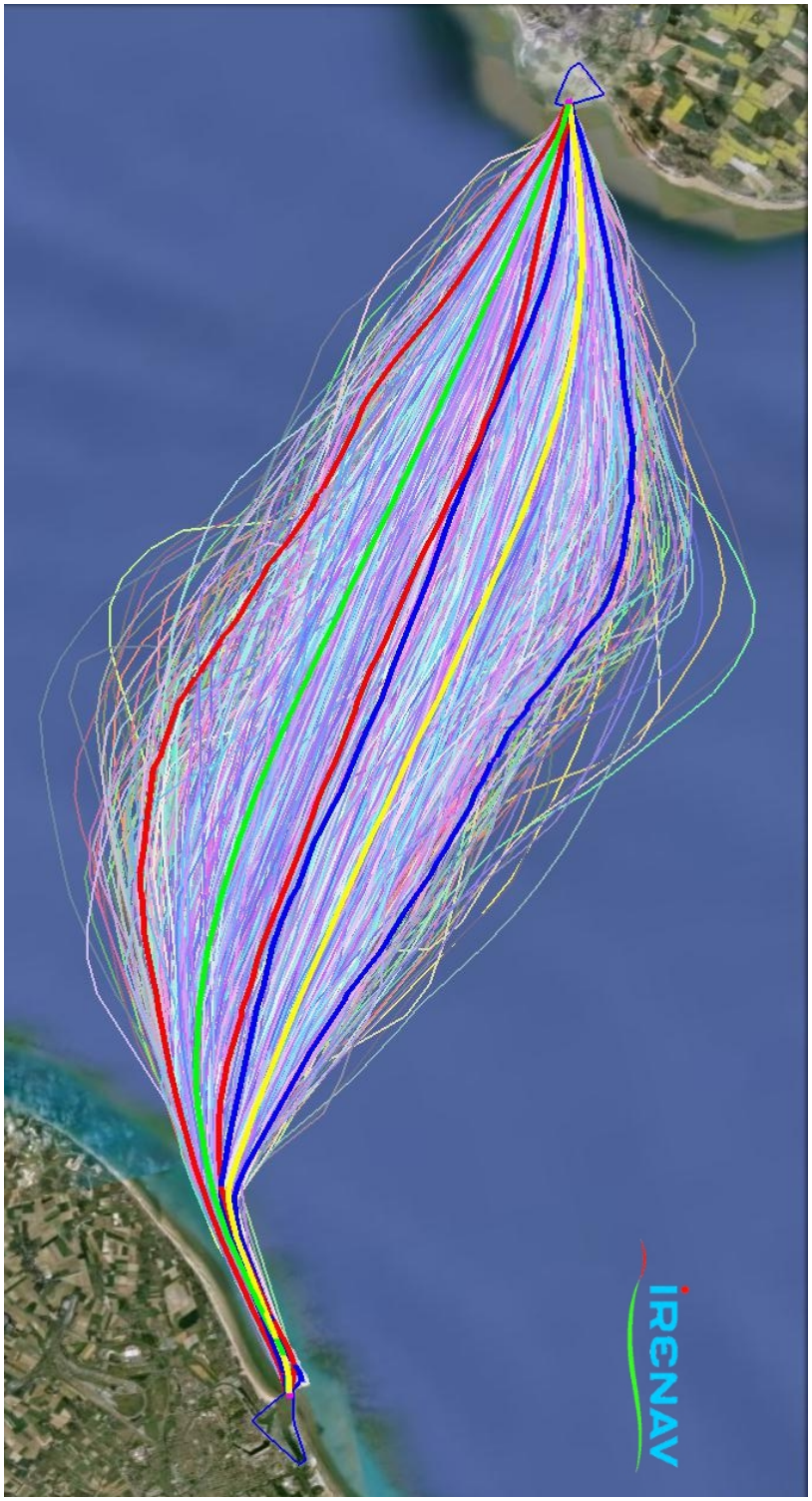


FIGURE A.19 – Routes types de l'itinéraire "Calais-Douvres" et "Douvres-Calais" (navires à passagers)

A.10 Route type de l'itinéraire "Pas de calais"

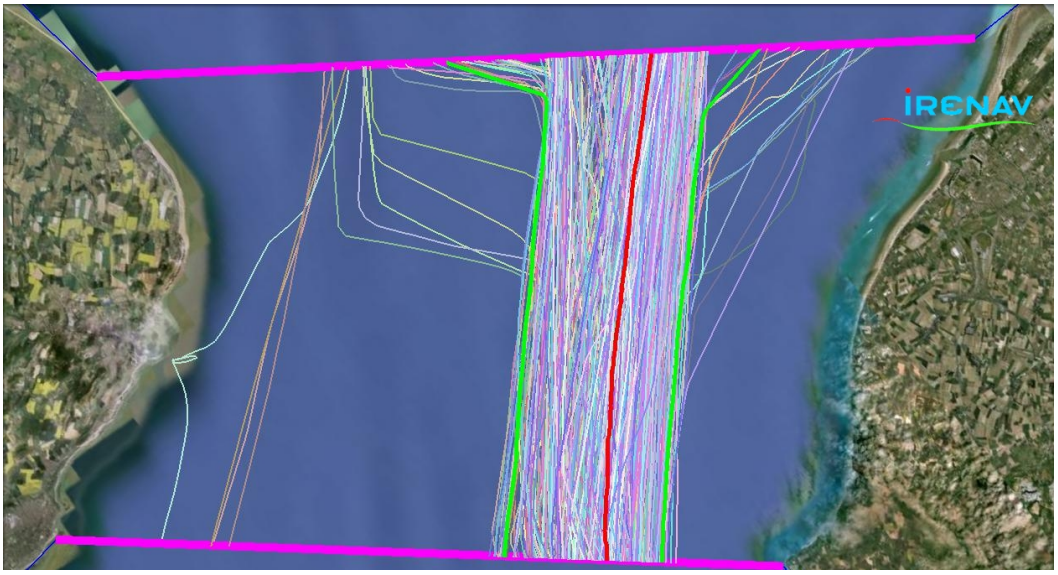


FIGURE A.20 – Route type de l'itinéraire "Manche-Mer du nord" (tous navires)

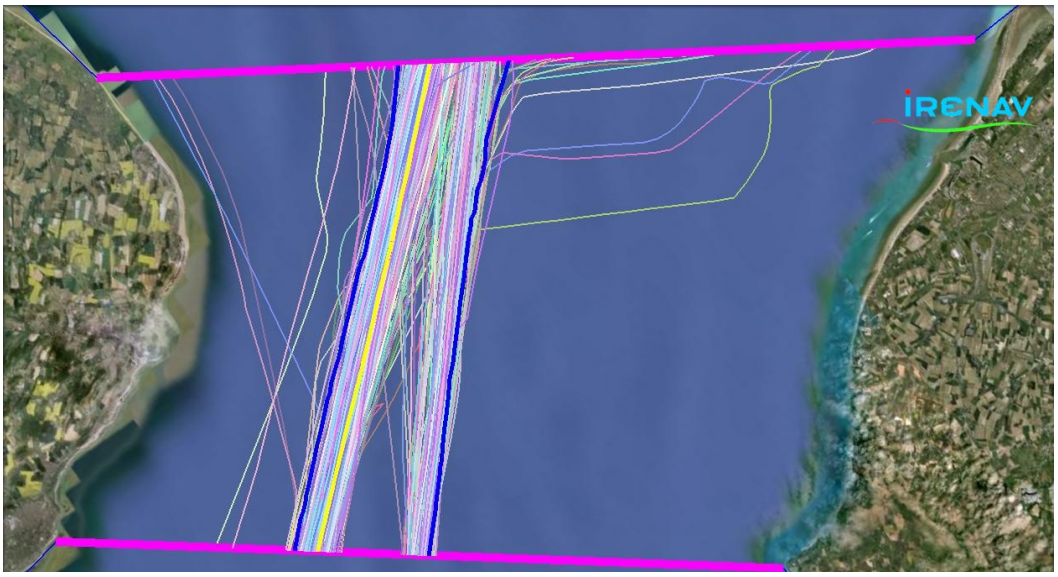


FIGURE A.21 – Route type de l'itinéraire "Mer du nord-Manche" (tous navires)



FIGURE A.22 – Routes types de l'itinéraire "Manche-Mer du nord" et "Mer du nord-Manche" (tous navires)

A.11 Route type 3D de l'itinéraire "Brest-Lanvéoc"

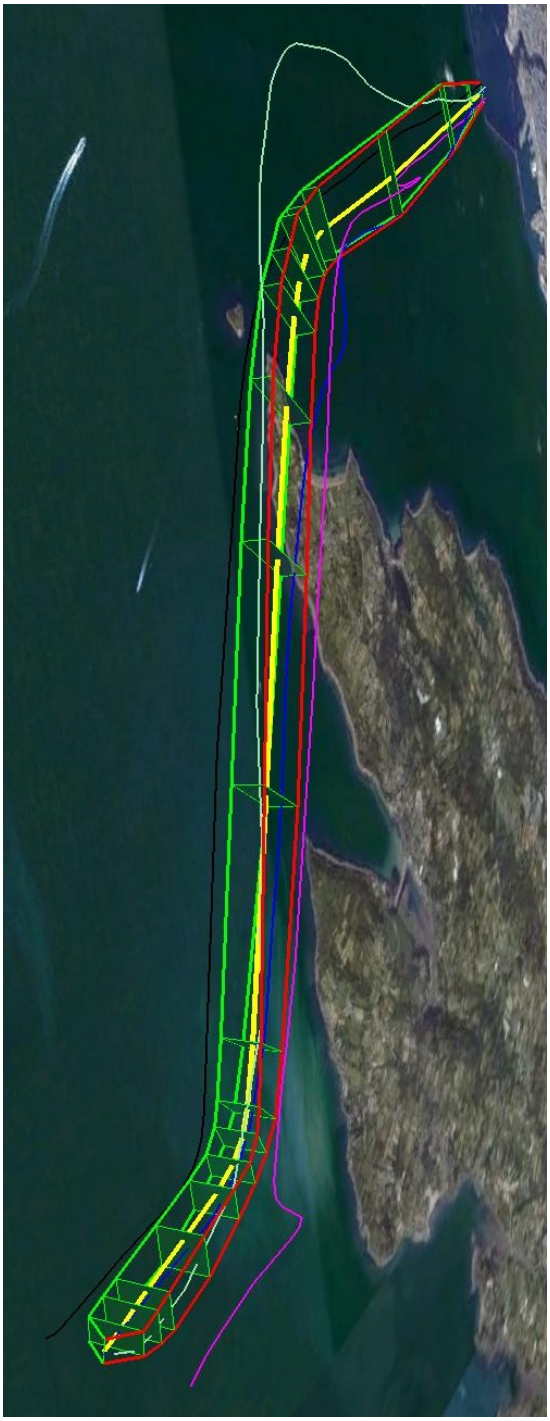


FIGURE A.23 – Représentation 3D de la route type de l'itinéraire "Brest-Lanvéoc" (navires à passagers)

A.12 Histogrammes des mesures de similarité

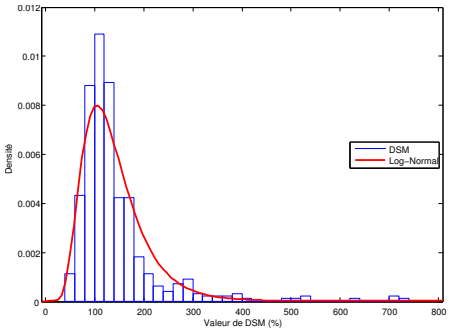


FIGURE A.24 – Histogramme des valeurs de la DSM

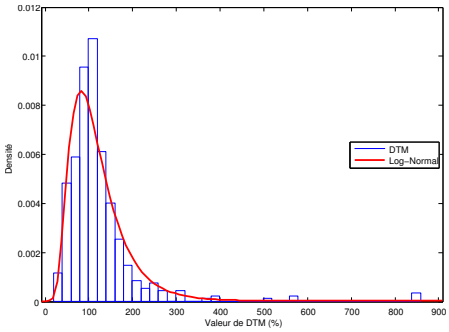


FIGURE A.25 – Histogramme des valeurs de la DTM

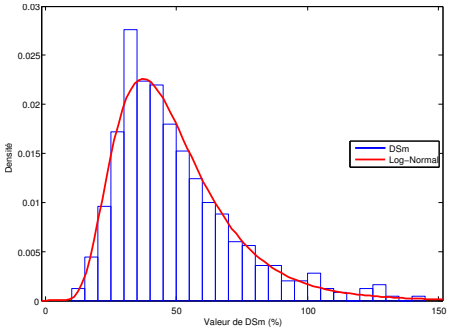


FIGURE A.26 – Histogramme des valeurs de la DSm

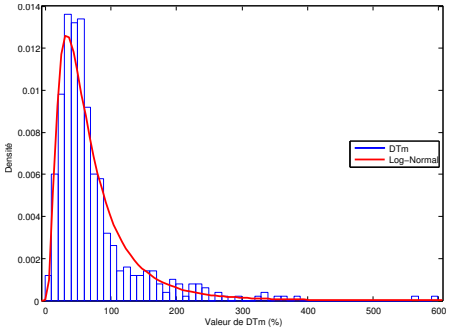


FIGURE A.27 – Histogramme des valeurs de la DTm

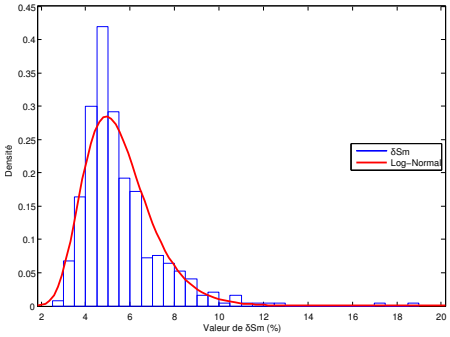


FIGURE A.28 – Histogramme des valeurs de la δSm

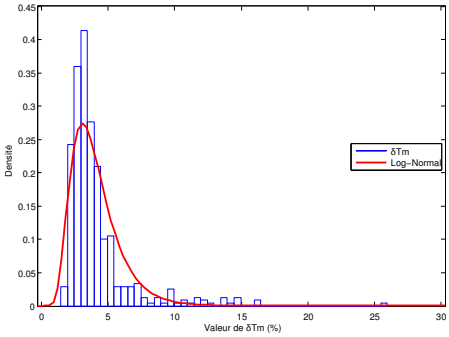


FIGURE A.29 – Histogramme des valeurs de la δTm

A.13 Probabilités cumulées des mesures de similarité

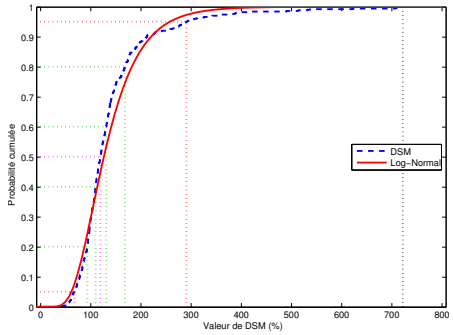


FIGURE A.30 – Probabilités cumulées des valeurs de la DSM

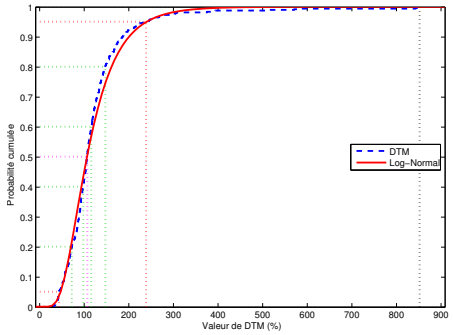


FIGURE A.31 – Probabilités cumulées des valeurs de la DTM

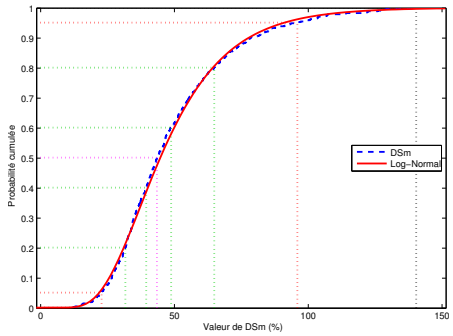


FIGURE A.32 – Probabilités cumulées des valeurs de la DSm

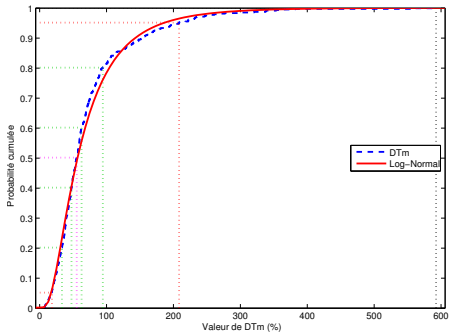


FIGURE A.33 – Probabilités cumulées des valeurs de la DTm

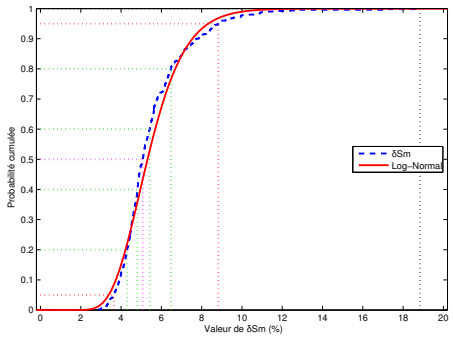


FIGURE A.34 – Probabilités cumulées des valeurs de la δSm

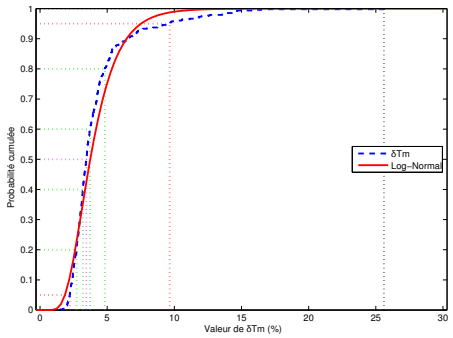


FIGURE A.35 – Probabilités cumulées des valeurs de la δTm

Motifs spatio-temporels de trajectoires d'objets mobiles, de l'extraction à la détection de comportements inhabituels. Application au trafic maritime.

Résumé :

Les systèmes de géolocalisation permettent la surveillance en temps réel des déplacements d'objets mobiles. Aujourd'hui, les données produites par ces capteurs sont reçues et stockées dans des bases de données spatio-temporelles. Un processus de fouille de données appliqué sur ces bases de données spatio-temporelles permet d'extraire le comportement des objets mobiles (patrons spatio-temporels) et d'analyser en temps réel les trajectoires d'objets mobiles suivant un même itinéraire. En utilisant ces modèles, des situations inhabituelles peuvent être détectés. Cette thèse définit à la fois des patrons spatio-temporels ainsi que des outils de comparaison et de qualification de trajectoires en utilisant un indice de similarité basée sur des mesures spatiales et temporelles et la logique floue. Ces outils peuvent être utilisés pour faciliter la surveillance du trafic maritime.

Mots clés :

Fouille de données, mesures de similarité, objets mobiles, patron spatio-temporel, logique floue, qualification de trajectoires.

Extraction of trajectories' spatio-temporal patterns in order to detect abnormal behaviour. Application to maritime traffic.

Abstract :

Mobile objects are now equipped with sensors allowing real time monitoring of their movements. Nowadays, the data produced by these sensors are received and stored in spatio-temporal databases in order to visually follow their movements. Data mining on this huge quantity of stored positions allows to extract the behaviour of these mobile objects (spatio-temporal patterns) and to analyze in real time trajectories of mobile objects following the same itinerary. Using these patterns, unusual situations can be detected. This research defines both spatio-temporel patterns and comparison tools to qualify trajectories using a similarity index based on spatial and temporal measures and fuzzy logic. These tools can be used to ease maritime traffic monitoring.

Keywords :

Data mining, similarity, mobile objects, spatio-temporel pattern, fuzzy logic, trajectory qualification.