



HAL
open science

Détection de données aberrantes appliquée à la localisation GPS

Salim Zair

► **To cite this version:**

Salim Zair. Détection de données aberrantes appliquée à la localisation GPS. Traitement du signal et de l'image [eess.SP]. Université Paris Saclay (COmUE), 2016. Français. NNT : 2016SACLS313 . tel-01516609

HAL Id: tel-01516609

<https://theses.hal.science/tel-01516609v1>

Submitted on 2 May 2017

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.

NNT : 2016SACLS313

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
L'UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE N°580
Sciences et Technologies de l'Information et de la Communication
Spécialité de doctorat : Traitement du Signal et des Images

Par

Mr Salim Zair

Détection de données aberrantes appliquée à la localisation GPS

Thèse présentée et soutenue à Gif-sur-Yvette, le 07 octobre 2016 :

Composition du Jury :

Mr François Peyret	Directeur de recherche émérite, IFSTTAR	Président du jury
Mr Jean-Yves Tourneret	Professeur, ENSEEIHT Toulouse	Rapporteur
Mr Philippe Bonnifait	Professeur, Univ. Tech. de Compiègne	Rapporteur
Mr Philippe Forster	Professeur, Univ. Paris Ouest Nanterre	Examineur
Mme Sylvie Le Hégarat-Masclé	Professeur, Univ. Paris Sud Orsay	Directrice de thèse
Mr Emmanuel Seignez	Maitre de conférences, Univ. Paris Sud Orsay	Co-encadrant

Remerciements

Je remercie l'ensemble de l'équipe MOSS SATIE et le laboratoire SATIE de m'avoir accueilli dans le laboratoire.

Je tiens à remercier ma directrice de thèse Sylvie Le Hégarat-Masclé de m'avoir accepté pour faire une thèse. Ses disponibilités, Ses précieux conseils ainsi que sa démarche scientifique m'ont énormément aidé à aborder les problématiques de la thèse et à effectuer mon travail dans les meilleures conditions, merci infiniment. Je remercie aussi mon Co-encadrant Emmanuel Seignez pour ses conseils, son expertise en localisation robuste et les différentes expérimentations réalisées qui ont contribué à la concrétisation de ce travail.

Je remercie Mr François Peyret d'avoir présidé le jury. Mes sincères remerciements vont à mes deux rapporteurs, le professeur Jean-Yves Tournier et le professeur Philippe Bonnifait pour le temps consacré à la lecture de la thèse ainsi que les différents commentaires qui ont permis d'améliorer la manuscrit. Je remercie également Le professeur Philippe Forster d'avoir accepté d'être membre du jury.

Je remercie particulièrement Roger Reynaud et Sergio Rodriguez pour leurs accueils chaleureux au sein de l'IUT d'Orsay et de leurs aides précieuses pour le bon déroulement de mon monitorat. Je remercie enfin tous les collègues et tous mes amis que je ne pourrai pas tous citer : Halim, Hamid, Lounis, Seif, Djawed, Zaki, Sofiane, Imad, Walid, Houssam, Karim,

Dédicaces

À ma sœur Katia et ma mère Hayet, mes deux anges gardiens, qu'elles reposent en paix.

À mon grand-père El-Hadj Abdelrahmane et mon oncle Ali, qui nous ont quittés durant ma dernière année de thèse, qu'ils reposent en paix.

À mon père Mohand Chérif, mon petit frère Amir Juba et à toute ma famille, qui ont toujours été là pour moi.

Une dédicace particulière à ma future épouse Hanane, celle qui partagera ma vie.

Table des matières

1	Introduction	1
2	Systèmes de navigation GNSS	5
2.1	Introduction	5
2.2	Les systèmes de navigation satellite	6
2.3	Les mesures GNSS	11
2.4	Localisation GNSS et navigation	20
2.5	Conclusion	25
3	Méthodes de détection et d'identification de données aberrantes	27
3.1	Introduction	27
3.2	Détection d' <i>outliers</i> par couplage avec d'autres informations	28
3.3	Détection d' <i>outliers</i> intra-mesures GNSS	30
3.4	Conclusion	48
4	Jeux de données et mesures de performance	49
4.1	Plate-forme expérimentale et jeux de données	49
4.2	Mesures de performance	53
4.3	Conclusion	56
5	Modélisation a-contrario pour la localisation GNSS	59
5.1	Introduction	59
5.2	Localisation à partir de données multi- <i>epochs</i>	60
5.3	Modélisation <i>a contrario</i>	62
5.4	Expérimentations et résultats	74

5.5	Conclusion	89
6	Filtrage particulière et détection d'<i>outliers</i> dans les mesures Doppler	91
6.1	Introduction	91
6.2	Formulation du problème	92
6.3	Approche proposée	96
6.4	Expérimentation et résultats	106
6.5	Conclusion	113
7	Approche crédibiliste appliquée à la localisation GNSS	115
7.1	Introduction	115
7.2	Fonctions de croyance	116
7.3	Approche proposée pour la détection d' <i>outliers</i>	122
7.4	Application à la localisation GPS	128
7.5	Résultats	131
7.6	Conclusion	137
8	Conclusion et perspectives	139
A	Annexe	143
A.1	Nombre de fausses alarmes : exemple des dates d'anniversaire	143
A.2	Formulation de la jacobienne de l'algorithme 6	145
A.3	Détails de l'équation 5.12 de la stratégie de sélection	146
A.4	Formulation de la jacobienne de l'algorithme 12	147
A.5	Éléments complémentaires de la théorie des fonctions de croyance	150
	Liste des symboles	155
	Bibliographie	163

Introduction

Les systèmes de positionnement par satellites, ou GNSS (*Global Navigation Satellite System*) connaissent un développement croissant et génèrent de nouvelles applications dans différents domaines des transports. Le GPS (*Global Positioning System*) de conception américaine est le système de navigation par satellites le plus utilisé car il offre la meilleure couverture mondiale. Les différents projets de modernisation de l'actuel système satellite russe GLONASS et le développement des systèmes de navigation européen GALILEO et chinois COMPASS soulèvent de nouveaux enjeux concernant la disponibilité et l'interopérabilité de ces constellations satellites. Avec le développement des technologies de l'information, des *smartphones* et des voitures connectées, les champs d'applications des GNSS se sont étendus et encouragent les constructeurs à proposer de plus en plus de services afin d'améliorer les performances des systèmes actuels.

La localisation est sans nul doute la fonction première d'un système GNSS. Ce dernier permet à un utilisateur de déterminer sa position sur le globe terrestre dans un repère absolu. Avec une flotte de systèmes de navigation pouvant atteindre 120 satellites, le service de localisation sera amélioré en offrant une meilleure précision ainsi qu'une fiabilité et une continuité de service accrues. Cependant, de par l'augmentation du nombre de satellites en considérant toutes les constellations, le nombre de pannes, ou de mesures erronées, augmentera nécessairement. Alors que les algorithmes présents dans les récepteurs GPS détectent au maximum une mesure erronée par *epoch*, avec une combinaison de différents systèmes de navigation, l'hypothèse d'une seule erreur à la fois n'est plus tenable et la détection et gestion des données erronées (défaillantes ou aberrantes selon les différentes terminologies) devient un nouveau défi technologique.

Ces mesures erronées, dites génériquement *outliers*, dépendent fortement de l'environnement où se trouve le récepteur. En zones urbaines, les acquisitions sont fortement dégradées par des phénomènes de multi-trajets ou de multiples réflexions des signaux avant d'arriver à l'antenne réceptrice, et en forêt de multiples obstacles bloquent les signaux satellites, ce qui diminue la redondance des mesures. Les signaux réfléchis, appelés 'multi-trajets', provoquent un allongement artificiel du temps de trajet de l'onde électromagnétique entre le satellite et le récepteur pouvant atteindre 100m en milieu urbain dense. Ces phénomènes provoqués par des grands bâtiments vitrés ou le feuillage des arbres sont difficiles à modéliser contrairement aux erreurs atmosphériques. Ainsi, les algorithmes de protection intégrés dans les récepteurs GNSS, initialement dédiés aux applications en aviation civile, ne sont pas adaptés à la navigation terrestre en milieu urbain avec la présence de multiples sources d'erreurs.

Problématique

Dans cette thèse, nous nous intéressons au problème de détection de mesures GPS rendues erronées notamment par de multiples réflexions en milieu contraint. La détection de ces erreurs, qu'on appellera 'données aberrantes' ou *outliers*, représente un enjeu majeur dans les applications de navigation autonome et de localisation robuste. Nous considérons notre problème sous un certain nombre de contraintes :

- Tout d'abord, les méthodes de localisation utilisées en milieu urbain, ou les multi-trajets sont les plus présents, doivent être robustes à un grand nombre d'*outliers*.
- Deuxièmement, nous nous limitons dans cette thèse à traiter les mesures exclusivement fournies par le récepteur GPS. En effet, les nombreuses méthodes proposées qui exploitent d'autres capteurs sont spécifiques à ces capteurs (*ad-hoc* au sens positif du terme). Notre problème sera alors d'améliorer la précision de la localisation GNSS seule, sans préjuger de son utilisation future et du fait qu'elle puisse éventuellement servir comme entrée d'un potentiel système de fusion multi-capteurs.
- Notre troisième contrainte se situe au niveau des propriétés ou caractéristiques de la méthode de détection. Bien qu'il existe une multitude de travaux en statistique classique pour traiter le problème, ces dernières impliquent des paramètres *a priori*, e.g. seuils de détection ou probabilité de fausse alarme. Notre objectif sera alors de définir une méthode avec le moins de paramètres à ajuster ou une méthode robuste à ses propres paramètres.
- Finalement, hormis les méthodes statistiques basées sur l'analyse des résidus supposant une représentation ponctuelle de la solution, la détection des *outliers* utilisant des approches ensemblistes a été beaucoup moins investiguée (hormis l'analyse par intervalles). Ainsi, nous nous intéresserons aussi aux approches qui prennent en compte l'imprécision des mesures dans la modélisation du problème, pour évaluer leur intérêt par rapport aux approches probabilistes plus classiques.

En résumé, le problème que nous choisissons de traiter a comme contraintes d'utiliser exclusivement des données GPS et d'être robuste, que ce soit à un nombre important d'*outliers* ou à la précision des paramètres du modèle. Nous faisons l'hypothèse que le critère pour séparer *inliers* et *outliers* est fondé sur la notion de cohérence des données entre elles, et ceci quelle que soit la représentation de la solution (ponctuelle ou ensembliste).

Le manuscrit est organisé comme suit :

- Le **Chapitre 1** présente les systèmes de navigation par satellites opérationnels et en cours de développement. La constitution des signaux GNSS ainsi que les différentes sources d'erreurs dégradant les mesures y sont brièvement décrites. Nous terminons ce chapitre en exposant le principe de la mesure de pseudo-distance et de celle Doppler ainsi que l'exploitation de ces mesures pour l'estimation de la position et/ou de la vitesse du récepteur en localisation.
- Le **Chapitre 2** présente différentes méthodes de détection de données aberrantes adaptées à la localisation GPS. Nous commençons par citer quelques méthodes dites 'avec information *a priori*', i.e. s'aidant de capteurs ou de carte 3D pour améliorer la localisation ou éliminer les satellites en non-visibilité directe. Les méthodes de détection de données aberrantes courantes en statistiques et ayant déjà des applications pour notre problème sont ensuite présentées ainsi

que leur applications en localisation GPS robuste. À la fin du chapitre, le choix d'une approche *a contrario*, qui sera au cœur des chapitres 4 et 5, est justifié.

- Le **Chapitre 3** décrit les différents jeux de données fournis par les capteurs GPS utilisés dans les différentes expériences. Nous donnons les caractéristiques des capteurs utilisés ainsi que des jeux de données réalisés en statique et en dynamique sous différentes conditions d'acquisition (environnement urbain, autoroute, forêt).
- Le **Chapitre 4** présente notre contribution pour la détection de données aberrantes dans les mesures pseudo-distances. Une approche *a contrario* est appliquée pour partitionner l'ensemble des mesures en deux sous-ensembles : les données cohérentes et les données aberrantes. Deux critères fondés sur l'espérance du nombre de fausses alarmes (NFA) seront utilisés pour mesurer la cohérence d'un ensemble de mesures sous l'hypothèse du modèle utilisé. Pour répondre aux spécificités de l'application GNSS, nous proposons une stratégie pour pallier le problème du manque de redondance après élimination des mesures aberrantes de même qu'une stratégie intelligente de sélection des satellites sur une fenêtre temporelle. La méthode est testée sur les différents jeux de données (statique et dynamique) et comparée aux méthodes classiques utilisées en localisation GPS.
- Le **Chapitre 5** aborde l'estimation de la localisation GPS sous l'hypothèse non-linéaire/non-gaussienne et en utilisant à la fois les mesures pseudo-distances et celles Doppler. La méthode développée dans le chapitre 4 est étendue au cas des observations Doppler et couplée avec un filtre particulière SIR et un filtre particulière de Rao-Blackwell. Une analyse des résultats en milieu urbain est effectuée ainsi qu'une évaluation qualitative des résultats en termes de détection de données aberrantes sur les pseudo-distances et les mesures Doppler.
- Le **Chapitre 6** s'intéresse à la représentation de l'information pour une meilleure gestion des incertitudes versus les imprécisions. Le formalisme des fonctions de croyance est utilisé pour l'estimation de la solution et la détection de données incohérentes (ou conflictuelles). Nous proposons une approche inspirée du RANSAC, adaptée dans le cadre de la théorie des fonctions de croyances pour partitionner les mesures en deux sous-ensembles : les mesures cohérentes et les mesures incohérentes, à l'aide d'une mesure spécifique à cette théorie, à savoir le conflit. Par la suite, les outils nécessaires à la localisation dynamique dans un cadre crédibiliste sont développés. Une comparaison de la théorie des fonctions de croyance avec des approches ensemblistes venant de l'analyse par intervalles est réalisée.

Systemes de navigation GNSS

Sommaire

2.1	Introduction	5
2.2	Les systèmes de navigation satellite	6
2.2.1	GPS	6
2.2.2	GLONASS	8
2.2.3	GALILEO	9
2.2.4	COMPASS	9
2.2.5	Les systèmes d'augmentation et DGPS	9
2.3	Les mesures GNSS	11
2.3.1	Message de navigation	11
2.3.2	Message d'observation	11
2.3.3	Sources d'erreur dégradant les mesures	15
2.3.4	Facteur de qualité pour la localisation	18
2.4	Localisation GNSS et navigation	20
2.4.1	Principe de trilatération	20
2.4.2	Résolution du problème de localisation	21
2.4.3	Détermination de la vitesse du récepteur	23
2.4.4	Filtre de Kalman étendu pour la navigation	24
2.5	Conclusion	25

2.1 Introduction

Dans ce premier chapitre, nous présentons les systèmes de navigation par satellites. Contrairement aux capteurs proprioceptifs qui nous fournissent une position relative par rapport à une position précédente ou connue, les systèmes GNSS sont des capteurs extéroceptifs fournissant une position absolue n'importe où sur la surface terrestre. Nous nous intéresserons dans une première partie aux différents systèmes de navigation GNSS avec une attention particulière portée au système américain GPS. Puis, nous expliquerons le principe des différentes mesures utilisées par la suite pour estimer la position et la vitesse d'un récepteur ainsi que les différents phénomènes pouvant dégrader ces mesures. Enfin nous terminerons ce chapitre en présentant quelques méthodes implémentées dans de nombreux récepteurs du commerce pour se localiser ou naviguer.

2.2 Les systèmes de navigation satellite

2.2.1 GPS

Le système de navigation américain GPS, aussi connu sous le nom de NAVSTAR GPS (*NAVigation Satellite Timing And Ranging Global Positioning System*), est un système de géolocalisation et de navigation introduit par le département de défense américain dans les années 1980. Le premier envoi de satellites GPS a eu lieu en 1978. Le système de navigation devient pleinement opérationnel le 17 juillet 1995 avec 24 satellites répartis sur 6 orbites quasi-circulaires avec une inclinaison de 55° par rapport au plan de l'équateur. Ce système permet de fournir instantanément une position et une vitesse dans n'importe quel endroit de la surface terrestre [Grewal et al., 2007]. Initialement conçu dans un but militaire, le système GPS est aujourd'hui largement utilisé dans des domaines aussi variés que : la navigation maritime, la surveillance d'ouvrage de génie civile, le guidage d'avion ou la synchronisation temporelle. À la première mise en service, le GPS avait une précision d'une centaine de mètres pour la navigation civile. En effet, l'algorithme de cryptage *Selective Availability* avait pour but, à l'époque, d'empêcher des attaques par des armes de précision type missile. En mai 2000, le gouvernement désactive le cryptage et permet un accès libre au signal civil L1. La précision en position est alors passée d'une centaine de mètres à une quinzaine de mètres, ce qui a permis d'étendre le champ d'application du GPS.

2.2.1.1 Signaux GPS

Les satellites GPS émettent sur deux fréquences de porteuses L1 et L2. Ces dernières sont construites à partir de la fréquence fondamentale $f_0 = 10.23\text{MHz}$, générée par l'horloge atomique des satellites, avec $f_{L1} = 154 \times f_0$ et $f_{L2} = 120 \times f_0$. La porteuse L1 est modulée avec deux codes spécifiques à chaque satellite appelées PRN (*Pseudo Random Noise*) qui permettent d'identifier le satellite émetteur et d'estimer la distance entre le satellite et le récepteur.

- **Le Code C/A (*Coarse/Acquisition*)** : qui est un *Gold code*, est généré à une fréquence de $f_0/10$ avec une période d'une milliseconde. Le code C/A n'est pas modulé avec la porteuse L2.
- **Le Code P (*Precise code*)** : qui est transmis à la fréquence fondamentale est réservé aux applications militaires. Ce code bénéficie d'un cryptage avec un autre code, appelé code Y, pour lutter contre les attaques de type *spoofing*.

2.2.1.2 Référentiel spatial et temporel

Le modèle géodésique utilisé par le système GPS est le WGS84 (*World Geodetic System 1984*). Il est défini par un ensemble de paramètres donnant, entre autres, la forme de l'ellipsoïde terrestre. Le repère WGS84 a son origine au centre de la terre et le plan OZ du référentiel se situe approximativement à 102m à l'Est du méridien de Greenwich. Notons cependant que les récepteurs GPS fournissent également, sous le protocole NMEA (*National Marine Electronics Association*), la position de l'utilisateur dans le repère géographique (latitude, longitude, altitude) mieux adapté à la navigation.

Le référentiel temporel utilisé par le système de navigation GPS est le temps GPS (*GPSTime*) qui a été synchronisé avec le référentiel UTC (*Coordinated Universal Time*) le 06 janvier 1980. Le temps GPS est divisé en semaines, avec une semaine qui débute le dimanche à 00h. Par exemple, un *GPSTime=482400s* et *893 weeks* se traduit par la $(893 + 1024)^{\text{ème}}$ semaine, par rapport à la date de la synchronisation avec le temps UTC, le 6^{ème} jour (vendredi) à la 50400^{ème} seconde, ce qui donne : le vendredi 07 octobre 2016 à 14h00. Le temps UTC est régulièrement modifié pour se synchroniser avec la rotation de la terre. La dernière modification a eu lieu le soir du 30 juin 2015 avec un ajustement d'une seconde (secondes intercalaires), ce qui donne au temps GPS une avance de 17 secondes par rapport au temps UTC.

2.2.1.3 Constitution du système de navigation GPS

Le système GPS est constitué de trois différents segments. Il en est de même pour les autres systèmes de navigation introduits plus loin dans ce chapitre. Nous détaillons ici le cas du système GPS :

- **Segment de contrôle** : Le segment de contrôle a pour rôle le pilotage des satellites. Les six stations au sol réparties tout au long de l'Équateur mettent à jour les données transmises par la constellation. La station principale, à Schriever dans le Colorado, se charge d'estimer les éphémérides et les paramètres d'horloge et de contrôler leur bon fonctionnement. Les autres stations (Ascension, Diego Garcia, Kwajalein, Cape Canaveral) relaient les nouveaux paramètres orbitaux.
- **Segment Spatial** : Le système GPS comptait initialement 24 satellites disposés sur 6 orbites. Aujourd'hui, il comprend 31 satellites en service et 3-4 satellites de réserve. Les satellites GPS suivent une orbite quasi-circulaire à une altitude approximative de 20200 km qu'ils parcourent en 11h58mn2s, soit un demi-jour sidéral. Ainsi, les satellites, vus du sol, reprennent la même position dans le ciel au bout d'un jour sidéral. Les satellites envoient en permanence des signaux électromagnétiques vers la Terre sur deux fréquences principales $L1 = 1575.42\text{MHz}$ et $L2 = 1227.60\text{MHz}$. Ces signaux contiennent des informations de navigation des satellites, de leur 'état de santé', des almanachs (position des différents satellites de la constellation) et différentes corrections (correction d'horloge et atmosphérique).

Les satellites sont conçus pour fournir un service fiable pour une durée de 7.5 à 11 ans, dépendant de la génération des blocs. Actuellement, la constellation GPS est un mélange d'anciens et de nouveaux satellites totalisant 31 satellites opérationnels. L'ancienne génération est composée de 12 satellites venant du bloc (de la génération de satellites) IIR, mis en orbite entre 1997 et 2004 pour une durée de vie de 7.5 ans. La nouvelle génération comprend 7 satellites du bloc IIR-M (le dernier lancement a eu lieu en 2009) et 12 satellites du bloc IIF (où le dernier satellite a été lancé le 05 février 2016). Le bloc IIR-M fait suite au programme IIR construit par Lockheed Martin. L'extension "M" vient du fait qu'un nouveau code militaire M est émis sur les deux fréquences pour améliorer la résistance au brouillage ainsi qu'un nouveau code civil (L2C) sur la fréquence L2. Construit par Boeing, le bloc IIF se caractérise par l'émission sur une nouvelle fréquence $L5 = 1176.45\text{ MHz}$, d'une horloge atomique de pointe et d'une durée de vie de 12 ans par satellite.

- **Segment utilisateur** : Ce segment représente les utilisateurs du système GPS qu'ils soient civils ou militaires. Un récepteur (antenne) permet d'accéder aux informations de position, ce qui

fait du système GPS un outil gratuit et illimité. Quatre satellites sont nécessaires au minimum pour une bonne géolocalisation. Il existe plusieurs types de récepteurs allant de ceux à faible coût, utilisant une bande de fréquence avec réception de trame réduite, aux récepteurs haut de gamme. Ces derniers réalisent une acquisition bi-fréquence, un traitement simultané de plusieurs satellites et de plusieurs constellations GNSS ou la correction d'erreurs permettant l'activation d'un mode différentiel ou DGPS.

Le système GPS fournit ainsi deux types de services aux utilisateurs pour la localisation et la mesure de temps :

- SPS (*Standard Positioning Service*) : ce service a été mis à disposition des utilisateurs civils et permet une précision en position de l'ordre d'une quinzaine de mètres.
- PPS (*Precise Positioning Service*) : ce service est réservé aux militaires américains et à certains utilisateurs autorisés par le département de la défense. Le signal permettant la localisation est crypté.

2.2.2 GLONASS

GLONASS (*Globalnaya navigatsionnaya sputnikovaya sistema*) est le système de navigation russe (développé par l'ex-URSS) équivalant au GPS. Le système a été développé initialement pour acquérir une indépendance technologique dans les systèmes de navigation par satellite. Cependant, la Russie a souffert de la crise entre 1989 et 1999 avec une réduction de 80% du budget pour l'entretien et la modernisation de la constellation si bien que cette dernière ne comptait que six satellites en 1999. Le programme de modernisation a été relancé dans les années 2000 pour atteindre un service entièrement opérationnel en octobre 2011. La constellation est constituée de 24 satellites gravitant à une altitude de 19100 km sur trois plans orbitaux inclinés de 68.4° , permettant une meilleure couverture dans les hautes latitudes (les deux pôles terrestres).

Contrairement au GPS, qui transmet des codes différents pour chaque satellite sur une seule fréquence (transmission CDMA *Code Division Multiple Access*), GLONASS utilise un seul code pour tous les satellites en les distinguant par leur fréquence d'émission (transmission FDMA *Frequency-Division Multiple Access*). La fréquence centrale de la bande L1 est de $1602.0 + n \times 0.5625$ MHz, où n est un entier variant entre -7 et 6 . La bande L2 partage la même formule avec un pas de 437.5 KHz et une fréquence centrale de 1246 Mhz. Une combinaison entre les deux constellations, GPS et GLONASS, pour la navigation peut être réalisée afin d'augmenter le nombre de satellites visibles par un récepteur en un point donné. Néanmoins, cela nécessite l'utilisation de deux architectures matérielles parallèles et la gestion de la désynchronisation temporelle entre les deux systèmes. Une modélisation ainsi qu'une évaluation de la combinaison de ces deux systèmes pour le positionnement est détaillée dans [Cai and Gao, 2012]. Les recherches et développements sur le GLONASS, avec la génération de satellites GLONASS-KM, se dirigent actuellement vers une interopérabilité totale entre les constellations GPS et GALILEO à l'horizon 2025.

2.2.3 GALILEO

GALILEO est le nouveau système de positionnement global européen. Ce système assure à l'Europe une certaine indépendance technologique envers le GPS américain et le GLONASS russe. Mais ce n'est pas la seule motivation car l'agence spatiale européenne veut que GALILEO soit aussi un système complémentaire des dispositifs déjà opérationnels. À sa mise en service complète, qui est prévue pour l'an 2020 avec une flotte de 30 satellites, GALILEO sera interopérable avec les constellations GPS et GLONASS. Ceci va doubler la disponibilité des satellites et permettre de se localiser sur Terre de façon plus exacte, et plus particulièrement en environnement urbain où les bâtiments occultent les satellites de basse élévation. De plus, GALILEO couvrira d'une meilleure manière que le GPS les zones de hautes latitudes telles que le Nord de l'Europe. Le GPS est moins incliné par rapport au plan équatorial que ne le sera GALILEO.

Chaque satellite de la constellation circulera sur trois plans orbitaux distincts avec une inclinaison de 56° par rapport au plan équatorial à une altitude moyenne de 23222 km. Les retombées économiques de la mise en service totale du système sont estimées à 90 milliard d'euro. Le système de navigation européen sera géré par deux centres de contrôle principaux (segment de contrôle) déjà implantés en Europe, et sera complété par une vingtaine de stations au sol réparties dans le monde. Après de multiples retards dus à des problèmes financiers et à l'échec de la mise en orbite de deux satellites GALILEO (les satellites 5 et 6) le 22 août 2014, la mise en service initiale (avec 16 satellites opérationnels) est prévue en octobre 2016 précédée par le lancement des satellites 13 et 14 le 24 mai 2016.

2.2.4 COMPASS

COMPASS, aussi appelé Beidou-2, est le système de navigation chinois constitué actuellement de 17 satellites en service. Précédé par le système régional à base de satellites géostationnaires Beidou-1 qui a pris fin en 2012, COMPASS a pour but d'assurer une couverture terrestre globale à l'horizon 2020. Ainsi, la combinaison des actuels systèmes GPS GLONASS et des futurs GALILEO et COMPASS portera le nombre de satellites de navigation à plus de 100, augmentant ainsi le nombre de satellites visibles y compris en milieu contraint comme en ville.

2.2.5 Les systèmes d'augmentation et DGPS

2.2.5.1 Le DGPS

Le terme DGPS (*Differential GPS*) s'applique à tous les systèmes qui envoient des corrections liées à une mesure de distance entre un satellite GNSS et une station émettrice de position connue. Il est ainsi possible de corriger une grande partie des erreurs liées à la mesure de la distance satellite/récepteur comme les effets ionosphériques et troposphériques ou les erreurs d'horloge. La transmission des corrections aux récepteurs peut se faire par des stations terrestres via une liaison radio comme le GBAS (*Ground-Based Augmentation System*), ou en utilisant un réseau de stations au sol qui transmettent les

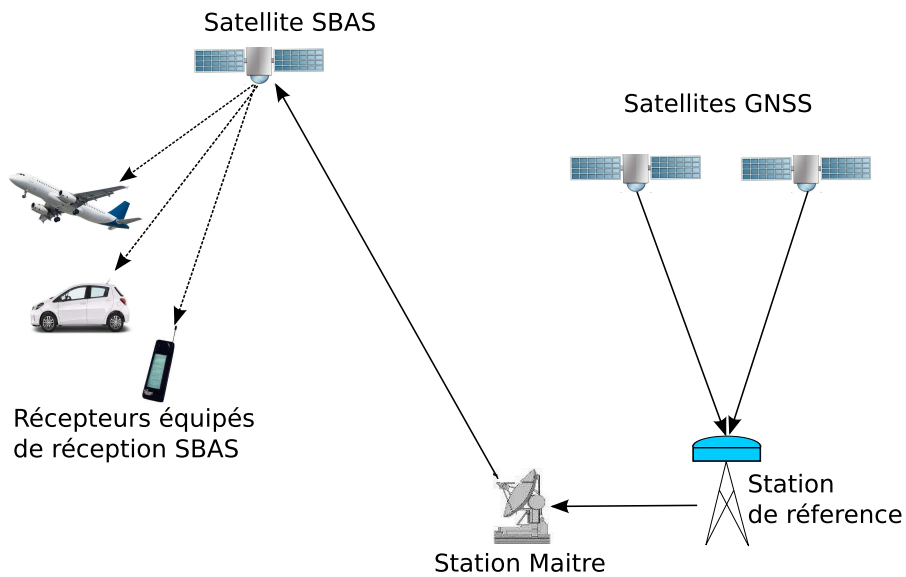


FIGURE 2.1 – Illustration du positionnement GNSS avec des corrections satellite ou base terrestre.

informations via des satellites géostationnaires qu'on appelle SBAS (*satellite-Based Augmentation System*), comme représenté sur la figure 2.1. La portée des GBAS est assez restreinte (environ 20 km) par rapport aux SBAS. La précision de ces systèmes varie suivant la qualité de récepteur, la position géographique et le type de données utilisées. Elle est de l'ordre du mètre en utilisant un récepteur mono fréquence GPS L_1 , et elle peut atteindre une précision millimétrique avec un récepteur bi-fréquence, multi-constellations utilisant les mesures de code et de phase. Cette qualité de localisation est appelée le mode RTK (*Real Time Kinematic*).

2.2.5.2 Les systèmes d'augmentations satellites SBAS

Le SBAS est un service régional constitué d'une flotte de satellites géostationnaires qui a pour but de fournir des corrections au récepteur GNSS afin d'améliorer sa précision notamment en localisation. Il est généralement utilisé en navigation maritime et en aviation civile pour la phase de décollage et d'approche des avions. Il permet d'envoyer des corrections différentielles permettant la correction de mesures de temps de trajet satellite/récepteur, mais aussi des paramètres de fiabilité des satellites GNSS et des données d'intégrité. Parmi les systèmes d'augmentation satellite, on peut citer :

- WAAS : (*Wide Area Augmentation System*) qui est le système régional américain qui couvre le continent nord américain.
- GAGAN : (*GPS Aided Geo Augmented Navigation*) qui est le système d'augmentation indien qui vient compléter le système de navigation par satellite régional indien IRNSS.
- EGNOS : (*European Geostationary Navigation Overlay Service*) qui est le système européen permettant de fournir un service ouvert (*Open service*) et un service *Safety-of-life* utilisé en aviation. La précision obtenue atteint deux mètres dans 99% des cas. EGNOS est principalement conçu pour les applications avioniques, mais il peut être utilisé pour les applications terrestres. Cependant, il est difficile à utiliser en milieu urbain en raison de la faible élévation des satellites situés à 30° , inadaptée pour les canyons urbains. De plus, il ne permet pas actuellement d'être

utilisé comme une source d'information (mesure de distance) supplémentaire pour la localisation, mais pourra l'être d'ici 2020 avec une transmission sur la fréquence L1 du système GPS.

2.3 Les mesures GNSS

L'utilisation et l'acquisition des données GPS se fait en temps-réel pour une localisation instantanée. On distingue deux types de données : le message transmis par la constellation, qui est indépendant du segment utilisateur, et la mesure faite par le récepteur à partir des signaux satellites.

Lors du traitement de données en hors-ligne, il est nécessaire de les stocker dans un format normalisé, indépendamment du constructeur. Le format RINEX (*Receiver Independent EXchange Format*) est un format d'échange de données GNSS développé par l'institut d'astronomie de l'Université de Berne dans le but de mettre dans un même format les données collectées en format propriétaire par des récepteurs de marques différentes lors de la campagne GPS EUREF 8. En ce qui concerne les systèmes de localisation, deux types de fichiers sont généralement utilisés : le message de navigation et le message d'observation.

2.3.1 Message de navigation

Le message de navigation est transmis à la vitesse de 50 bits par seconde et se décompose en cinq sous-messages (*subframes*) de 300 bits. Les deux premiers mots de chaque portion sont identiques : ils contiennent un préambule fixe, le *time-of-week*, qui permet de passer du code C/A au code P, et un bit d'alerte indiquant un fonctionnement erroné. Le message de navigation contient des informations sur les éphémérides des satellites (qui indiquent les paramètres képlériens de l'orbite), des paramètres d'horloge ou des indicateurs sur la santé du satellite. Les éphémérides sont mises à jour toutes les 2 heures pour le GPS et toutes les 30 minutes pour le GLONASS [Samama, 2008].

Le reste des sous-messages contient des almanachs, décrivant de façon grossière les orbites de tous les satellites de la constellation. Les messages sont transmis alternativement de telle sorte que la réception complète du message de navigation nécessite 25 cycles, soit 12 minutes et 30 secondes.

2.3.2 Message d'observation

Le message d'observation contient des mesures de temps de transmission du signal, de décalage de fréquences et des mesures de phases. Suivant le type de récepteur, ces informations sont soit mesurées sur une porteuse (généralement L1) ou sur les deux porteuses, L1 et L2, et/ou sur plusieurs constellations pour des récepteurs plus coûteux. Nous introduisons dans ce qui suit les mesures utilisées dans nos travaux.

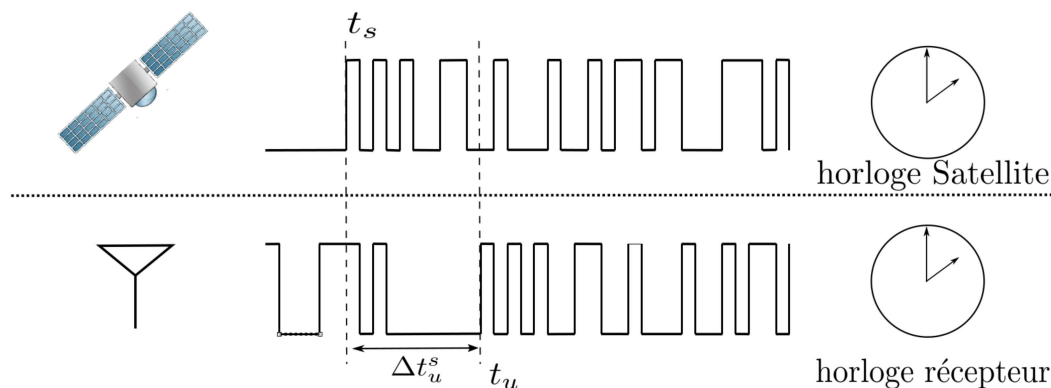


FIGURE 2.2 – Mesure du temps de trajet satellite/récepteur par comparaison de PRN.

2.3.2.1 Mesure de pseudo-distance

La mesure de distance entre un satellite et un récepteur s'effectue en comparant le signal (code) envoyé par un satellite avec sa réplique au sol générée par le récepteur. Ce temps est mesuré de deux manières différentes.

La première maximise la corrélation entre les codes. Le signal émis est un code pseudo-aléatoire BPSK envoyé à une fréquence de 1.023 Mhz par chaque satellite. Le même code, différent d'un satellite à un autre, est généré en même temps par le récepteur GPS. Pour mesurer le temps d'arrivée du signal, ou TOA (*Time of arrival*), un procédé de corrélation est réalisé entre le signal généré localement et le signal reçu du satellite. Plusieurs pics sont détectés, dont un seul correspond au temps de transmission du signal (généralement compris entre 0.068s et 0.088s). La distance satellite/récepteur est obtenue en multipliant la différence temporelle par la vitesse de la lumière c . Ce principe s'applique pour tous les systèmes de navigation existants. Cependant, la mesure obtenue n'est pas une distance mais une pseudo-distance. En effet, les satellites et le récepteur n'ont pas le même référentiel de temps ce qui ne permet pas d'avoir une mesure de temps précise.

La figure 2.2 nous montre le principe de la mesure de pseudo-distance. Le satellite émet le code à l'instant t_s dans le référentiel de temps du satellite et l'antenne du récepteur reçoit le code à l'instant t_u dans le référentiel de temps du récepteur. Le temps de transmission s'écrit :

$$\Delta t_u^s = t_u - t_s. \quad (2.1)$$

Dans un même référentiel appelé 'temps GPS' ou *GPStime*, cette différence temporelle s'écrit :

$$\begin{cases} t_e &= t_s + dt_s, \\ t_r &= t_u + dt_u, \end{cases} \quad (2.2)$$

où t_e représente l'instant d'émission du signal dans le référentiel *GPStime*, dt_s le décalage entre le référentiel temporel satellite et *GPStime*, t_r représente l'instant de réception du signal dans le référentiel *GPStime* et dt_u est le décalage entre le référentiel temporel récepteur et *GPStime*.

La mesure de pseudo-distance correspond à l'équation suivante :

$$\begin{aligned}
 \rho &= c\Delta t_u^s + \varepsilon, \\
 &= c((t_r - dt_u) - (t_e - dt_s)) + \varepsilon, \\
 &= c((t_r - t_e) + (dt_s - dt_u)) + \varepsilon, \\
 &= d_u^s + c(dt_s - dt_u) + \varepsilon,
 \end{aligned} \tag{2.3}$$

où d_u^s désigne la distance euclidienne entre le satellite et le récepteur et ε représente les différents phénomènes impactant la vitesse de propagation du signal. La valeur dt_s est obtenue depuis le message de navigation et celle dt_u est une inconnue à estimer en plus de la position 3D du récepteur.

La deuxième façon de mesurer le temps de trajet est d'utiliser la mesure de phase, ou plus précisément la différence de phase entre les signaux reçus et émis. Cela revient à compter le nombre de cycles ou périodes entre les deux instants. Pour les signaux GPS, un cycle mesure environ 24cm dans le cas de la porteuse L1 et 19cm pour L2 (la longueur d'un cycle correspond à la longueur d'onde).

La pseudo-distance s'écrit :

$$\rho = \lambda \left(\frac{\Delta\phi}{2\pi} + N_{cycles} \right), \tag{2.4}$$

où $\Delta\phi$ est le déphasage entre les deux signaux reçus et générés, λ la longueur d'onde et N_{cycles} un entier inconnu désignant le nombre de cycles et qu'on appelle l'ambiguïté entière. Cette technique est beaucoup plus précise que la mesure de pseudo-distance précédente. Elle peut atteindre une précision millimétrique en statique. Cependant, à cause des différents phénomènes dégradant la mesure (effet atmosphérique, perte du signal), la résolution de l'ambiguïté avec un seul récepteur est impossible. Aussi, cette technique est-elle employée préférentiellement pour des applications avec un récepteur fixe.

2.3.2.2 Mesure Doppler

La mesure de la fréquence du signal de la porteuse peut nous renseigner sur la vitesse instantanée d'un mobile. En effet, la différence de vitesses relatives entre un émetteur (le satellite) et un récepteur (l'antenne GNSS) induit une distorsion du signal qui peut être exploitée pour estimer la vitesse du récepteur. Ce phénomène, appelé effet Doppler, induit une compression ou une dilatation du signal et fait donc varier la fréquence observée. L'effet Doppler doit notamment être pris en compte dans le processus d'autocorrélation du code C/A. La mesure peut aussi s'effectuer sur le code, mais l'utilisation du signal de porteuse est bien plus précise. La mesure Doppler se calcule en comptant le nombre de cycles de la porteuse sur un intervalle de temps donné. La fréquence Doppler maximum mesurée pour le cas du GPS avec un récepteur en mode statique est de 4 KHz.

La figure 2.3 illustre les différents cas de figure de l'effet Doppler en navigation GNSS. La convention pour le signe de la mesure Doppler est que si la projection de la vitesse du satellite sur l'axe reliant la position du satellite et le récepteur est positive (le satellite se rapproche du récepteur), pour

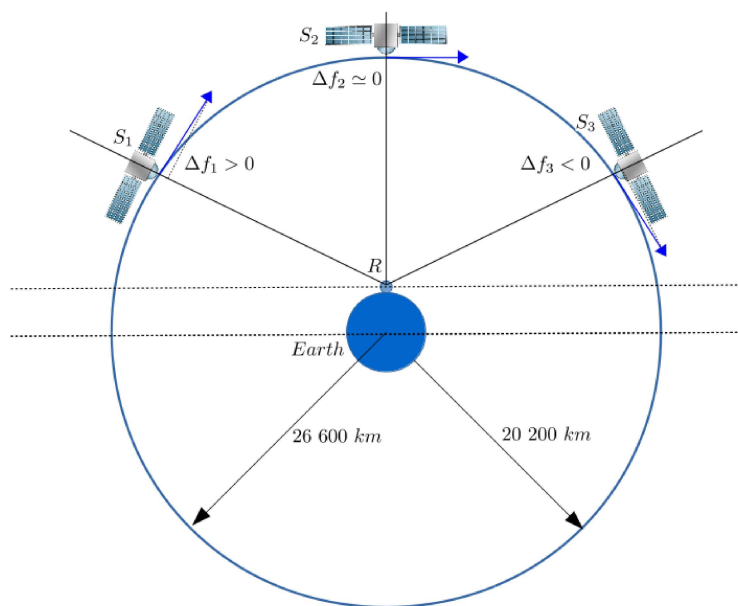


FIGURE 2.3 – Effet Doppler en navigation GNSS.

le satellite S_1 de la figure 2.3, la différence de fréquence est positive. Dans le cas contraire (cas où le satellite s'éloigne du récepteur), comme pour le satellite S_3 , la mesure est négative. Le satellite S_2 est le plus proche du récepteur et la mesure Doppler associée est presque nulle.

Au niveau de l'antenne, la fréquence de réception du signal de porteuse peut être modélisée par l'équation 2.5 :

$$f_r = f_t \left(1 - \frac{\mathbf{v}_r \cdot \mathbf{a}}{c} \right), \quad (2.5)$$

où f_r et f_e sont respectivement la fréquence de réception et de transmission du signal, \mathbf{v}_r est la différence de vitesse entre le satellite et le récepteur, qui s'écrit $\mathbf{v}_r = \dot{\mathbf{x}}_s - \dot{\mathbf{x}}_r$ où $\dot{\mathbf{x}}_s$ et $\dot{\mathbf{x}}_r$ sont respectivement la vitesse du satellite et du récepteur dans le repère terrestre ECEF, et \mathbf{a} est le vecteur unitaire parallèle à l'axe passant par l'émetteur et le récepteur. Le produit $\mathbf{v}_r \cdot \mathbf{a}$ représente le produit scalaire entre les deux vecteurs sur l'axe du vecteur unitaire \mathbf{a} .

Le décalage Doppler dû au mouvement relatif du récepteur et de l'émetteur s'exprime sous la forme :

$$\Delta f = f_r - f_t = -f_t \frac{(\dot{\mathbf{x}}_s - \dot{\mathbf{x}}_r) \cdot \mathbf{a}}{c}. \quad (2.6)$$

L'équation 2.6 ne peut pas être directement utilisée pour estimer la vitesse du récepteur. En effet, comme pour la mesure de pseudo-distance où la synchronisation entre les horloges satellite et récepteur est indispensable, la dérive des deux horloges impacte aussi la mesure Doppler. On réécrit l'équation 2.5 pour un récepteur qui reçoit un signal à une fréquence f_{r_j} ayant été transmis par le satellite S_j à une fréquence f_{t_j} :

$$f_{r_j} = f_{t_j} \left(1 - \frac{(\dot{\mathbf{x}}_{S_j} - \dot{\mathbf{x}}_r) \cdot \mathbf{a}_j}{c} \right). \quad (2.7)$$

La fréquence f_{t_j} correspond plus précisément à la fréquence de référence de la porteuse plus un terme de correction Δf_t qui permet de s'aligner sur le référentiel de temps GPS. Ce terme de correction est généré périodiquement par le segment de contrôle.

La mesure de la fréquence du signal reçu du satellite S_j , notée f_j , diffère de la valeur f_{r_j} par un biais de fréquence. Ce décalage est relié à la dérive d'horloge du récepteur relativement au temps GPS. La relation entre ces deux fréquences s'écrit :

$$f_{r_j} = f_j (1 + \delta t), \quad (2.8)$$

où δt est la dérive d'horloge du récepteur (unité seconde/seconde) et prend une valeur positive ou négative selon si l'horloge du récepteur va plus vite ou moins vite que le temps GPS.

En combinant les équations 2.8 et 2.7, on obtient :

$$c \frac{(f_j - f_{t_j})}{f_{t_j}} + \dot{\mathbf{x}}_{S_j} \cdot \mathbf{a}_j = \dot{\mathbf{x}}_r \cdot \mathbf{a}_j - c \frac{f_j \delta t}{f_{t_j}}. \quad (2.9)$$

La fréquence Doppler mesurée par le récepteur est représentée par la différence de fréquences $(f_j - f_{t_j})$. L'estimation de la vitesse sera abordée plus en détail dans la section 2.4.3.

2.3.3 Sources d'erreur dégradant les mesures

Lors de son parcours, le signal GPS est dégradé (bruité) par plusieurs phénomènes [Grewal et al., 2007, Kaplan and Hegarty, 2006]. Ces phénomènes peuvent être introduits dans l'équation 2.3 :

$$\rho_i = d_u^s + c (dt_s - dt_u) + \varepsilon_{noise} + \varepsilon_{ion} + \varepsilon_{trop} + \varepsilon_{rel} + \varepsilon_{env}. \quad (2.10)$$

Les sections suivantes présentent ces différentes erreurs.

2.3.3.1 Bruits électroniques : ε_{noise}

Lorsque le code du signal est généré dans les satellites, il passe par une multitude de circuits électroniques et de câbles avant d'être émis en onde électromagnétique, le même phénomène se produit à l'arrivée dans l'antenne réceptrice. Le temps de parcours du signal dans la partie satellite est nommé

δ_{TGD} (*Total Group Delay*), ce temps de propagation est propre à chaque satellite et dépend de sa configuration matérielle. Il est inclus dans le message de navigation, et sert à corriger le décalage temporel ainsi que l'effet ionosphérique. Dans la partie antenne réceptrice, les signaux passent par le même câble et sont donc retardés de la même valeur. Ce retard ne perturbe pas le calcul de position, mais l'estimation du décalage d'horloge du récepteur.

2.3.3.2 Erreurs troposphériques : ε_{trop}

La partie inférieure de l'atmosphère de la Terre est composée de gaz sec et de vapeur d'eau, ce qui allonge le trajet de propagation du fait de la réfraction. L'amplitude du retard du signal résultant dépend de l'indice de réfraction de l'air le long du trajet de propagation et varie typiquement de 2,5 m dans la direction zénith à 10-15 m à angle faible altitude du satellite. La troposphère est un milieu non dispersif, de sorte que le retard n'est pas dépendant de la fréquence des signaux GPS. Contrairement à l'ionosphère, le retard de trajet troposphérique est donc le même pour le code et les composants du signal porteur. Par conséquent, ce délai ne peut pas être mesuré en utilisant à la fois des mesures de pseudo-distance L1 et L2.

Le calcul de la correction dépend du taux d'humidité de l'air qui varie aussi selon la température et de l'élévation du satellite par rapport au récepteur. L'équation 2.11 nous montre le calcul de la correction à effectuer :

$$L_{trop} = \frac{1.001 (d_{dry} + d_{wet})}{\sqrt{0.002001 + \sin^2(elev)}}, \quad (2.11)$$

avec :

- d_{dry} = délai vertical des composants secs de l'atmosphère (modélisés comme des gaz parfaits).
- d_{wet} = délai vertical des composants humides de l'atmosphère (vapeur d'eau, difficile à modéliser).
- $elev$ = élévation du satellite (nécessite une connaissance *a priori* de la position du récepteur).

Pour pallier le problème de modélisation de délais verticaux, des modèles saisonniers sont utilisés qui approximent au mieux leurs valeurs.

2.3.3.3 Erreurs ionosphériques : ε_{ion}

La couche ionosphérique s'étend approximativement entre 50 et 1000 km à partir de la surface du globe. Elle se compose de gaz ionisés par le rayonnement solaire. L'ionosphère agit comme un milieu dispersif pour les signaux GPS à cause du nuage d'électrons ionisés qui fait varier la vitesse des signaux en fonction de leur fréquence. L'ionisation varie en fonction des périodes de la journée, elle est maximale le jour et minimale la nuit.

La traversée de la couche ionosphérique par le code GPS aura tendance à allonger la mesure de pseudo-distance. Cela est dû à la nature électromagnétique de l'onde qui est retardée par une diminu-

tion de la vitesse de groupe. Dans le cas d'un récepteur bi-fréquence, on peut quantifier avec précision l'erreur de mesure en utilisant les deux signaux GPS L1 et L2. Ces deux ondes ont des fréquences différentes, et donc se propagent dans l'ionosphère à une vitesse différente ce qui implique une différence de mesure de la pseudo-distance. La formule 2.12 nous donne la correction ionosphérique en utilisant les deux mesures de codes :

$$\epsilon_{ion} = \frac{\rho_{L1} - \rho_{L2}}{1 - \frac{f_{L1}}{f_{L2}}} - c\tau_{TGD}, \quad (2.12)$$

avec :

- ρ_{L1} = pseudo-distance de la fréquence L1,
- ρ_{L2} = pseudo-distance de la fréquence L2,
- f_{L1} = fréquence de la porteuse L1,
- f_{L2} = fréquence de la porteuse L2.
- τ_{TGD} = *Total Delay Group*, délai instrumental du satellite.

2.3.3.4 Erreurs relativistes : ϵ_{rel}

La grande vitesse de déplacement du satellite sur son orbite, et son éloignement par rapport au centre de gravité terrestre induisent des effets liés à la théorie de la relativité. Ainsi, l'écoulement de l'horloge atomique embarquée dans le satellite varie selon sa position et sa vitesse sur l'orbite (à l'apogée, la vitesse est minimale et la distance maximale, tandis qu'au périégée, la vitesse est maximale et la distance minimale). Soit ϵ_{rel} la correction relativiste liée à l'horloge, qui s'écrit ;

$$\epsilon_{rel} = 2 \frac{\sqrt{\mu a}}{c} e \sin(E), \quad (2.13)$$

avec :

- μ = constante de gravité universelle (= $3.986005 \cdot 10^{14}$),
- a = demi axe de l'ellipse orbitale,
- e = excentricité de la trajectoire elliptique,
- E = anomalie excentrique.

2.3.3.5 Erreurs de multi-trajets et de non visibilité directe des satellites : ϵ_{env}

Les multi-trajets font partie des erreurs les plus courantes en zones contraintes. Les milieux urbains sont particulièrement sujets à ce type de phénomènes à cause des immeubles et des arbres environnants. Ces erreurs apparaissent lorsque le récepteur GNSS acquiert différents signaux provenant du même émetteur (ayant le même code) et que ces signaux empruntent des trajectoires différentes. Ils sont réfléchis par des obstacles entourant le récepteur tels que un point d'eau, une construction ou le feuillage des arbres, et empêchent le récepteur de distinguer le signal du satellite en visibilité directe,

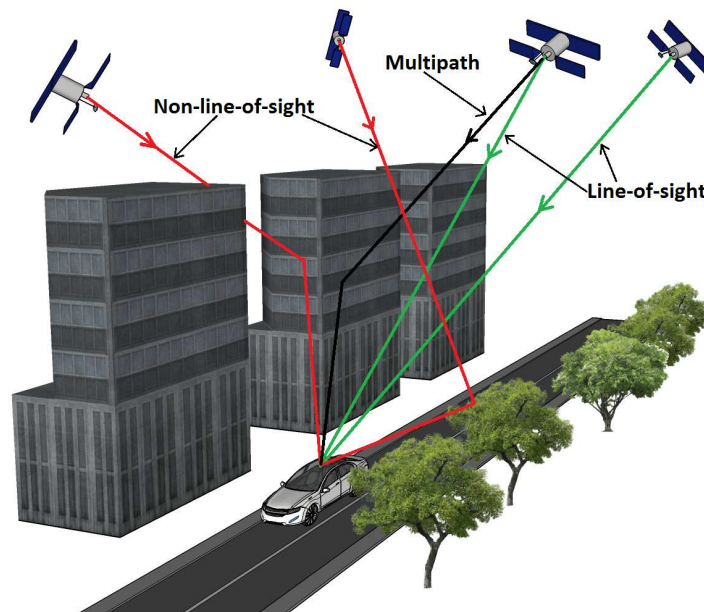


FIGURE 2.4 – Illustration d’un signal en multi-trajet et une réception d’un signal en non-visibilité direct (NLOS).

appelé LOS pour ‘*Line Of Sight*’, et les signaux réfléchis. La nature très intermittente de ce phénomène, lié aux changements d’environnement lors d’un déplacement en milieu urbain, le rend très difficile à modéliser. L’effet des multi-trajets se traduit généralement par un allongement de la mesure de TOA, directement liée à la mesure de pseudo-distance mais aussi à la mesure de phase pour une estimation plus précise de la position.

La figure 2.4 illustre ce phénomène dans le cas d’une réception de signaux en milieu urbain : l’antenne GNSS reçoit le signal venant du satellite à droite et deux signaux du satellite à sa gauche comportant un signal LOS. Les deux satellites restant ne sont pas en visibilité directe par rapport au récepteur, les signaux reçus sont dits en réception NLOS ou ‘*Non-Line Of Sight*’. Généralement, les réceptions NLOS proviennent des satellites en élévation basse, sauf en canyon urbain comme illustré la figure 2.4. Ces erreurs peuvent atteindre des distances supérieures à 150m.

2.3.4 Facteur de qualité pour la localisation

Pour détecter les cas problématiques pour la localisation dus à ces phénomènes impactant les différentes mesures, différents critères ont été proposés qui déterminent, *a priori*, si une estimation de position est de bonne qualité. Nous présentons les critères les plus utilisés.

2.3.4.1 Configuration des satellites (DOP)

La configuration des satellites, appelée DOP pour ‘*Dilution Of Precision*’, constitue un critère purement géométrique. Cet indicateur dépend de la disposition des satellites (visibles) par rapport au

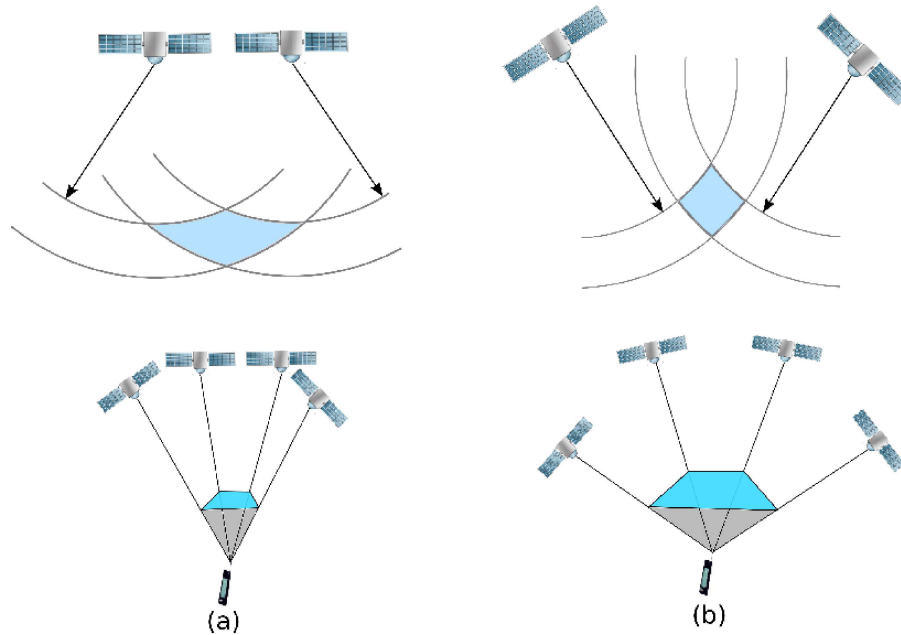


FIGURE 2.5 – (a) Mauvaise et (b) bonne configuration des satellites .

récepteur.

Comme le montre la figure 2.5, plus les satellites sont éloignés les uns des autres, plus grand est le volume occupé par les satellites et le récepteur, et meilleur est l'indicateur DOP. Dans le cas contraire, plus les satellites sont rapprochés, moins ils occupent de volume, ce qui induit un mauvais DOP. On remarque aussi que l'incertitude associée à la position est plus réduite lorsque les satellites sont bien répartis, ce qui fait du DOP un indicateur *a priori* utile pour la mesure de la qualité de l'estimation de la position.

D'une façon générale, l'indicateur DOP est particulièrement médiocre en environnement contraint, où les différents obstacles servent de masque, bloquant ainsi la transmission directe du signal et dégradant la configuration des satellites.

Le calcul de la position utilise la racine de la somme au carré de la trace de la matrice $M = (H_4^T H_4)^{-1}$ où la matrice H_4 est définie dans la section 2.4.2. Différents indicateurs peuvent être considérés selon le type de données à traiter : Le GDOP (*Geometric DOP*) qui évalue la position 3D et le biais, le PDOP (*Position DOP*) qui ne considère que la position 3D ou le HDOP (*Horizontal DOP*) qui ne considère que les deux coordonnées au sol.

2.3.4.2 Élévation des satellites

L'élévation d'un satellite est l'angle mesuré entre l'horizon et le satellite. Dans la plupart des récepteurs GPS, un masque d'élévation de 15° est appliqué. Il y a deux raisons à cela : la forte probabilité qu'un récepteur acquiert un signal en NLOS depuis un satellite en basse élévation, et le fait que le signal transmis par ces satellites subit des perturbations importantes liées à une plus longue traversée

de la couche atmosphérique par rapport aux autres satellites.

2.3.4.3 Nombre de satellites

La redondance de données est un facteur important pour améliorer la précision en position. Un minimum de quatre satellites en LOS est nécessaire pour avoir une position GPS. Cette condition est vérifiée lorsque le récepteur se situe dans un environnement dégagé ou sans obstacle. Le nombre de satellites dans ces conditions peut atteindre 12 en basse latitude, contrairement au cas autour des pôles à cause de la faible couverture GPS à ces latitudes.

Le nombre de satellites est aussi fortement réduit dans les canyons urbains et en forêt, ou parfois moins de quatre signaux satellites sont acquis, rendant impossible l'estimation d'une position instantanée.

2.3.4.4 C/N_0

Dans les milieux urbains, la puissance du signal GNSS est particulièrement atténuée. Le rapport porteuse-sur-bruit C/N_0 (*carrier-to-noise ratio*), qui se situe entre 40 et 50 dB en environnement dégagé, est fortement atténué en présence d'obstacles pouvant absorber une partie de son énergie. Le récepteur considère généralement que la qualité des mesures est satisfaisante pour un C/N_0 en dessus de 30 ou 35dB. Notons cependant qu'un tel C/N_0 ne garantit pas pour autant que les signaux acquis soient LOS.

2.4 Localisation GNSS et navigation

Tout d'abord précisons le vocabulaire utilisé dans la suite de ce manuscrit. Bien que la terminologie "positionnement" (de l'anglais *positioning*) soit aussi utilisée, nous préférons "localisation". Par localisation, nous entendons la position 3D dans le système de coordonnées choisi (ECEF dans notre cas). Cette position est une position ponctuelle qui n'inclut pas l'orientation du récepteur. La méthode pour déterminer la position consiste, comme présentée dans la section 2.3.2.1, en une autocorrélation entre le signal satellite reçu par le récepteur et sa réplique locale. Cette mesure, qui est une pseudo-distance, est considérée par le module qui traite ces données comme étant la distance euclidienne entre le satellite et le récepteur. La solution obtenue n'est donc pas exacte.

2.4.1 Principe de trilatération

La trilatération consiste à estimer la localisation du récepteur comme la solution de l'intersection de sphères, telle que le centre de ces sphères représente la position d'un satellite. La distance est obtenue en multipliant la mesure de TOA par la vitesse de l'onde équivalente à la vitesse de la lumière. La figure 2.6 illustre le principe de trilatération.

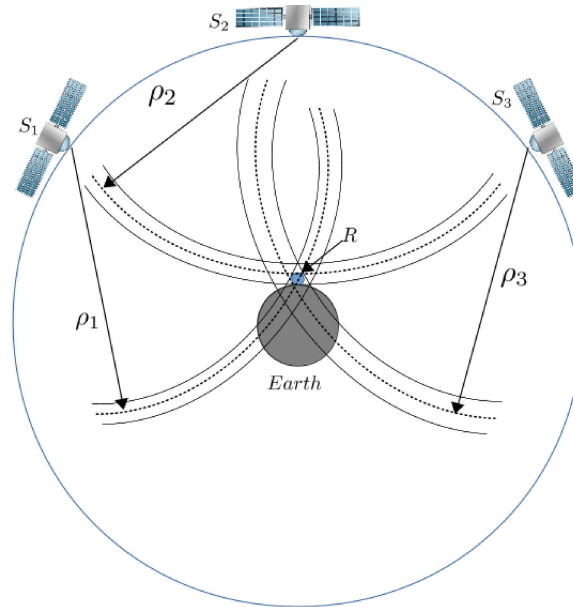


FIGURE 2.6 – Principe de la trilatération par satellite.

L'intersection de ces trois sphères nous donne deux positions possibles, parmi lesquelles nous choisissons celle à proximité de la surface terrestre. Sur la figure 2.6, la valeur mesurée de pseudo-distance pour le satellite S_1 , appelée ρ_1 , est représentée en pointillés. Les positions des satellites sont obtenues en utilisant des paramètres orbitaux fournis par le message de navigation. Les courbes en trait plein de part et d'autre de cette mesure représentent l'imprécision issue des différentes erreurs pouvant impacter la mesure de pseudo-distance. Cependant, dans le cas des systèmes GNSS, la contrainte de proximité de la surface terrestre n'est pas une contrainte suffisante qui fournit une réelle simplification du problème. En effet, l'ajout de la contrainte terrestre est indépendante du processus de transmission du signal et n'apporte pas d'information sur le biais d'horloge. Il est donc nécessaire d'ajouter un quatrième satellite pour déterminer le biais d'horloge entre l'horloge du satellite et celle du récepteur.

2.4.2 Résolution du problème de localisation

La résolution du système d'équations non linéaires GNSS peut être effectuée de différentes manières, selon :

- une méthode directe sans *a priori* sur la position du récepteur, e.g. [Bancroft, 1985];
- une méthode de résolution itérative, e.g. [Kaplan and Hegarty, 2006];
- une méthode de filtrage, e.g. le filtre de Kalman.

La résolution itérative consiste à linéariser le système d'équations en ayant une connaissance approximative de la position du récepteur. Soit (x_r, y_r, z_r) la position réelle inconnue du récepteur, $(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r)$ la position approximative connue du récepteur, et $(\Delta x_r, \Delta y_r, \Delta z_r)$ le déplacement entre ces deux positions. L'équation 2.3 peut s'écrire sans prendre en compte le bruit :

$$f_i(x_r, y_r, z_r, \delta t) = \rho_{S_i} = \sqrt{(x_{S_i} - x_r)^2 + (y_{S_i} - y_r)^2 + (z_{S_i} - z_r)^2} + c\delta t, \quad (2.14)$$

où $(x_{S_i}, y_{S_i}, z_{S_i})$ est la position du satellite S_i dans le repère ECEF et δt est le biais temporel. En utilisant la position approximative du récepteur ainsi qu'une approximation du biais temporel δt dans l'équation 2.14, une estimation de la pseudo-distance est obtenue

$$f_i(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r, \tilde{\delta}t) = \sqrt{(x_{S_i} - \tilde{x}_r)^2 + (y_{S_i} - \tilde{y}_r)^2 + (z_{S_i} - \tilde{z}_r)^2} + c\tilde{\delta}t. \quad (2.15)$$

En ayant défini l'incrément pour chaque coordonnée $(\Delta x_r, \Delta y_r, \Delta z_r)$ ainsi que $\Delta\delta t = \delta t - \tilde{\delta}t$, l'équation 2.14 peut se développer en fonction de $(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r, \tilde{\delta}t)$ par une série de Taylor du premier ordre :

$$\begin{aligned} f_i(x_r, y_r, z_r, \delta t) &= f_i(\tilde{x}_r + \Delta x_r, \tilde{y}_r + \Delta y_r, \tilde{z}_r + \Delta z_r, \tilde{\delta}t + \Delta\delta t), \\ &= f_i(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r, \tilde{\delta}t) + \frac{\partial f_i(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r, \tilde{\delta}t)}{\partial \tilde{x}_r} \Delta x_r + \frac{\partial f_i(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r, \tilde{\delta}t)}{\partial \tilde{y}_r} \Delta y_r \\ &\quad + \frac{\partial f_i(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r, \tilde{\delta}t)}{\partial \tilde{z}_r} \Delta z_r + \frac{\partial f_i(\tilde{x}_r, \tilde{y}_r, \tilde{z}_r, \tilde{\delta}t)}{\partial \tilde{\delta}t} \Delta\delta t, \\ &= \tilde{\rho}_{S_i} - \frac{x_{S_i} - \tilde{x}_r}{R_i} \Delta x_r - \frac{y_{S_i} - \tilde{y}_r}{R_i} \Delta y_r - \frac{z_{S_i} - \tilde{z}_r}{R_i} \Delta z_r + c\Delta\delta t, \end{aligned} \quad (2.16)$$

où $R_i = \sqrt{(x_{S_i} - \tilde{x}_r)^2 + (y_{S_i} - \tilde{y}_r)^2 + (z_{S_i} - \tilde{z}_r)^2}$ est la distance euclidienne entre le satellite S_i et le récepteur. En considérant quatre satellites, la localisation revient à la résolution d'un système de quatre équations linéaires à quatre inconnues de la forme :

$$\Delta\rho_4 = H_4 \Delta X, \quad (2.17)$$

où

$$\Delta\rho_4 = \begin{bmatrix} \rho_{S_1} - \tilde{\rho}_{S_1} \\ \rho_{S_2} - \tilde{\rho}_{S_2} \\ \rho_{S_3} - \tilde{\rho}_{S_3} \\ \rho_{S_4} - \tilde{\rho}_{S_4} \end{bmatrix}; H_4 = \begin{bmatrix} -a_{x1} & -a_{y1} & -a_{z1} & 1 \\ -a_{x2} & -a_{y2} & -a_{z2} & 1 \\ -a_{x3} & -a_{y3} & -a_{z3} & 1 \\ -a_{x4} & -a_{y4} & -a_{z4} & 1 \end{bmatrix}; \Delta X = \begin{bmatrix} \Delta x_r \\ \Delta y_r \\ \Delta z_r \\ c\Delta t \end{bmatrix}, \quad (2.18)$$

où (a_{xi}, a_{yi}, a_{zi}) sont les termes du premier ordre de la série de Taylor de l'équation 2.16 pour le satellite S_j .

Il faut alors déterminer ΔX :

$$\Delta X = H_4^{-1} \Delta\rho_4. \quad (2.19)$$

La solution obtenue est peut-être inexacte à cause des erreurs entachant la mesure de pseudo-distance, ainsi la somme des résidus (différences entre une mesure et son estimation) est égale à zéro $(\sum_{i=1}^4 r_i^2 = \sum_{i=1}^4 [\rho_{S_i} - \tilde{\rho}_{S_i}]^2 = 0)$.

En utilisant plus de quatre satellites, le problème se transforme en minimisation de la somme quadratique des résidus. Soit $\Delta\rho$ le vecteur des erreurs d'observations et H la jacobienne pour $N > 4$ mesures, la résolution du problème se fait alors de façon itérative en utilisant l'algorithme de Gauss-Newton :

$$\begin{cases} \Delta \mathbf{X}_k &= (\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \Delta \boldsymbol{\rho} \\ \mathbf{X}_{k+1} &= \mathbf{X}_k + \Delta \mathbf{X}_k \end{cases}, \quad (2.20)$$

où \mathbf{X}_k et $\Delta \mathbf{X}_k$ représentent respectivement l'estimation du vecteur position avec le biais temporel et le vecteur incrément à l'itération k . Le critère d'arrêt de l'algorithme est atteint quand $\Delta \mathbf{X}_k$ est inférieur à un seuil *a priori*. Les composantes du vecteur d'état \mathbf{X}_0 sont initialisées à 0.

2.4.3 Détermination de la vitesse du récepteur

La vitesse du récepteur peut être déterminée par différence finie entre les positions en différents instants. C'est ainsi que procèdent la plupart des récepteurs bas coût que l'on trouve sur le marché. Sur les récepteurs GPS modernes, la mesure de la fréquence de la porteuse est utilisée pour estimer le décalage Doppler entre la fréquence de transmission et de réception. Elle permet d'avoir une estimation instantanée de la vitesse du récepteur avec une grande précision, indépendamment de la mesure de pseudo-distance.

L'équation 2.9, peut être réécrite :

$$c \frac{D_i}{f_{t_i}} + \dot{\mathbf{x}}_{S_i} \cdot \mathbf{a}_i = \dot{\mathbf{x}}_r \cdot \mathbf{a}_i - c \dot{\delta} t, \quad (2.21)$$

où D_i représente la mesure Doppler. La fraction $\frac{f_i}{f_{t_i}}$ a été simplifiée car elle est très proche de 1. Le vecteur unitaire $\mathbf{a}_i = (a_{x_i}, a_{y_i}, a_{z_i})^\top$ est celui défini dans la section précédente.

Le terme à droite de l'équation 2.21 est connu si on possède un *a priori* sur la position du récepteur, par exemple issu des mesures de pseudo-distances. Nous remplaçons ce terme par $d_j = c \frac{D_i}{f_{t_i}} + \dot{\mathbf{x}}_{S_i} \cdot \mathbf{a}_i$.

Comme pour la partie localisation, en considérant quatre mesures Doppler, un système d'équations linéaires est obtenu :

$$\mathbf{d} = \dot{\mathbf{H}}_4 \dot{\mathbf{X}}, \quad (2.22)$$

avec :

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix}; \dot{\mathbf{H}}_4 = \begin{bmatrix} a_{x_1} & a_{y_1} & a_{z_1} & 1 \\ a_{x_2} & a_{y_2} & a_{z_2} & 1 \\ a_{x_3} & a_{y_3} & a_{z_3} & 1 \\ a_{x_4} & a_{y_4} & a_{z_4} & 1 \end{bmatrix}; \dot{\mathbf{X}} = \begin{bmatrix} \dot{x}_r \\ \dot{y}_r \\ \dot{z}_r \\ -c \dot{\delta} t \end{bmatrix}, \quad (2.23)$$

La résolution s'effectue simplement en inversant le système ($\dot{\mathbf{X}} = \dot{\mathbf{H}}_4^{-1} \mathbf{d}$). De la même manière que dans la précédente sous section, une méthode itérative est utilisée pour estimer la vitesse instantanée

pour $N > 4$. Remarquons tout de même que le système ne peut être résolu sans avoir une position approximative du récepteur. Étant donnée la distance importante qui sépare un récepteur d'un satellite, cette position n'a pas besoin d'être très précise pour avoir une estimation correcte de la vitesse.

2.4.4 Filtre de Kalman étendu pour la navigation

La méthode des moindres carrés vue précédemment nous donnent une estimation instantanée de la position ou de la vitesse d'un récepteur. Outre la simplicité de cette technique, elle nécessite aux minimum quatre mesures de pseudo-distance ou mesures Doppler à chaque instant pour fournir une solution, ce qui n'est pas le cas en milieu contraint. Le filtre de Kalman est connu dans de nombreuses disciplines dont la navigation GPS [Kalman, 1960]. Il est utilisé pour pallier les problèmes d'estimation d'état et de filtrage. Un modèle d'évolution, aussi appelé modèle de prédiction, de l'état est défini. Le filtre de Kalman est un filtre récursif, c'est-à-dire, qu'il utilise la connaissance de l'état précédent pour estimer l'état courant, ce qui a pour avantage de lisser la trajectoire du récepteur. Cependant, deux hypothèses sont nécessaires pour que le filtre soit optimal selon le critère de le minimum de l'erreur quadratique moyenne : les modèles d'évolution et d'observation doivent être linéaires et les bruits doivent suivre une loi gaussienne.

Le système d'état à résoudre est alors de la forme suivante :

$$\begin{cases} \mathbf{X}_k &= f(\mathbf{X}_{k-1}) + \mathbf{v}_{k-1}, \\ \mathbf{Y}_k &= h(\mathbf{X}_k) + \boldsymbol{\omega}_{k-1}, \end{cases} \quad (2.24)$$

avec \mathbf{X}_k le vecteur d'état, \mathbf{Y}_k le vecteur des observations, $f(\mathbf{X}_{k-1})$ le modèle d'évolution du système, $h(\mathbf{X}_k)$ le modèle d'observation des mesures, et \mathbf{v}_{k-1} et $\boldsymbol{\omega}_{k-1}$ respectivement les bruits de prédiction (ou du modèle) et d'observation. En cas de non-linéarité du modèle d'évolution et/ou d'observation, on utilise le filtre de Kalman étendu (EKF).

Comme le filtre de Kalman classique, l'EKF comprend deux étapes : une étape de prédiction et une étape d'estimation. Nous le détaillerons dans le cas d'un vecteur d'état incluant position et vitesse, et en considérant exclusivement les mesures de pseudo-distance [Cooper and Durrant-Whyte, 1994].

Soit le vecteur d'état $\mathbf{X} = (x_r, y_r, z_r, \delta t, \dot{x}_r, \dot{y}_r, \dot{z}_r, \dot{\delta t})$, l'étape de prédiction de l'EKF s'écrit :

$$\begin{cases} \tilde{\mathbf{X}}_{k|k-1} &= f(\tilde{\mathbf{X}}_{k-1|k-1}), \\ \mathbf{P}_{k|k-1} &= \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}, \end{cases} \quad (2.25)$$

avec $\tilde{\mathbf{X}}_{k-1|k-1}$ l'estimation du vecteur d'état à l'instant $k-1$, $\tilde{\mathbf{X}}_{k|k-1}$ l'estimation du vecteur d'état à l'instant k sachant l'estimation précédente, $\mathbf{P}_{k-1|k-1}$ la matrice de covariance de l'erreur sur la prédiction du vecteur d'état et \mathbf{Q}_{k-1} la matrice de covariance du bruit de prédiction supposé Gaussien. Le modèle d'évolution de notre vecteur d'état est non-linéaire. La fonction $f(\mathbf{X}_{k-1})$ peut directement être remplacée par le modèle linéarisé autour de $(\tilde{\mathbf{X}}_{k-1})$ ce qui nous donne : $f(\tilde{\mathbf{X}}_{k-1|k-1}) = \mathbf{F}_{k-1} \tilde{\mathbf{X}}_{k-1|k-1}$.

L'étape d'estimation, aussi appelée étape de correction, prend en compte les observations acquises à l'instant k , selon les équations suivantes :

$$\begin{cases} z_k &= Y_k - h(\tilde{\mathbf{X}}_{k|k-1}), \\ \mathbf{K}_k &= \mathbf{P}_{k|k-1} \mathbf{H}_k^\top (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^\top + \mathbf{R}_k)^{-1}, \\ \tilde{\mathbf{X}}_{k|k} &= \tilde{\mathbf{X}}_{k|k-1} + \mathbf{K}_k z_k, \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}, \end{cases} \quad (2.26)$$

où z_k représente le terme d'innovation, \mathbf{K}_k le gain de Kalman et \mathbf{R}_k la matrice de covariance des bruits de mesures. On remarque qu'en utilisant que les mesures de pseudo-distances, seuls les quatre premiers états du vecteur $\tilde{\mathbf{X}}_{k|k}$ sont observables.

Dans la partie estimation, le modèle d'observation non-linéaire, introduit dans l'équation 2.3, est utilisé pour le calcul de l'innovation. La matrice linéarisée autour de $\tilde{\mathbf{X}}_{k|k-1}$, introduite dans l'équation 2.18, est utilisée pour le calcul du gain de Kalman et l'estimation de la matrice de covariance *a posteriori* de l'erreur d'état.

La même procédure est suivie pour estimer la vitesse à partir des mesures Doppler, avec un *a priori* à chaque instant sur la position 3D du récepteur. Il est aussi possible de combiner les mesures de pseudo-distances et Doppler pour estimer conjointement la position et la vitesse. L'accélération est ajoutée comme état caché pour prédire la vitesse au cours du temps. Une modélisation de ce cas a été introduite par [Mao et al., 2002] en utilisant un filtre de Kalman appelé 'sans parfum' : *Unscented Kalman Filter* (UKF).

2.5 Conclusion

Dans ce chapitre, nous avons abordé les principes de base de la navigation GNSS. La problématique de la thèse s'oriente vers la localisation en milieu contraint, particulièrement en ville où la précision fournie par le GPS est médiocre.

La localisation dans les canyons urbains est primordiale dans de nombreuses applications, que ce soit pour se localiser avec un téléphone équipé d'un GPS ou pour la navigation autonome appliquée aux véhicules automobiles. Nous avons choisi de n'utiliser que les mesures fournies par un récepteur GPS et de les traiter dans le but d'améliorer la précision de la localisation. La solution obtenue pourra être utilisée seule ou comme entrée d'un système de fusion de données multicapteurs.

Les données à traiter seront entachées des sources d'erreurs les plus courantes en milieu urbain à savoir les multi-trajets et les NLOS. La représentation de l'information sera tout aussi importante, et d'elle aussi, dépendra la manière dont ces sources d'erreurs seront manipulées.

Nous avons aussi présenté le problème de la localisation et les deux méthodes les plus classiques pour estimer la position du récepteur à partir des données GPS. La solution recherchée (position ou plus généralement vecteur d'état) est alors 'ponctuelle' au sens où elle est incertaine mais précise.

D'autres approches (notamment l'analyse par intervalles présentée au chapitre 3) proposent une représentation de la solution certaine mais imprécise. En effet, les informations (ou données) fournies par les capteurs sont imparfaites, notamment à la fois imprécises et incertaines [Khaleghi et al., 2011]. Une représentation adaptée (adéquate) de ces imperfections est nécessaire à tout processus de décision ou d'estimation. Par exemple, la représentation vecteur d'état ponctuel (filtre de Kalman) permet de gérer les incertitudes sur l'état du système, alors qu'une représentation sous forme d'ensembles permet d'assurer ou 'garantir' la présence de la solution dans l'ensemble sans donner de solution précise. Nous insistons sur le fait que la représentation choisie (ponctuelle ou ensembliste) conditionne les méthodes de traitement des données, que ce soit pour la détection des *outliers* ou la localisation.

Méthodes de détection et d'identification de données aberrantes

Sommaire

3.1	Introduction	27
3.2	Détection d' <i>outliers</i> par couplage avec d'autres informations	28
3.3	Détection d' <i>outliers</i> intra-mesures GNSS	30
3.3.1	Analyse par intervalles	30
3.3.2	Approches statistiques	32
3.3.3	Méthodes robustes pour la localisation	37
3.4	Conclusion	48

3.1 Introduction

Comme vu au chapitre précédent, la localisation est une tâche primordiale pour des applications telles que la navigation autonome en ville. Par exemple, un véhicule doit savoir précisément sur quelle voie il se trouve ; un piéton attend du système une localisation avec une précision nettement inférieure à celle des objets qui l'entourent (bâtiments, rue...). Cette tâche n'est pas évidente en raison des différents masquages que rencontrent les récepteurs GPS en ville et qui biaisent l'estimation de position. Comme vu au chapitre précédent, les erreurs provoquées par les obstacles environnant tels que les bâtiments qui bloquent les signaux satellites ou les réfléchissent, se manifestent sous forme de multi-trajets et/ou de réceptions NLOS. Dans ce chapitre, nous nous intéressons plus généralement à la notion de données aberrantes ou selon la terminologie anglaise *outliers*.

Les définitions d'un *outlier* dans la littérature sont aussi variées que les noms des méthodes utilisées pour les détecter : techniques de détection de défauts, de changements brusques, d'anomalies ou de données aberrantes, dont l'appellation varie aussi en fonction des domaines d'application, aussi divers que la finance, l'astrophysique, l'économétrie, les bases de données, les réseaux [Gupta et al., 2014], l'apprentissage statistique [Hodge and Austin, 2004], l'automatique ou le traitement du signal et des images.

Parmi les nombreuses définitions qui ont été proposées dans la littérature, citons textuellement :

Définition 1 : "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs" [Grubbs, 1969],

Définition 2 : "An observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism" [Hawkins, 1980],

Définition 3 : "An observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data" [Barnett and Lewis, 1994],

Définition 4 : "An outlier in a set of data or an observation or a point that is considerably dissimilar or inconsistent with the remainder of the data" [Ramaswamy et al., 2000].

Toutes ces définitions s'accordent sur le fait que la notion d'*outlier* est relative à un ensemble d'observations où la donnée aberrante n'est pas compatible ou cohérente avec le reste de cet ensemble d'observations. Cette anomalie peut être due à un mécanisme extérieur au processus qui génère ces données ou à une défaillance du système d'acquisition. Dans la majorité des applications, la détection d'*outliers* est une tâche hautement critique. Par exemple, une fraude à la carte bancaire, une détection de fuite dans un *pipeline* ou l'intrusion d'un inconnu dans une région observée par un système de vidéosurveillance nécessite une réaction rapide du système de détection d'anomalies.

Classifier les méthodes de détection d'*outliers* n'est donc pas une tâche aisée et dépend fortement du domaine d'application. Dans ce chapitre, la présentation que nous donnons des différentes techniques de détection de données aberrantes a été partiellement dictée par la problématique de la localisation GPS. Précisément, nous commençons par citer des méthodes spécifiques à la localisation GPS au sens où elles exploitent d'autres capteurs et leurs spécificités (ce sont des méthodes *ad hoc* au sens positif du terme). Puis nous citons les approches historiques qui se sont focalisées sur la pondération des mesures aberrantes ou sur leur identification et élimination. Enfin, nous présentons, parmi les approches probabilistes, la détection *a contrario* qui sera l'un des cadres choisis pour nos travaux.

3.2 Détection d'*outliers* par couplage avec d'autres informations

Pour la localisation en zone urbaine appliquée aux véhicules automobiles, différentes approches ont été proposées [Skog and Handel, 2009]. Un premier type d'approche enrichit les mesures GPS avec des données proprioceptives issues de capteurs embarqués dans le véhicule [Cui and Ge, 2003] [Cappelle et al., 2012]. Dans [Abdel-Hafez, 2010], l'auteur estime la covariance des bruits de mesures GPS en s'aidant d'un capteur inertiel et en adoptant une approche *Ultra-tightly coupling*. L'estimation de la solution est réalisée par la technique *Autocovariance Least-Squares*. [Obst et al., 2012] détecte et atténue les multi-trajets en utilisant une carte 3D, un couplage avec un capteur inertiel est réalisé avec le filtre de Kalman sans parfum (UKF) pour la localisation robuste. Dans [Chiang et al., 2013], une implémentation temps-réel d'un système INS/GPS avec élimination des mesures GPS aberrantes ou "défaillantes" (selon le terme des auteurs) est réalisée. L'identification des mesures défaillantes est possible grâce à une modification du terme d'innovation dans le filtre de Kalman pour la prédiction des mesures de pseudo-distances (PR) avec le capteur inertiel. Dans [Won et al., 2014], l'auteur utilise,

en plus du récepteur GPS, un système de navigation inertiel et une caméra pour la localisation en visibilité restreinte.

Dans l'esprit de l'exploitation de capteurs complémentaires/supplémentaires, d'autres approches proposent de détecter les satellites en non-visibilité directe. Par exemple, dans [Marais et al., 2013], une combinaison des données GPS et des images du ciel est effectuée pour la détection des satellites NLOS. Une segmentation de l'image en visée verticale en direction du ciel et percevant les obstacles est réalisée. Les positions des satellites, fournies par les *almanachs*, sont ensuite projetées sur la carte du ciel afin d'identifier les réceptions NLOS. La même approche est suivie dans [Meguro et al., 2009] en utilisant une caméra infrarouge pour la vision nocturne et dans [Ramalingam et al., 2010] en ajoutant une carte 3D de l'environnement. Une autre approche, proposée par [Peyraud et al., 2013], consiste à détecter si un satellite est directement visible (LOS) ou pas (NLOS) en créant une image virtuelle de chaque satellite avec une caméra virtuelle orientée en direction du satellite considéré. La caméra est simulée en utilisant une carte 3D et un capteur d'orientation du véhicule.

Toujours en s'appuyant sur des données extérieures mais ici non issues de capteur embarqué, dans [Miura and Kamijo, 2014], la correction des erreurs PR est réalisée en estimant les biais avec une carte 3D créée à partir d'une carte OSM (*Open Street Map*) et d'un modèle numérique de terrain. Un filtre particulière est utilisé pour estimer la position. Pour la localisation piétonne, [Hsu et al., 2015] utilise la méthode du lancé de rayon (*Ray tracing*) combinée avec une carte 3D. Une approche similaire a été proposée par [Bétaille et al., 2013, Bétaille et al., 2016] en exploitant les modèles numériques de terrain sous forme d'un modèle de tranchée urbaine, pour détecter et corriger les signaux NLOS et ainsi améliorer l'estimation.

La combinaison de carte 3D avec les mesures de rapport signal sur bruit (SNR) issues du GPS a été exploitée par différents auteurs [Groves et al., 2012] [Obst and Wanielik, 2013]. Les auteurs se fondent sur une distribution de probabilité des signaux LOS et NLOS en fonction du SNR. La même hypothèse est adoptée par [Iwase et al., 2013] pour sélectionner des satellites de référence. Cependant, au lieu de s'intéresser aux résidus des mesures GPS, ce dernier estime directement les erreurs de multi-trajets connaissant la position du véhicule. Cette méthode nécessite néanmoins de sélectionner des satellites de référence en se basant sur leur élévation et le SNR. Pour simplifier le problème, l'altitude a été supposée comme *a priori*, ainsi seulement deux satellites de référence sont nécessaires. Néanmoins, le critère sur la distribution du SNR n'est pas toujours discriminant pour différencier les signaux LOS et NLOS. Une approche basée sur la modélisation des distributions LOS et NLOS combinée à un suivi de l'état du système a été proposée par [Huerta et al., 2009].

Une approche récente pour améliorer la localisation des piétons dans les carrefours par couplage avec une carte 3D de l'environnement est le *Shadow matching* [Groves, 2011]. Dans cette approche, les conditions de visibilité des signaux satellites à différents endroits d'une carte 3D sont prédites et comparées avec les mesures de satellites acquises par le récepteur pour déterminer une position. Seules les réceptions en LOS sont prédites bien que les signaux NLOS soient souvent nombreux en ville. [Wang et al., 2013] ont alors proposé d'améliorer l'approche en introduisant un système de scores pour chaque cellule de la carte pour apprendre à reconnaître les signaux directs des signaux réfléchis. [Groves et al., 2015] présente un état de l'art sur ce type d'approche.

Dans la suite de ce chapitre, nous nous concentrons sur les méthodes utilisant exclusivement les mesures GPS. Soulignons, cependant, que ce n'est nullement en contradiction avec un couplage ultérieur avec des données annexes (issues d'autres capteurs ou cartographiques) puisqu'une amélioration des algorithmes de localisation GNSS devrait améliorer les résultats des approches exploitant ces données, par exemple en fusion de données.

3.3 Détection d'*outliers* intra-mesures GNSS

3.3.1 Analyse par intervalles

L'arithmétique par intervalles a été introduite par Moore dans les années 70 [Moore et al., 1966]. Elle a pour but de représenter un nombre, entier ou réel, ou une densité de probabilité par un intervalle ou un ensemble contenant la solution. Par exemple, dans le cas d'un nombre, cet intervalle est délimité par une borne supérieure et une borne inférieure. Une telle représentation permet de manipuler l'imprécision de façon déterministe. Cette arithmétique a été appliquée à la résolution de problèmes non-linéaires complexes de par sa capacité à représenter l'imprécision d'un nombre et 'garantir' un résultat si les hypothèses sur les bornes initiales sont satisfaites. En ce sens, ces approches se distinguent des méthodes probabilistes qui se concentrent sur la modélisation de l'incertitude (plutôt que l'imprécision). En revanche, l'arithmétique par intervalles ne donne pas d'information sur la solution précise, sauf qu'elle est quelque part à l'intersection des sources.

Un intervalle dans \mathbb{R} est défini comme un sous-ensemble connexe, borné et fermé de réels :

$$[x] = [\underline{x}, \bar{x}] \text{ avec } \underline{x} \leq x \leq \bar{x}. \quad (3.1)$$

D'une façon plus générale, un vecteur de nombres réels est représenté par un 'pavé' ou vecteur d'intervalles, défini comme un produit cartésien d'intervalles tel que :

$$\begin{aligned} [x] &= [x_1, \bar{x}_1] \times [x_2, \bar{x}_2] \times \cdots \times [x_n, \bar{x}_n], \\ &= ([x_1, \bar{x}_1], [x_2, \bar{x}_2], \cdots, [x_n, \bar{x}_n])^\top. \end{aligned} \quad (3.2)$$

Les opérateurs d'inclusion, d'intersection et d'union sont directement applicables au contexte ensembliste. Les opérateurs classiques utilisés pour les nombres réels (+, -, \times , 2) doivent quant à eux suivre les règles de l'arithmétique par intervalles [Moore et al., 1966].

En localisation GPS, les pseudo-distances sont des mesures imprécises de part les erreurs de modèle et les perturbations. Ces observations peuvent alors être représentées sous forme d'intervalles centrés sur la valeur de pseudo-distance et d'une largeur correspondant à une imprécision *a priori*. De la même manière, les positions des satellites peuvent-être représentées sous forme de pavés. L'incertitude sur la position d'un satellite dans une direction donnée est représentée par la largeur de l'intervalle dans la direction correspondante de la boîte englobante. Le modèle d'observation ensembliste des pseudo-distances appliqué à la localisation GPS s'écrit :

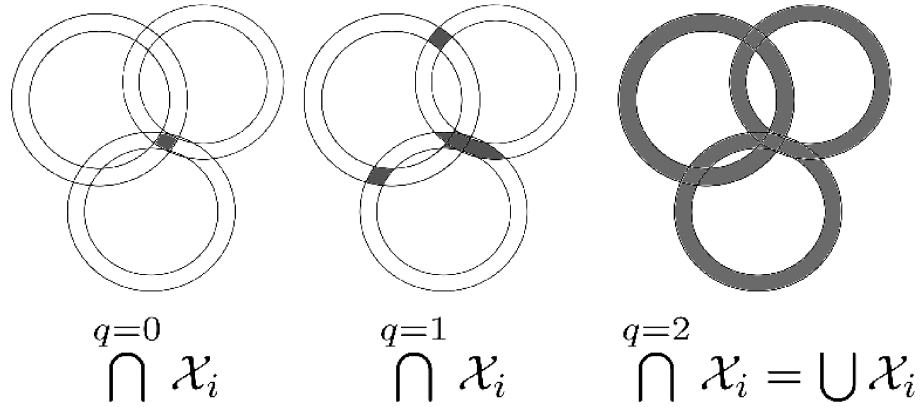


FIGURE 3.1 – Illustration de la q-relaxation avec 3 balises.

$$[\rho]_i = [f_i]([x]) = \sqrt{([x_{S_i}] - [x_r])^2 + ([y_{S_i}] - [y_r])^2 + ([z_{S_i}] - [z_r])^2} + c[\delta t]. \quad (3.3)$$

La fonction $[f_i]$ est appelée ‘fonction d’inclusion naturelle’ du modèle d’observation. La fonction qui fournit la plus petite boîte englobant toutes les solutions est appelée ‘fonction d’inclusion minimale’ [Jaulin et al., 2001b].

L’estimation a pour objet de caractériser l’ensemble $[x]$ à partir des observations. L’algorithme SIVIA (*Set Inversion Via Interval Analysis*) effectue des tests d’inclusion qui permettent de calculer un pavage intérieur et extérieur pour encadrer le sous-pavé solution [Jaulin and Walter, 1993]. Un exemple d’application temps-réel de la méthode SIVIA pour la localisation d’un robot mobile est présenté dans [Seigneur et al., 2009].

Une autre manière d’estimer le pavé le plus précis encadrant la solution est d’utiliser les contracteurs. Ils consistent à remplacer $[x]$ par un sous-pavé $[x']$ plus réduit en satisfaisant un certain nombre de contraintes linéaires ou non-linéaires [Jaulin et al., 2001a]. Le contracteur le plus utilisé est le FBP (*Forward-Backward Propagation*) qui considère les contraintes une à une et applique une série de propagations et rétropropagations jusqu’à obtenir le pavé le plus petit possible. Cette méthode a été appliquée par [Gning and Bonnifait, 2006] pour la localisation d’un véhicule.

L’analyse par intervalles (AI) a pour hypothèse que le bruit est borné afin d’obtenir une solution dite ‘garantie’. En cas de données aberrantes ou de forte perturbation sur les observations, la solution peut être l’ensemble vide. Pour pallier ce problème, la technique RSIVIA (*Relaxed Set Inverter Via Interval Analysis*) a été développée. C’est une méthode d’inversion ensembliste rendue robuste aux données aberrantes par l’usage de la q-relaxation [Jaulin, 2009]. La q-relaxation consiste à ne considérer que q observations parmi m . La figure 3.1 illustre la q-relaxation avec des ensembles correspondant à 3 anneaux obtenus comme solutions de la projection sur un plan de la sphère centrée sur un satellite GPS de rayon égal à la pseudo-distance. L’exemple choisi nous donne une intersection très petite (correspondant peut être à une estimation très précise de la solution) si aucune contrainte n’est relâchée, i.e. avec $q = 0$. Les solutions pour $q = 1$ et $q = 2$ sont également indiquées sur la figure. L’algorithme 1 présente le principe général de la méthode RSIVIA. Pour satisfaire aux contraintes temps-réel, un temps de calcul *a priori* peut être pris en compte au niveau de la condition *until* (ligne 9). En relâ-

Algorithm 1: Analyse par intervalles avec q-relaxation ; entrées : $[x]$ un ensemble de N pavés, $[y]$ l'ensemble des m observations, f les fonctions d'inclusions et q le nombre de contraintes à relâcher et en sorties : $[\tilde{x}]$ le pavé solution.

```

1 Initialiser la liste  $\mathcal{L} = \{[x]\}$ .
2 while  $\mathcal{L} \neq \emptyset$  do
3   Tirer( $[x], \mathcal{L}$ )
4   repeat
5     for  $i = 1 \leftarrow m$  do
6       Calculer  $[x_i] = [x] \cap f_i^{-1}([y_i])$ 
7     end
8      $[x] = \left[ \bigcap_{i \in \{1 \dots m\}} [x_i] \right]$ 
9   until aucune contraction supplémentaire possible;
10  if  $[x] \neq \emptyset$  and  $w([x]) < \varepsilon$  then
11    Appliquer une bisection de  $([x]_1, [x]_2) \leftarrow [x]$ 
12     $\mathcal{L} = \mathcal{L} + \{[x]_1, [x]_2\}$ 
13  end
14   $[\tilde{x}] = \cup(\mathcal{L})$ 
15 end

```

chant $q = m - 1$ contraintes, la solution est l'union des ensembles initiaux. La méthode d'inversion ensembliste par q-relaxation a été appliquée en robotique sous-marine [Sliwka et al., 2011] et pour la localisation GPS robuste [Drevelle and Bonnifait, 2011].

La méthode peut être étendue au cas où les ensembles initiaux n'ont pas la même 'probabilité' d'inclure la solution. Dans ce cas, l'analyse par intervalles se définit comme un cas particulier d'un formalisme plus général appelé théorie des fonctions de croyance. Une application pour la localisation dynamique d'un véhicule a été introduite par [Nassreddine et al., 2010] combinant les deux théories.

3.3.2 Approches statistiques

Nous allons présenter dans cette section les approches statistiques les plus utilisées pour la détection d'*outliers* ou de défauts dans un ensemble de mesures issues d'un même capteur. Les résidus (qui seront présentés dans la section suivante) sont parmi les indicateurs les plus utilisés pour détecter la présence de mesures erronées dans un ensemble d'observations.

3.3.2.1 Détection d'*outliers* par analyse des résidus

Soient N mesures regroupées dans le vecteur Y_t et soit le modèle d'observation des mesures $h(X_t)$ avec X_t le vecteur d'état à estimer de taille m . Le vecteur X_t est composé, en localisation GPS, de la position 3D et du biais d'horloge. Les observations sont reliées aux mesures par la relation :

$$\mathbf{Y}_t = h(\mathbf{X}_t) + \mathbf{E}_t + \mathbf{v}_{Y_t}, \quad (3.4)$$

avec \mathbf{E}_t le vecteur des défauts (les défauts sont ici synonymes de biais) et \mathbf{v}_{Y_t} le bruit des observations supposé suivre une distribution normale (gaussienne) centrée et de covariance \mathbf{Q}_{Y_t} : $\mathbf{v}_{Y_t} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{Y_t})$ et $\mathbf{Q}_{Y_t} = \sigma_{Y_t}^2 \mathbf{I}_{N \times N}$ avec $\mathbf{I}_{N \times N}$ la matrice identité de taille N . Ainsi, le vecteur \mathbf{Y}_t suit une loi gaussienne non-centrée $\mathbf{Y}_t \sim \mathcal{N}(\boldsymbol{\mu}_{Y_t}, \mathbf{Q}_{Y_t})$ avec $\boldsymbol{\mu}_{Y_t} = h(\mathbf{X}_t) + \mathbf{E}_t$.

Pour une meilleure analyse de l'impact des défauts sur l'estimation du vecteur d'état, l'équation 3.4 est linéarisée autour d'un état initial \mathbf{X}_0 :

$$\Delta \mathbf{Y}_t = \mathbf{H}_t \Delta \mathbf{X}_t + \mathbf{E}_t + \mathbf{v}_{Y_t}, \quad (3.5)$$

avec \mathbf{H}_t la matrice jacobienne du modèle d'observation autour du point de fonctionnement, $\Delta \mathbf{Y}_t = \mathbf{Y}_t - \mathbf{Y}_0 = \mathbf{Y}_t - h(\mathbf{X}_0)$ et $\Delta \mathbf{X}_t = \mathbf{X}_t - \mathbf{X}_0$. On a ainsi $\Delta \mathbf{Y}_t \sim \mathcal{N}(\boldsymbol{\mu}_{\Delta Y_t}, \mathbf{Q}_{Y_t})$ avec $\boldsymbol{\mu}_{\Delta Y_t} = \boldsymbol{\mu}_{Y_t} - \mathbf{Y}_0$.

Si $N > m$, le système 3.5 est résolu de façon à minimiser l'erreur aux moindres carrés :

$$\Delta \hat{\mathbf{X}}_t = (\mathbf{H}_t^\top \mathbf{H}_t)^{-1} \mathbf{H}_t^\top \Delta \mathbf{Y}_t = \mathbf{H}_t^+ \Delta \mathbf{Y}_t, \quad (3.6)$$

où \mathbf{H}_t^+ représente la pseudo-inverse de la matrice \mathbf{H}_t . Nous avons omis la matrice \mathbf{Q}_{Y_t} pour simplifier l'équation 3.6.

Les résidus correspondent aux différences entre les observations et les estimations à un instant donné à partir d'un état connu. Ils sont souvent utilisés en statistiques comme indicateurs de défaut dans les mesures. Le résidu associé à une observation donnée s'exprime sous la forme :

$$\begin{aligned} \mathbf{r}_t &= \Delta \mathbf{Y}_t - \mathbf{H}_t \Delta \hat{\mathbf{X}}_t, \\ &= \Delta \mathbf{Y}_t - \mathbf{H}_t \mathbf{H}_t^+ \Delta \mathbf{Y}_t, \\ &= (\mathbf{I} - \mathbf{H}_t \mathbf{H}_t^+) \Delta \mathbf{Y}_t, \\ &= \mathbf{S} \Delta \mathbf{Y}_t. \end{aligned} \quad (3.7)$$

$\Delta \mathbf{Y}_t$ étant une variable aléatoire gaussienne, les résidus sont également gaussiens $\mathbf{r}_t \sim \mathcal{N}(\boldsymbol{\mu}_{r_t}, \mathbf{Q}_{r_t})$ avec $\boldsymbol{\mu}_{r_t} = \mathbf{S} \boldsymbol{\mu}_{\Delta Y_t}$ et $\mathbf{Q}_{r_t} = \mathbf{S} \mathbf{Q}_{Y_t} \mathbf{S}^\top$.

Si le vecteur des défauts \mathbf{E}_t est nul, alors les résidus suivent une distribution centrée :

$$\begin{aligned} \boldsymbol{\mu}_{r_t} &= \mathbf{S} \boldsymbol{\mu}_{Y_t}, \\ &= (\mathbf{I} - \mathbf{H}_t \mathbf{H}_t^+) \mathbf{H}_t \Delta \mathbf{X}_t, \\ &= \mathbf{H}_t \Delta \mathbf{X}_t - \mathbf{H}_t \Delta \mathbf{X}_t = 0. \end{aligned} \quad (3.8)$$

En revanche, si \mathbf{E}_t est non nul, la moyenne des résidus est non nulle :

$$\begin{aligned}\boldsymbol{\mu}_{r_t} &= \mathbf{S} (\mathbf{H}_t \Delta \mathbf{X}_t + \mathbf{E}_t), \\ &= \mathbf{S} \mathbf{E}_t.\end{aligned}\quad (3.9)$$

Les résidus sont très sensibles aux *outliers*, multi-trajets ou réceptions NLOS pour notre application, via la matrice \mathbf{S} . Notons que dans le cas où les mesures sont affectées par des *outliers* sans que cela soit détectable, on parle de compensation d'erreurs. Pour plus de détails, le lecteur pourra se référer à [Kaplan and Hegarty, 2006].

3.3.2.2 Test d'hypothèses

Un test d'hypothèses consiste à : soit accepter l'hypothèse \mathcal{H}_0 , soit la rejeter, ce qui équivaut à accepter l'hypothèse \mathcal{H}_1 . Par exemple, si la première hypothèse correspond au cas où il n'existe pas de défaut dans le jeu de données et/ou les variables suivent une loi gaussienne centrée en zéro, alors une loi gaussienne non-centrée sera associée à la deuxième hypothèse. Classiquement, le test défini pour le cas de la détection de mesures PR anormales ou erronées est :

$$\begin{aligned}\mathcal{H}_0 &: \boldsymbol{\mu}_{Y_t} = h(\mathbf{X}_t), \\ \mathcal{H}_1 &: \boldsymbol{\mu}_{Y_t} = h(\mathbf{X}_t) + \mathbf{E}_t.\end{aligned}\quad (3.10)$$

Une fois les hypothèses définies, il s'agit de choisir une variable représentant les mesures et un seuil de décision *a priori*. Généralement, un test sur la somme des carrés des résidus normalisés par leur variance qu'on appelle NSSE (*Normalized Sum Squared Errors*) est effectué. Soit σ_{r_i} l'écart-type du résidu r_i de chaque mesure. La mesure NSSE, notée \mathbf{R} , est :

$$\mathbf{R} = \sum_{j=1}^N \left(\frac{r_j}{\sigma_j} \right)^2. \quad (3.11)$$

En l'absence de défaut, la mesure NSSE suit une distribution du χ^2 à $N - m$ degrés de liberté, notée χ_{N-m}^2 . Dans le cas de notre application GPS, le nombre de variables est $m = 4$. En présence de défauts, la NSSE suit une distribution de χ^2 non-centrée, noté χ_{N-4, δ_T}^2 , où δ_T est le paramètre dit de 'non-centralité'. Les hypothèses testées sont alors :

$$\begin{aligned}\mathcal{H}_0 &: \mathbf{R} \sim \chi_{N-4}^2, \\ \mathcal{H}_1 &: \mathbf{R} \sim \chi_{N-4, \delta_T}^2.\end{aligned}\quad (3.12)$$

La densité de probabilité d'une loi du χ^2 à k degrés de liberté est définie par :

$$\begin{aligned}f_{\chi_k^2}(x) &= \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}, \\ \Gamma(k) &= \int_0^{\infty} t^{k-1} e^{-t} dt.\end{aligned}\quad (3.13)$$

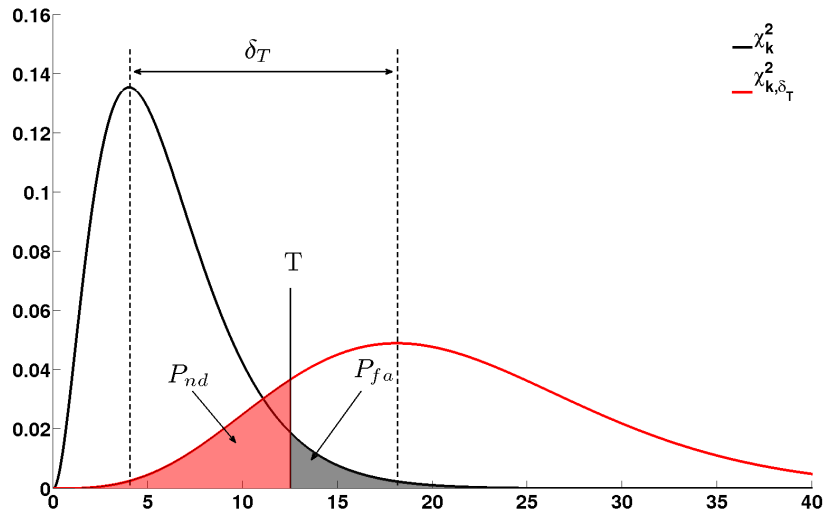


FIGURE 3.2 – Test du χ^2 avec deux distributions : centrée χ_k^2 et non-centrée χ_{k, δ_T}^2 avec $k = N - 4 = 6$.

La densité de probabilité d'une loi du χ^2 non-centrée à k degrés de liberté est définie par :

$$f_{\chi_{k, \delta_T}^2}(x) = \frac{e^{-\frac{x+\delta_T}{2}}}{2^{\frac{k}{2}}} \sum_{j=0}^{\infty} \frac{\delta_T^j x^{j-1+\frac{k}{2}}}{2^{2j} j! \Gamma(j+\frac{k}{2})}. \quad (3.14)$$

La figure 3.2 représente deux distributions du χ^2 à 6 degrés de liberté, respectivement centrée et non-centrée de paramètre δ_T . Le seuil de décision T est ajusté de façon à respecter la probabilité de fausse alarme P_{fa} (en noir). La probabilité de non-détection P_{nd} est représentée par la région sous la courbe à droite du seuil (en rouge). Ces probabilités peuvent être exprimées sous la forme :

$$\begin{aligned} P_{nd} &= P(R < T | \mathcal{H}_1) = \int_0^T \frac{e^{-\frac{x+\delta_T}{2}}}{2^{\frac{k}{2}}} \sum_{j=0}^{\infty} \frac{\delta_T^j x^{j-1+\frac{k}{2}}}{2^{2j} j! \Gamma(j+\frac{k}{2})} dx, \\ p_{fa} &= P(R > T | \mathcal{H}_0) = \int_T^{\infty} \frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} dx. \end{aligned} \quad (3.15)$$

Il est à noter que ce sont souvent les probabilités de fausse alarme et de non-détection qui sont imposées, et que le seuil T et le paramètre de non-centralité s'en déduisent. Ces probabilités sont définies à partir des exigences en termes d'intégrité, pour le cas du GPS, ou d'un niveau de protection PL (*Protection Level*) requis dans les applications critiques telles que l'aviation.

Présentons à présent quelques applications de ces tests d'hypothèses pour détecter les défauts et améliorer la localisation.

3.3.2.3 RAIM + FDE

Plusieurs algorithmes ont été proposés pour détecter les ‘fautes’ dans les mesures, particulièrement en aviation civile. Ils exploitent généralement la redondance attendue (supposée) des mesures pour détecter un à plusieurs *outliers*. L’algorithme RAIM (*Receiver Autonomous Integrity Monitoring*) [Brown, 1987] est implanté dans les récepteurs GPS pour la détection de fautes (pannes de satellites) en utilisant les mesures du récepteur, et optionnellement des capteurs additionnels.

Le RAIM a pour objet de détecter une défaillance ou une anomalie dans les mesures et de protéger l’utilisateur d’erreurs de position excessives, notamment sur les composantes horizontales [Parkinson, 1996]. Pour son fonctionnement, le RAIM nécessite l’acquisition d’un minimum de cinq mesures satellites. La détection d’une mesure erronée se traduit par l’envoi d’une alerte par l’algorithme indiquant que la solution fournie par le récepteur n’est pas utilisable [Hewitson and Wang, 2006].

La technique FDE (*Fault Detection and Exclusion*) est une étape qui vient en aval de la méthode RAIM. Elle consiste à estimer la mesure défaillante parmi les N mesures acquises. La technique FDE est un cas particulier des méthodes d’intégrité reposant sur une démarche en trois étapes : *Detection, Identification, Adaptation* (DIA) [Teunissen, 1990]. Cette technique nécessite pour son fonctionnement un minimum de six satellites. L’identification de la mesure défaillante permet de l’exclure de l’ensemble des mesures destinées à estimer la position du récepteur.

Un calcul simple pour déterminer la ou les mesures défaillantes est d’appliquer un w -test [Baarda, 1968] sur les résidus et d’identifier les mesures aberrantes comme étant celles qui ont les coefficients w_i les plus élevés :

$$w_i = \frac{r_i}{\sigma\sqrt{S_{ii}}}, \quad (3.16)$$

où les termes S_{ii} de l’équation 3.16 représentent les éléments de la diagonale de la matrice S .

Contrairement aux applications en aviation civile, où le seul risque est qu’une importante défaillance du système GPS se produise, en milieu urbain un récepteur GNSS est sujet à de multiples phénomènes dus notamment à la réception de signaux en non-visibilité directe [Le Marchand et al., 2008]. L’hypothèse du RAIM sur la présence d’une seule mesure erronée par instant est donc difficilement applicable en milieu contraint, où la probabilité de non-détection sera élevée.

Différentes améliorations de la méthode FDE ont été apportées. La technique *Forward Backward* (FDE) [Kuusniemi, 2005] consiste à exclure des mesures identifiées comme erronées de façon récursive, tout en permettant de réintroduire des mesures précédemment exclues (par erreur du fait qu’il est difficile d’identifier les mesures défaillantes). La combinaison d’un modèle de déplacement basé sur le filtre de Kalman avec la technique FDE a été proposée par [Hewitson and Wang, 2007]. Ce modèle a pour but d’améliorer l’intégrité de la solution en intégrant des mesures proprioceptives au processus de fusion de contrôle d’intégrité. [Yang et al., 2013] propose comme FDE optimal une méthode pour calculer correctement les probabilités de non-détection et de fausse exclusion, ou typiquement des erreurs dites de type III.

3.3.2.4 Test à partir du rapport de vraisemblance

Le test GLR (*Generalized Likelihood Ratio*) est un test statistique introduit par [Willsky and Jones, 1976]. Cette méthode est largement utilisée par la communauté du traitement du signal et de l'automatique pour le diagnostic de fautes. [Basseville et al., 1993] et [Gustafsson and Gustafsson, 2000] donnent une présentation générale des différentes variantes de la méthode GLR et de ses applications. Ce test a été initialement développé comme une extension de filtre de Kalman pour détecter les changements brusques dans les systèmes dynamiques [Willsky and Jones, 1976]. Le GLR opère au niveau des innovations du filtre de Kalman et tire profit du fait qu'une erreur de mesure sur ces innovations peut s'exprimer d'une manière analytique. [Giremus and Escher, 2007] a proposé une adaptation du test GLR au contrôle autonome d'intégrité appliquée à l'aviation civile (AAIM).

Le principe du GLR est de comparer deux fonctions de vraisemblance associées à l'absence (hypothèse H_0) ou la présence (hypothèse H_1) de changements brusques dans les mesures. Le rapport des vraisemblances est un indicateur de la présence ou l'absence de changements brusques. Les deux hypothèses sont alors :

- H_0 : aucun saut détecté jusqu'au temps t ,
- H_1 : un saut d'une amplitude ν apparaît à un instant $t \leq k$.

Le log du rapport de vraisemblance de ces deux hypothèses s'écrit :

$$l_t(k, \nu) = \log \frac{P(\mathbf{Z}_{1:t} | H_1(k, \nu))}{P(\mathbf{Z}_{1:t} | H_0)}, \quad (3.17)$$

avec $\mathbf{Z}_{1:t}$ l'ensemble des observations. Le niveau de changement ν est un paramètre inconnu. Une estimation de ce paramètre est réalisée par l'estimateur de maximum de vraisemblance de la densité de probabilité $P(\mathbf{Z}_{1:t} | H_1(k, \nu))$.

Une autre méthode, introduite par [Gustafsson and Gustafsson, 2000], consiste à éliminer le paramètre ν en marginalisant la fonction de vraisemblance sous l'hypothèse H_1 appelée MLRT (*Marginalized Likelihood Ratio Test*). Une méthode de détection et de correction des multi-trajets en localisation GPS a été proposée par [Cheng et al., 2016] utilisant la MLRT.

3.3.3 Méthodes robustes pour la localisation

3.3.3.1 Estimateurs robustes

Quand des observations sont affectées par des erreurs systématiques ou que le modèle utilisé est approximatif, l'estimateur des moindres carrés fournit une solution très éloignée de la réalité. Ces erreurs grossières mènent à des erreurs en localisation très large. Une des approches populaires est alors d'utiliser une estimation robuste. L'objectif d'un estimateur robuste n'est pas de détecter les données erronées mais d'en atténuer l'influence sur l'estimation des paramètres. Les données aberrantes dans les observations seront en effet associées à un poids faible au cours de la procédure d'estimation.

Tout le problème est en effet de trouver le meilleur estimateur robuste pour une application donnée qui permet d'atténuer au maximum l'effet des *outliers* sans pour autant nuire à la qualité de l'estimation de la solution (notamment liée au nombre d'observations). Par exemple, la majorité des estimateurs par régression linéaire sont très sensibles aux *outliers*, même en la présence d'une seule donnée aberrante.

Nous allons détailler quelques estimateurs robustes utilisés en statistique.

Danish Method :

La méthode Danish [Krarup et al., 1980] est utilisée dans les applications géodésiques depuis les années 80 [Wieser and Brunner, 2002]. Elle consiste à estimer itérativement les poids accordés à chaque mesure en utilisant une matrice de covariance d'erreur *a priori*. Plus précisément, les résidus obtenus par une première résolution par moindres carrés sont examinés. Le poids des résidus supérieurs à un seuil sont diminués de façon exponentielle. La technique est ainsi répétée jusqu'à convergence de l'algorithme.

Soit σ_{k,r_i}^2 le poids associé au résidu r_i à l'itération k , la mise à jour du poids associé au résidu r_i à l'itération $k + 1$ est donnée par :

$$\sigma_{k+1,r_i}^2 = \sigma_{k,r_i}^2 \times \begin{cases} 1 & \text{si } |\omega_{i,k}| \leq T_{r_i} \\ e\left(-\frac{\omega_{i,k}}{T_{r_i}}\right) & \text{si } |\omega_{i,k}| > T_{r_i} \end{cases} . \quad (3.18)$$

Le seuil T_{r_i} est généralement compris entre 2 et 3 [Knight and Wang, 2009], et la valeur $\omega_{i,k} = \left| \frac{r_i}{\sqrt{C_{r_i,1}}} \right|$, où $C_{r_i,1}$ est la $i^{\text{ème}}$ composante diagonale à la première itération de la matrice C_r représentant la matrice de covariance des résidus qui s'écrit :

$$C_r = Q_{Y_t} - H_t \left(H_t^T Q_{Y_t}^{-1} H_t \right)^{-1} H_t^T . \quad (3.19)$$

La matrice de covariance des résidus est statique et ne change pas au cours des itérations, ce qui conditionne les performances de la méthode au choix *a priori* des valeurs de cette matrice de covariance des mesures. La méthode *Danish* est une méthode robuste aux *outliers* qui peut dégénérer en une méthode d'élimination de mesures défaillantes dans le cas où les poids associés aux résidus convergent vers 0 [Kuusniemi et al., 2004]. Cette dernière a montré son efficacité d'un point de vue pratique et son intérêt en temps de calcul mais c'est une technique empirique sans réel fondement théorique en statistique. De plus, comme la plupart des méthodes robustes, elle n'est efficace que si le nombre d'*inliers* est supérieur au nombre d'*outliers*.

Filtre médian :

En statistique, la médiane minimise le biais maximum dû à des mesures aberrantes [Huber, 1987].

Elle représente donc un choix intéressant lorsque le contrôle du biais est important. L'échantillon médian est le plus simple estimateur disponible avec un point de rupture élevé (défini comme le plus petit nombre d'outliers qui peut faire chuter la qualité de l'estimateur). Par conséquent, il est souvent utilisé comme point de départ avant toute procédure itérative visant à obtenir une estimation plus précise. En effet, un mauvais point de départ peut entraîner la convergence vers une solution locale éloignée de la vraie solution.

M-estimateurs :

Les M-estimateurs sont parmi les méthodes les plus utilisées pour traiter les données aberrantes. Dans la méthode des moindres carrés, l'objectif est de minimiser la somme des carrés des résidus $\sum_{i=1}^n r_i^2$. Cette optimisation est assez sensible à la présence d'outliers dans les observations. Le principe d'un M-estimateur est de remplacer le carré du résidu r^2 par une fonction $\phi(r)$, permettant d'atténuer l'effet des outliers sur l'estimation des paramètres.

Les M-estimateurs sont une généralisation des estimateurs de maximum de vraisemblance (MLE pour *Maximum Likelihood Estimator*) où la fonction de vraisemblance est remplacée par la fonction de pénalité $\phi(r)$. En effet, la performance des MLE dépend fortement de la distribution supposée des données. La fonction $\phi(r)$ est conçue de telle sorte que les résidus générés par les outliers perdent leur influence sur l'estimation des paramètres.

Le M-estimateur, proposé par [Huber, 1964], consiste à résoudre l'équation :

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \phi(r_i, \theta), \quad (3.20)$$

où $\hat{\theta}$ est la solution optimale du problème. La fonction de pénalité $\phi(r)$ est symétrique, positive et possède un minimum unique en zéro. Le tableau 3.1 énumère les M-estimateurs les plus connus [Valero and Moreno, 2005]. Les fonctions de pénalité peuvent être différentes suivant les valeurs des résidus. Par exemple, l'estimateur d'Huber partitionne les résidus en deux groupes : les résidus inférieurs à un seuil k et les résidus supérieurs à cette valeur. Le seuil k est un paramètre *a priori* ou à apprendre et doit être adapté suivant la distribution des données. Pour une distribution gaussienne, le seuil est égal à $1.345\hat{\sigma}$ [Knight and Wang, 2009] où $\hat{\sigma}$ est un paramètre *a posteriori* estimé à partir des résidus tel que :

$$\hat{\sigma} = \frac{1}{0.6745} \text{Mediane}(|\mathbf{r} - \text{Mediane}(\mathbf{r})|). \quad (3.21)$$

Au lieu de résoudre l'équation 3.20 directement, une procédure itérative appelée IRLS (*Iterative Reweighted Least-Square*) est appliquée pour estimer itérativement les poids associés à chaque observation. La fonction de pondération équivalente est déduite de la fonction de pénalité $\phi(r)$ par :

$$w(r) = \frac{1}{r} \psi(r) = \frac{1}{r} \frac{d}{dr} \phi(r). \quad (3.22)$$

La fonction d'influence $\psi(r)$ décrit la sensibilité de l'estimation globale sur les données avec une

Estimateur	Domaine	$\phi(r)$	$w(r)$
$Norm_1$	\mathbb{R}	$ r $	$1/ r $
$Norm_1 - Norm_2$	\mathbb{R}	$2\sqrt{1+r^2/2}$	$1/\sqrt{1+r^2/2}$
$Norm_p$	\mathbb{R}	$ r ^p / p$	$ r ^{p-2}$
Cauchy	\mathbb{R}	$(k^2/2) \log(1 + (r/k)^2)$	$1/(1 + (r/k)^2)$
Fair	\mathbb{R}	$k^2 (r /k - \log(1 + r /k))$	$1/(1 + r /k)$
Huber	$ r \leq k$	$r^2/2$	1
	$ r > k$	$k(r - k/2)$	$k/ r $
Tuckey	$ r \leq k$	$k^2/6 (1 - (1 - (r/k)^2))^3$	$(1 - (r/k)^2)^2$
	$ r > k$	$k^2/6$	0
Andrew	$ r \leq 1$	$(1/\pi^2) (1 - \cos(\pi r))$	$(1/\pi r) \sin(\pi r)$
	$ r > 1$	$2/\pi^2$	0
Welsch	\mathbb{R}	$(k^2/2) (1 - e^{-(r/k)^2})$	$e^{-(r/k)^2}$
Geman-McClure	\mathbb{R}	$r^2 / (2(1+r^2))$	$1/(1+r^2)^2$

TABLE 3.1 – Exemples de M-estimateurs.

erreur r . Les fonctions de pondération pour chaque estimateur sont données dans la dernière colonne du tableau 3.1.

La procédure IRLS se déroule comme suit :

- Estimer les paramètres du modèle par moindres carrés utilisant des poids unitaires sur chaque observation ;
- À partir des résidus, calculer les valeurs des poids suivant la fonction de pondération de l'estimateur choisi ;
- Effectuer une nouvelle estimation des paramètres du modèle en utilisant les nouveaux poids ;
- Répéter le processus à partir de l'étape 2 jusqu'à convergence.

3.3.3.2 Filtrage robuste

Outlier Robust Kalman Filter (ORKF) :

Dans la version classique du filtre de Kalman introduit au chapitre 2, le bruit affectant les mesures est considéré comme un bruit blanc gaussien. Les paramètres des matrices de covariance des bruits de mesures sont souvent constants et fixés *a priori* suivant les spécifications des capteurs. Ce type de modélisation n'est pas robuste aux observations qui sont de distribution non gaussienne.

Plusieurs approches ont été proposées pour améliorer la robustesse du filtre de Kalman aux *outliers* et aux bruits non-gaussiens. Elles sont généralement basées sur des modèles de bruit à queue de distribution lourde [Sorenson and Alspach, 1971] ou des t-distributions [Meinhold and Singpurwalla, 1989].

Algorithm 2: Algorithme du filtre de Kalman robuste aux *outliers*.

```

1  $m \leftarrow F_{k-1}x_{k-1}$ .
2  $\Sigma \leftarrow F_{k-1}P_{k-1}F_{k-1}^\top + Q_{k-1}$ .
3  $\Theta \leftarrow R_k$ .
4 repeat
5    $K_k \leftarrow (H_k^\top \Sigma H_k + \Theta)^{-1} H_k^\top \Sigma$ .
6    $x_k \leftarrow m + K_k^\top (y_k - H_k^\top m - d)$ .
7    $P_k \leftarrow K_k^\top \Theta K_k + (I - K_k^\top H_k^\top) \Sigma (I - H_k K_k)$ .
8    $\delta \leftarrow y_k - H_k^\top x_k - d$ .
9    $\Theta \leftarrow \frac{sR_k + \delta \delta^\top + H_k^\top P_k H_k}{s+1}$ .
10 until convergence;
11 En sortie :  $\{\hat{x}_k, P_k\}$ 

```

Récemment, dans [Ting et al., 2007a] [Ting et al., 2007b], les auteurs ont introduit un filtre de Kalman capable d'effectuer un suivi temps-réel en présence de données aberrantes. Ceci se fait de manière non-supervisée, et ne nécessite pas de paramétrage manuel. L'apprentissage est réalisé sur les poids associés à chaque observation en utilisant la méthode EM (*Expectation-Maximization*). De façon similaire, [Särkkä and Nummenmaa, 2009] ont défini un modèle heuristique pour les paramètres du bruit afin de suivre simultanément l'état du système ainsi que le niveau du bruit des capteurs.

Une généralisation des précédentes approches a été proposée par [Agamennoni et al., 2011] et [Agamennoni et al., 2012]. Le modèle proposé ne fait aucune hypothèse sur le niveau de bruit (non constant et non gaussien), autorisant ainsi une variation dans le temps du niveau de bruit et des distributions à queues plus lourdes que la distribution gaussienne. Cette approche, présentée dans l'algorithme 2, permet de modifier de façon adaptative la matrice de covariance des bruits d'observations pour atténuer l'effet des *outliers* sur l'estimation d'état. L'ORKF a été appliqué pour la localisation d'un véhicule en milieu urbain avec un GPS (les données considérées alors sont les solutions de localisation fournies par le GPS) et une centrale inertielle.

Filtre particulière :

Dans l'estimation d'état dynamique non-linéaire avec une hypothèse d'un bruit non gaussien, le filtre particulière, et ses variantes, se sont imposés comme des méthodes efficaces pour succéder au traditionnel filtre de Kalman [Arulampalam et al., 2002]. Le filtre particulière est une méthode de Monte Carlo séquentielle qui consiste à représenter la densité de probabilité *a posteriori* en un ensemble de particules pondérées par un poids qui détermine l'importance d'une particule parmi l'ensemble des échantillons.

L'objectif d'un filtre est de calculer la probabilité conditionnelle de l'état x_k sachant les mesures $y_{0:k}$. Cette probabilité est approximée par [Doucet et al., 2000] :

$$p(x_k | y_{0:k}) = \sum_{i=1}^N w_k^i \delta(x_k = x_k^i), \quad (3.23)$$

avec $\delta()$ la fonction Dirac, w_k le poids associé à chaque particule i et N le nombre total d'échantillons. Le poids w_k peut-être estimé à l'aide de la technique d'échantillonnage d'importance [Doucet, 1998].

En plus des étapes de prédiction et d'estimation classiques pour une technique de filtrage, le filtre particulaire le plus populaire, le filtre séquentiel d'importance avec rééchantillonnage (SIR), intègre une étape d'élimination de particules.

Dans l'étape de prédiction, le but est de générer les particules de l'étape k sachant la loi de transition $p(x_k | x_{k-1})$ qui représente le modèle d'évolution du système d'état. L'étape d'observation, ou de correction consiste à prendre en compte les observations pour calculer les poids des particules w_k^i . L'estimation peut se faire récursivement :

$$w_k^i \propto w_{k-1} \frac{p(y_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{0:k-1}^i, y_{1:k})}, \quad (3.24)$$

avec $q()$ la densité d'importance qui génère les particules x_k . Pour simplifier l'estimation du poids, la densité d'importance $q(x_k^i | x_{0:k-1}^i, y_{1:k})$ est choisie égale à la loi de transition $p(x_k | x_{k-1})$. Le poids est donc proportionnel au poids de l'itération précédente et à la vraisemblance $p(y_k | x_k^i)$. L'étape de rééchantillonnage consiste à éliminer les particules de poids très faible. Cette étape est optionnelle et n'est appelée que si le nombre d'échantillons effectifs N_{eff} est inférieur à un seuil qui dépend de l'amplitude de la variance des observations en fonction de l'état :

$$N_{eff} = \frac{1}{\sum_{i=1}^N (w_k^i)^2}. \quad (3.25)$$

Le filtre SIR peut être inefficace et est assez sensible aux données aberrantes. [Pitt and Shephard, 1999] présente l'ASIR (*Auxillary SIR*), une amélioration du SIR, qui introduit une variable auxiliaire dans la densité d'importance $q(x_k, j | y_{1:k})$ permettant à l'échantillonnage de s'effectuer sur la paire $\{x_k^i, j^i\}$ telle que j^i se réfère à l'indice de la particule i à l'étape $k-1$. Cette technique permet de conditionner la propagation des particules sachant leur poids à l'itération précédente. Néanmoins, l'ASIR fournit des performances dégradées quand la variance des bruits de mesures est importante [Arulampalam et al., 2002].

En ce qui concerne les applications à la détection de multi-trajets en localisation GPS, une approche utilisant le filtre particulaire pour gérer/estimer les multi-trajets est proposée par [Giremus et al., 2007]. L'auteur utilise un filtre Rao-Blackwellisé de sorte que la détection des multi-trajets soit réalisée par le filtre particulaire, alors que l'estimation des paramètres de navigation et les erreurs de multi-trajet sont déterminées analytiquement par un banc de filtres de Kalman étendu. Une comparaison avec la méthode GLR et la méthode des modèles multiples a été réalisée sur des données simulées. [Rabaoui et al., 2012] propose une estimation conjointe des erreurs d'observations par un mélange de

processus de Dirichlet et de l'état caché du système par une filtre particulière Rao-Blackwellisé. La méthode a été testée avec des données réelles en milieu urbain en utilisant les mesures de pseudo-distances.

3.3.3.3 Méthodes d'estimations par consensus

Méthode RANSAC :

La plus populaire des méthodes par consensus est probablement le RANSAC (*RANdom SAMple Consensus*). C'est une méthode d'estimation itérative robuste aux *outliers* proposée par FISCHLER dans les années 80 [Fischler and Bolles, 1981]. Elle est dérivée de la méthode MAX-CONSENSUS, mais à la différence de cette dernière, le RANSAC n'effectue pas une recherche exhaustive sur tous les d -uplets pour trouver le modèle idéal (un d -uplet est un ensemble ordonné de d observations). Ainsi, RANSAC est non-déterministe au sens où nous échantillons de manière stochastique l'espace des solutions. Un nombre d'itérations est fixé, ou déterminé théoriquement suivant une probabilité p connue. Soit p la probabilité que le RANSAC sélectionne aléatoirement, en une quelconque itération, uniquement des *inliers* parmi les N observations disponibles dans le jeu de données. Alors, ces *inliers* permettront une estimation correcte des paramètres du modèle. Ainsi p représente la probabilité que le RANSAC fournisse une estimation exacte. Soit p' la probabilité qu'une observation soit correcte, alors la probabilité qu'il y ait au moins une donnée aberrante parmi les N observations sur chaque tirage parmi k itérations est $(1 - p'^N)^k$. Donc : $1 - p = (1 - p'^N)^k$. Le nombre minimum d'itérations pour obtenir un résultat correct s'écrit selon les deux probabilités p et p' :

$$k_{min} = \frac{\log(1 - p)}{\log(1 - p'^N)}. \quad (3.26)$$

Le nombre d'itérations k est donc un paramètre variable de l'algorithme et dépend de la probabilité p généralement inconnue. Toutefois, pour une probabilité p fixée, le paramètre k est mis à jour à chaque nouveau consensus.

L'algorithme RANSAC repose sur quatre étapes effectuées de façon itérative, comme présenté dans l'algorithme 3. La mise en œuvre de cet algorithme nécessite d'adapter les points suivants :

- **La génération d'hypothèses** : Elle consiste à effectuer le tirage aléatoire des d -uplets, à partir du nombre minimum d'observations d pour estimer les paramètres du modèle. Chaque tirage est une hypothèse \mathcal{H}_i parmi l'ensemble des hypothèses possibles \mathcal{H} pour un modèle \mathcal{M} .
- **Mesure de consensus** : il s'agit de la métrique qui va mesurer les erreurs de chaque échantillon-observation par rapport au modèle de l'hypothèse \mathcal{H}_i .
- **Critère de validation** : il est utilisé pour sélectionner les échantillons dont l'erreur est inférieure à un seuil de précision (δ).
- **Critère d'arrêt** : généralement choisi en fonction du nombre d'itérations k .

Prenons l'exemple de la détection d'une droite dans un ensemble de points 2D de taille N (voir la figure 3.3a). Le but est d'estimer la pente a et le biais b de la droite selon l'équation $y = ax + b$, soit deux paramètres à estimer : $d = 2$ (d -uplet) car nous avons 2 inconnues (a, b). Nous allons illustrer une itération de l'algorithme RANSAC en commençant par la ligne 3 avec le tirage aléatoire du 2-uplet (à partir de 2 points tirés aléatoirement parmi les N points). Nous obtenons une première estimation des

Algorithm 3: Algorithme RANSAC ; entrées : $\mathcal{O} = \{o_1, \dots, o_N\}$ l'ensemble des observations, δ le seuil de précision, la mesure de consensus utilisée notée ici $\|\cdot\|$ et k le nombre maximum de tirages aléatoires ; sortie : \mathcal{D} l'ensemble des *inliers* et les paramètres du modèle de l'hypothèse \mathcal{H} retenue.

```

1 Initialiser  $i = 0$  et  $\mathcal{D} = \emptyset$ .
2 while  $i < k$  do
3   Tirer aléatoirement  $d$  observations parmi les  $N$ , et estimer le  $d$ -uplet (paramètres du modèle)
   noté hypothèse  $\mathcal{H}_i$  que les observations tirées soient des inliers
4   Sélectionner l'ensemble des inliers tels que  $\mathcal{D}_{\mathcal{H}_i} = \{\|\mathcal{H}_i, o_i\| < \delta\}$ 
5   if  $|\mathcal{D}_{\mathcal{H}_i}| > |\mathcal{D}|$  then
6      $\mathcal{D} = \mathcal{D}_{\mathcal{H}_i}$ 
7      $\mathcal{H} = \mathcal{H}_i$ 
8   end
9    $i = i + 1$ 
10 end

```

paramètres (a, b) sous l'hypothèse notée \mathcal{H}_i que les deux points tirés aléatoirement appartiennent à la droite. Ensuite (ligne 4 de l'algorithme 3), nous sélectionnons les points *inliers* tels que le résidu (ici la mesure de consensus est la valeur du résidu) du point considéré sous l'hypothèse \mathcal{H}_i soit inférieur au seuil δ . Si le nouvel ensemble des *inliers* obtenu à l'itération courante est de taille (cardinal) plus grand que celui de \mathcal{D} , qui est notre meilleur résultat sur les itérations précédentes, alors on actualise \mathcal{D} et l'hypothèse lui correspondant \mathcal{H} . Ainsi l'ensemble final \mathcal{D} correspondra à l'hypothèse qui maximise le nombre d'*inliers* (voir figure 3.3b).

Le RANSAC s'est rapidement imposé dans la communauté des traiteurs d'images pour sa simplicité d'implémentation et son efficacité à traiter et identifier les *outliers*. Une application en GPS a été introduite par [Schroth et al., 2008] pour détecter la présence de défaillances dans les signaux satellites. Le principe du RANSAC a été suivi par l'auteur en tirant aléatoirement 4 satellites parmi N_{sat} . Dans l'expérience rapportée par l'auteur, le nombre de satellites est en moyenne 18, car il considère aussi la constellation GALILEO (par simulation) afin d'améliorer la disponibilité des signaux. Une étape de test du conditionnement et d'estimation du GDOP a été ajoutée pour présélectionner les meilleures configurations de satellites. Le critère de validation dépend d'une probabilité fixée *a priori* sur les erreurs de mesures.

Le RANSAC possède donc deux paramètres critiques, la probabilité p qui limite le nombre d'itérations et qui peut conduire à une solution sous optimale, et le seuil δ qui valide le nombre de mesures cohérentes entre elles. Un mauvais choix de ce seuil de décision peut entraîner une perte de performance importante. Mais si δ est judicieusement choisi, le RANSAC peut atteindre des niveaux de détections allant jusqu'à 50% d'*outliers*. Un paramètre δ sous-estimé (voir figure 3.3c) peut mener à un ensemble \mathcal{D} trop petit, et donc non fiable pour l'estimation. Un paramètre δ sur-estimé (voir figure 3.3d) va inclure des données aberrantes, pouvant significativement dégrader l'estimation de la solution finale.

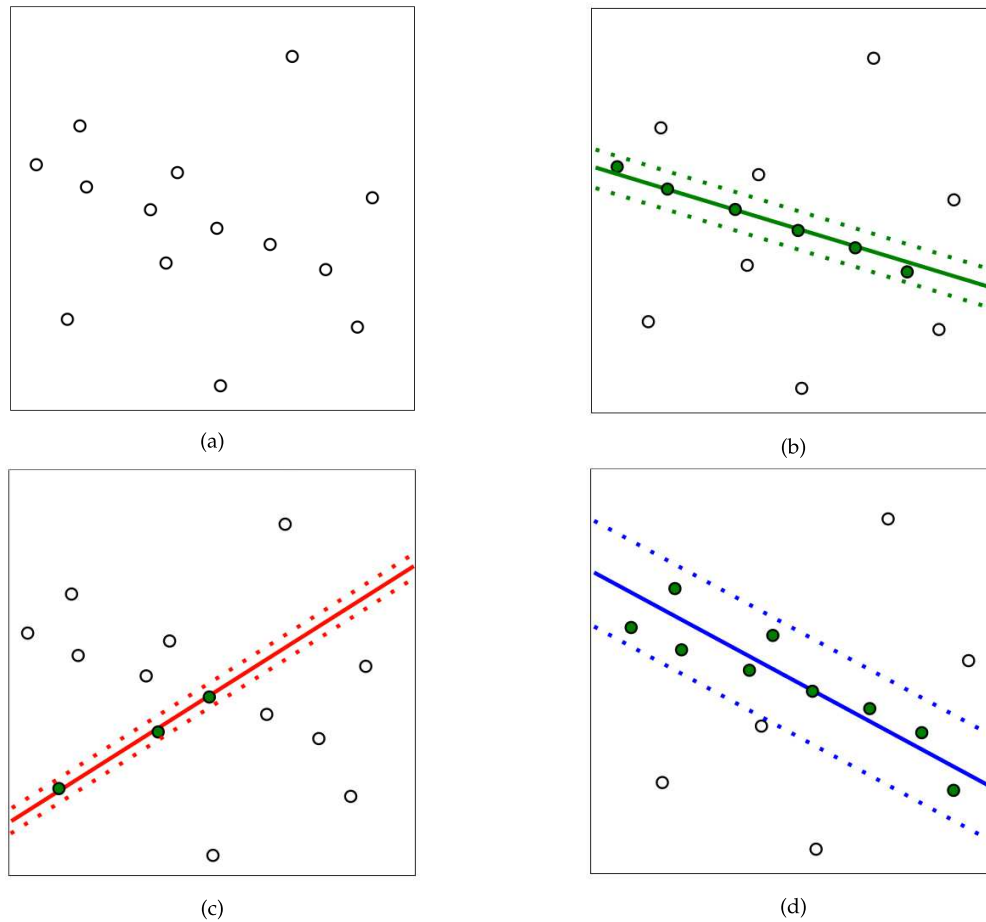


FIGURE 3.3 – Impact du paramètre δ sur l'estimation de la droite par RANSAC [Moulon, 2014] : (a) Ensemble de points, (b) bon choix du δ , (c) δ sous-évalué, (d) δ sur-évalué.

Pour améliorer le temps d'exécution, [Chum and Matas, 2005] a proposé une heuristique effectuant un tirage des échantillons suivant un indice de confiance associé à chaque observation. Pour améliorer la robustesse, [Choi and Medioni, 2009] propose de tester différentes valeurs de δ de façon exhaustive, et ensuite de déterminer la meilleure en estimant la variance des paramètres du modèle. Enfin, pour améliorer la précision, [Chum et al., 2003] propose d'optimiser localement chaque hypothèse sur le point d'être acceptée. Le tirage se fait donc sur les données étiquetées comme *inliers*. L'hypothèse retenue est celle qui aura la plus petite erreur moyenne.

A Contrario RANSAC :

La modélisation *a contrario* est une approche introduite par Desolneux [Desolneux et al., 2000] pour détecter des structures géométriques dans un image sans avoir de connaissance *a priori*. Cette approche repose sur le principe d'Helmholtz qui postule que "toute structure trop régulière pour apparaître par hasard, attire l'attention et devient une perception". Une autre interprétation a été apportée par L. Moisan [Moisan, 2003] en complément du principe d'Helmholtz : "Il est beaucoup plus simple de définir un modèle que l'on souhaite réfuter (typiquement un modèle uniforme) qu'un modèle précis des objets que l'on souhaite détecter".

En théorie de la décision, les tests d'hypothèses nécessitent la définition d'une hypothèse nulle H_0 et une hypothèse H_1 représentant le cas où H_0 est rejetée (voir la section 3.3.2.2). L'approche *a contrario* ne suppose qu'un seul modèle à réfuter H_0 , qu'on appelle modèle de fond ou modèle naïf, qui est défini de façon suffisamment générique pour que l'idée de L. Moisan s'applique. Par exemple, en traitement d'image, pour une image binaire, le modèle de fond est généralement la distribution uniforme qui modélise l'absence de structure formée par les points (pixels à 1 dans l'image binaire). Cette méthode ne contrôle que la probabilité de fausse alarme.

L'approche *a contrario* repose, dans son application, sur la définition de quatre critères [Moulon, 2014] :

- **Un modèle de fond** : aussi appelé modèle naïf (pour rappeler qu'il doit être très basique), il désigne le modèle à réfuter, généralement un bruit aléatoirement distribué. Ce modèle est analogue à l'hypothèse H_0 utilisée dans les tests statistiques (voir la section 3.3.2.2).
- **Une mesure de similarité** : Il s'agit typiquement d'une mesure qui quantifie le degré d'adéquation d'une observation avec le modèle à contredire. Pour nombre d'applications (dont la localisation), ce pourra être les résidus.
- **Une mesure de significativité** : C'est une mesure qui quantifie le degré d'adéquation d'un groupe d'observations sous l'hypothèse du modèle naïf à contredire. Il s'agit de quantifier le fait qu'un ensemble d'observations soit assez significatif pour former une structure détectable.
- **Un critère d'optimisation** : L'optimisation permet de trouver le groupe d'observations qui réfute le plus le modèle de départ suivant la mesure de significativité définie précédemment.

La démarche utilisée ici peut être reformulée de la façon suivante : si nous avons n observations quelconques dont k observations ayant une caractéristique en commun, la question posée est : est-ce que le groupement des k observations parmi n est significatif ? (l'exemple classique des dates d'anniversaires est donné en annexe A.1). Le degré de significativité est mesuré par la 'force' avec laquelle les k observations contredisent le modèle naïf. Par la suite, l'espérance du nombre d'occurrence du groupe des k observations sous l'hypothèse du modèle naïf est appelé 'nombre de fausses alarmes' et est défini par

$$NFA(e) = |E| P(e), \quad (3.27)$$

où e est un événement de E l'ensemble des événements détectables et P la probabilité qu'un événement soit détectable.

Pour illustrer le principe d'Helmholtz, prenons l'exemple de la figure 3.4, extraite de l'article de [Desolneux et al., 2003]. La figure 3.4a présente un jeu de données de 27 points avec alignement parfait de 7 points et une distribution aléatoire de 20 points sur la zone considérée. Ces 20 points représentent du bruit ou des données aberrantes pour la détection de droite. Qualitativement, l'alignement est d'autant plus visible qu'il est peu probable qu'il soit apparu par hasard. Ainsi, sur la figure 3.4b, la détection automatique de la droite est possible car l'alignement est improbable au regard du nombre de points présents sur l'image. En revanche, dans la figure 3.4c où 80 points ont été ajoutés et distribués uniformément sur l'image, un alignement de 7 points parmi l'ensemble n'est pas si improbable qu'il soit détectable. Nous n'avons ici donné que le principe des approches *a contrario* et nous ren-

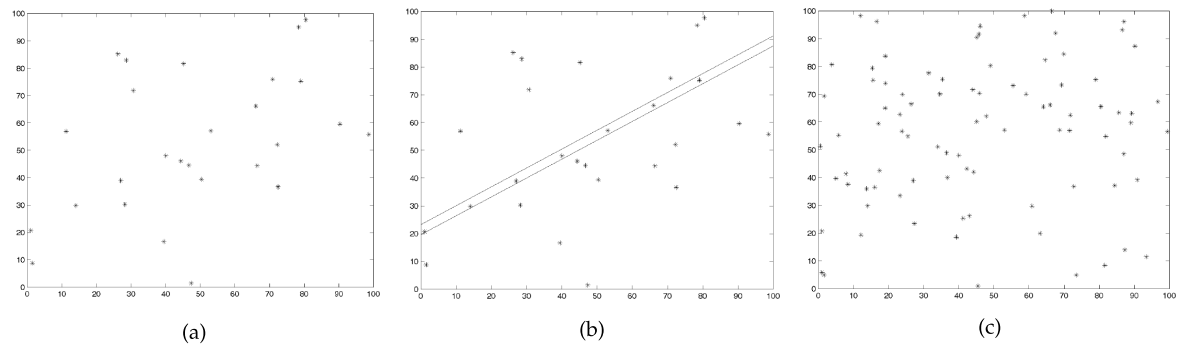


FIGURE 3.4 – Illustration du principe d’Helmholtz sur la détection d’alignement [Desolneux et al., 2003]; (a) alignement de 7 points sur un ensemble augmenté de 20 points uniformément répartis;(b) détection de l’alignement de 7 points;(c) non-détection de l’alignement en ajoutant 80 points tirés aléatoirement, il faudrait alors au minimum 11 points alignés.

voyons le lecteur vers des ouvrages spécialisés pour plus de détails [Desolneux et al., 2007].

L’algorithme AC-RANSAC est un mélange du RANSAC et d’un critère *a contrario*. Cet algorithme permet de pallier le problème du choix du seuil dans le RANSAC. Précisément, [Moisan and Stival, 2004] a remplacé le seuil de décision du RANSAC par une mesure de significativité (Nombre de Fausses Alarmes) et la décision est prise en choisissant le sous-ensemble d’échantillons qui minimise cette mesure. C’est le fait de rechercher un optimum qui permet de s’affranchir du seuil présent dans le RANSAC, mais pour cela il faut un critère qui ne varie pas de façon monotone en fonction du nombre d’inliers considérés. Les auteurs ont alors montré que l’AC-RANSAC (*a contrario* RANSAC) est robuste jusqu’à 90% d’outliers.

L’approche *a contrario* a été appliquée dans différents domaines du traitement d’image. Nous pouvons citer de façon non-exhaustive :

- [Robin et al., 2010] qui applique l’approche *a contrario* à la détection de changement sous-pixélique en imagerie satellite. Un modèle gaussien est utilisé comme modèle naïf à contredire pour représenter les éléments non-structurés.
- [Sabater et al., 2012] qui applique l’approche *a contrario* à la détection de bonne correspondance en stéréovision. Les images utilisées présentent des objets en mouvement qui faussent les résultats et doivent être correctement identifiés. D’une façon plus générale, [Rabin et al., 2010] a proposé une approche *a contrario* pour l’appariement de points clés entre images à mettre en correspondance.
- [Ammar et al., 2013] qui adapte la mesure NFA au problème de détection d’objets dans un contexte de vision embarquée pour détecter des objets inattendus. Deux modèles naïfs sont utilisés successivement d’abord pour détecter des points, puis pour les regrouper.

3.4 Conclusion

Dans ce chapitre, nous avons présenté les principales méthodes de détection et/ou gestion des données aberrantes. Nous nous sommes particulièrement intéressés aux méthodes adaptées aux mesures issues du capteur GPS (les pseudo-distances et les mesures Doppler), qui présentent des erreurs importantes dues aux signaux multi-trajets et aux NLOS.

Parmi les méthodes statistiques détaillées dans ce chapitre, les approches par consensus présentent de réels avantages tant au niveau détection d'*outliers* que pour l'estimation robuste de la solution. L'approche *a contrario* couplée au RANSAC permet de modéliser le bruit non structuré afin d'obtenir les paramètres du modèle et l'ensemble des observations qui réfutent le plus significativement sous l'hypothèse du modèle naïf. Dans notre cas, et dans le chapitre 5, nous allons adapter la méthode *a contrario* au contexte de la détection de données aberrantes dans les mesures de pseudo-distances. Deux modèles naïfs sont testés, à savoir, une distribution uniforme et une distribution gaussienne. Le but est de comparer ces deux modèles en termes de détection d'*outliers* mais aussi de performance en localisation.

Jeux de données et mesures de performance

Sommaire

4.1 Plate-forme expérimentale et jeux de données	49
4.1.1 Capteurs GPS	50
4.1.2 Jeu de données 1	51
4.1.3 Jeu de données 2	51
4.1.4 Jeu de données 3	52
4.2 Mesures de performance	53
4.2.1 Performance de la détection des <i>outliers</i>	54
4.2.2 Performance de la localisation	55
4.3 Conclusion	56

4.1 Plate-forme expérimentale et jeux de données

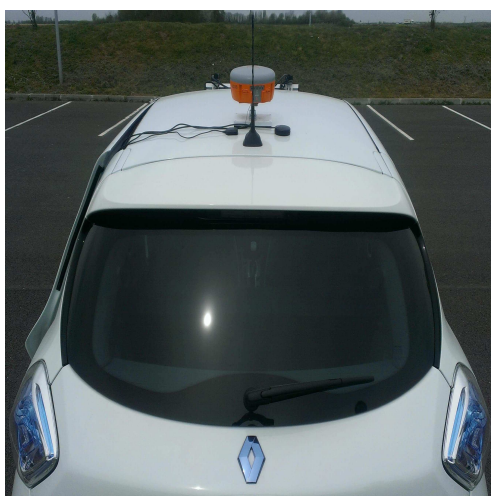
Pour évaluer les algorithmes proposés ultérieurement, trois jeux de données ont été réalisés dans différents environnements (dégagés et contraints) avec des variations de la dynamique du véhicule (en



FIGURE 4.1 – Récepteurs GNSS utilisés durant les expérimentations : (a) récepteur RTK ALTUS APS3, (b) récepteur GARMIN 18x, (c) récepteur UBLOX EVK-5T.

TABLE 4.1 – Spécification des récepteurs GNSS utilisés durant nos expérimentations.

Récepteur	ALTUS APS3	GARMIN 18x	UBLOX EVK-5T
fréquence GPS	L1/L2/L5	L1	L1
GLONASS	Oui	Non	Non
SBAS	Oui	Non	Non
Données brutes	Oui	Non	Oui (ρ)
Doppler	Oui	(Oui)	Non
précision position	1cm (RTK)/40cm (DGPS)	5m	3m
précision vitesse	-	5cm/s	-



(a)



(b)

FIGURE 4.2 – (a) Plateforme expérimentale avec les trois récepteurs GPS visibles sur le toit de la voiture et (b) la base qui envoie les corrections au GPS ALTUS fixé sur le toit de la voiture.

statique, en autoroute, en zone urbaine et en forêt). Ces jeux de données seront utilisés dans différents chapitres, nous les présentons indépendamment.

4.1.1 Capteurs GPS

Présentons les récepteurs GPS utilisés pour constituer nos jeux de données et spécifions dès à présent leur rôle. La figure 4.1 montre les récepteurs GNSS utilisés. Le tableau 4.1 résume les spécifications de chaque récepteur. Le tiré (-), dans le tableau 4.1, signifie que l'information n'apparaît pas dans les spécifications d'usine. Le récepteur UBLOX ne fournit que les pseudo-distances pour estimer la position. En ce qui concerne le récepteur GARMIN, nous avons mis l'utilisation du Doppler entre parenthèses (au sens littéral) car elle n'est pas spécifiée par le constructeur, mais nous pensons qu'une précision de 5cm/s en vitesse ne peut être atteinte qu'en utilisant des mesures Doppler.

Le récepteur Altus APS3, en mode RTK (*Real Time Kinematic*), est utilisé pour estimer la vérité de terrain concernant la localisation durant les expérimentations dynamiques. Le même modèle de

récepteur nous sert de base en figure 4.2b. Il permet l'envoi de corrections vers la *rover* (récepteur vérité terrain) par le réseau informatique, une transmission radio ou en post-traitement. La disposition des trois GPS sur le toit de la voiture est présentée en figure 4.2a. L'expérimentation présentée en figure 4.2 est utilisée pour constituer le jeu de données 3, détaillé dans la sous-section 4.1.4.

Les fréquences d'acquisition des divers récepteurs peuvent différer d'un jeu de données à un autre (1Hz ou 2Hz) (1Hz ou 2Hz), sauf en ce qui concerne le GARMIN 18x, pour lequel les données sont acquises à 1Hz.

4.1.2 Jeu de données 1

Le premier jeu de données est réalisé en statique en environnement dégagé. Durant une heure, près de l'école Supélec (plateau de Saclay), le récepteur GPS UBLOX EVK-5T a acquis des données avec une fréquence d'acquisition de 1Hz. Ce jeu de données nous a permis d'extraire des mesures PR peu bruitées auxquelles nous avons rajouté en simulation différentes distributions d'erreurs pour chaque *epoch*. Ce jeu de données sera utilisé par la suite comme un jeu de données simulé pour quantifier la détection de données aberrantes par les critères NFA_1 et NFA_2 (chapitre 5, paragraphe 5.4.1.1).

Précisément, l'objectif dans le premier jeu de données est de simuler l'existence de certaines valeurs aberrantes tout en connaissant la vérité terrain concernant les données aberrantes. Les observations simulées sont générées comme suit : pour chaque *epoch* d'une seconde, trois observations sont tirées au hasard parmi les données réelles de l'expérience en environnement dégagé, et pour chaque observation, un biais tiré aléatoirement entre 0 et 150m est ajouté. En termes de vérité terrain, ces observations tirées au hasard correspondent aux valeurs aberrantes simulées alors que les autres observations sont les *inliers*. En nous appuyant soit sur des travaux effectués également sur des simulations de réception de signaux multi-trajets/ NLOS [Giremus et al., 2007, Abdel-Hafez, 2010] soit concernant l'acquisition de mesures réelles [Le Marchand et al., 2009], nous avons considéré trois distributions statistiques pour les biais simulés : une distribution uniforme entre 0 et 150m, une distribution gaussienne centrée $\mathcal{N}(0, 50^2)$ et une distribution gaussienne non-centrée $\mathcal{N}(50, 15^2)$.

4.1.3 Jeu de données 2

Le deuxième jeu de données est également réalisé en statique mais en milieu urbain dans le quartier de La Défense à Paris. Cet environnement est contraignant pour la réception de signaux GPS de part la présence de tours environnantes générant des signaux multi-trajets et des réceptions en NLOS (même en provenance de satellites d'élévation moyenne aux alentours de 45°). Le même récepteur avec les mêmes spécifications est utilisé pour cette expérience.

La figure 4.3 présente une vue du ciel de l'environnement urbain où les données ont été acquises et un histogramme du nombre de satellites visibles simultanément durant cette expérience (figure 4.3b). Le récepteur est resté statique pendant toute l'expérience (environ 40 minutes), tandis que les sept satellites GPS se déplaçaient légèrement. Sur la figure 4.3a, leurs pistes sont tracées en orange

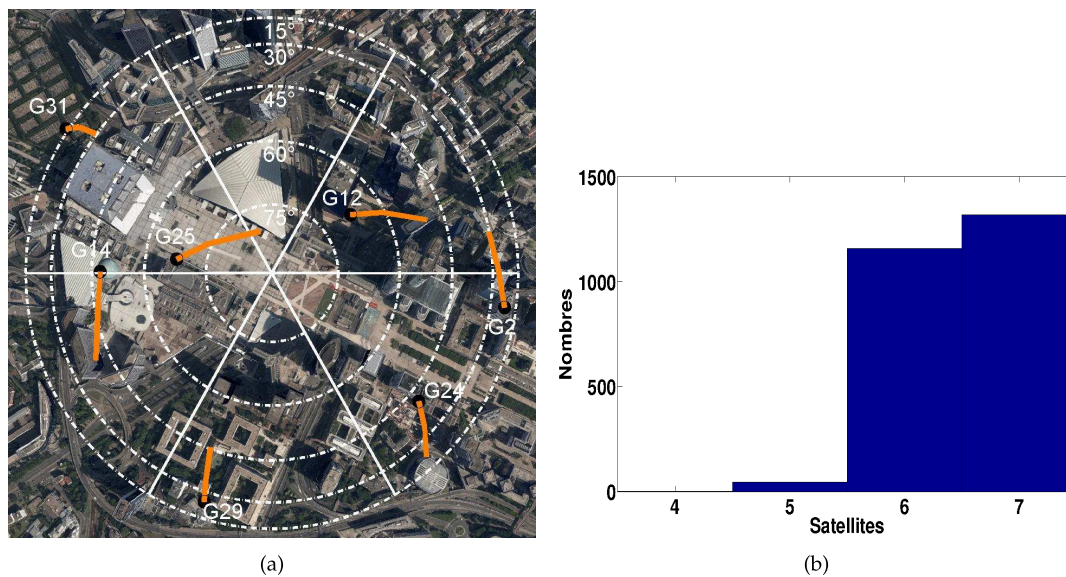


FIGURE 4.3 – Vue du ciel de l’environnement urbain de l’expérience, superposée aux trajectoires des satellites (tracées en orange).

avec des positions de début de l’expérience indiquées par les points noirs près du nom du satellite. Ce jeu de données est utilisé dans le chapitre 5, paragraphe 5.4.2.1 et le chapitre 7, paragraphe 7.5.

4.1.4 Jeu de données 3

Le troisième jeu de données est réalisé en mode dynamique. Sur un trajet de plus de 42km, trois différents environnements ont été traversés : une partie autoroute caractérisée par un environnement dégagé et une vitesse élevée du véhicule, une partie urbaine caractérisée par un environnement propice au multi-trajets et aux réceptions de signaux NLOS et une vitesse du véhicule moyenne, voire basse, et une dernière partie qui traverse une forêt caractérisée par des blocages des signaux GPS diminuant la redondance des données. La trajectoire de l’ensemble de l’expérience est représentée sur la figure 4.4.

Pour cette expérience, notre véhicule expérimental ZOE était équipé des deux GPS bas-coûts, UBLOX et GARMIN, et du GPS haut de gamme APS-3. Les données provenant de ce dernier ont été utilisées en mode RTK avec une base APS-3 qui envoie les corrections via le réseau pour fournir la vérité terrain en vue de comparer les performances des différentes méthodes de localisation. Ce mode de transmission peut engendrer des problèmes de latence et de congestion du réseau 3G pouvant invalider la solution RTK fournie par le récepteur. Les deux GPS, le UBLOX EVK-5T et le GPS haute précision ALTUS APS-3, avaient tous deux une fréquence d’acquisition de 2Hz. Le GPS UBLOX est utilisé pour valider notre approche dans les chapitres 5, paragraphe 5.4.2.2 et 7, paragraphe 7.5. Le récepteur GARMIN est utilisé pour être comparé à notre approche dans le chapitre 6, paragraphe 6.4.1.

La partie autoroute est d’environ 35km (pour une durée de 22mn), les figures 4.5a et 4.5b présentent la trajectoire des satellites pendant l’expérience et l’histogramme des nombres de satellites visibles. La

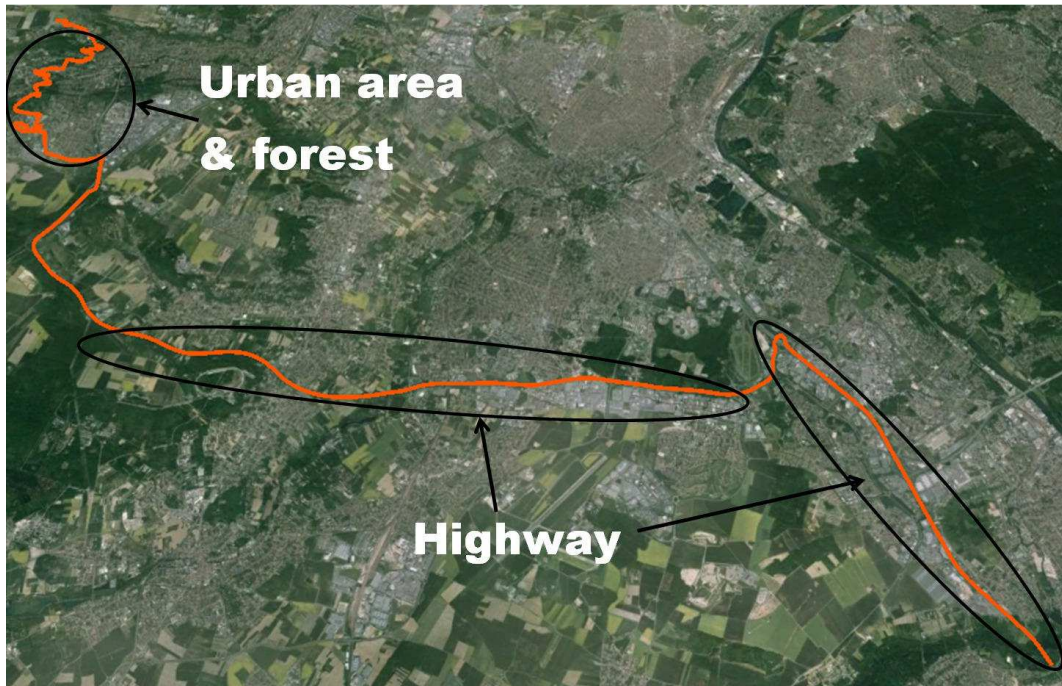


FIGURE 4.4 – Vue d’ensemble (© Google Earth) de la trajectoire. La durée totale de l’expérience est de 48mn.

partie en zone urbaine est d’environ 2.5km (pour une durée de 7mn environ) et la partie en forêt est d’environ 5.5km (pour une durée de 10mn environ). La figure 4.7a présente un zoom sur la partie de l’expérience en forêt et en zone urbaine. Les figures 4.6a et 4.6b donnent la configuration des satellites pendant la traversée en zone urbaine et en forêt et l’histogramme des nombres de satellites visibles durant la même période.

Nous avons fixé la fréquence d’échantillonnage des données brutes du récepteur ALTUS APS-3 à 1Hz. La précision de localisation est égale à 1cm, selon les spécifications d’usine dans le cas du mode *fixed solution*. Cette solution était disponible, en environnements contraints, durant 41% du temps de l’expérimentation (voir la figure 4.7b), alors que les deux autres modes, *RTK float* et *differential*, dont la précision peut chuter jusqu’à 40cm, sont disponibles durant 39% et 20%, respectivement.

Pendant toute l’expérience, le récepteur GPS UBLOX (UBX) fournit à la fois les données brutes (mesures PR et paramètres de navigation des satellites) et des trames NMEA (*National Marine Electronics Association*) contenant un calcul de la position du récepteur. La fréquence choisie (2Hz) a été augmentée par rapport à la précédente expérience (en mode statique) afin de compenser, en partie, la diminution de la redondance des données en raison du mouvement du récepteur.

4.2 Mesures de performance

Dans cette thèse, notre problème principal est la détection des données GNSS aberrantes. Pour quantifier la qualité de nos algorithmes, nous devons définir des métriques capables de quantifier

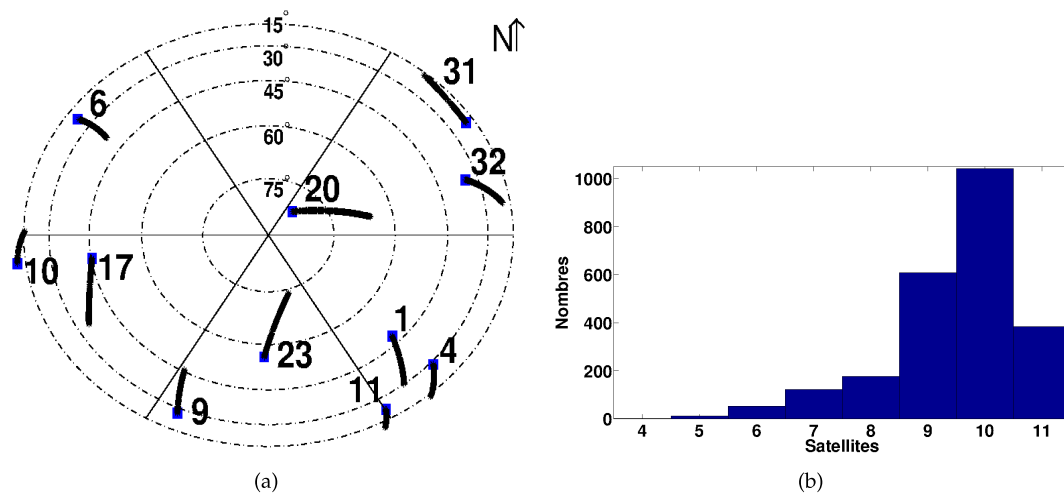


FIGURE 4.5 – (a) Configuration et trajectoire des satellites durant toute l’expérimentation, position initiale représenté par le carré bleu et (b) histogramme des nombres de satellites visibles durant le trajet en autoroute.

la qualité de cette détection, à savoir les métriques classiques associées aux nombres de fausses détections, non détections, etc. Cependant, la vérité de terrain concernant ces métriques est quasiment impossible à établir hormis lorsque nous considérons des données simulées. Aussi, nous quantifions également indirectement la détection des *outliers* via notre application, à savoir la localisation. L’idée est que si les données aberrantes sont bien détectées, la précision de localisation sera accrue (ou alors, soit la détection des données aberrantes est mauvaise, soit elle est inutile). Pour quantifier la qualité de la localisation, nous utilisons alors les métriques classiques associées aux différentes normes d’erreurs.

Spécifions à présent ces deux types de mesures de performance : en détection et en localisation.

4.2.1 Performance de la détection des *outliers*

Pour évaluer les résultats en termes de partitionnement des *inliers/outliers*, nous calculons le nombre de vrais positifs (*VP*), faux positifs (*FP*), faux négatifs (*FN*) et vrais négatifs (*VN*), en adoptant les définitions suivantes :

- les *VP* sont des *inliers* (mesures PR sans biais) qui appartiennent à \mathcal{D} ;
- les *FP* sont des *outliers* (mesures PR avec biais) qui appartiennent à \mathcal{D} ;
- les *FN* sont des *inliers* classifiés comme *outliers* ($\notin \mathcal{D}$) ;
- les *VN* sont des *outliers* classifiés comme *outliers* ($\notin \mathcal{D}$) ;

avec \mathcal{D} l’ensemble des mesures détectées/classifiées par l’algorithme comme des *inliers*. À partir de ces valeurs, nous déduisons les mesures classiques d’exactitude de la détection généralement appelées selon la terminologie anglaise *Accuracy* et de précision de la détection appelée en anglais *Precision*. Dans la suite de ce manuscrit, nous adoptons la terminologie anglaise pour éviter les ambiguïtés. Ces

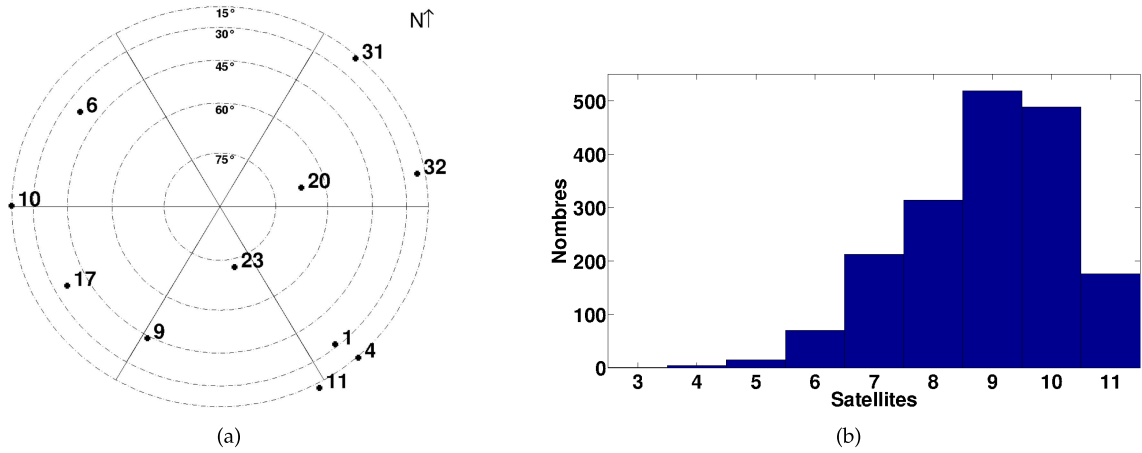


FIGURE 4.6 – (a) Configuration des satellites durant l'expérimentation en milieu urbain et (b) histogramme des nombres de satellites visibles durant le trajet en milieu urbain et en forêt.

mesures sont calculées en fonction des nombres VP , TP , VN et FP définis précédemment :

$$\begin{aligned} Accuracy &= \frac{(VP+TN)}{(VP+VN+FP+FN)}, \\ Precision &= \frac{(VP)}{(VP+FP)}. \end{aligned} \quad (4.1)$$

4.2.2 Performance de la localisation

Les performances d'un algorithme de localisation sont mesurées par les indicateurs : norme 1, norme infinie et le *RMSE* (*Root Mean Square Error*) ainsi que la valeur moyenne et la variance des erreurs. Dans notre cas, étant donné la prévalence des coordonnées à la surface de la Terre, ces mesures sont calculées pour chacune des deux coordonnées Est et Nord. Soit ε_i l'erreur en position à l'instant i le long d'une direction donnée, Est ou Nord, les normes associées à cette erreur sont évaluées à partir de k échantillons par :

$$\begin{aligned} Norm_1 &= \frac{1}{k} \sum_{i=1}^k abs(\varepsilon_i), \\ Norm_{inf} &= max(abs(\varepsilon_i)), \\ RMSE &= \frac{1}{k} \sqrt{\sum_{i=1}^k \varepsilon_i^2}. \end{aligned} \quad (4.2)$$

Soit E_i la distance euclidienne entre l'estimation de la position et la vérité de terrain à l'instant i , la moyenne et la variance de cette distance évaluées à partir de k échantillons sont données par :

$$\begin{aligned} \mu_{loc} &= \frac{1}{k} \sum_{i=1}^k E_i, \\ \sigma_{loc}^2 &= \frac{1}{k} \sum_{i=1}^k (E_i - \mu_{loc})^2. \end{aligned} \quad (4.3)$$

Une moyenne non nulle indique la présence d'un biais statistique dans les estimations des posi-

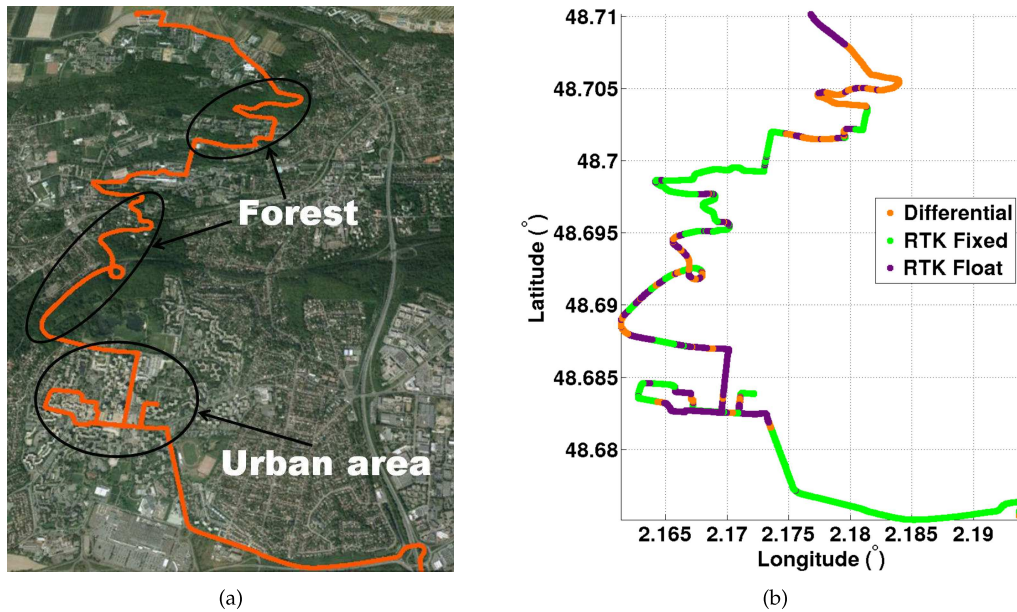


FIGURE 4.7 – (a,b) Trajectoire durant l'expérimentation, soit (a) tracée sur une carte Google Earth[®] soit (b) labellisée en termes de qualité de la solution *Real-Time Kinematic* (RTK) ('vérité de terrain').

tions. Une moyenne nulle pouvant correspondre à une compensation des erreurs de positions, nous avons besoin de la variance (ou l'écart type) pour connaître l'importance des erreurs, via leur dispersion autour de la moyenne.

4.3 Conclusion

Dans ce chapitre, nous avons présenté les jeux de données utilisés par la suite dans l'évaluation des algorithmes proposés. Les chapitres de contribution dans la thèse utiliseront différemment les jeux de données et les récepteurs utilisés durant les expérimentations. En conclusion, nous donnons la répartition des jeux de données dans les différents chapitres :

- **Chapitre 5** : les trois jeux de données sont utilisés dans ce chapitre. Seules les mesures de pseudo-distances, issues du récepteur UBLOX, sont analysées et utilisées pour la localisation robuste. De ce fait, les solutions fournies par le récepteur GARMIN ne sont pas considérées dans ce chapitre pour assurer une comparaison juste des méthodes de localisation développées.
- **Chapitre 6** : seul le jeu de données 3 est utilisé dans ce chapitre où un intérêt particulier est accordé à la localisation dynamique en milieu urbain intégrant les mesures Doppler dans les algorithmes de localisation. Le récepteur UBLOX ne fournissant pas de données Doppler, les mesures brutes (non traitées) fournies par le GPS ALTUS APS 3 sont les données d'entrées des méthodes de localisation. Les solutions fournies par les récepteurs UBLOX et GARMIN (via les trames NMEA) seront utilisées à titre d'illustration de la performance des algorithmes proposés par rapport aux méthodes embarquées dans les récepteurs.
- **Chapitre 7** : les jeux de données 2 et 3 sont utilisés dans ce chapitre. Dans le jeu de données 3, le test des méthodes de localisation nécessitera l'utilisation des données brutes issues des ré-

cepteurs ALTUS ou UBLOX en fonction de l'utilisation des mesures Doppler dans le processus d'estimation.

Modélisation a-contrario pour la localisation GNSS

Sommaire

5.1	Introduction	59
5.2	Localisation à partir de données multi-epochs	60
5.3	Modélisation a contrario	62
5.3.1	Modèle naïf gaussien	62
5.3.2	Modèle naïf uniforme	65
5.3.3	Exemple de la détection de droite	67
5.3.4	Spécificité pour la localisation GNSS	69
5.4	Expérimentations et résultats	74
5.4.1	Résultats obtenus avec les données simulées	75
5.4.2	Résultats obtenus avec les données réelles	81
5.5	Conclusion	89

5.1 Introduction

Nous avons présenté dans les chapitres précédents un aperçu des différentes approches permettant de détecter les données aberrantes dans différents domaines de la fusion de données, du traitement du signal et des images. Le problème traité dans ce chapitre est la détection de données aberrantes dans les mesures GNSS (notamment, les pseudo-distances estimées par le récepteur) en vue d'élaborer une méthode de localisation robuste pour tout environnement. En effet, avec la mise en service des futurs systèmes de navigation satellite, le nombre de satellites disponibles à une *epoch* donnée augmentera. De ce fait, l'hypothèse d'avoir au plus une mesure aberrante par instant ne sera plus valable dans le cas multi-constellations, et elle sera encore moins réaliste en milieu urbain où les obstacles environnants dégradent les signaux reçus en introduisant une surestimation du trajet satellite/récepteur.

Dans ce chapitre, nous présentons une modélisation *a contrario* du problème de localisation GNSS pour simultanément détecter les données aberrantes (qu'on appellera *outliers*) et fournir une estimation robuste de la position du récepteur. L'approche *a contrario* est connue pour être robuste à un taux élevé de données aberrantes. Contrairement à l'implémentation proposée par [Schroth et al., 2008]

inspirée de l'algorithme RANSAC, qui utilise un seuil déduit d'une probabilité de fausse alarme pour partitionner l'ensemble des mesures en données pertinentes (correctes) et aberrantes, l'approche *a contrario* permet de définir un critère sans paramètre, à savoir le nombre de fausses alarmes (NFA) *a contrario* [Desolneux et al., 2000, Desolneux et al., 2007].

La première partie du chapitre introduit l'estimation d'une position à partir de données GNSS acquises sur une fenêtre temporelle pour pallier le manque d'observations après élimination des *outliers*. Dans un second temps, la modélisation *a contrario* du problème est développée selon deux critères NFA modélisant l'incohérence des mesures PR. À partir du cas statique, nous développons l'extension au cas dynamique, qui ajoute en particulier un filtrage temporel. La dernière partie du chapitre présente des résultats de simulations (en statique et dynamique) et une expérimentation en environnements variés (autoroute, canyon urbain et forêt) avec une analyse des résultats et une comparaison avec des méthodes de localisation classiques.

5.2 Localisation à partir de données multi-epochs

La mesure la plus utilisée pour la localisation GNSS est le temps de trajet ou '*time of arrival*' (TOA). Pour la constellation GPS, les récepteurs bas-coût acquièrent les signaux de la porteuse L1 de fréquence $f_{L1} = 1575,42\text{MHz}$ provenant d'un satellite S_i pour estimer la pseudo-distance ρ_i à partir du TOA (en multipliant le temps de trajet par la vitesse de la lumière). On notera dans ce chapitre, la pseudo-distance 'PR' en référence à '*pseudo-range*' en anglais.

Le modèle d'observation utilisé aura pour double objectif d'estimer la position du récepteur et de détecter et d'identifier les observations incohérentes (entre elles). Le critère de cohérence sera présenté dans la prochaine section. Soit un récepteur situé à la position (x_r, y_r, z_r) dans le référentiel cartésien ECEF (*Earth Centered Earth Fixed*) et soit un satellite S_i situé à la position $(x_{S_i}, y_{S_i}, z_{S_i})$ dans le même référentiel, alors nous définissons notre modèle d'observation comme :

$$\rho_i = \sqrt{(x_r - x_{S_i})^2 + (y_r - y_{S_i})^2 + (z_r - z_{S_i})^2} + c\delta t + \varepsilon, \quad (5.1)$$

où δt est le biais temporel entre les horloges satellite et récepteur, c est la vitesse de la lumière, et ε est un bruit gaussien centré qui représente les erreurs entachant la mesure (effets ionosphérique et troposphérique, bruit électronique, multi-trajets et réception de signaux en non-visibilité directe dits NLOS pour '*Non-Line Of Sight*').

Dans les problèmes de localisation GNSS, les observations sont donc constituées des positions des satellites à un instant donné et de la mesure PR entre un satellite S_i , soit le quadruplet $(x_{S_i}, y_{S_i}, z_{S_i}, \rho_i)$. En considérant que le vecteur d'état à estimer se compose des coordonnées du récepteur et du biais temporel $\mathbf{X} = (x_r, y_r, z_r, \delta t)$, alors, l'estimation du vecteur d'état \mathbf{X} requiert au moins quatre observations. Toutefois, en présence d'erreurs dégradant la mesure PR, un nombre supérieur d'observations est généralement utilisé pour exploiter la redondance des mesures et filtrer le bruit.

Afin de relier les observations au vecteur d'état \mathbf{X} à l'instant t , noté \mathbf{X}_t , considérons l'estimation *a priori* de la pseudo-distance $\tilde{\rho}_i(t|\mathbf{X}_t)$ pour un satellite S_i dont les coordonnées à l'instant t sont $(x_{S_i}(t), y_{S_i}(t), z_{S_i}(t))$:

$$\tilde{\rho}_i(t|\mathbf{X}_t) = \left[(x_r(t) - x_{S_i}(t))^2 + (y_r(t) - y_{S_i}(t))^2 + (z_r(t) - z_{S_i}(t))^2 \right]^{1/2} + c\delta t(t). \quad (5.2)$$

La notation $|\mathbf{X}_t$ signifie que l'état \mathbf{X}_t est connu. La valeur optimale de \mathbf{X}_t , notée $\hat{\mathbf{X}}_t$, est estimée par minimisation de la somme des erreurs quadratiques entre les PR mesurées et estimées : $\hat{\mathbf{X}}_t = \arg \min_{\mathbf{X}_t} \sum_i [\tilde{\rho}_i(t|\mathbf{X}_t) - \rho_i]^2$, ou i appartient à l'ensemble des indices des satellites fournissant les mesures PR à l'instant t .

La solution peut être obtenue en employant un algorithme itératif de résolution de système d'équation non-linéaire tel que l'algorithme de Gauss-Newton [Kaplan and Hegarty, 2006].

La formulation précédente du problème s'applique pour l'estimation d'une position ponctuelle. Lorsque le récepteur est en mouvement, les estimations successives qui sont effectuées de façon indépendante à des instants successifs t fournissent généralement des résultats médiocres. Une première façon d'améliorer ces estimations ponctuelles est de les filtrer temporellement. Ceci peut être réalisé en utilisant le filtre de Kalman (KF) ou le filtre particulaire qui utilisent l'(les)estimation(s) précédente(s) pour améliorer l'estimation courante. Pour ce faire, le vecteur d'état précédent \mathbf{X}_t (4-uplet) est étendu en introduisant les coordonnées de vitesse du récepteur qui permettent de relier les mesures à différents instants et la dérive d'horloge comme suit : soit $\mathbf{X}_t^m = (x_r, \dot{x}_r, y_r, \dot{y}_r, z_r, \dot{z}_r, \delta t, \dot{\delta}t)$ (8-uplet) pour le cas dynamique avec \dot{u} désignant la dérivée de la variable u et (t) désignant la dépendance en temps. La vitesse du récepteur, $(\dot{x}_r, \dot{y}_r, \dot{z}_r)$, est exprimée dans le même système de coordonnées que la position et $(\delta t, \dot{\delta}t)$ représente respectivement le biais d'horloge du récepteur et sa dérive selon le modèle d'horloge proposé dans [Radon-Nykodim, 1997].

L'équation du premier ordre reliant le vecteur d'état \mathbf{X}_t^m à l'instant t avec le vecteur d'état \mathbf{X}_{t-dt}^m à l'instant $t - dt$ est :

$$\begin{aligned} \mathbf{C}_{dt} &= \begin{pmatrix} 1 & dt \\ 0 & 1 \end{pmatrix}, \mathbf{0}_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \\ \mathbf{M}_{dt} &= \begin{pmatrix} \mathbf{C}_{dt} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{C}_{dt} & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{C}_{dt} & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{C}_{dt} \end{pmatrix}, \\ \mathbf{X}_t^m &= \mathbf{M}_{dt} \cdot \mathbf{X}_{t-dt}^m + o(dt^2), \end{aligned} \quad (5.3)$$

où $o(dt^2)$ représente les termes d'ordre supérieur à un. Dans de nombreuses applications (e.g. le filtre de Kalman [Jwo and Wang, 2007]), l'erreur due à l'approximation de premier ordre est assimilée à un bruit blanc gaussien identiquement distribué. En résumé (et dans la suite de nos développe-

ments), l'équation (5.3) prédit le vecteur d'état à l'instant courant (i.e. au moment où les mesures sont fournies par le récepteur) sachant l'estimation précédente.

Pour estimer \mathbf{X}_t^m de façon robuste, plusieurs instants (ou *epochs*) sont considérés sur une période au cours de laquelle une approximation du premier ordre est acceptable (en termes d'erreur). Notons $\mathcal{T} = \{t_k = t - kdt, k \in \{0, \dots, n\}\}$ l'ensemble des *epochs* considérées pour l'estimation de \mathbf{X}_t . Pour chaque instant $t_k \in \mathcal{T}$, la mesure PR $\tilde{\rho}_i(t_k)$ correspondant au satellite S_i à la position $[\mathbf{X}_{S_i}(t_k)]^\top = (x_{S_i}(t_k), y_{S_i}(t_k), z_{S_i}(t_k))$ peut être estimée sachant \mathbf{X}_t^m :

$$\tilde{\rho}_i(t_k | \mathbf{X}_t^m) = \| [\mathbf{M}_{kdt} \cdot \mathbf{X}_t^m]_{1,3,5} - \mathbf{X}_{S_i}(t_k) \| + c(\delta t(t) - \dot{\delta}t(t) kdt), \quad (5.4)$$

où l'exposant \top désigne la transposée d'une matrice ou d'un vecteur, la notation indicée $[\]_{l_1, l_2, l_3}$ désigne la restriction de la matrice ou du vecteur aux lignes l_1, l_2 et l_3 , et $\|v\|$ est la norme du vecteur v .

Alors, la valeur optimale $\hat{\mathbf{X}}_t^m$ est celle qui minimise l'erreur quadratique :

$$\hat{\mathbf{X}}_t^m = \arg \min_{\mathbf{X}_t^m} \sum_{t_k \in \mathcal{T}} \sum_{i \in \mathcal{I}(t_k)} [\tilde{\rho}_i(t_k | \mathbf{X}_t^m) - \rho_i(t_k)]^2, \quad (5.5)$$

où $\mathcal{I}(t_k)$ désigne l'ensemble des indices des mesures satellites acquises à l'instant t_k .

5.3 Modélisation a contrario

Selon une approche *a contrario*, les données pertinentes (qu'on appellera "*inliers*") sont détectables comme trop régulières pour apparaître 'par hasard'. Pour instancier ce concept de perception à travers un critère de nombre de fausses alarmes (NFA), deux éléments doivent être définis. Premièrement, nous définissons une (ou plusieurs) mesure(s) caractérisant les *inliers* et deuxièmement, nous définissons un modèle 'naïf' qui représente les statistiques des valeurs aberrantes (l'hypothèse H_0 en théorie statistique de la décision). Dans le cas des données GNSS, la principale caractéristique décrivant les *inliers* est la cohérence des mesures de pseudo-distance. Dans cette étude, nous examinons deux mesures de leur cohérence (ou cohérence par anglicisme).

5.3.1 Modèle naïf gaussien

Supposons un sous-ensemble d'*inliers* noté \mathcal{D} : \mathcal{D} est un ensemble d'indices associé à l'ensemble des observations effectuées au cours d'une fenêtre temporelle notée \mathcal{T} . La première mesure de la cohérence \mathcal{D} est la somme des carrés des résidus (appelés erreurs de sorte que leur somme est notée SSE pour *Sum of Squared Errors*) entre le PR observé et le PR calculé. Les sous-ensembles les plus cohérents d'*inliers* devraient présenter des valeurs particulièrement faibles de somme des résidus au carré. Soit

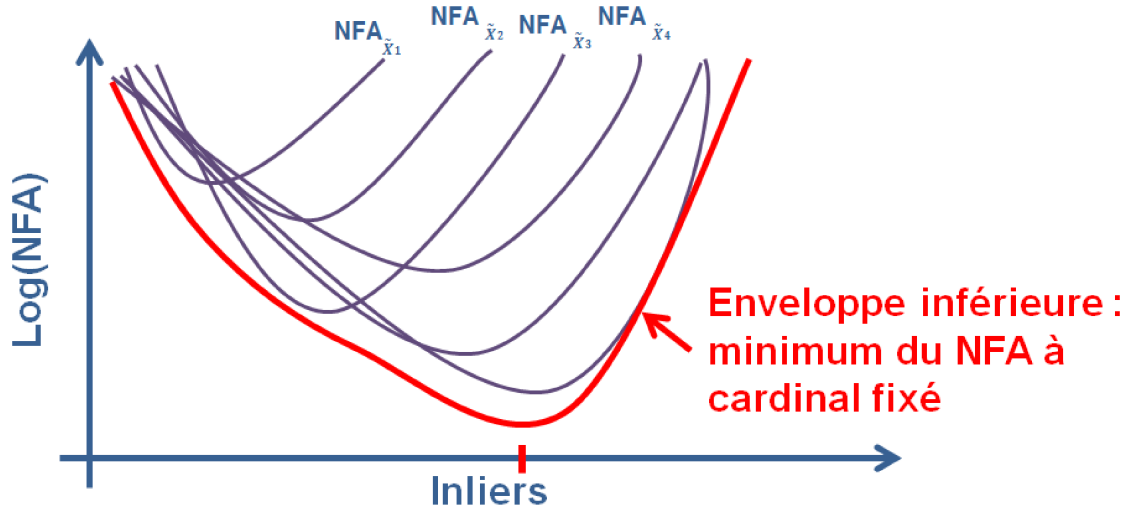


FIGURE 5.1 – Minimisation de la courbe NFA.

alors le modèle \mathcal{M}_1 qui modélise la loi des résidus en cas d'incohérence. Il se présente comme suit : l'ensemble des résidus est un processus aléatoire de $|\mathcal{D}|$ variables gaussiennes indépendantes centrées, $\mathcal{N}(0, \sigma)$ avec σ l'écart-type de l'ensemble complètement déstructuré et $|\cdot|$ désignant le cardinal d'un ensemble.

En supposant une valeur pour le vecteur d'état \mathbf{X}_t^m (ou \mathbf{X}_t dans le cas statique), le résidu $r_i(t)$ associé à l'observation PR du satellite S_i à l'instant t , $\rho_i(t)$, est :

$$r_i(t) = [\tilde{\rho}_i(t | \mathbf{X}_t^m) - \rho_i(t)], \quad (5.6)$$

où $\tilde{\rho}_i(t | \mathbf{X}_t^m)$ est calculé en utilisant soit l'équation (5.4) dans le cas dynamique soit l'équation (5.2) dans le cas statique.

Afin de mesurer la cohérence du sous-ensemble \mathcal{D} , nous calculons $\delta_{\mathcal{D}}^2$ la somme des valeurs de $r_i^2(t)$ dont les indices i appartiennent à l'ensemble \mathcal{D} et t appartient à la fenêtre temporelle \mathcal{T} . Pour interpréter les valeurs $\delta_{\mathcal{D}}^2$, nous considérons la probabilité d'observer $\delta_{\mathcal{D}}^2$ par hasard, notée $\mathbb{P}_{\mathcal{M}_1}(\delta_{\mathcal{D}}^2, \sigma)$. Le nombre de fausses alarmes (NFA) du sous-ensemble \mathcal{D} est alors calculé selon l'équation présentée et démontrée dans la thèse de Robin [Robin, 2007] :

$$\begin{aligned} NFA_1(\mathcal{D}) &= \eta_1 \mathbb{P}_{\mathcal{M}_1}(\delta_{\mathcal{D}}^2, \sigma), \\ &= \eta_1 \frac{1}{\Gamma\left(\frac{|\mathcal{D}|}{2}\right)} \int_0^{\delta_{\mathcal{D}}^2/2\sigma^2} e^{-t} t^{\frac{|\mathcal{D}|}{2}-1} dt, \end{aligned} \quad (5.7)$$

où Γ est la fonction gamma et η_1 est un terme de normalisation équivalant à un nombre de tests contrôlant le nombre moyen de fausses alarmes [Desolneux et al., 2007].

Rappelons que, comme vu au chapitre 2, la somme des carrés des erreurs et le test du χ^2 ont déjà été utilisés comme critère pour détecter les valeurs aberrantes comme dans la méthode RAIM [Brown, 1992,

Algorithm 4: Fonction NFA basée sur \mathcal{M}_1 avec comme entrées : M observations, σ , t , N_{test} et comme sorties : \mathbf{X} et \mathcal{D}

```

1 Initialiser les vecteurs  $\delta_{min}[i]$  et  $NFA[i]$  et le scalaire  $NFA_{min}$  à  $+\infty$ ;
2 for 1 to  $N_{test}$  do
3   Tirer aléatoirement  $d$  observations  $(x_{S_i}, y_{S_i}, z_{S_i}, \rho_i)$  notées  $\{o_1, \dots, o_d\}$  (Algo. 7 page 73);
4   Estimer  $\tilde{\mathbf{X}}$  le  $d$ -uplet des paramètres inconnus à partir de  $\{o_1, \dots, o_d\}$  (Algo. 6 page 71);
5   Calculer le vecteur des résidus suivant l'équation (5.6);
6   Trier le vecteur de résidus par ordre croissant dans le vecteur  $\mathbf{W}_r = (w_i)_{i \in \{1 \dots M\}}$  avec  $j$  la
   fonction des indices de permutation telle que  $\forall i \in \{1 \dots M\}, w_i = r_{j(i)}$ ;
7   Initialiser  $\delta_{\mathcal{D}}^2 = 0$ ;
8   for  $i \leftarrow d + 1$  to  $M$  do
9      $\delta_{\mathcal{D}}^2 \leftarrow \delta_{\mathcal{D}}^2 + w_i^2$ ;
10    if  $\delta_{\mathcal{D}}^2 < \delta_{min}[i]$  then
11       $\delta_{min}[i] = \delta_{\mathcal{D}}^2$ ;
12      Calculer  $NFA[i]$  selon l'équation (5.7);
13      if  $NFA[i] < NFA_{min}$  then
14         $NFA_{min} = NFA[i]$ ;  $\mathcal{D} = \{j(1) \dots j(i)\}$ ;
15         $\mathbf{X} = \tilde{\mathbf{X}}$ ;
16      end
17    end
18  end
19 end

```

[Walter and Enge, 1995, Hewitson and Wang, 2006]. Cependant, dans ces approches classiques, un seuil, qui est généralement fixé selon une probabilité de fausse alarme P_{FA} , est utilisé pour décider quelles valeurs de SSE sont acceptables. Dans notre cas, nous nous affranchissons de ce seuil en recherchant la solution (qui comprend à la fois le sous-ensemble d'*inliers* \mathcal{D} et le vecteur d'état \mathbf{X}_t^m) qui minimise le critère NFA (équation (5.7)). Cette absence de seuil apparaît clairement dans les paramètres d'entrée des algorithmes présentés plus loin.

Une autre différence réside dans le couplage avec la technique RANSAC pour échantillonner les valeurs de \mathbf{X}_t^m . Nous avons vu (chapitre 3) que [Schroth et al., 2008] a proposé d'adapter le RANSAC à la détection de mesures aberrantes provenant de systèmes multi-constellations (GPS+GLONASS). Néanmoins, alors que dans l'implémentation du RANSAC de [Schroth et al., 2008] l'ensemble des mesures considérées comme *inliers* est obtenu en maximisant le cardinal de l'ensemble des résidus inférieurs à un seuil fixé (qui est un paramètre *a priori*), dans notre cas, comme il a été précédemment mentionné, l'ensemble *inlier* est déduit par une minimisation du critère NFA (voir l'équation (5.7)). La figure 5.4 (dans la section 5.3.3) illustrera, en termes de robustesse par rapport au seuil, la différence entre le RANSAC associé à un paramètre de seuil ou celui associé au critère NFA.

L'implémentation du NFA sous l'hypothèse du modèle naïf gaussien est présentée par l'algorithme 4. La figure 5.1 illustre le processus de minimisation de la courbe $\log(NFA)$ en fonction du nombre d'*inliers*. Tout d'abord, on suppose une position $\tilde{\mathbf{X}}_1$ estimée via les d observations tirées aléatoirement

parmi M observations (ligne 4 de l'algorithme 4). Ensuite, nous calculons les résidus puis la somme des carrés des résidus triés de façon croissante des $M - d$ observations restantes. Enfin nous calculons la fonction NFA correspondant à la position \tilde{X}_1 qui est donnée par la courbe $NFA_{\tilde{X}_1}$ de la figure 5.1. Le processus de minimisation se poursuit pour les positions estimées $\tilde{X}_2, \tilde{X}_3, \tilde{X}_4$ jusqu'à $\tilde{X}_{N_{test}}$. Nous obtenons en sortie de l'algorithme une courbe des valeurs de NFA minimales qui est l'enveloppe inférieure de toutes les courbes NFA calculées (voir la figure 5.1) au cours des itérations. Soulignons que chaque point de cette courbe peut avoir été obtenu à partir d'une estimation de \tilde{X}_i différente. Le cardinal de l'ensemble recherché des *inliers* \mathcal{D} correspond à la valeur qui minimise cette courbe des valeurs de NFA minimum.

5.3.2 Modèle naïf uniforme

La deuxième mesure de cohérence proposée découle d'une interprétation géométrique des mesures de pseudo-distance. Supposons une position du récepteur (x_r, y_r, z_r) , et considérons une ellipsoïde (dans notre cas, ou une surface 3D fermée pour un cas plus général) centrée sur le récepteur modélisant ainsi son imprécision sur sa position supposée. Nous dirons que cette ellipsoïde valide une mesure de pseudo-distance ρ_i provenant du satellite S_i si et seulement si la sphère S_i centrée sur la position S_i (notée $(x_{S_i}, y_{S_i}, z_{S_i})$) l'intersecte. Le volume de l'ellipsoïde d'imprécision a un impact direct sur le nombre de pseudo-distances validées.

Algorithm 5: Fonction NFA basée sur le modèle \mathcal{M}_2 avec comme entrées : M observations, \mathcal{V} , t , N_{test} et comme sorties : X et \mathcal{D}

```

1 Initialiser le vecteur  $NFA [i]$  et le scalaire  $NFA_{min}$  vers  $+\infty$ ;
2 for 1 to  $N_{test}$  do
3   tirer aléatoirement de  $d$  observations  $(x_{S_i}, y_{S_i}, z_{S_i}, \rho_i)$  notées par  $\{o_1, \dots, o_d\}$  (Algo. 7 page 73);
4   Estimer  $\tilde{X}$  les  $d$ -uplet des paramètres inconnues à partir de  $\{o_1, \dots, o_d\}$  (Algo. 6 page 71);
5   Calculer le vecteur d'ellipsoïde selon l'équation (5.9);
6   Trier le vecteur de résidus par ordre croissant dans le vecteur  $W_r = (w_i)_{i \in \{1 \dots M\}}$  avec  $j$  la
   fonction des indices de permutation tel que  $\forall i \in \{1 \dots M\}, w_i = u_{j(i)}$ ;
7   for  $i \leftarrow d + 1$  to  $M$  do
8     Calculer  $\mathcal{V}_{\mathcal{D}}$  depuis  $w_i$ 
9      $p_{\mathcal{D}} \leftarrow \frac{\mathcal{V}_{\mathcal{D}}}{\mathcal{V}}$ 
10     $|\mathcal{D}| \leftarrow i$ 
11    Calculer  $NFA [i]$  selon l'équation (5.8);
12    if  $NFA [i] < NFA_{min}$  then
13       $NFA_{min} = NFA [i]$ ;
14       $\mathcal{D} = \{j(1) \dots j(i)\}$ ;
15       $X = \tilde{X}$ ;
16    end
17  end
18 end

```

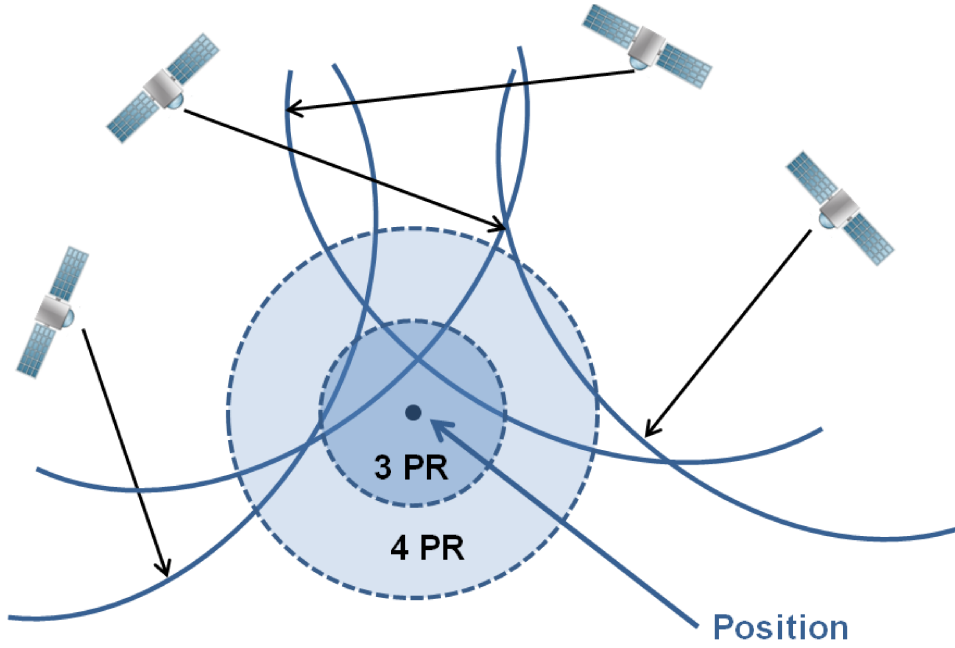


FIGURE 5.2 – Exemple d’illustration du critère NFA basée sur le modèle \mathcal{M}_2 .

Étant donné un sous-ensemble \mathcal{D} de mesures de pseudo-distance, l’ellipsoïde d’imprécision $\mathcal{E}_{\mathcal{D}}$ validant les éléments de \mathcal{D} est le plus petit ellipsoïde (dans notre cas) qui intersecte avec toutes les sphères \mathcal{S}_i d’indice i appartenant à \mathcal{D} . Dans ce qui suit, le volume de $\mathcal{E}_{\mathcal{D}}$ est noté $\mathcal{V}_{\mathcal{D}}$. Les sous-ensembles d’inliers les plus cohérents devraient présenter des volumes $\mathcal{V}_{\mathcal{D}}$ particulièrement petits. Notons aussi que, dans le cas dynamique, l’ellipsoïde est centré sur la position supposée du récepteur au moment t suivant le vecteur d’état X_t^m .

Le modèle naïf \mathcal{M}_2 modélise la distribution spatiale des contraintes pseudo-distance dans le cas de bruit pur. Il se présente comme suit : l’ensemble des contraintes de pseudo-distance est un processus aléatoire de M (le nombre total d’observations) variables indépendantes uniformément distribuées sur un volume borné \mathcal{V} .

Selon ce modèle, la probabilité *a priori* qu’une mesure de pseudo-distance soit compatible (validée) par un ellipsoïde $\mathcal{E}_{\mathcal{D}}$ est le ratio $\frac{\mathcal{V}_{\mathcal{D}}}{\mathcal{V}}$, noté $p_{\mathcal{D}}$. Ainsi, la probabilité $\mathbb{P}_{\mathcal{M}_2}(|\mathcal{D}|)$ d’avoir au moins $|\mathcal{D}|$ mesures de pseudo-distance validées est donnée par la queue de distribution de la fonction binomiale. Ainsi, la mesure NFA sur l’ellipsoïde $\mathcal{E}_{\mathcal{D}}$ s’écrit :

$$\begin{aligned} NFA_2(\mathcal{D}) &= \eta_2 \mathbb{P}_{\mathcal{M}_2}(p_{\mathcal{D}}, M, |\mathcal{D}|), \\ &= \eta_2 \sum_{k=|\mathcal{D}|}^M \binom{M}{k} p_{\mathcal{D}}^k (1 - p_{\mathcal{D}})^{M-k}, \end{aligned} \quad (5.8)$$

où η_2 est le nombre de tests.

En pratique, les ellipsoïdes $\mathcal{E}_{\mathcal{D}}$ considérés sont des ellipsoïdes homothétiques centrés sur la position supposée du récepteur à l’instant donné. Les dimensions des ellipsoïdes homothétiques dépendent de la position du satellite par rapport au récepteur et sont calculées comme suit : la longueur de l’axe semi-principal z est égal à b fois la longueur des axes semi-principaux x et y qui sont égaux.

En connaissant la distance entre le centre de l'ellipsoïde et la surface de la sphère \mathcal{S}_i (qui est le résidu r_i) et l'angle d'élévation e_i du centre de \mathcal{S}_i (qui est la position du satellite), la longueur de l'axe semi-principale x et y est :

$$u_i = r_i \sqrt{\cos^2 e_i + (\sin^2 e_i) / b^2}. \quad (5.9)$$

Par exemple, avec un satellite d'élévation 0° , la longueur des axes x et y est égale à r_i , et celle de l'axe z est égale à br_i ; tandis que pour un satellite d'élévation 90° au-dessus du récepteur, l'axe semi-principal selon z et les deux autres axes auront une longueur de r_i et de r_i/b respectivement. En d'autres termes, un même résidu engendre une imprécision plus grande s'il est déduit d'un satellite d'élévation basse que s'il est associé à un satellite d'élévation haute. Rappelons que b est un paramètre empirique entre les longueurs des axes semi-principaux z et x . Dans notre cas $b = 3$.

La figure 5.2 illustre le principe du NFA sous l'hypothèse du modèle naïf uniforme dans le cas 2D. L'algorithme 5 présente l'implémentation du deuxième critère NFA qui est en grande partie similaire à l'algorithme 4. Nous commençons, comme pour l'implémentation du premier critère, par estimer le d -uplet (position ou vecteur d'état) à partir de d observations tirées aléatoirement. Nous générons à partir de cette position des ellipsoïdes de volume $\mathcal{V}_{\mathcal{D}}$. Les valeurs de $\mathcal{V}_{\mathcal{D}}$ sont calculées de façon à augmenter de 1 à chaque fois le nombre d'observations intersectant (i.e. compatibles avec) $\mathcal{V}_{\mathcal{D}}$. Ainsi, comme précédemment, on utilise l'ensemble trié des résidus calculés sur les $M - d$ observations non utilisées par l'estimation du d -uplet. Le nombre d'observations, interprétées comme des contraintes, de pseudo-distances intersectant le volume $\mathcal{V}_{\mathcal{D}}$ correspond à l'ensemble des *inliers* \mathcal{D} utilisé pour le calcul de $NFA_2(\mathcal{D})$. La procédure de minimisation itérative (obtention de l'enveloppe inférieure des courbes $NFA_{\bar{X}_i}$ du critère NFA_2 est identique au premier critère NFA telle que décrite par la figure 5.1.

5.3.3 Exemple de la détection de droite

Nous allons illustrer la performance des deux critères NFA sur un exemple bien connu, à savoir l'estimation d'une droite [Desolneux et al., 2000]. Le d -uplet des paramètres recherchés est alors une paire de points du plan (deux paramètres $d = 2$). Trois ensembles de deux cents échantillons (2D) sont générés comme suit. Soit la droite d'équation $y = x$ qui sera la droite recherchée. Pour chaque ensemble d'échantillons, les *inliers* sont simulés en tirant aléatoirement des points appartenant à la droite tout en ajoutant un bruit gaussien indépendant sur chaque coordonnée de ces points, et les valeurs aberrantes sont simulées par tirage uniforme dans le plan. La figure 5.3 montre les points simulés et les résultats obtenus pour trois pourcentages de valeurs aberrantes : 20% (figure 5.3a), 50% (figure 5.3b) et 80% (figure 5.3c). La figure 5.3 permet la comparaison des résultats obtenus en utilisant l'estimateur des moindres carrés (LS), le M-estimateur [Knight and Wang, 2009] et les deux critères NFA proposés, respectivement. Pour le critère NFA_2 , la figure montre les limites de la courbe 2D définissant $\mathcal{V}_{\mathcal{D}}$ qui est dans ce cas une bande. Les performances de ces quatre estimateurs s'ordonnent de la façon suivante : l'estimateur LS peut estimer correctement la droite recherchée jusqu'à 20% de données aberrantes, le M-estimateur peut obtenir une estimation correcte jusqu'à 50% de valeurs aberrantes, et

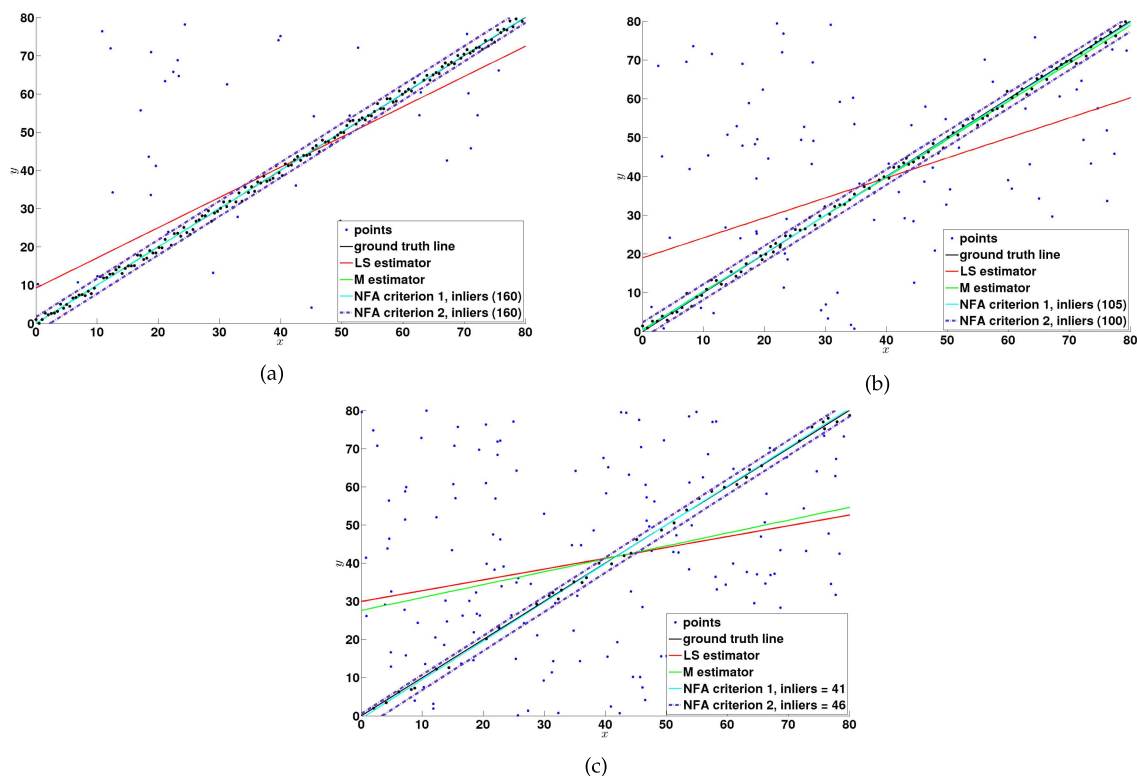


FIGURE 5.3 – Estimation d’une droite à partir d’un ensemble de 200 échantillons incluant un taux d’outliers de (a) 20% d’outliers, (b) 50% d’outliers et (c) 80% d’outliers ; les critères NFA sont robustes aux outliers jusqu’à un taux de 80% contrairement au M-estimateur et aux moindres carrés.

les deux critères NFA réussissent à estimer une solution correcte même en présence de 80% de valeurs aberrantes. Ces derniers critères fournissent des estimations très proches et exactes ou presque exactes des sous-ensembles d’inliers pour 20% et 50% de valeurs aberrantes et diffèrent légèrement pour 80% de valeurs aberrantes.

Comparons à présent les algorithmes proposés (critères NFA) avec le RANSAC dont nous avons vu (chapitre 3) qu’il est très populaire principalement en raison de sa simplicité. Ce dernier utilise un seuil pour partitionner l’ensemble des échantillons en un sous-ensemble d’inliers et un sous-ensemble d’outliers. Dans le cas d’un bruit gaussien, le seuil est généralement choisi égal à $1 \times \sigma$ où la valeur de σ est l’écart-type du bruit. Étant donné que le premier critère NFA implique également l’écart type du bruit (modèle naïf \mathcal{M}_1) comme paramètre d’entrée, nous étudierons la robustesse des deux méthodes à σ . La figure 5.4a retrace l’ensemble des échantillons incluant 80% d’outliers et les figures 5.4b, 5.4c et 5.4d montrent les résultats obtenus en faisant varier (axe des abscisses) σ . Nous observons que le RANSAC est robuste jusqu’à $\sigma \approx 10$, alors que le critère NFA_1 reste robuste pour toutes les valeurs de σ testées. Dans la figure 5.4b, nous voyons une ‘cassure’ dans la courbe indiquant la frontière entre le sous-ensemble d’inliers et le sous-ensemble d’outliers. Nous notons aussi que le nombre d’inliers trouvés par le critère NFA_1 est supérieur au nombre d’inliers simulés. Ceci est dû au fait que quelques outliers (environ 10%), qui ont été tirés aléatoirement selon une distribution uniforme sur l’ensemble du cadre, sont ‘tombés’ très proches de la droite simulée et sont vus comme inliers. Les paramètres de la droite étant notés (a, b) (avec comme ‘vérité’ $(1, 0)$), les figures 5.4c et 5.4d nous montrent que l’estimation de ces paramètres a et b par le RANSAC diverge à partir de $\sigma \approx 10$, alors qu’avec le critère

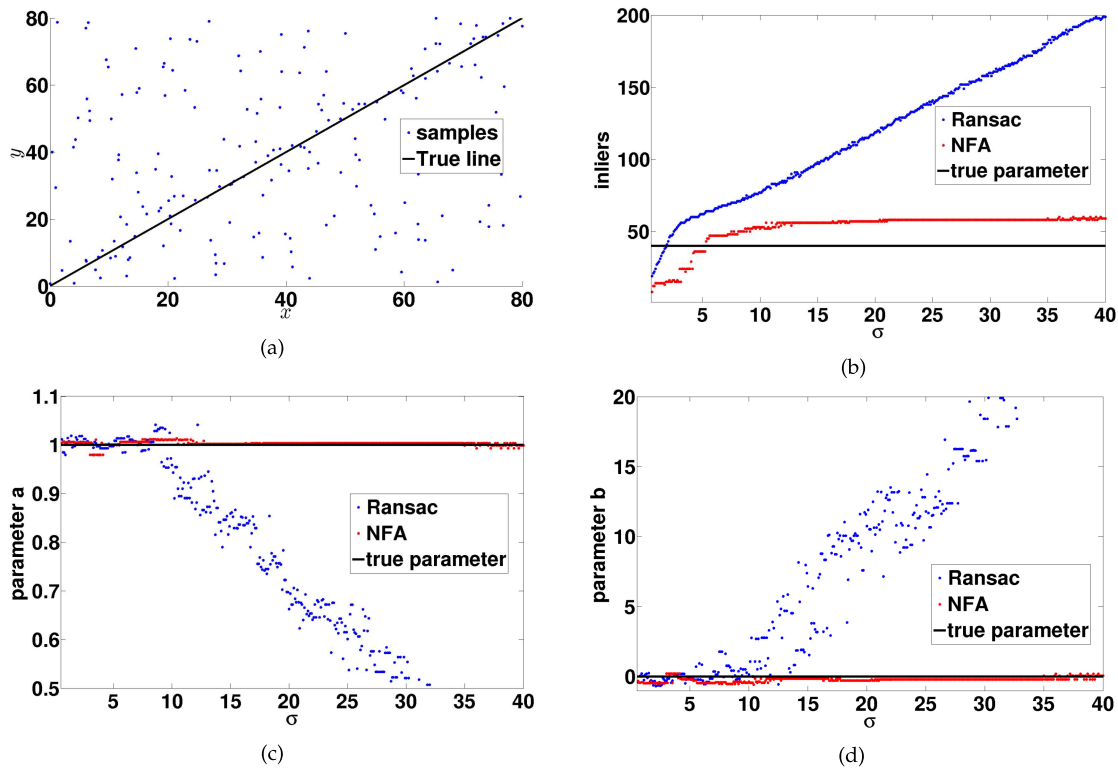


FIGURE 5.4 – Comparaison entre le RANSAC et le NFA dans le cas d’une estimation d’une droite $y = ax + b$ avec 80% d’outliers : (a) les échantillons, (b) cardinal du sous-ensemble d’inlier estimé, (c) estimation de la pente a , (d) estimation de l’ordonnée à l’origine b .

NFA_1 , leur valeur reste correctement estimée.

5.3.4 Spécificité pour la localisation GNSS

En vue de l’instanciation pratique des algorithmes proposés aux données GNSS, nous précisons trois points. Le premier concerne la spécificité du biais d’horloge, le second porte sur les aspects d’intégration temporelle et le troisième sur la sélection des mesures PR. Ces points sont particulièrement importants car, contrairement à l’exemple présenté dans la section 5.3.3 pour l’estimation de droite, le nombre de mesures PR disponibles à chaque *epoch* est faible, en particulier en milieu urbain.

Pour faire face à ce manque de données, certains auteurs utilisent une prédiction à partir du vecteur d’état de l’instant précédent. Par exemple, [Tzoreff and Bobrovsky, 2012] propose un modèle de trajectoire pour véhicule afin d’améliorer la localisation dans les zones urbaines. La méthode fournit des estimations précises même en cas de blocage des signaux satellites sur une courte durée, même avec un seul satellite en visibilité directe. Dans [Chang et al., 2009], un filtrage avec contrainte est proposé en considérant une vitesse et une altitude constante du véhicule lorsque il y a moins de deux satellites en visibilité directe. Dans ces travaux, deux méthodes de prédiction en fonction du comportement du véhicule ont été mises en œuvre, pour des mesures PR et des biais d’horloge respectivement, afin d’assurer la continuité de service du système de localisation.

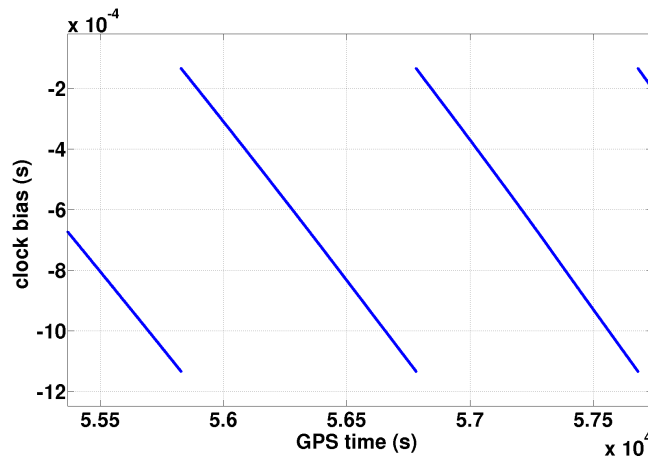


FIGURE 5.5 – Biais d’horloge en fonction du temps, cas du GPS Ublox EVK-5T.

Dans notre cas, plusieurs *epochs* sont considérées. Une fonction linéaire du premier ordre est utilisée pour prédire la position du récepteur à différentes *epochs*. En utilisant l’équation de prédiction (5.3), l’information fournie par une mesure acquise à une *epoch* précédente peut être utilisée conjointement à d’autres mesures du vecteur d’état à l’instant t (voir les équations (5.4) et (5.5)). En outre, l’utilisation de mesures acquises à différentes *epochs* agit comme un filtre temporel passe-bas et nous permet de bénéficier de la redondance temporelle des données.

5.3.4.1 Les sauts d’horloge

Comme annoncé au paragraphe précédent, nous allons considérer une fenêtre temporelle afin d’augmenter le nombre de mesures PR disponibles pour l’estimation du vecteur d’état X_t (cas statique) ou X_t^m (cas dynamique). En effet, même dans le cas statique où la position du récepteur est supposée constante, la quatrième coordonnée de X_t , à savoir le biais d’horloge δt , varie avec le temps. Dans ce travail, comme dans [Chang et al., 2009], nous supposons que le biais d’horloge suit un modèle linéaire avec saut comme illustré sur la figure 5.5. Les paramètres d’horloge doivent être estimés car ils sont spécifiques à chaque récepteur GNSS. En outre, des sauts d’horloge se produisent lorsque le récepteur réinitialise son horloge pour éviter que l’horloge interne du récepteur ne diffère trop de l’horloge satellite (ou *GPStime*).

Nous détectons un saut d’horloge à t comme une discontinuité par rapport au modèle linéaire supposé : en utilisant une régression classique, une estimation approximative du biais d’horloge à t est calculée afin d’être comparée aux valeurs précédentes de biais d’horloge. Si un saut est détecté, le biais d’horloge est réinitialisé en recalculant sa valeur. En effet, même si elle est grossière, cette estimation est utilisée dans l’ensemble des *epochs* où le saut d’horloge se produit. En effet, étant donné que les sauts d’horloge se produisent relativement rarement (toutes les 20 minutes dans notre cas), une telle approximation n’a pas d’incidence sur les résultats. En résumé, pour estimer les sauts d’horloge, nous ajoutons la dérive de l’horloge en plus du biais d’horloge dans le vecteur d’état, et ainsi la dimension de ce dernier (d) devient égal à 5 (position 3D, biais et dérive d’horloge) pour le cas statique et 8 (position 3D, vitesse 2D, biais et dérive d’horloge) pour le cas dynamique.

5.3.4.2 Régularisation sur une fenêtre temporelle

Dans le cas dynamique, pour chaque mesure PR acquise à t , le résidu est calculé à partir du vecteur d'état prédit et de l'équation (5.4) en utilisant le modèle de prédiction décrit par l'équation (5.3). À la fois en raison du fait que le vecteur d'état (et donc de l'espace des solutions) est de cardinal élevé et afin de bénéficier de la redondance temporelle, nous avons introduit un terme de régularisation. Par ailleurs, la régularisation permet à l'algorithme de localisation de ne pas diverger dans le cas où l'ensemble des *inliers* contiendrait des mesures erronées.

La technique de régularisation permet un meilleur conditionnement du problème et un lissage de la trajectoire du récepteur. Dans les algorithmes 4 ou 5, le vecteur d'état testé (d -uplet des paramètres inconnus) est estimé à partir du sous-ensemble des observations tiré aléatoirement pour le test considéré. En introduisant un terme de régularisation, l'estimation recherchée est obtenue en minimisant

$$\sum_{t_k \in \mathcal{T}} \sum_{i \in \mathcal{I}(t_k)} r_i^2(t) + \lambda^\top \left[\text{abs} \left(\tilde{\mathbf{X}}_t^m - \mathbf{X}_{t|t-1} \right) \right], \quad (5.10)$$

où $r_i(t)$ est le résidu fourni par l'équation (5.6), $\mathbf{X}_{t|t-1}$ est l'état prédit du récepteur selon l'équation (5.3), $\text{abs}(v)$ est le vecteur des valeurs absolues des composantes du vecteur v , et λ est le vecteur des coefficients de régularisation (λ est un vecteur de dimension d dont les coefficients contrôlent l'importance de la régularisation entre l'estimation de $\tilde{\mathbf{X}}_t$ et le vecteur d'état prédit $\mathbf{X}_{t|t-1}$). Comme dans la plupart des applications pratiques, le paramètre de régularisation est ajusté (ou appris) en utilisant un premier ensemble de données. Dans notre cas, nous n'avons pas régularisé le biais d'horloge et la dérive d'horloge, δt et $\dot{\delta t}$, de sorte que les coefficients correspondant dans λ sont fixés à 0. L'algorithme 6 décrit le processus d'estimation avec régularisation.

Algorithm 6: Estimateur non-linéaire avec terme de régularisation, avec comme entrées : la matrice \mathbf{Y} où les lignes sont les observations $(x_{S_i}, y_{S_i}, z_{S_i}, \rho_i)$, $\mathbf{X}_{t|t-1}$, λ et pour sortie $\tilde{\mathbf{X}}$

```

1 Fixer  $\varepsilon$  au seuil de précision ;
2 if  $\lambda = \mathbf{0}$  then
3   | Estimer  $\tilde{\mathbf{X}}$  avec l'algorithme de Gauss-Newton ;
4 else
5   | Initialiser  $\mathbf{X}$  et  $\tilde{\mathbf{X}}$  à  $\mathbf{X}_{t|t-1}$  ;
6   | Définir  $\rho$ , le vecteur des  $d$  mesures PR à partir de  $\mathbf{Y}$  ;
7   while  $\text{abs}(\tilde{\mathbf{X}} - \mathbf{X}) > \varepsilon$  do
8     |  $\mathbf{X} \leftarrow \tilde{\mathbf{X}}$  ;
9     | Calculer la jacobienne  $\mathbf{J}(\mathbf{X}, \mathbf{Y})$  (voir l'annexe A.2) ;
10    |  $\tilde{\mathbf{X}} \leftarrow (\mathbf{J}^\top \mathbf{J} + \lambda \mathbf{I})^{-1} \mathbf{J}^\top (\rho(\mathbf{X}, \mathbf{Y}) - \rho) + \mathbf{X}$  ;
11  end
12 end

```

Enfin, nous supposons que, pour l'ensemble des *epochs* de \mathcal{T} , l'altitude ne change pas de façon significative, hypothèse raisonnable pour la trajectoire d'un véhicule terrestre. Ainsi en posant la vitesse

en altitude \dot{z}_r à zéro, nous estimons sept paramètres dynamiques et nous effectuons la régularisation sur cinq de ces paramètres.

5.3.4.3 Stratégie de sélection des signaux satellites

Dans les deux algorithmes 4 et 5, la complexité de calcul est linéaire en N_{test} . Théoriquement, explorer l'ensemble des solutions du sous-espace nécessite $N_{test} = \binom{M}{d}$. Ce nombre augmente de façon exponentielle (selon l'approximation de Stirling) lorsque M augmente, c'est-à-dire suivant la taille de la fenêtre temporelle. Rappelons que cette dernière est d'autant plus élevée que le vecteur d'état est de grande dimension. Afin de diminuer ce nombre de tests, nous mettons en œuvre une stratégie de sélection des mesures PR qui évite le tirage de sous-ensembles d'observations mal conditionnée et permet de se concentrer sur les configurations favorables (beaucoup moins nombreuses). Cette stratégie repose sur plusieurs hypothèses.

La première hypothèse est que les *inliers* de l'estimation précédente ont une forte probabilité de rester des *inliers* pour l'estimation actuelle. En effet, si les mesures aberrantes sont principalement dues à la configuration géométrique de l'environnement par rapport au récepteur et aux satellites (par exemple des bâtiments qui produisent des NLOS, multi-trajets et/ou du blocage), ce dernier changeant assez lentement, les *inliers* sont partiellement corrélés dans le temps. Cette non-indépendance des *inliers* (ou des *outliers*) a été soulignée et exploitée par certains auteurs : par exemple [Sunderhauf et al., 2012] a proposé une approche basée sur les graphes pour modéliser et atténuer les erreurs de multi-trajets dans les mesures GNSS. Dans notre cas, si l'estimation est effectuée compte tenu de n_{ep} *epochs*, nous supposons que nous avons besoin d'une observation à la dernière *epoch* et les autres mesures PR sont choisies parmi les *inliers* correspondant aux *epochs* précédentes. D'un point de vue algorithmique, nous introduisons la matrice binaire \mathbf{O}_t , avec n_{ep} colonnes $\times N_{PR}$ lignes où N_{PR} est le nombre total de mesures PR disponibles pour le tirage aléatoire. Ainsi, $\mathbf{O}_t(i, j) = 1$ si, à l'instant $j \in \{1, \dots, n_{ep} - 1\}$, la mesure fournie par le satellite i appartient à \mathcal{D}_{t-1} (précédente série d'*inliers*), ou si $j = n_{ep}$, le satellite i fournit une mesure PR ; sinon $\mathbf{O}_t(i, j) = 0$.

La deuxième hypothèse porte sur les configurations permettant un meilleur conditionnement du système à résoudre (ligne 4 des algorithmes 4 et 5). Nous supposons que, dans les configurations les plus favorables, toutes les *epochs* et tous les satellites sont représentés. Nous voulons donc effectuer le tirage au sort de façon à avoir au moins une mesure par *epoch* et en favorisant la sélection des différents satellites. Plus précisément, soit π_0 et π_1 deux probabilités *a priori* avec $1 > \pi_0 > \pi_1 > 0$. Ces probabilités sont ajustées de façon à favoriser le tirage des différents satellites de la façon suivante : à une *epoch* donnée j , chaque mesure disponible (selon \mathbf{O}_t) est associée à une probabilité proportionnelle à π_0 ou π_1 selon si elle correspond à un satellite déjà sélectionné ou non. Précisément, avant d'être sélectionné un satellite a la probabilité $\alpha(j) \times \pi_0$ et après sélection, sa probabilité de sélection devient $\alpha(j) \times \pi_1$ (où $\alpha(j)$ est un facteur de normalisation de sorte que la somme des probabilités par colonne soit égale à un).

L'algorithme 7 donne la façon dont les mesures d'*inliers* sont tirées en fonction des contraintes citées. Pour des raisons pratiques, certaines variables intermédiaires ont été introduites : $\boldsymbol{\pi}$ le vecteur de dimension N_{PR} de la probabilité non normalisée de chaque satellite, $\boldsymbol{\psi}$ le vecteur de dimension $(n_{ep} - 1)$

Algorithm 7: Sélection des observations ; entrées : n_{ep} taille de la fenêtre temporelle, N_{PR} nombre maximum de mesures PR, O_t matrice des observations valides, d nombre d'inconnues ; sortie : $\{o_1, \dots, o_d\}$ l'ensemble des mesures sélectionnées.

```

1 Initialiser  $\pi : \forall i \in \{1, \dots, N_{PR}\}, \pi(i) = \pi_0$ ;
2 Initialiser le vecteur  $\psi$  du nombre d'observations par epoch, selon l'équation (5.11)  $\forall i \in \{1, n_{ep}\}$ ;
3  $\forall i \in \{1, n_{ep} - 1\}$ , mélanger les éléments de  $\psi$ ;
4  $l = 0$ ;
5 for  $j \leftarrow n_{ep}$  to 1 do
6   for  $k \leftarrow 1$  to  $\psi(j)$  do
7      $i \leftarrow$  tirage aléatoire non-uniforme ( $\pi$ ) parmi 1-valeur de  $O_t(\cdot, j)$ ;
8      $o_l \leftarrow$  observation  $i$  à l'instant  $j$ ;
9      $\pi(i) = \pi_1$ ;
10     $l \leftarrow l + 1$ ;
11  end
12 end

```

du nombre de mesures PR par *epoch* pour le tirage aléatoire, et $O_t(\cdot, j)$ la restriction de la matrice O_t à sa colonne j . Le vecteur ψ est calculé :

$$\psi(j) = 1, \forall j \in \{1, \dots, n_{ep} - 1\},$$

$$\psi(j) = \lfloor \frac{d-1}{n_{ep}-1} \rfloor + H_{ea}(j - ((d-1) \bmod (n_{ep}-1))), \quad (5.11)$$

où $\lfloor x \rfloor$ est la partie entière (le plus grand entier plus petit que x), $(a \bmod b)$ avec a et b deux entiers est le reste de la division euclidienne de a sur b et $H_{ea}(x)$ est la fonction *Heaviside* ($H(x) = 1 \forall x \geq 0, H(x) = 0 \forall x < 0$).

Afin d'illustrer le gain en termes de sélection aléatoire, supposons que la matrice O_t est remplie de 1 (ce qui veut dire que toutes les mesures PR sont disponibles). Supposons également que $N_{PR} > d$ et choisissons $\pi_1 \ll \pi_0$ de telle sorte qu'un satellite ne soit choisi qu'une seule fois (ou zéro). À chaque *epoch* j , $\psi(j)$ observations sont choisies aléatoirement, et le nombre de satellites restants (pour le prochain tirage) décroît de la même valeur $\psi(j)$. Finalement, le nombre de tests N_{test} , i.e. les sélections aléatoires possibles de l'ensemble des observations, est :

$$\phi(j) = N_{PR} - 1 - \sum_{i=1}^{j-1} \psi(i),$$

$$N_{test} = \binom{N_{PR}}{1} \binom{n_{ep}-1}{(d-1) \bmod (n_{ep}-1)} \prod_{j=1}^{n_{ep}-1} \binom{\phi(j)}{\psi(j)}. \quad (5.12)$$

Les détails des calculs sont donnés en annexe A.3. Par exemple, avec $d = 7, N_{PR} = 9$ et $n_{ep} = 4$, en absence de stratégie de sélection, le nombre total de combinaisons est $\binom{n_{ep} \times N_{PR}}{d} = 8347680$, alors

qu'avec notre stratégie $N_{test} = \binom{9}{1} \binom{8}{2} \binom{6}{2} \binom{4}{2} = 22680$, ce qui correspond à un rapport de 368. Notons aussi que cette valeur est généralement une limite supérieure du nombre de tests possible puisque la matrice O_t est généralement creuse.

5.3.4.4 Algorithme global et limitations pratiques

Dans les algorithmes proposés, nous avons besoin d'au moins $d = 7$ mesures (dans le cas dynamique sans estimation de la vitesse en z) par période d'intégration temporelle. Cependant, le blocage et le changement rapide des satellites en vue sont fréquents dans les zones urbaines. Dans ces cas, certaines stratégies doivent être mises en œuvre. Citons notamment :

- Si la fréquence d'échantillonnage est suffisamment élevée (à savoir $2Hz$ ou plus), il est possible d'augmenter la taille de la période d'intégration temporelle ;
- Nous pouvons supposer certaines composantes du vecteur d'état constantes, par exemple la dérive d'horloge, la vitesse du récepteur, l'altitude du récepteur, de sorte que l'estimation du reste des composantes du vecteur d'état (supposées non constantes) nécessite moins de mesures ;
- Enfin, si l'estimation du vecteur d'état n'est pas possible (par exemple pas de satellite en visibilité directe comme dans les tunnels), il demeure possible d'approcher le vecteur d'état par sa prédiction (ne serait-ce que ne pas arrêter le programme et assurer la continuité du service).

Dans notre cas, nous avons mis en œuvre la dernière stratégie.

Dans l'approche proposée, fondée sur la détection des valeurs aberrantes, nous supposons que le modèle de prédiction du premier ordre est suffisant pour 'aligner' (en temps) les mesures effectuées à différentes *epochs* (au cours de la période d'intégration temporelle). Cette hypothèse constitue la limite de notre approche. Si cette hypothèse s'était avérée incorrecte (ce qui ne fût pas notre cas pour les données acquises), il aurait fallu utiliser un modèle de prédiction plus élaboré (par exemple du second ordre qui, malheureusement, nécessite un plus grand nombre de paramètres) [Zhou and Li, 2015].

Une vue générale de notre algorithme est représentée en figure. 5.6. L'initialisation du vecteur X , notée X_0 , est estimée avec $\lambda = 0$, i.e. sans terme de régularisation.

5.4 Expérimentations et résultats

Dans ce paragraphe, nous allons valider nos algorithmes sur des données simulées et des données réelles. Ces données correspondent aux données présentées dans le chapitre 4, paragraphes 4.1.2, 4.1.3 et 4.1.4.

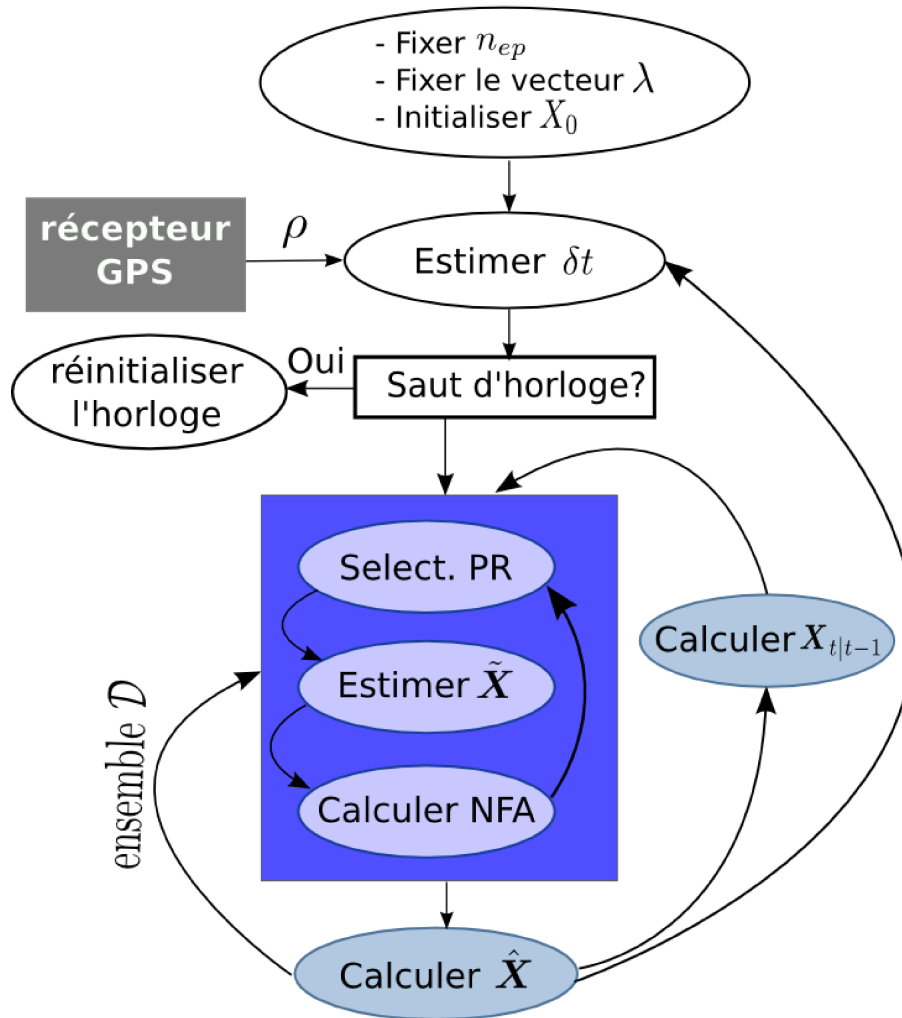


FIGURE 5.6 – Vue globale de l’algorithme.

5.4.1 Résultats obtenus avec les données simulées

Pour montrer l’intérêt du critère NFA, nous considérons les données acquises au cours d’une période de cinq secondes (voir la sous-section 5.3.4.1) Le d -uplet est alors un 5 -uplet puisque deux paramètres doivent maintenant être estimés pour la synchronisation temporelle entre le récepteur et les satellites.

5.4.1.1 Cas statique

Les résultats que nous présentons ici ont été obtenus avec le jeu de données 1. Étant en mode statique, nous avons introduit les contraintes $\dot{x}_r = \dot{y}_r = \dot{z}_r = 0$ dans le vecteur d’état ce qui réduit notre vecteur à $\mathbf{X}_i^s = (x_r, y_r, z_r, \delta t, \dot{\delta t})^\top$. La période considérée est égale à cinq secondes, ainsi chaque estimation de \mathbf{X}_i^s est réalisée avec un nombre maximum d’observations $M = 40$ PR (maximum 8 PR par *epoch*).

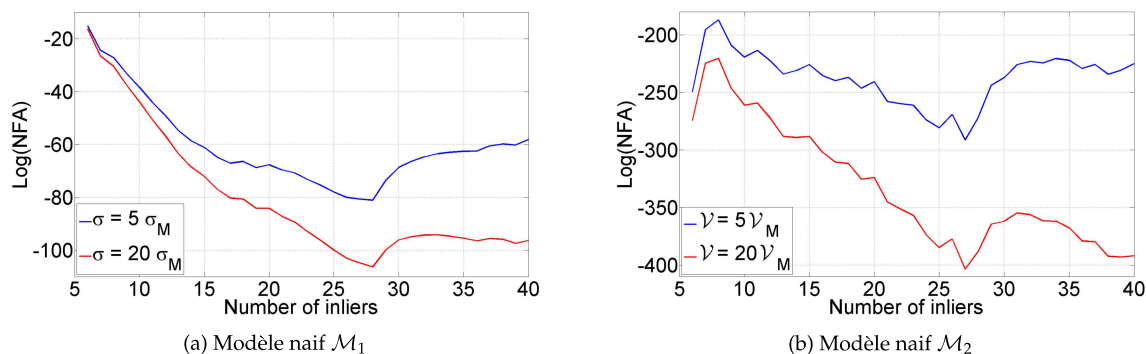


FIGURE 5.7 – Courbes $\log(NFA)$ en fonction du nombre d'inliers : (a) critère NFA_1 (b) critère NFA_2 ; le minimum global est clair et robuste au paramètre du modèle naïf (σ ou ν).

La figure 5.7 montre des exemples typiques de la fonction NFA ($NFA[i]$) dans les algorithmes 4 et 5) en fonction du cardinal de l'ensemble des *inliers* ($\in [d + 1, M]$). Il apparaît clairement que les courbes NFA décroissent jusqu'à un minimum global avant d'augmenter. Ce minimum global est obtenu pour le sous-ensemble d'observations PR les plus cohérentes d'après le critère NFA considéré. Les courbes NFA ont un minimum global clair et précis à l'inverse des fonctions croissantes de façon monotone représentant soit la somme des résidus aux carrés, soit le nombre de contraintes satisfaites (fonctions également tracées en fonction du cardinal de l'ensemble des *inliers*). Dans notre étude, le paramètre du modèle naïf, à savoir σ pour \mathcal{M}_1 et ν pour \mathcal{M}_2 , a été fixé de façon proportionnelle à σ_M et ν_M qui sont respectivement la variance de l'ensemble des données et le volume comprenant toutes les contraintes PR. Pour vérifier la robustesse du modèle par rapport au paramètre du modèle naïf, les courbes NFA obtenues en utilisant deux valeurs différentes de σ pour \mathcal{M}_1 et de ν pour \mathcal{M}_2 ont été tracées. Les figures 5.7a et 5.7b montrent que, dans les deux cas, le cardinal de l'ensemble des *inliers* qui permet d'atteindre le minimum global est le même. Il est de $|\mathcal{D}| = 28$, avec NFA_1 et de $|\mathcal{D}| = 27$ avec NFA_2 tandis que le nombre réel d'*inliers* est de 25. Cette petite différence est due au fait que certaines valeurs aberrantes simulées présentent un très faible biais et sont donc statistiquement compatibles avec le reste des données.

L'ensemble de simulation comprend 210000 échantillons (131250 *inliers* et 78750 données aberrantes simulées) partitionnés en 5250 périodes de 5s. Pour chaque période, 40 observations sont classées entre *inliers* et *outliers*. Le tableau 5.1 montre les valeurs obtenues suivant les mesures de performance des détections d'*inliers* et d'*outliers* présentées dans le chapitre 4.2.1. En comparant les deux critères, les résultats sont très proches, mais le critère NFA_2 apparaît légèrement plus prudent que le critère NFA_1 puisqu'il possède un nombre inférieur de faux positifs (*FP*, cf. chapitre 4) et un plus grand nombre de faux négatifs (*FN*). Dans le tableau 5.1, les valeurs d'*Accuracy* et de *Precision* (définies au chapitre 4) sont également données, confirmant la très bonne capacité des critères NFA à détecter les données aberrantes. Les valeurs du tableau 5.1 ont été calculées en fonction de la vérité terrain établie comme suit : toute observation tirée (pour ajouter un biais) est une valeur aberrante simulée, quelle que soit la valeur de biais ajouté. Selon cette définition, certains multi-trajets simulés ou mesures NLOS peuvent entrer dans le domaine des mesures LOS et ne pas être détectés comme des valeurs aberrantes avec un test statistique de cohérence. Cependant, en considérant ces statistiques comme des bornes inférieures, les valeurs obtenues sont plutôt satisfaisantes.

TABLE 5.1 – Performance des critères NFA_1 et NFA_2 pour le diagnostic des données aberrantes : cas statique. Les deux critères NFA conduisent à des taux élevés d'*Accuracy* et de *Precision*.

	<i>VP</i>	<i>FP</i>	<i>FN</i>	<i>VN</i>	<i>Accuracy</i>	<i>Precision</i>
NFA_1	122008	4528	9242	74222	93.44%	96.42%
NFA_2	118976	4220	12274	74530	92.15%	96.57%

TABLE 5.2 – Erreurs de localisation (en m) en coordonnées (Est,Nord,Up) dans le cas de données acquises en milieu dégagé avec des *outliers* simulés par l'ajout de biais (tirés aléatoirement) provenant de trois distributions différentes. Les meilleurs résultats sont indiqués en gras. Le critère NFA_1 réalise les meilleures performances.

		$Norm_1$	$RMSE$	$Norm_{inf}$
Uniforme	LS esti.	(13, 15, 32)	(16, 19, 40)	(52, 70, 153)
	M esti.	(9, 11, 24)	(13, 17, 34)	(64, 90, 199)
	NFA_1	(1.9, 2.9, 3.4)	(2.7, 3.7, 5.6)	(18, 41, 59)
	NFA_2	(2.5, 3.3, 3.8)	(3.5, 4.3, 7.1)	(37, 43, 101)
$\mathcal{N}(0, 50^2)$	LS esti.	(6.0, 11, 17)	(7.6, 14, 21)	(35, 63, 83)
	M esti.	(4.3, 6.5, 7.8)	(5.2, 8.4, 11)	(27, 48, 54)
	NFA_1	(5.5, 3.0, 8.5)	(5.8, 3.2, 9.3)	(8.5, 9.1, 14)
	NFA_2	(5.7, 2.9 , 8.9)	(6.0, 3.4, 10)	(8.2, 39, 75)
$\mathcal{N}(50, 15^2)$	LS esti.	(5.9, 11, 16)	(7.4, 14, 21)	(26, 45, 64)
	M esti.	(6.5, 12, 18)	(8.3, 15, 23)	(38, 48, 74)
	NFA_1	(5.6, 2.7, 8.7)	(5.9, 2.8, 9.7)	(13, 10, 84)
	NFA_2	(5.9, 4.8, 11)	(6.4, 7.4, 14)	(22, 46, 84)

Concernant la précision de la localisation, le tableau 5.2 montre les performances obtenues pour le jeu de données $5250 \times 5s$, en termes de $Norm_1$ (moyenne de la valeur absolue des résidus, cf. chapitre 4), $RMSE$ (racine carrée de la moyenne des résidus aux carrés) et $Norm_{inf}$ (maximum des résidus), en distinguant les trois cas de distribution de biais simulé. Les algorithmes proposés sont comparés au M-estimateur classique et à l'estimateur des moindres carrés (LS pour *Least squares*). Dans le tableau 5.2, nous voyons clairement que les meilleurs résultats sont obtenus en utilisant le critère NFA_1 (puis NFA_2), démontrant ainsi l'intérêt d'écarter les données aberrantes de l'ensemble des observations afin d'obtenir une position plus précise du récepteur. L'écart de performance entre les différentes approches varie en fonction de la distribution considérée du biais. En particulier, lorsque la distribution des biais est plus proche d'une distribution de bruit classique ($\mathcal{N}(0, 50^2)$), la performance du M-estimateur et l'estimateur LS est plus proche de la performance des critères NFA que dans les autres cas de statistique d'erreurs.

Enfin, la figure 5.8 nous montre les 1750 emplacements 2D, limité au cas d'un tirage uniforme du

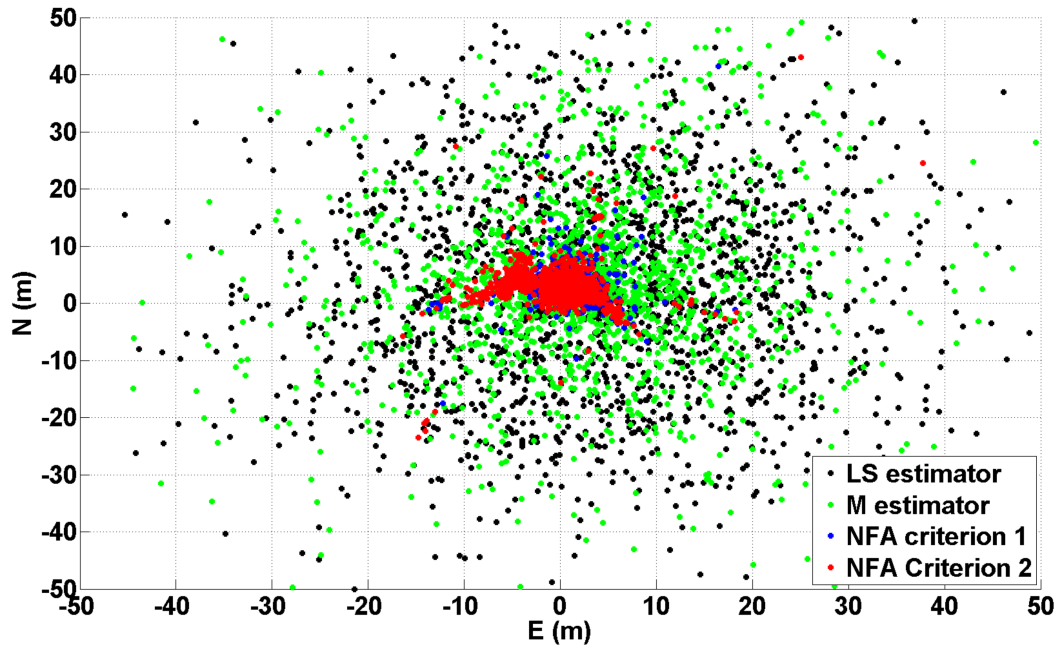


FIGURE 5.8 – Cluster de 1750 points représentant les positions estimées avec différentes méthodes. Le cluster des critères NFA est plus compact que celui de l’estimateur LS et celui du M-estimateur.

biais (vue du ciel, en se centrant autour de la vérité terrain en $(0,0)$) estimés par l’estimateur LS, le M-estimateur et par les algorithmes proposés. Nous notons que les clusters de l’estimateur LS et du M-estimateur sont beaucoup plus étalés que les clusters correspondant aux critères NFA_1 et NFA_2 . Plus précisément 97,0% des erreurs de localisation sont inférieures à 10m pour NFA_1 et de 96,5% pour NFA_2 . Les quelques points présentant une large erreur qui subsiste sont dus à une mauvaise configuration des satellites, par exemple un GDOP (*Geometric Dilution Of Precision*) élevé.

5.4.1.2 Cas dynamique

L’objectif de la seconde simulation est double. Il s’agit premièrement de valider l’approche proposée dans le cas dynamique et en particulier d’observer l’impact de l’historique des données sur la qualité de la localisation en termes d’*Accuracy*. Deuxièmement, nous souhaitons déterminer les paramètres de régularisation qui seront ensuite utilisés dans le cas de données réelles.

Nous cherchons d’abord la ‘meilleure’ longueur de fenêtre temporelle. Ce paramètre dépend étroitement de l’échantillonnage en temps (dans notre cas $T = 0.5s$) et de la redondance des données. Techniquement, une fenêtre temporelle de 1 ou 2 secondes est suffisante pour notre cas, que ce soit en ligne droite ou en virage serré. Nous proposons ici de faire une comparaison du taux d’exclusion des données aberrantes entre les NFA_1 et NFA_2 en fonction de la longueur de la fenêtre temporelle (une ou deux secondes) et trois différents biais sur les pseudo-distances, simulant ainsi des trajets multiples. Rappelons que la simulation pour le cas dynamique est réalisée en utilisant des données réelles corrigées, en changeant le biais d’horloge et la dérive et en ajoutant différentes réalisations d’un bruit gaussien sur les pseudo-distances, cf. chapitre 4.

TABLE 5.3 – Erreurs de localisations (en m) sur les données dynamiques simulées en coordonnées (Est,Nord).

norm	$Norm_1$	$RMSE$	moy, var
estimateur LS	(11.00, 13.55)	(13.76, 16.59)	(19.45, 9.28)
NFA_1	(1.87, 1.87)	(4.34, 3.22)	(2.94, 4.53)
NFA_2	(2.08, 2.08)	(6.95, 3.36)	(3.29, 6.98)

La figure 4.6, du chapitre 4, représente la position des satellites au cours des 2 minutes de test. Le tableau 5.4 montre le taux d'exclusion de mesures aberrantes pour chaque satellite en fonction des deux fenêtres temporelles ($T_1 = 1s$ correspondant à 3 *epochs* et $T_2 = 2s$ correspondant à 5 *epochs*) et de trois multi-trajets simulés. Nous notons que les satellites 9 et 10 ne sont pas testés en raison de certaines pertes de signal. Nous observons alors un taux d'exclusion élevé pour 10m de biais et une fenêtre temporelle de 3 *epochs* (supérieur à 75 % et 80 % pour les critères NFA_1 et NFA_2 respectivement) ainsi que pour une fenêtre temporelle de 5 *epochs* (> 85 % pour NFA_1 et NFA_2). Par ailleurs, les satellites sont plus ou moins fréquemment exclus selon la configuration, et nous observons qu'augmenter la longueur de la fenêtre temporelle augmente les taux d'exclusion des satellites. Dans la suite de nos expériences (y compris les données réelles), le temps d'intégration $T = 2s$ sera utilisé. En outre, nous observons que le NFA_2 a un plus grand taux d'exclusion que le NFA_1 , ce qui indique que le NFA_2 est plus contraignant (prudent) que le NFA_1 .

Le second test a pour objectif d'évaluer les performances en localisation des deux critères NFA_1 et NFA_2 . Nous avons simulé une trajectoire lissée en ligne droite où les observations sont bruitées ($\mathcal{N}(0, 3^2)$) avec trois observations par *epoch* entachées d'un bruit biaisé ($\mathcal{N}(50, 10^2)$), ce qui revient à intégrer quinze données aberrantes dans l'ensemble des données issues d'une fenêtre temporelle comportant cinq *epoch* (pour une fréquence d'acquisition de 2Hz). Le tableau 5.3 nous montre les statistiques de localisation pour l'estimateur des moindres carrés avec un terme de régularisation (voir l'algorithme 6), le NFA_1 et le NFA_2 . Le premier NFA présente une meilleure performance que le second dans tous les critères statistiques. D'après le tableau 5.4, nous avons déduit que le NFA_2 est moins précis que le NFA_1 et que le taux d'exclusion de données aberrantes du NFA_2 est plus important.

Le vecteur de régularisation λ est estimé en testant différents paramètres sur plusieurs trajectoires connaissant la vérité terrain. La figure 5.9 illustre un exemple de trajectoire avec en légende les paramètres de régularisation testés. Le meilleur vecteur de régularisation est choisi suivant la procédure suivante : les trajectoires correspondant à chaque vecteur de régularisation sont calculées, puis l'erreur euclidienne entre chaque estimation de position et la vérité terrain est calculée. Enfin pour chaque *epoch*, nous déterminons le 'meilleur' vecteur de régularisation en lui attribuant un score binaire (1 ou 0). Le vecteur qui obtient le meilleur score sur toutes les trajectoires testées sera choisi comme paramètre du vecteur λ durant l'expérimentation.

biais	10m				25m				40m			
	3		5		3		5		3		5	
satellites	NFA 1	NFA 2	NFA 1	NFA 2	NFA 1	NFA 2	NFA 1	NFA 2	NFA 1	NFA 2	NFA 1	NFA 2
1	77.20	81.67	88.67	91.07	94.27	96.56	97.22	98.10	99.88	100	100	100
4	77.43	80.98	86.36	87.75	95.07	96.79	98.02	97.89	99.54	99.77	100	100
6	80.29	82.82	88.80	89.77	89.58	96.56	97.01	97.72	99.65	99.88	100	100
11	78.35	82.47	89.94	91.92	96.33	96.67	97.30	98.02	99.77	100	99.87	100
17	76.63	83.96	88.12	91.45	91.75	95.30	98.40	99.03	100	100	100	100
20	77.78	81.90	88.76	90.94	94.96	96.68	97.64	98.48	99.65	99.88	99.80	100
23	76.40	80.98	87.83	89.98	94.84	96.79	96.55	97.81	99.31	99.54	99.75	100
31	78.01	82.93	85.18	87.37	93.81	95.65	97.68	98.15	99.65	99.88	100	100
32	75.83	81.90	87.37	89.14	94.16	96.68	97.31	98.02	99.31	99.77	100	100

TABLE 5.4 – Taux d'exclusion en % pour différents biais et différentes longueurs de fenêtre temporelle.

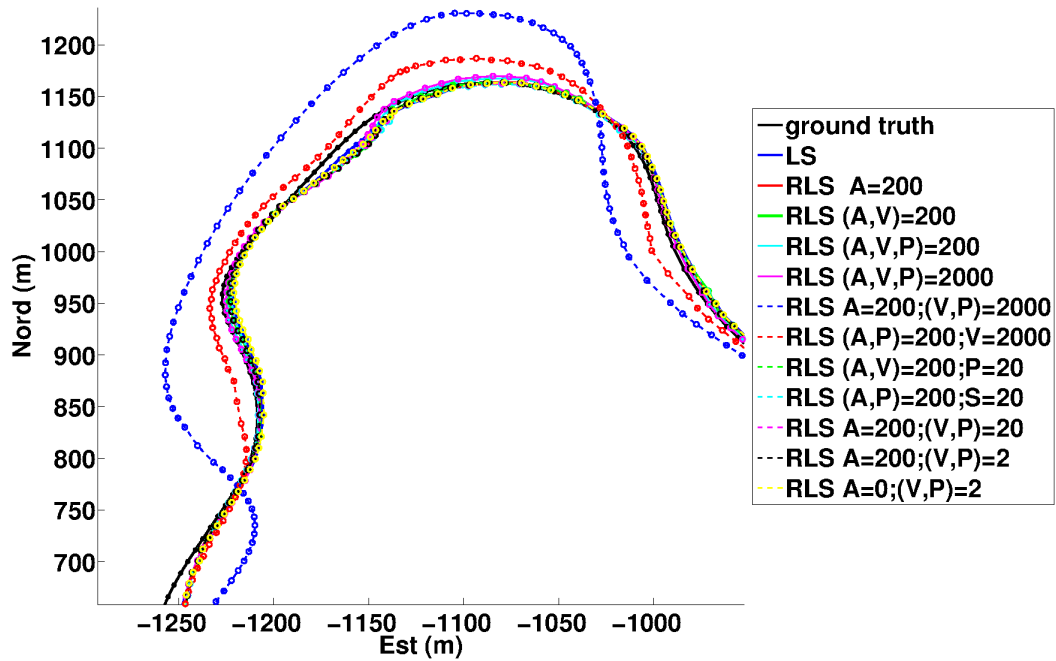


FIGURE 5.9 – Exemple de trajectoire utilisée pour estimer le vecteur de régularisation.

Afin de vérifier notre raisonnement, le tableau 5.5 nous montre quelques statistiques sur les données simulées dans le cas dynamique. Pour le taux d'exclusion ($1 - \frac{FP}{FP+VN}$), le NFA_2 est légèrement meilleur que le NFA_1 ($99.54\% > 98.36\%$). Par contre, le taux de faux négatifs (FN) du NFA_2 est 23 fois plus grand que celui du NFA_1 , ce qui induit une diminution de l'exactitude du NFA_2 .

Les paramètres fixés durant les simulations pour le cas dynamique (paramètres de régularisation, fréquence d'échantillonnage, longueur de la fenêtre temporelle) seront ceux utilisés pour les données réelles.

TABLE 5.5 – Performance des critères NFA_1 et NFA_2 pour le diagnostic des données aberrantes : cas dynamique. Les deux critères NFA conduisent à des taux élevés d'Accuracy et de Precision.

	VP	FP	FN	VN	Accuracy	Precision
NFA_1	791151	5824	4349	350426	99.11%	99.26%
NFA_2	693005	1630	102495	354620	90.95%	99.76%

5.4.2 Résultats obtenus avec les données réelles

Ayant validé notre approche sur des données simulées, nous appliquons notre méthode au cas des données réelles présentées au chapitre 4.

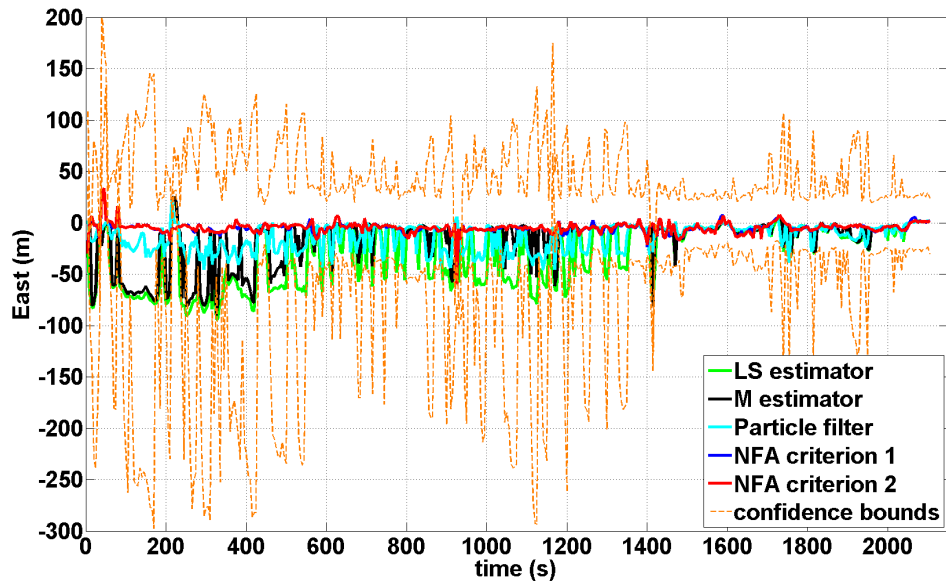
5.4.2.1 Cas statique

Le tableau 5.6 présente les taux d'exclusion pour chaque satellite calculés pendant deux périodes de 2 minutes respectivement au début et à la fin de l'expérience ($T_1 = [100 - 220]$ et $T_2 = [1750 - 1870]$, avec un échantillonnage d'une seconde). Le taux d'exclusion d'une mesure PR d'un satellite donné S_i est égal au nombre de mesures aberrantes parmi le nombre total de mesures PR de S_i sur la période considérée. En fonction du satellite, les taux d'exclusion sont plus ou moins variables au cours de l'expérience. Par exemple, au cours de T_1 , le satellite G2 est très souvent exclu (à la fois en fonction du critère NFA_1 et de celui NFA_2), mais pas les autres satellites. Pendant T_2 , les taux d'exclusion les plus élevés sont obtenus pour les satellites G12 et G24. Nous notons également que les taux d'exclusion de mesures provenant du satellite G25 diffèrent suivant le critère NFA_1 ou NFA_2 . En effet, il est à une élévation de $80,5^\circ$ induisant une faible précision sur la coordonnée Up et le critère NFA_2 a une plus grande tolérance envers cette coordonnée, car nous avons défini les ellipsoïdes homothétiques avec la dimension selon z égale à trois fois la dimension selon x ou y .

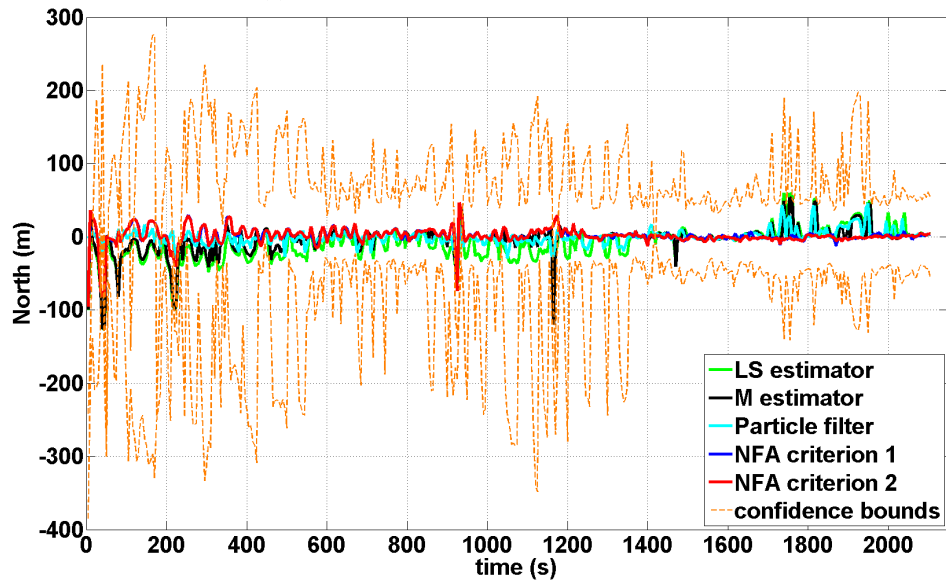
TABLE 5.6 – Taux d'exclusion des mesures satellites sur les périodes T_1 et T_2 .

Satellite	G2	G12	G14	G24	G25	G29	G31
T_1, \mathcal{M}_1	0.91	0	0	0	0.0083	0	0.058
T_1, \mathcal{M}_2	0.73	0.14	0.04	0.066	0.015	0.12	0.16
T_2, \mathcal{M}_1	0.155	0.425	0.135	0.32	0.34	0.05	0.01
T_2, \mathcal{M}_2	0.145	0.375	0.135	0.38	0.095	0.14	0.015

La figure 5.10 montre les erreurs pour les coordonnées Est et Nord en fonction du temps, en utilisant des approches différentes : l'estimateur LS, le M-estimateur, les deux critères NFA (1 et 2), le filtre particulaire et l'analyse par intervalles [Drevelle, 2011] combinée à la q -relaxation ($q = 2$). L'emplacement du récepteur est estimé à chaque période de 5s. Le fait que la précision de la localisation soit variable dans le temps est clairement illustré sur la figure 5.10a : jusqu'à $t = 500s$, le LS et le M-estimateur présentent un fort biais (environ $80m$) sur la coordonnée Est, puis les résultats du M-estimateur montrent une forte amélioration. Les critères NFA (1 ou 2) fournissent des estimations très précises de l'emplacement du récepteur, sauf à certains instants particuliers (par exemple $t = 30s, t = 950s$) où les autres méthodes échouent aussi à cause d'une configuration des satellites très défavorable. Le filtre particulaire estime les trois coordonnées ENU avec 3000 particules, la courbe de la coordonnée Nord est lisse et dépasse toutes les courbes des méthodes testées, à l'inverse la courbe de la coordonnée est bruitée la plupart du temps. Pour obtenir ces résultats, les paramètres du filtre (seuil de ré-échantillonnage, fonction distribution de la probabilité *a priori*) ont été ajustés manuellement connaissant la vérité terrain. L'analyse par intervalles [Drevelle, 2011] ne fournit qu'une estimation de la boîte (ensemble compact) dans laquelle la solution se trouve. Les bornes de ces boîtes sont tracées sur la figure 5.10. Puisque pour augmenter l'imprécision, il faut garder un nombre suffisant de données, nous définissons q égal à 2 en q -relaxation, même si dans ce jeu de données le nombre réel de valeurs aberrantes peut être supérieur (selon l'analyse *a posteriori* des résultats de l'algorithme NFA). Même si la vérité terrain (erreur égale à zéro) est incluse dans les bornes de confiance (ce qui est généralement interprété en termes de 'solution garantie'), dans la plupart des cas, la solution proposée par l'analyse par intervalles a des bornes larges et ne donne pas une solution précise.



(a) Erreur de localisation selon coordonnée Est.



(b) Erreur de localisation selon coordonnée Nord.

FIGURE 5.10 – Erreur de localisation en fonction du temps dans le cas du jeu de données acquis en milieu urbain. Les courbes associées aux critères NFA_1 et NFA_2 se superposent la plupart du temps et sont proches de l'axe $y = 0$ (représentant vérité terrain).

TABLE 5.7 – Erreurs de localisation (en m) des données réelles selon les coordonnées (Est, Nord, Up)

norm	$Norm_1$	$RMSE$	$Norm_{inf}$
LS estimator	(30, 15, 50)	(41, 22, 104)	(94, 104, 1684)
M estimator	(17, 10, 33)	(29, 18, 95)	(89, 119, 1672)
Particle Filter	(15, 6, 14)	(19, 9, 19)	(45, 45, 54)
NFA_1	(5, 6, 17)	(7, 11, 30)	(59, 55, 159)
NFA_2	(5, 6, 15)	(6, 11, 27)	(58, 61, 164)

TABLE 5.8 – Erreur de localisation en m en coordonnées (Est, Nord) sur le trajet en autoroute avec le jeu de données réelles. Comparaison avec les méthodes classiques de localisation.

	$Norm_1$	$RMSE$	μ_{dGT}	σ_{dGT}	$P^{-1}(.95)$
RLS	(5.2, 1.2)	(6.7, 1.6)	5.5	4.1	13.3
PF	(5.1, 2.3)	(7.0, 3.3)	6.1	4.8	13.8
Kalman	(5.5, 1.8)	(6.4, 3.5)	6.2	3.9	15
NFA_1	(2.6, 2.0)	(3.0, 2.9)	3.6	2.0	6.5
NFA_2	(2.7, 4.3)	(3.2, 6.0)	5.6	4.0	12.3

Comme pour le premier jeu de données et comme explicité au chapitre 4, la précision globale de la localisation est mesurée en termes de $Norm_1$, $RMSE$ et $Norm_{inf}$. Le tableau 5.7 montre les valeurs obtenues pour les coordonnées (Est, Nord, Up) pour les différentes méthodes : estimateur LS, M-estimateur, critères NFA (1 et 2) et filtre particulaire. Il semble qu'en fixant les paramètres adéquats, le filtre particulaire fournisse de meilleures estimations pour les coordonnées Nord et Up suivant les trois normes, les critères NFA fournissent la meilleure estimation en coordonnées Est ($Norm_1$ et $RMSE$) et une meilleure précision globale en 2D (13m pour les critères NFA_1 et NFA_2 , et 21m pour filtre particulaire).

5.4.2.2 Cas dynamique

Afin de quantifier l'intérêt de l'approche *a contrario*, nous comparons nos résultats avec le filtre de Kalman étendu (EKF), le filtre particulaire (PF) et l'estimateur RLS (moindres carrés régularisés). Le PF n'estime que six paramètres : deux coordonnées de position (Est et Nord), deux coordonnées de vitesse, et deux paramètres d'horloge : le biais et la dérive, car l'altitude est supposée connue afin de limiter la dimension de l'espace des solutions. L'EKF est également mis en œuvre avec les mêmes six paramètres. L'estimateur RLS est mis en œuvre tel que décrit par l'algorithme 6. Un inconvénient du PF est qu'un certain nombre de paramètres tels que le seuil de ré-échantillonnage, le nombre de particules ou la fonction de poids doivent être ajustés manuellement avec force tests pour obtenir de bons résultats. Dans cette étude, tous les paramètres de la méthode ont été ajustés sur les données simulées. Dans le cas du PF, cela a conduit à un choix de 3000 particules et l'utilisation de la technique de rééchantillonnage d'importance (SIR) [Arulampalam et al., 2002]. Dans le cas de la méthode NFA proposée, cela a conduit au choix d'une période de 5s et au vecteur de régularisation suivant : $\lambda^T = (200 \ 20 \ 200 \ 20 \ 10000 \ 20 \ 0 \ 0)$ (à savoir, l'altitude devrait très peu varier tandis que le biais d'horloge et la dérive ne sont pas régularisés du tout).

Étant donné que les contraintes sont très différentes en autoroute, dans la zone urbaine et dans la forêt, l'analyse des résultats est effectuée séparément.

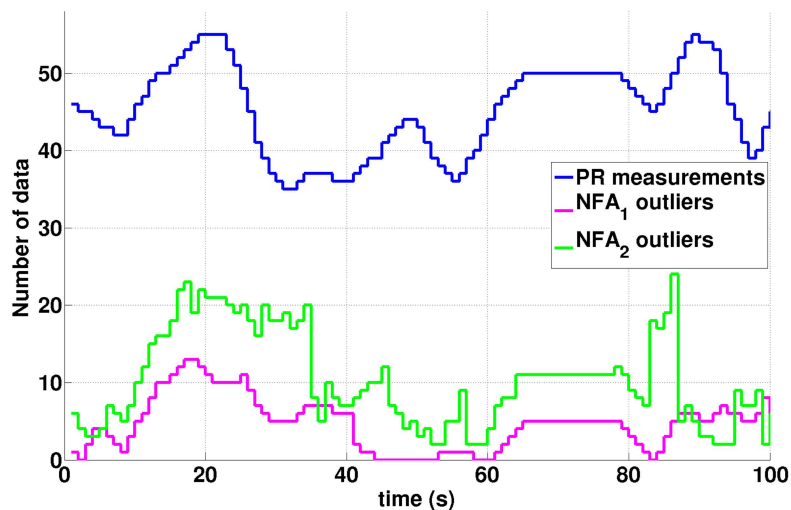
5.4.2.2.1 Trajet sur autoroute Commençons par examiner la partie de la route. Le tableau 5.8 montre les résultats statistiques de localisation selon les coordonnées (Est, Nord). Outre les deux normes $Norm_1$ et $Norm_2$, nous avons ajouté la moyenne (appelée μ_{DGT}), l'écart type (appelé σ_{DGT}) et la p -

valeur obtenue à partir de la fonction de distribution cumulative inverse (P^{-1}) pour 95% (p égal à 5 %) des distances euclidiennes entre la position estimée et la vérité terrain. Les conclusions sont les suivantes. Le critère NFA_1 offre de meilleures performances que toutes les autres méthodes testées que ce soit en termes d'*Accuracy* ou de *Precision*. Le critère NFA_2 montre des performances proches de celle du NFA_1 sur la coordonnée Est, mais la coordonnée Nord est estimée avec une moindre précision. Les résultats du PF sont assez proches de l'EKF et de l'estimateur RLS. Cette partie confirme le comportement correct des critères NFA par rapport aux méthodes de localisation traditionnelles, même dans un environnement ouvert.

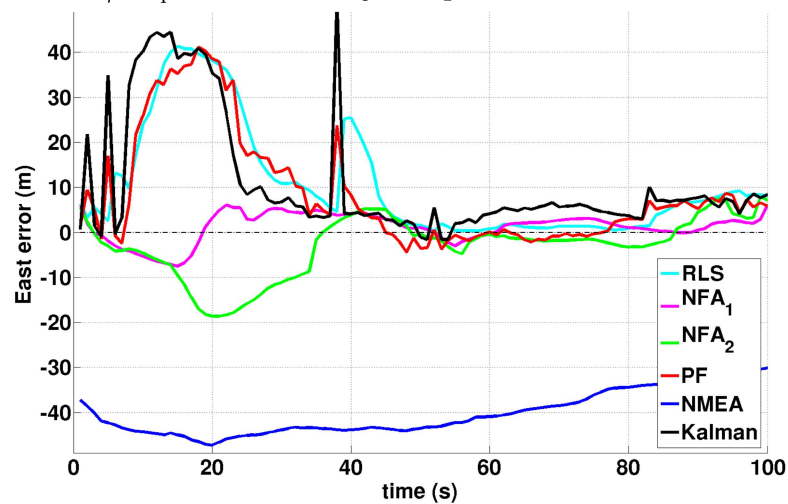
5.4.2.2.2 Trajet en zone urbaine Les canyons urbains se caractérisent par : (i) la présence d'un taux élevé de signaux satellites en NLOS qui devront alors être détectés et éliminés, (ii) un changement fréquent dans la configuration des satellites visibles et (iii) une mauvaise géométrie des satellites ou GDOP [Yarlagadda et al., 2000]. Dans notre expérience, le nombre de mesures PR disponibles pour une fenêtre temporelle de 5s se situe entre 35 et 55 (7 à 11 satellites visibles par *epoch*).

La figure 5.11 présente les résultats en termes de (a) nombre de données aberrantes, (b) précision de la coordonnées Est de la position du récepteur et (c) la précision de la coordonnée Nord. Tout d'abord, nous notons que les deux courbes représentant le nombre de mesures aberrantes selon les critères NFA_1 ou à NFA_2 (figure 5.11a) sont fortement corrélées. Toutefois, le critère NFA_2 apparaît plus prudent que le critère NFA_1 avec environ deux fois plus de mesures PR classées comme des *outliers*. Nous notons également que le nombre d'*outliers* apparaît partiellement corrélé à l'erreur (figure 5.11b et 5.11c) des méthodes d'estimation sans exclusion des mesures PR, en particulier, pour l'estimateur RLS. Cela souligne clairement l'intérêt de la suppression de ces mesures pour une localisation plus précise. L'amélioration est particulièrement sensible sur l'estimation de la coordonnée Est (voir la figure 5.11b) au début de la séquence temporelle présentée (instants entre 0 et 17 pour les deux critères NFA et jusqu'à 25 pour le critère NFA_1). Les satellites fournissant des mesures erronées ont été identifiés *a posteriori* comme les satellites 6, 10 et 17 (voir la figure 4.6). Les deux critères NFA ont réussi à les détecter. Toutefois, le critère NFA_2 supprime également d'autres mesures PR (les signaux du satellite 9 et les premières données des satellites 31 et 32) réduisant de manière significative le GDOP (voir la figure 4.6) et donc la fiabilité de l'estimation de la coordonnée Est. Un comportement similaire se produit pour l'erreur de la coordonnée Nord présentée sur la figure 5.11c entre les instants 60 et 80, où le critère NFA_1 rejette les données du satellite 11 et le critère NFA_2 élimine aussi les données du satellite 6.

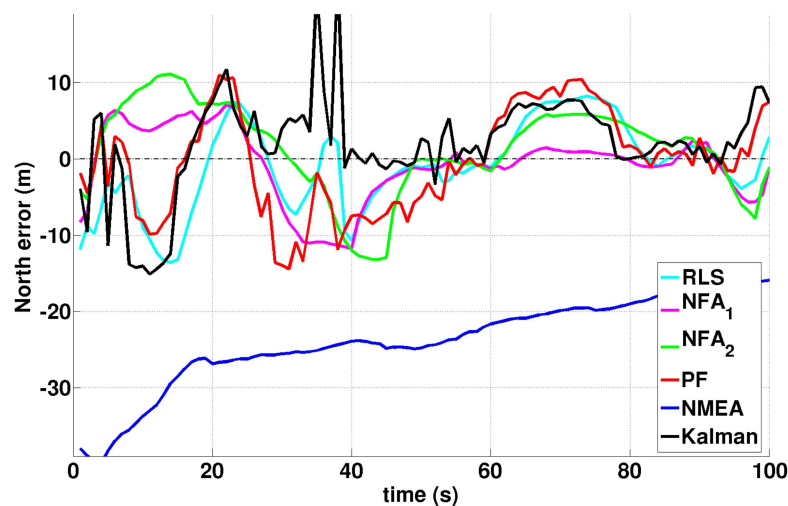
Pour l'évaluation quantitative, le tableau 5.10 nous montre les statistiques des erreurs de localisation en zone urbaine et permet la comparaison avec les méthodes classiques de localisation. Les erreurs obtenues avec les approches proposées sont plus précises que les méthodes classiques et beaucoup plus proches des erreurs obtenues en autoroute (c'est-à-dire en environnement dégagé où la plupart des méthodes fournissent une localisation de bonne précision, voir le tableau 5.8) que les autres méthodes testées (estimateur RLS, filtre PF et l'EKF). Le critère NFA_1 obtient les meilleures performances, le critère NFA_2 prend la deuxième place sans doute de par la surestimation des ensembles des *outliers* conduisant à un manque de redondance dans certains cas. Les méthodes classiques fournissent les résultats les plus piètres avec un écart-type plutôt élevé (supérieur à 10m). Enfin, l'intervalle de confiance à 95% est beaucoup plus faible (divisé par un facteur 2 ou 3) pour des résultats obtenus



(a) Nombre de mesures PR et d'*outliers* parmi les observations sur la fenêtre temporelle de 5 *epochs* pour les critères NFA_1 et NFA_2 .



(b) Erreur de localisation selon la coordonnée Est.



(c) Erreur de localisation selon la coordonnée Nord.

FIGURE 5.11 – Résultats expérimentaux en milieu urbain.

à l'aide de l'algorithme 4 ou 5 par rapport aux approches classiques.

Sur ce jeu de données en zone urbaine, nous avons aussi vérifié l'intérêt de la stratégie de sélection des mesures PR en comparant les résultats du tableau 5.10 avec ceux obtenus appliquant nos algorithmes sans aucune stratégie de sélection (tableau 5.9). Les performances réalisées sans stratégie de sélection avec $N_{Test} = 100000$ sont proches de nos résultats mais légèrement inférieures à celles obtenues en utilisant la stratégie de sélection des mesures PR et $N_{Test} = 5000$; par exemple, la valeur de la $Norm_1$, est respectivement égale à (3.21, 4.45) et (4.99, 5.05) pour les critères NFA_1 et NFA_2 au lieu de (3.11, 4.40) et (4.64, 4.99) (voir le tableau 5.10). Cela montre que, en plus des arguments sur le temps de calcul (qui est linéairement proportionnel à N_{Test}), la stratégie de sélection nous permet également d'obtenir des résultats précis et plus robustes, en se concentrant sur des configurations *a priori* plus favorables.

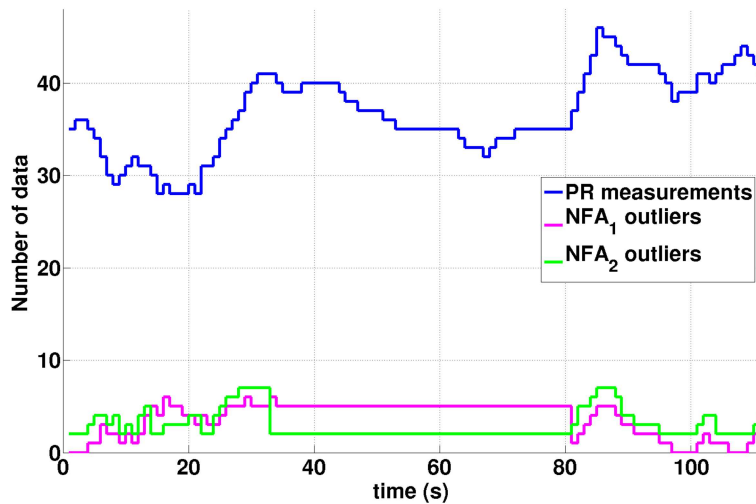
TABLE 5.9 – Erreur de localisation (en m) pour les coordonnées (Est, Nord) sur le trajet en milieu urbain avec le jeu de données réelles. Comparaison avec et sans la stratégie de sélection.

	norm	$Norm_1$	$RMSE$	μ_{dGT}	σ_{dGT}
5000 it.	NFA_1	(3.24, 4.46)	(4.11, 5.94)	6.07	3.93
	NFA_2	(4.99, 5.64)	(6.05, 7.80)	8.40	5.18
100000 it.	NFA_1	(3.21, 4.45)	(4.06, 5.94)	6.03	3.93
	NFA_2	(4.99, 5.05)	(6.03, 6.58)	7.84	4.26
PR sel., 5000 it.	NFA_1	(3.11, 4.40)	(3.97, 5.87)	5.93	3.88
	NFA_2	(4.64, 4.99)	(5.73, 6.55)	7.56	4.31

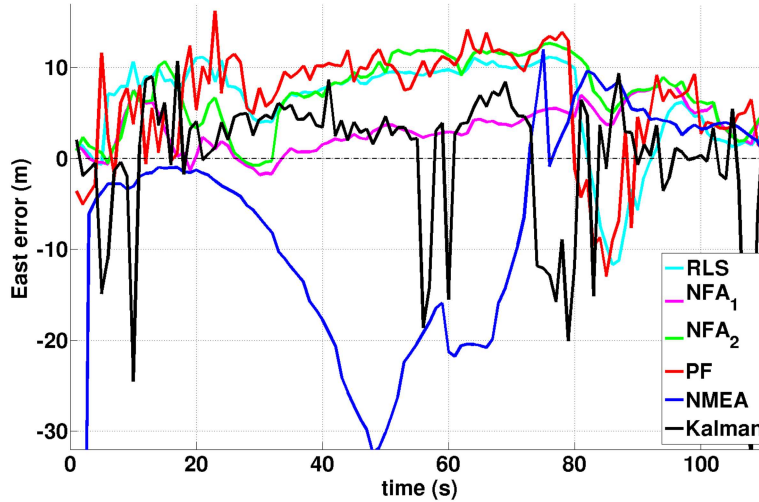
TABLE 5.10 – Erreur de localisation (en m) pour les coordonnées (Est, Nord) sur le trajet en milieu urbain avec le jeu de données réelles. Comparaison avec les méthodes classiques de localisation.

	$Norm_1$	$RMSE$	μ_{dGT}	σ_{dGT}	$P^{-1}(.95)$
UBX	(12.6, 10.1)	(21.1, 14.5)	17.0	19.17	53
RLS	(7.5, 7.5)	(10.7, 11.7)	11.62	10.76	36.1
PF	(7.0, 7.9)	(9.9, 11.3)	11.67	9.52	32.5
Kalman	(5.8, 5.1)	(8.4, 10.1)	8.76	9.80	40
NFA_1	(3.1, 4.4)	(4.0, 5.9)	5.93	3.88	14.05
NFA_2	(4.6, 5.0)	(5.7, 6.5)	7.56	4.31	15.7

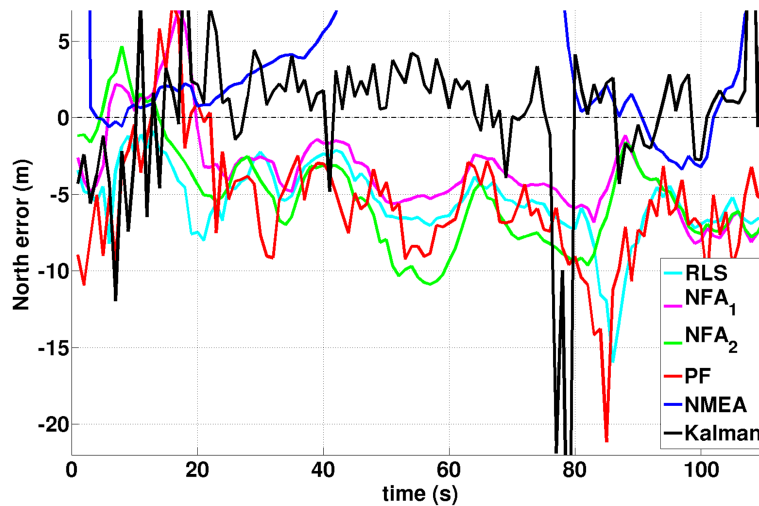
5.4.2.2.3 Trajet en forêt La figure 5.12 montre une sous-partie de l'expérience effectuée sur la zone forestière. Le nombre moyen de mesures PR disponibles par *epoch* varie entre 5 et 9, ou en d'autres termes, entre 28 et 46 mesures PR sont utilisées pour la localisation comme indiqué sur la figure 5.12a. En ce qui concerne les données aberrantes détectées (figure 5.12a), le critère NFA_1 et le critère NFA_2 sont beaucoup plus en accord entre eux qu'en zone urbaine. En ce qui concerne les erreurs de localisation sur les coordonnées Est et du Nord (respectivement, figure 5.12b et 5.12c), les conclusions sont proches de celles en zone urbaine : (i) les meilleures performances ont été réalisées par les critères NFA, en particulier sur la coordonnée Est avec le critère NFA_1 , et (ii) les approches classiques



(a) Nombre de mesures PR et d'outliers parmi les observations sur la fenêtre temporelle de 5 epochs pour les critères NFA_1 et NFA_2



(b) Erreur de localisation selon la coordonnée Est.



(c) Erreur de localisation selon la coordonnée Nord.

FIGURE 5.12 – Résultats expérimentaux sur la zone forestière.

présentent des pics élevés d'erreur qui peuvent s'expliquer en observant l'existence d'une corrélation partielle avec le nombre d'*outliers*.

TABLE 5.11 – Erreur de localisation (en m) pour les coordonnées (Est, Nord) sur le trajet en forêt avec le jeu de données réelles. Comparaison avec les méthodes classiques de localisation.

	$Norm_1$	$RMSE$	μ_{dGT}	σ_{dGT}	$P^{-1}(.95)$
UBX	(16.9, 17.5)	(26.3, 28.5)	25.08	29.63	> 100
RLS	(5.5, 4.5)	(6.7, 5.9)	7.84	4.31	15.4
PF	(6.2, 6.7)	(7.9, 8.4)	9.76	6.19	21.7
Kalman	(4.6, 4.1)	(6.4, 6.4)	7.07	5.90	18.3
NFA_1	(3.7, 4.4)	(5.3, 5.8)	6.38	4.61	16
NFA_2	(3.9, 5.0)	(5.4, 6.3)	7.11	4.28	17

En ce qui concerne l'évaluation quantitative de la précision de la localisation, les résultats sont présentés dans le tableau 5.11. Les conclusions sur la meilleure performance de l'approche à base de critère NFA par rapport à celles classiques restent vraies. Cependant, l'écart est plus faible. Cela peut être dû au fait que, dans les zones forestières, le phénomène dominant est le blocage des signaux satellite plutôt que la réception NLOS. De plus, les mesures PR sont moins nombreuses (voir la figure 5.12a courbe bleue) et le problème est moins un problème de détection des données aberrantes que de filtrage du bruit. Plus précisément, le RLS, PF et l'EKF fournissent des résultats légèrement meilleurs qu'en zone urbaine ce qui les rend proches des performances des critères NFA. Cependant, les résultats les plus surprenants sont ceux des trames NMEA qui sont très médiocres (plus médiocres que pour la zone urbaine) et l'estimateur RLS, qui fournit plutôt de bons résultats (beaucoup mieux qu'en zone urbaine). Ceci, en plus du fait que seulement quelques mesures PR sont détectées comme des données aberrantes, nous laisse penser que le processus de régularisation peut être une alternative intéressante à un filtrage classique.

5.5 Conclusion

Dans ce chapitre, nous avons considéré le problème de localisation exclusivement à partir des mesures PR des données GNSS. Une nouvelle approche robuste à un nombre important de valeurs aberrantes a été présentée. Elle s'appuie sur une modélisation *a contrario* de sorte que l'algorithme détecte le sous-ensemble de mesures pertinentes comme 'trop' cohérentes ou régulières pour apparaître par hasard (selon le modèle naïf représentant les données non structurées). Deux modèles naïfs ont été proposés. Le premier suppose une distribution gaussienne des résidus. Le second suppose une distribution binomiale pour le nombre de contraintes géométriques cohérentes. À partir de ces modèles, les critères NFA ont été formulés. Minimiser ces critères NFA nous a permis d'obtenir simultanément le sous-ensemble d'*outliers* et la position du récepteur.

Les premiers essais ont été effectués dans le cas d'un récepteur statique, en utilisant les données présentant un très faible bruit (environnement ouvert) avec les données aberrantes simulées afin de vérifier l'efficacité de l'approche proposée. Une expérimentation en milieu urbain en mode statique

a été menée en comparant notre approche avec des méthodes de localisation classiques telles que les M-estimateurs. Une deuxième expérimentation a été réalisée avec un récepteur en mouvement dans différents environnements (en autoroute, canyon urbain et en forêt), avec différentes conditions de réception des signaux et de configurations satellites, et différentes dynamiques de véhicule (en termes de vitesse et d'accélération). Les résultats montrent clairement que les meilleures performances sont atteintes par les algorithmes proposés par rapport aux approches classiques telles que le filtre de Kalman ou filtre particulaire, et soulignent l'intérêt de la suppression des valeurs aberrantes pour améliorer la précision en localisation.

Dans le chapitre suivant, nous introduirons les mesures Doppler dans le processus de localisation pour permettre une estimation précise et instantanée de la vitesse du récepteur. Un processus de fusion non linéaire des mesures de pseudo-distances et Doppler sera présenté. Les mesures Doppler sont plus robustes aux multi-trajets que les mesures TOA, cependant une analyse de la présence d'*outliers* dans les mesures Doppler est effectuée. Ainsi, après l'identification des mesures erronées, l'effet de retirer les mesures Doppler *outliers* du processus de localisation est présenté.

Filtrage particulaire et détection d'*outliers* dans les mesures Doppler

Sommaire

6.1	Introduction	91
6.2	Formulation du problème	92
6.2.1	Modèle d'observation	92
6.2.2	Problème de localisation	93
6.2.3	Le modèle dynamique	94
6.3	Approche proposée	96
6.3.1	Algorithme de détection d' <i>outliers</i>	97
6.3.2	Algorithme de localisation non-linéaire/non-gaussien	101
6.4	Expérimentation et résultats	106
6.4.1	Résultats de localisations	106
6.4.2	Validation des données aberrantes estimées	110
6.5	Conclusion	113

6.1 Introduction

Dans le chapitre précédent, une méthode de localisation robuste aux données aberrantes a été présentée. Nous avons utilisé exclusivement des mesures de pseudo-distance estimées à partir des données brutes GNSS. Les mesures Doppler, estimées à partir des données brutes GNSS, peuvent aussi être utilisées pour l'estimation instantanée de la vitesse en vue d'améliorer la localisation. Un algorithme de localisation basé sur l'UKF (*Unscented Kalman Filter*) utilisant les deux types observations PR et Doppler a été proposé [Mao et al., 2002]. Ces mesures Doppler sont particulièrement utiles dans des environnements contraints où le nombre d'observations disponibles est faible. La mesure Doppler est plus précise et plus fiable que la mesure de pseudo-distance du fait qu'elle est obtenue à partir du signal de la porteuse L1 de fréquence 1,575GHz alors que la mesure de pseudo-distance est issue du code C/A de fréquence 1,023MHz. Même si ces observations sont *a priori* plus robustes que les pseudo-distances, [Le Marchand et al., 2009] et [Xie and Petovello, 2015] ont montré expérimentalement que les mesures Doppler sont aussi affectées, dans une moindre mesure que les mesures PR, par les signaux multi-trajets et les réceptions en non-visibilité directe, ou NLOS.

La principale contribution de ce chapitre est de proposer un algorithme de localisation robuste qui utilise à la fois les mesures PR et les mesures Doppler. Il s'appuie sur des méthodes de traitement de signal connues et ayant été appliquées à d'autres problèmes ou d'autres données. Il comporte deux objectifs : (i) le partitionnement entre *inliers/outliers* sans paramètre de seuil ; (ii) le filtrage qui va permettre une localisation GPS robuste aux bruits et aux blocages intermittents des satellites.

Le premier est réalisé en formulant (comme au chapitre précédent) le problème en termes de minimisation d'un critère. Dans le précédent chapitre 5, nous avons défini et comparé deux critères NFA, et nous avons montré qu'ils sont plus efficaces que les tests statistiques classiques pour partitionner les mesures PR entre un ensemble cohérent de données (les *inliers*) et un ensemble de données incohérentes (les *outliers*). Toutefois, dans le chapitre précédent, seules des mesures PR avaient été considérées. Dans ce chapitre, nous proposons d'étendre notre approche en incluant les mesures Doppler.

Notre deuxième objectif (le filtrage robuste) est atteint en utilisant un filtre particulière pour filtrer le bruit présent dans les *inliers* (mesures PR et Doppler) en exploitant la redondance temporelle des observations. Cependant, dans sa version de base, ce type de filtre ne peut gérer les vecteurs d'état de grandes dimensions. Pour y remédier, nous proposons d'utiliser le filtre particulière Rao-Blackwellisé (RBPF) [Schon et al., 2005]. Son principe est de diviser le système en deux sous-systèmes, une partie linéaire et une non-linéaire, de sorte que la partie linéaire puisse être analytiquement résolue, tandis que la partie non-linéaire est estimée en utilisant la technique d'échantillonnage d'importance (comme dans le PF classique). Le RBPF a été appliqué avec succès à la navigation [Rabaoui et al., 2012], au suivi [Gustafsson et al., 2002] et à l'estimation des multi-trajets GPS [Giremus et al., 2007]. Les auteurs dans [Rabaoui et al., 2012], qui l'utilisent pour la localisation par GPS dans les canyons urbains, ne considèrent que les mesures PR, alors que nous proposons d'étendre ce filtre aux mesures PR et Doppler et de le coupler avec la détection des données aberrantes en utilisant le critère NFA issu de la modélisation *a contrario*.

La section 6.2 introduit les notations et les équations de base inhérentes au problème de la localisation en utilisant les données GNSS, pseudo-distance et Doppler. La section 7.3 décrit la méthode proposée qui implique une détection de données aberrantes dans l'ensemble du jeu de données en utilisant le critère NFA suivi par le processus de localisation GPS en utilisant les données brutes, soit en fonction du PF, soit du RBPF. Les résultats expérimentaux sont présentés dans la section 6.4.

6.2 Formulation du problème

6.2.1 Modèle d'observation

Nous considérons deux types d'informations fournies par chaque satellite S_j (de la constellation GPS). La première est la mesure de pseudo-distance ρ_j qui est liée à la distance entre le récepteur et le satellite S_j . Soit $\mathbf{x}_r^{ENU} = (e_r, n_r, u_r)^T$ la position du récepteur dans le système de coordonnées local ENU (Est, Nord, Up), avec en exposant T l'opérateur de transposition, et la position S_j est notée $\mathbf{x}_{S_j}^{ENU} = (e_{S_j}, n_{S_j}, u_{S_j})^T$. Le choix du système de coordonnées ENU, fait pour sa large utilisation en

navigation terrestre, nous permet de traiter la coordonnée 'Up', qui représente l'altitude, de façon indépendante. La pseudo-distance ρ_j dépend de x_r^{ENU} , $x_{S_j}^{ENU}$, δt qui est le biais temporel (différence) entre les deux horloges désynchronisées du satellite et du récepteur, c la vitesse de la lumière et du bruit aléatoire ε_j . Le modèle d'observation de la pseudo-distance s'écrit :

$$\begin{aligned}\rho_j &= \|x_r^{ENU} - x_{S_j}^{ENU}\| + c\delta t + \varepsilon_j, \\ &= \sqrt{(e_r - e_{S_j})^2 + (n_r - n_{S_j})^2 + (u_r - u_{S_j})^2} + c\delta t + \varepsilon_j\end{aligned}\quad (6.1)$$

L'équation (6.1) est une version simplifiée du modèle d'observation de pseudo-distance. Elle ne représente pas explicitement les multi-trajets ni les réceptions NLOS, de sorte que ces derniers puissent être détectés comme des écarts par rapport à ce modèle.

Le second type d'information est la mesure Doppler qui est liée à la vitesse du récepteur $\dot{x}_r^{ENU} = (\dot{e}_r, \dot{n}_r, \dot{u}_r)^\top$. En notant $\dot{x}_{S_j}^{ENU} = (\dot{e}_{S_j}, \dot{n}_{S_j}, \dot{u}_{S_j})^\top$, la vitesse du satellite S_j est déterminée en utilisant les émissions, d'éphémérides [Zhang et al., 2006]. Le modèle d'observation de la fréquence Doppler s'écrit :

$$\dot{\rho}_j = (\dot{x}_r^{ENU} - \dot{x}_{S_j}^{ENU}) \cdot a_j - c\dot{\delta t} + \varepsilon'_j, \quad (6.2)$$

où $\dot{\rho}_j$, aussi appelé "PR rate", est égal à $c \frac{D_j}{f_{L1}}$, avec $f_{L1} = 1575,42.10^6 \text{ Hz}$ et où les D_j sont les mesures Doppler (en Hz) fournies par S_j , a_j est le vecteur unitaire colinéaire avec la droite passant par le récepteur et le satellite S_j ($a_j = \frac{x_{S_j}^{ENU} - x_r^{ENU}}{\|x_{S_j}^{ENU} - x_r^{ENU}\|}$), "." désigne le produit scalaire, $\dot{\delta t}$ la dérive d'horloge et ε'_j un bruit aléatoire.

6.2.2 Problème de localisation

Pour estimer la position du récepteur, différents systèmes d'équations peuvent être considérés selon les données utilisées : les mesures PR, les mesures Doppler ou les deux à la fois.

En n'utilisant que les mesures PR, quatre observations au minimum sont nécessaires pour estimer le vecteur x_r^{ENU} et le biais temporel δt via la résolution du système d'équations (6.1), déjà traité dans le chapitre précédent.

Si on ne considère que les mesures Doppler, théoriquement, le vecteur \dot{x}_r^{ENU} , la dérive d'horloge $\dot{\delta t}$ et le vecteur \dot{x}_r^{ENU} pourraient être estimés, car ils apparaissent dans l'équation (6.2). Cependant, il existe deux obstacles à cette approche : (i) pour une estimation à un instant donné, au moins sept observations provenant de différents satellites seraient nécessaires, ce qui est incompatible avec la robustesse au blocage satellite dans des environnements contraints ; (ii) la sensibilité de l'équation (6.2) au vecteur \dot{x}_r^{ENU} est faible, car \dot{x}_r^{ENU} intervient seulement dans le vecteur unitaire a_j . Ainsi, en pratique, les mesures PR sont également utilisées pour calculer une estimation du vecteur \dot{x}_r^{ENU} et par la suite une estimation de a_j , notée \tilde{a}_j , qui est introduite dans l'équation (6.2) :

$$\dot{\rho}_j + \dot{\mathbf{x}}_{S_j}^{ENU} \cdot \tilde{\mathbf{a}}_j = \dot{\mathbf{x}}_r^{ENU} \cdot \tilde{\mathbf{a}}_j - c\delta t + \varepsilon'_j \quad (6.3)$$

La troisième et dernière approche consiste à considérer simultanément des mesures PR et Doppler. Le vecteur $\xi_1 = (e_r, \dot{e}_r, n_r, \dot{n}_r, u_r, \dot{u}_r, \delta t, \dot{\delta t})^\top$ contient l'ensemble des paramètres impliqués dans les équations (6.1) et (6.2). En ayant linéarisé les équations (6.1) et (6.2), la résolution du système qui s'en déduit (ayant au moins huit équations) peut être réalisée à l'aide de l'algorithme itératif de Gauss-Newton. Plus précisément, si \mathbf{Y}_S et $\dot{\mathbf{Y}}_S$ désignent les matrices regroupant les vecteurs $\mathbf{x}_{S_j}^{ENU}$ et $\dot{\mathbf{x}}_{S_j}^{ENU}$, respectivement, l'incrément $\delta \hat{\xi}_1^{(k)}$ à additionner à l'estimation précédente $\xi_1^{(k-1)}$ (k étant l'indice de l'itération) est :

$$\delta \hat{\xi}_1^{(k)} = \arg \min_{\delta \xi_1} \|(z - z^{(k-1)}) - \mathbf{H}(\mathbf{Y}_S, \dot{\mathbf{Y}}_S, \xi_1^{(k-1)}) \delta \xi_1\|, \quad (6.4)$$

où $z = (\rho_1 \cdots \rho_n, \dot{\rho}_1 \cdots \dot{\rho}_n)^\top$ est le vecteur des observations, $z^{(k-1)}$ est l'estimation de z calculée depuis l'estimation précédente (itération $k-1$) du vecteur d'état $\xi_1^{(k-1)}$ et \mathbf{H} est la matrice jacobienne du système d'équations formé par les équations (6.1) et (6.2) (se référer à [Li et al., 2011] pour plus de détails).

6.2.3 Le modèle dynamique

Afin d'augmenter la robustesse de l'estimation, il est possible de prendre en compte plusieurs *epochs* (comme dans le chapitre 5). Les données acquises à différentes *epochs* doivent être liées par le biais d'un modèle. Dans [Zhou and Li, 2015], les auteurs proposent un modèle dynamique polynomial lissé sur un intervalle de temps. En se limitant au premier ordre, les mesures PR sont reliées à l'aide d'un modèle dynamique impliquant la position et la vitesse du récepteur, de sorte que le vecteur ξ_1 , présenté auparavant, coïncide avec le modèle dynamique (voir le chapitre 5.2). Toutefois, de part la présence de mesures Doppler à différents instants, l'accélération doit être introduite dans le modèle dynamique, et le vecteur d'état considéré devient $\xi_2 = (e_r, \dot{e}_r, \ddot{e}_r, n_r, \dot{n}_r, \ddot{n}_r, u_r, \dot{u}_r, \delta t, \dot{\delta t})^\top$.

Les vecteurs d'état, que ce soit ξ_1 ou ξ_2 , à des instants différents sont linéairement reliés par des matrices de transition $\mathbf{M}_{i,dt}$ des modèles dynamiques considérés, définies comme suit :

$$\begin{aligned} \mathbf{C}_{dt} &= \begin{pmatrix} 1 & dt \\ 0 & 1 \end{pmatrix}, \mathbf{0}_{2 \times 2} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \mathbf{D}_{dt} &= \begin{pmatrix} \frac{dt^2}{2} & 0 \\ dt & 0 \end{pmatrix}, \mathbf{I}_{2 \times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned} \quad (6.5)$$

$$\mathbf{M}_{1,dt} = \begin{pmatrix} \mathbf{C}_{dt} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{C}_{dt} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{C}_{dt} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{C}_{dt} \end{pmatrix} \quad (6.6)$$

$$\mathbf{M}_{2,dt} = \begin{pmatrix} \mathbf{C}_{dt} & \mathbf{D}_{dt} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{I}_{2 \times 2} & \mathbf{D}_{dt}^\tau & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{C}_{dt} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{C}_{dt} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{0}_{2 \times 2} & \mathbf{C}_{dt} \end{pmatrix} \quad (6.7)$$

où l'exposant τ désigne l'opérateur de transposition anti diagonal (la transposée de la matrice par rapport à l'anti-diagonale). Soit $\xi_{i,t}$, $i \in \{1, 2\}$, le vecteur d'état à l'instant t qui est donné par :

$$\xi_{i,t+dt} = \mathbf{M}_{i,dt} \xi_{i,t} + o(dt^{i+1}), \quad (6.8)$$

où $o(dt^{i+1})$ est l'erreur (approximation) du modèle dynamique considéré.

En utilisant le modèle dynamique de l'équation (6.8), nous sommes maintenant en mesure de calculer les estimations des différentes observations (PR ou Doppler) à des instants différents. Plus précisément, soit $\mathcal{T} = \{t_k = t + kdt, k \in \{0, \dots, n_{ep} - 1\}\}$ l'ensemble des *epochs* considérées pour l'estimation de la solution, $\mathbf{x}_{S_j}^{ENU}(t_k) = (e_{S_j}(t_k), n_{S_j}(t_k), u_{S_j}(t_k))^\top$. La position du satellite S_j à l'instant t_k et l'estimation du PR $\tilde{\rho}_j(t_k)$ associé à S_j à t_k pouvant être calculées à partir de ξ_1 :

$$\tilde{\rho}_j(t_k | \xi_{1,t}) = \left\| \left[\mathbf{M}_{1,kdt} \cdot \xi_{1,t} \right]_{1,3,5} - \mathbf{x}_{S_j}^{ENU}(t_k) \right\| + c(\delta t - \dot{\delta} t k dt), \quad (6.9)$$

où l'indice dans la notation $[]_{l_1, l_2, l_3}$ représente la restriction de la matrice ou du vecteur pour les lignes l_1 , l_2 et l_3 et $\|v\|$ est la norme du vecteur v . L'estimation du PR peut aussi être déduite du deuxième vecteur d'état ξ_2 par l'équation :

$$\tilde{\rho}_j(t_k | \xi_{2,t}) = \left\| \left[\mathbf{M}_{2,kdt} \cdot \xi_{2,t} \right]_{1,4,7} - \mathbf{x}_{S_j}^{ENU}(t_k) \right\| + c(\delta t - \dot{\delta} t k dt). \quad (6.10)$$

De façon similaire, l'estimation de la mesure Doppler à t_k issue du satellite S_j peut être déduite de ξ_2 :

$$\begin{aligned} \mathbf{a}_{j,t_k} &= - \frac{\left[\mathbf{M}_{2,kdt} \cdot \xi_{2,t} \right]_{1,4,7} - \mathbf{x}_{S_j}^{ENU}(t_k)}{\left\| \left[\mathbf{M}_{2,kdt} \cdot \xi_{2,t} \right]_{1,4,7} - \mathbf{x}_{S_j}^{ENU}(t_k) \right\|}, \\ \tilde{\rho}_j(t_k | \xi_{2,t}) &= \left(\left[\mathbf{M}_{2,kdt} \cdot \xi_{2,t} \right]_{2,5,8} - \mathbf{x}_{S_j}^{ENU}(t_k) \right) \cdot \mathbf{a}_{j,t_k} - c \dot{\delta} t. \end{aligned} \quad (6.11)$$

En utilisant la régression classique, les vecteurs d'états optimaux $\hat{\xi}_{1,t}$ et $\hat{\xi}_{2,t}$ sont ceux qui minimisent les erreurs quadratiques :

$$\hat{\xi}_{1,t} = \arg \min_{\xi_{1,t}} \sum_{t_k \in \mathcal{T}} \sum_{i \in \mathcal{I}(t_k)} \left[\tilde{\rho}_j(t_k | \xi_{1,t}) - \rho_j(t_k) \right]^2, \quad (6.12)$$

$$\hat{\xi}_{2,t} = \arg \min_{\xi_{2,t}} \sum_{t_k \in \mathcal{T}} \sum_{i \in \mathcal{I}(t_k)} \left[\left[\tilde{\rho}_j(t_k | \xi_{2,t}) - \rho_j(t_k) \right]^2 + \beta \left[\tilde{\rho}_j(t_k | \xi_{2,t}) - \dot{\rho}_j(t_k) \right]^2 \right]. \quad (6.13)$$

où $\mathcal{I}(t_k)$ est l'ensemble des indices des satellites fournissant des mesures à t_k , et β est un facteur de pondération entre les résidus associés aux mesures PR et Doppler, respectivement.

Dans les équations précédentes, la minimisation est effectuée en prenant en compte toutes les mesures (PR et/ou Doppler) disponibles pour l'ensemble d'epochs \mathcal{T} considéré. Cependant, certaines de ces mesures peuvent correspondre à des valeurs aberrantes, et pourraient alors biaiser l'estimation. Dans la section suivante, en plus de l'acronyme PR, nous utilisons les abréviations "Dp" pour "mesure Doppler" et "(PR, Dp)" pour "à la fois des mesures de pseudo-distance et Doppler".

6.3 Approche proposée

Comme vu au chapitre 3, en présence de données aberrantes, plusieurs stratégies ont été proposées dans la littérature. Les méthodes robustes visent à atténuer automatiquement le poids de ces données aberrantes dans le processus d'estimation. Notamment, et puisque nous l'utiliserons dans ce chapitre, rappelons que le filtre particulaire ou ses variantes appartenant à la classe des estimateurs robustes peuvent théoriquement faire face aux valeurs aberrantes simplement en faisant varier les poids des particules générées. Toutefois, si ce filtre a prouvé son efficacité contre le bruit, nous allons voir que trop de données aberrantes compromettent la stabilité du filtre. Dans le cas du traitement des données GPS, des tests statistiques ont été proposés pour détecter les données aberrantes, e.g. [Knight and Wang, 2009]. La façon la plus simple pour faire face à ces données aberrantes est de les retirer de l'ensemble des mesures acquises (comme si les satellites correspondants étaient bloqués). Telle est la stratégie employée dans la technique de détection et d'exclusion de défaut ou *Fault Detection and Exclusion* (FDE) mise en œuvre dans les récepteurs GPS (avec cependant comme limite, dans la version implémentée, de ne considérer qu'au plus une mesure erronée [Le Marchand et al., 2008]). Des stratégies plus sophistiquées ont également été proposées, par exemple [Giremus et al., 2007, Cheng et al., 2016], qui visent à corriger les valeurs aberrantes.

Dans ce chapitre, nous considérons les questions de base suivantes :

- Pour le problème de localisation, est-ce que les mesures Doppler sont moins sujettes aux données aberrantes que les mesures PR ?
- La présence de données aberrantes impacte-t-elle également les algorithmes de localisation robustes, tels que le filtre particulaire ou le PF Rao-Blackwellisé ?
- Dans le cas affirmatif, cela vaut-il la peine de détecter et de rejeter ces données aberrantes ?

Dans l'algorithme de localisation, il est possible d'ajouter une étape de détection de données aberrantes qui va sélectionner les données, parmi celles disponibles, à prendre en compte dans l'estimation du vecteur d'état. Plus précisément, les algorithmes de filtrage s'articulent en deux étapes, la prédiction et l'estimation. L'étape de détection des données aberrantes est alors insérée avant l'étape d'estimation.

Pour notre étude (en particulier, le partitionnement entre *inliers/outliers*), les données aberrantes sont recherchées, soit seulement dans l'ensemble des mesures PR, soit aussi dans l'ensemble de mesures Dp, selon l'hypothèse faite que les mesures Dp sont plus robustes, ou ne le sont pas plus, que les PR :

- Si les mesures Doppler sont supposées fiables comme dans [Kaplan and Hegarty, 2006], elles sont directement utilisées pour calculer \dot{x}_r , et la détection des données aberrantes est effectuée uniquement dans le jeu de données PR ;
- Sinon, nous supposons, comme [Le Marchand et al., 2009], que même si les mesures Doppler sont moins affectées par la réception NLOS que les PR (et donc, plus fiables), les deux observations Doppler et PR sont affectées par des multi-trajets. La détection des données aberrantes est donc appliquée à (PR, Dp), de sorte que seuls les *inliers* Dp sont utilisés pour calculer \dot{x}_r (et seulement les *inliers* PR sont pris en compte pour l'étape d'estimation suivante).

6.3.1 Algorithme de détection d'*outliers*

Dans le chapitre précédent, nous avons proposé et comparé deux modèles "naïfs" conduisant à deux critères NFA pour séparer les données entre *inliers* et *outliers*. Cependant, ces modèles ne traitent que les mesures PR. Dans cette étude, nous étendons le premier critère NFA aux mesures (PR, Dp). Le choix de ce premier critère (plutôt que le second) a été motivé par le fait qu'il conduisait expérimentalement (d'après les tests effectués au chapitre précédent) à des résultats légèrement meilleurs.

Avant de présenter l'algorithme étendu, précisons les équations utilisées et les notations. En supposant une valeur de ξ_i notée $\tilde{\xi}_i$ et les caractéristiques des satellites (position et vitesse), nous sommes en mesure de calculer, en utilisant les équations (6.9) et (6.11), une estimation des valeurs PR ou Doppler à chaque *epoch*. Nous pouvons comparer ces estimations aux valeurs réellement observées. Par définition, les résidus sont les différences entre les mesures calculées (sous l'hypothèse de $\tilde{\xi}_i$) et celles observées : les $r_j(t_k)$, associés aux mesures PR à t_k , $\rho_j(t_k)$, sont :

$$r_j(t_k) = [\tilde{\rho}_j(t_k|\tilde{\xi}_i) - \rho_j(t_k)] . \quad (6.14)$$

où $\tilde{\rho}_j(t_k|\tilde{\xi}_i)$ est calculé avec l'équation (6.9) ou (6.10) et $\dot{r}_j(t_k)$, associé aux mesures Dp à t_k , $\dot{\rho}_j(t_k)$, est donné par :

$$\dot{r}_j(t_k) = [\tilde{\dot{\rho}}_j(t_k|\tilde{\xi}_i) - \dot{\rho}_j(t_k)] . \quad (6.15)$$

où $\tilde{\dot{\rho}}_j(t_k|\tilde{\xi}_i)$ est calculé avec l'équation (6.11).

Afin de donner le même poids à ces deux types de mesures, PR et Dp, $r_j(t_k)$ et $\dot{r}_j(t_k)$ sont normalisés par leurs écart-types respectifs, σ_{PR} et σ_{Dp} , et concaténés dans le vecteur \mathbf{R} :

$$\mathbf{R} = \left(\frac{r_1(t_{n_{ep}-1})}{\sigma_{PR}}, \dots, \frac{r_m(t)}{\sigma_{PR}}, \frac{\dot{r}_{m+1}(t_{n_{ep}-1})}{\sigma_{Dp}}, \dots, \frac{\dot{r}_M(t)}{\sigma_{Dp}} \right), \quad (6.16)$$

avec M le cardinal de l'ensemble (PR,Dp).

Pour les équations (6.12) et (6.13), plusieurs *epochs* sont considérées. Ceci nous permet d'augmenter le nombre de données disponibles, ainsi que la qualité de l'estimation, à condition que le modèle dynamique (les équations (6.9), (6.10) et (6.11)) utilisé pour 'aligner' les données acquises à différentes *epochs* soit suffisamment précis. Le nombre d'*epochs* à considérer, n_{ep} , est alors un compromis entre la disponibilité des données et l'approximation du modèle dynamique. Dans ce qui suit, \mathcal{S}_{PR} et \mathcal{S}_{Dp} désignent les ensembles d'observations disponibles (PR et Dp, respectivement) sur l'intervalle d'*epochs* considéré \mathcal{T} .

Considérons maintenant un sous-ensemble de mesures noté \mathcal{D} dans l'ensemble des observations $\{\mathcal{S}_{PR}, \mathcal{S}_{Dp}\}$. Sachant \mathcal{D} et \mathbf{R} (voir équation (6.16)), $\delta_{\mathcal{D}}^2$ est défini comme étant la somme des carrés des composantes de \mathbf{R} pour les indices j appartenant à \mathcal{D} (en effet, $\{\mathcal{S}_{PR}, \mathcal{S}_{Dp}\}$ étant indexés, \mathcal{D} peut être aussi bien vu comme un ensemble d'indices). Comme au chapitre précédent et dans [Zair et al., 2016a], $\delta_{\mathcal{D}}^2$ nous permet de quantifier la cohérence de \mathcal{D} en termes de mesure NFA (associée au modèle naïf gaussien $\mathcal{N}(0, \sigma)$) :

$$NFA(\mathcal{D}) = \eta_1 \frac{1}{\Gamma\left(\frac{|\mathcal{D}|}{2}\right)} \int_0^{\delta_{\mathcal{D}}^2/2\sigma^2} e^{-t} t^{\frac{|\mathcal{D}|}{2}-1} dt, \quad (6.17)$$

où Γ est la fonction gamma, $|\cdot|$ est le cardinal d'un ensemble et η_1 est un terme de normalisation qui nous permet de contrôler le nombre moyen de fausses alarmes [Desolneux et al., 2007].

L'algorithme 8 présente la version étendue de l'algorithme 4 du chapitre 5 qui nous permet de trouver le sous-ensemble \mathcal{D} minimisant le critère NFA. Ici, les paramètres d'entrée sont les ensembles d'observation \mathcal{S}_{PR} et \mathcal{S}_{Dp} (avec \mathcal{S}_{Dp} éventuellement vide si les mesures Dp ne sont pas présent en compte), le nombre d'itérations N_{test} , le paramètre σ du modèle naïf \mathcal{M} , les écarts-types, σ_{PR} et σ_{Dp} , pour la normalisation des résidus et le paramètre binaire I_{Dp} qui est égal à 0 ou 1, selon les données traitées : mesures PR seulement ou mesures (PR,Dp), respectivement. Les sorties de l'algorithme sont l'ensemble \mathcal{D} des *inliers* et le vecteur $\tilde{\xi}_i$ estimé.

Suivant le principe de *l'a contrario* RANSAC (e.g., [Rabin et al., 2010]), l'algorithme effectue différentes estimations et tests (boucle jusqu'à N_{test}) afin de choisir la meilleure solution suivant le critère NFA. Précisément, pour chaque test, il effectue les trois étapes suivantes. Tout d'abord, l'étape de sélection des données consiste en un tirage aléatoire de $d = 8$ éléments dans \mathcal{S}_{PR} (l'ensemble des observations PR) ou, si $I_{Dp} \neq 0$, $d = 10$ éléments dans \mathcal{S}_{PR} et \mathcal{S}_{Dp} (l'ensemble des observations Dp). Les nombres 8 et 10 correspondent aux nombres minimum d'observations pour estimer $\tilde{\xi}_1$ ou $\tilde{\xi}_2$. \mathcal{S}_{PR} et \mathcal{S}_{Dp} incluent toutes les observations disponibles effectuées au cours de la fenêtre temporelle des

Algorithm 8: Estimation du vecteur $\tilde{\xi}_i$ basée sur le critère NFA avec comme entrées $\mathcal{S}_{PR}, \mathcal{S}_{Dp}, \sigma, \sigma_{PR}, \sigma_{Dp}, I_{Dp}, t, N_{test}$ et comme sorties $\tilde{\xi}_i$ et \mathcal{D} .

```

1 if  $I_{Dp} \neq 0$  then
2   |  $d \leftarrow 10; M \leftarrow 2|\mathcal{S}_{PR}|;$ 
3 else
4   |  $d \leftarrow 8; M \leftarrow |\mathcal{S}_{PR}|;$ 
5 end
6 Initialiser les vecteurs  $\delta_{min}[j]$  et  $NFA[j]$  et le scalaire  $NFA_{min}$  à  $+\infty$ ;
7 for 1 to  $N_{test}$  do
8   | if  $I_{Dp} \neq 0$  then
9     | Depuis  $\mathcal{S}_{PR}$  et  $\mathcal{S}_{Dp}$ , respectivement, tirer aléatoirement  $\frac{d}{2}$  observations  $(x_{S_j}, y_{S_j}, z_{S_j}, \rho_j)$ 
10    | et  $\frac{d}{2}$  observations  $(\dot{x}_{S_j}, \dot{y}_{S_j}, \dot{z}_{S_j}, \dot{\rho}_j)$  notées  $\{o_1, \dots, o_{\frac{d}{2}}\}$  et  $\{\dot{o}_1, \dots, \dot{o}_{\frac{d}{2}}\}$ , respectivement;
11    | A partir de  $\{o_1, \dots, o_{\frac{d}{2}}, \dot{o}_1, \dots, \dot{o}_{\frac{d}{2}}\}$ , estimer  $\tilde{\xi}_2$  avec l'équation (6.13) ou l'équation (6.19)
12    | pour une solution régularisée;
13    | Avec l'équation (6.14), (6.15) et (6.16), calculer le vecteur des résidus normalisé noté  $\mathbf{R}$ ;
14   | else
15     | Depuis  $\mathcal{S}_{PR}$ , tirer aléatoirement  $d$  observations  $(x_{S_j}, y_{S_j}, z_{S_j}, \rho_j)$ , notées par  $\{o_1, \dots, o_d\}$ ;
16     | A partir de  $\{o_1, \dots, o_d\}$ , estimer  $\tilde{\xi}_1$  avec l'équation (6.12) ou l'équation (6.18) pour une
17     | solution régularisée;
18     | Avec l'équation (6.14), calculer le vecteur des résidus noté  $\mathbf{R}$ ;
19   | end
20   | Trier les valeurs du vecteur  $\mathbf{R}$  de façon croissante dans  $\mathbf{R}' = (r'_j)_{j \in \{1 \dots M\}}$  avec  $\pi(\cdot)$  la
21   | fonction des permutations d'indice, tel-que  $\forall j \in \{1 \dots M\}, r'_j = R_{\pi(j)}$ ;
22   | Initialiser  $\delta_{\mathcal{D}}^2 = 0$ ;
23   | for  $j = d + 1$  to  $M$  do
24     |  $\delta_{\mathcal{D}}^2 \leftarrow \delta_{\mathcal{D}}^2 + (r'_j)^2$ ;
25     | if  $\delta_{\mathcal{D}}^2 < \delta_{min}[j]$  then
26       |  $\delta_{min}[j] = \delta_{\mathcal{D}}^2$ ;
27       | Calculer  $NFA[j]$  selon l'équation (6.17);
28       | if  $NFA[j] < NFA_{min}$  then
29         |  $NFA_{min} = NFA[j]; \mathcal{D} = \{\pi(1) \dots \pi(j)\}$ ;
30       | end
31     | end
32   | end
33 end
34 À partir de l'ensemble des inliers  $\mathcal{D}$ , estimer  $\hat{\xi}_1$  ou  $\hat{\xi}_2$  avec l'équation (6.20).

```

n_{ep} dernières *epochs*. Comme dans le chapitre 5, le tirage aléatoire des observations est effectué de manière à favoriser le tirage des configurations de satellites favorables. Comme nous utilisons une fenêtre glissante sur des *epochs*, il y a un chevauchement entre les ensembles d'*epochs* considérées pour l'estimation à deux instants successifs. Par conséquent, à partir du traitement de l'instant précédent,

nous connaissons les *inliers* correspondants aux précédentes $n_{ep} - 1$ *epochs*. Comme dans le précédent chapitre, le tirage aléatoire est alors contraint de sorte que : (i) il existe au moins une mesure par *epoch*; (ii) pour les $n_{ep} - 1$ premières *epochs* de \mathcal{T} , les mesures PR/Doppler sont choisies parmi les *inliers* déjà détectés; (iii) la sélection de différents satellites est favorisée comme cela a été présenté dans la sous-section 5.3.4.3 du chapitre précédent.

Ces d observations sont utilisées pour obtenir une solution préliminaire $\tilde{\xi}_1$ ou $\tilde{\xi}_2$ (suivant la valeur de I_{Dp}). Pour estimer cette solution, un terme de régularisation peut être ajouté aux équations (6.9), (6.10) ou (6.11), permettant à la fois un meilleur conditionnement du problème et une trajectoire du récepteur plus lissée. Compte tenu du terme de régularisation, au lieu de l'équation (6.12), nous devons résoudre l'équation (6.18) :

$$\tilde{\xi}_1 = \arg \min_{\xi_1} \sum_{j \in \{1, \dots, d\}} (\tilde{o}_j(\xi_1) - o_j)^2 + \lambda_1^T \left[\text{abs} \left(\xi_1 - \xi_{1,t|t-1} \right) \right], \quad (6.18)$$

et au lieu de l'équation (6.13), nous résolvons l'équation (6.19) :

$$\tilde{\xi}_2 = \arg \min_{\xi_2} \sum_{j \in \{1, \dots, \frac{d}{2}\}} \left[(\tilde{o}_j(\xi_2) - o_j)^2 + \beta (\tilde{o}_j(\xi_2) - \hat{o}_j)^2 \right] + \lambda_2^T \left[\text{abs} \left(\xi_2 - \xi_{2,t|t-1} \right) \right]. \quad (6.19)$$

Dans l'équation (6.18) et l'équation (6.19), $\xi_{i,t|t-1}$, $i \in \{1, 2\}$, est le vecteur d'état prédit selon le modèle dynamique (équation (6.8)), $\text{abs}(v)$ retourne le vecteur des valeurs absolues des composantes de v et λ_i , $i \in \{1, 2\}$, est le vecteur des paramètres de régularisation (λ pondère l'importance de l'écart entre l'estimé $\tilde{\xi}_i$ et le vecteur d'état prédit $\xi_{i,t|t-1}$). L'annexe A.4 détaille le calcul de $\xi_{i,t|t-1}$, $i \in \{1, 2\}$.

Finalement, le sous-ensemble d'*inliers* est l'ensemble \mathcal{D} correspondant à la valeur NFA_{min} , et le vecteur d'état $\tilde{\xi}_1$ ou $\tilde{\xi}_2$ est estimé à partir de \mathcal{D} et de l'équation (6.20) :

$$\tilde{\xi}_i = \arg \min_{\xi_i} \sum_{j \in \mathcal{D}} R_j + \lambda_i^T \left[\text{abs} \left(\xi_i - \xi_{i,t|t-1} \right) \right], \quad (6.20)$$

où R_j est le résidu fourni par l'équation (6.16).

L'algorithme 8 a une complexité linéaire en fonction du nombre d'itérations N_{test} . Pour chaque itération, la complexité est essentiellement due à l'étape d'estimation du vecteur d'état (algorithme 12 donné en annexe). La complexité de cette dernière dépend de d puisque l'inversion de matrice et la multiplication de matrices sont en $O(d^3)$. La complexité du tri des termes de \mathbf{R} est en $O(M \log(M))$. Pour NFA(PR,Dp), $d = 10$ et M varie entre [12, 33] pour une fenêtre temporelle de 3 *epochs*. Par conséquent, pour contrôler le temps de calcul, il faut adapter le paramètre N_{test} .

Enfin, notons que, même si l'algorithme 8 fournit des estimations des paramètres de localisation du récepteur GNSS, le couplage proposé entre l'algorithme 8 et l'algorithme de localisation robuste (PF/ RBPF présenté dans la section suivante) se fait uniquement en termes de sélection des observations. En effet, dans l'algorithme 8, l'estimation fournie vise uniquement à évaluer la cohérence d'un

sous-ensemble de données, alors que le PF/RBPF permet le filtrage des données non-linéaires/non gaussiennes en exploitant une hypothèse *a priori* sur la régularité des trajectoires. Empiriquement, nous avons constaté que cette indépendance entre l'étape de détection (algorithme 8) et l'étape de filtrage (PF/RBPF) augmentait la robustesse de l'algorithme de localisation global.

6.3.2 Algorithme de localisation non-linéaire/non-gaussien

Comme vu au chapitre 3, le filtre particulaire est un procédé numérique qui consiste en une approximation de la probabilité *a posteriori* $p(\mathbf{x}_{pf,t} | z_t)$ (probabilité de l'état $\mathbf{x}_{pf,t}$ sachant l'ensemble des observations z_t) en utilisant un nombre suffisant de particules $\mathbf{x}_{pf,t}^{(i)}$. Une particule représente une solution du vecteur d'état, et le poids associé $w_t^{(i)}$ représente la valeur de cette particule par rapport à la loi d'intérêt. Une telle représentation basée sur des échantillons/particules nous permet d'approcher toute distribution statistique d'erreur, notamment les distributions non-paramétriques et non-gaussiennes.

6.3.2.1 SIR-PF

L'échantillonnage avec rééchantillonnage par importance ou *Sampling Importance Resampling* (SIR), aussi connu sous le nom de "*bootstrap filter*" [Arulampalam et al., 2002], est la méthode la plus populaire pour résoudre le problème du filtrage non linéaire.

Pour le SIR-PF, le nombre de particules requis est directement lié à la dimension du vecteur d'état. Afin de garder un nombre raisonnable de particules (limité à quelques milliers), nous supposons que l'altitude est soit constante soit connue, comme souvent en milieu urbain. Dans notre cas, l'estimation de l'altitude est disponible dans le vecteur $\tilde{\mathbf{x}}_i$ et n'est pas introduite dans le vecteur d'état. Pour réduire encore plus le nombre de particules, la vitesse est également exclue du vecteur d'état (contrairement à ce qui sera fait avec le vecteur d'état RBPF présenté dans la section suivante). Ainsi, les particules du SIR-PF sont données par $\mathbf{x}_{pf,t}^{(i)} = (e_{r,t}^{(i)}, n_{r,t}^{(i)}, \delta t_t^{(i)})^\top$, où l'exposant i désigne la particule et l'indice t l'instant.

A chaque *epoch*, le SIR-PF itère les trois étapes : "prédiction", "estimation" et "rééchantillonnage".

6.3.2.1.1 Étape de prédiction Cette étape, parfois appelée étape de mise à jour temporelle, vise à fournir une estimation du vecteur d'état pour la prochaine étape. Notons que cette étape peut être placée de façon équivalente soit au début de l'itération, *epoch* t , soit à la fin de l'itération de l'*epoch* $t - 1$.

Pour prédire la prochaine position d'une particule, nous avons besoin d'une estimation de la vitesse $\dot{\mathbf{x}}_r^{ENU}$. Puisque, $\dot{\mathbf{x}}_r^{ENU}$ ne fait pas partie du vecteur d'état, il doit être fourni par des données externes. En se contraignant à n'exploiter que des données GPS, nous considérons les mesures D_p pour en déduire $\dot{\mathbf{x}}_r^{ENU}$: les mesures D_p à $t - 1$ fournissent le *PR rate* d'où nous déduisons la vitesse du récepteur $\dot{\mathbf{x}}_r^{ENU} = (\dot{e}_r, \dot{n}_r, \dot{u}_r)$ en utilisant l'équation (6.3). Afin de se conformer aux notations communes

dans la communauté de la localisation et de la navigation, $\hat{\mathbf{x}}_r^{ENU}$ peut être représenté de manière équivalente en termes de norme et d'orientation de la vitesse : $V_{t-1}^{(i)} = \text{sqrt}(\dot{e}_r^2 + \dot{n}_r^2)$ et $\theta_{t-1}^{(i)} = \arctan\left(\frac{\dot{n}_r}{\dot{e}_r}\right)$, respectivement. L'état à l'instant t de la $(i)^{me}$ particule est alors obtenu selon :

$$\begin{cases} e_{r,t}^{(i)} &= e_{r,t-1}^{(i)} + V_{t-1}^{(i)} \cos\left(\theta_{t-1}^{(i)}\right) dt + \nu(\sigma_e), \\ n_{r,t}^{(i)} &= n_{r,t-1}^{(i)} + V_{t-1}^{(i)} \sin\left(\theta_{t-1}^{(i)}\right) dt + \nu(\sigma_n), \\ \delta t_t^{(i)} &= \delta t_{t-1}^{(i)} + \dot{\delta} t_{t-1} dt + \nu(\sigma_{\delta t}), \end{cases} \quad (6.21)$$

où le pas de temps dt est égal à 1 et ν est le bruit de prédiction associé à chacune des composantes du vecteur d'état. En effet, en tant qu'approche stochastique, le filtre particulaire est fondé sur des simulations stochastiques fournies ici par l'addition (aux vecteurs d'état des prévisions déterministes) d'un bruit gaussien avec une moyenne nulle et un écart-type $(\sigma_e, \sigma_n, \sigma_{\delta t})$.

Notons que dans notre cas, la vitesse utilisée pour la prédiction est estimée à partir de (\dot{e}_r, \dot{n}_r) à $t - 1$. Au lieu d'utiliser les mesures Doppler à $t - 1$, nous aurions pu utiliser celles acquises à l'instant t . Cependant, étant donné que la prédiction est comprise entre $t - 1$ et t , il n'y a pas objectivement de raison pour que la prédiction soit plus précise. En comparaison avec le RBPF (présenté dans la section 6.3.2.2), soulignons que l'estimation de la vitesse est réalisée comme un processus externe au SIR-PF lui-même, étant donné que la vitesse ne fait pas partie du vecteur d'état.

6.3.2.1.2 Étape d'estimation Cette étape, parfois appelée mise à jour des mesures, vise à corriger l'estimation de l'étape de prédiction connaissant les observations. Puisque la vitesse n'est pas représentée dans le vecteur d'état $\mathbf{x}_{pf,t}^{(i)}$, la probabilité *a posteriori* du SIR-PF est calculée relativement aux mesures PR uniquement. Elle est notée $p(\boldsymbol{\rho}_t | \mathbf{x}_{pf,t}^{(i)})$ avec $\boldsymbol{\rho}_t$ le vecteur des mesures ρ_j observées à t .

Le processus de mise à jour des poids $w_t^{(i)}$ est une fonction de pondération de leurs valeurs à l'itération précédente [Doucet et al., 2001] par la fonction de vraisemblance des observations $p(\boldsymbol{\rho}_t | \mathbf{x}_{pf,t}^{(i)})$: $w_t^{(i)} \propto w_{t-1}^{(i)} p(\boldsymbol{\rho}_t | \mathbf{x}_{pf,t}^{(i)})$. Dans la plupart des cas, en raison des contraintes de calcul, la fonction de vraisemblance $p(\boldsymbol{\rho}_t | \mathbf{x}_{pf,t}^{(i)})$ est approximée par une densité gaussienne multivariée. La vraisemblance est alors calculée selon :

$$p(\boldsymbol{\rho}_t | \mathbf{x}_{pf,t}^{(i)}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det \boldsymbol{\Sigma}_t^{\frac{1}{2}}} e^{-\frac{1}{2}(\tilde{\boldsymbol{\rho}}_t^{(i)} - \boldsymbol{\rho}_t)^T \boldsymbol{\Sigma}_t^{-1} (\tilde{\boldsymbol{\rho}}_t^{(i)} - \boldsymbol{\rho}_t)}, \quad (6.22)$$

avec N le nombre de mesures, $\tilde{\boldsymbol{\rho}}_t^{(i)}$ le vecteur des PR estimé suivant la particule $\mathbf{x}_{pf,t}^{(i)}$ et $\boldsymbol{\Sigma}_t$ la matrice de covariance des mesures. À la fin, une normalisation des poids est effectuée de telle sorte que $\sum_{i=1}^k w_t^{(i)} = 1$.

Après avoir mis à jour les poids, le vecteur d'état 'optimal' $\hat{\mathbf{x}}_{pf,t}$ est calculé comme la somme pondérée de toutes les particules :

$$\hat{\mathbf{x}}_{pf,t} = \sum_{i=1}^k w_t^{(i)} \mathbf{x}_{pf,t}^{(i)} \quad (6.23)$$

6.3.2.1.3 Étape de rééchantillonnage L'étape de rééchantillonnage n'est appliquée que si un critère sur les particules est satisfait. Cette étape vise à prévenir la dégénérescence de l'algorithme, en particulier pour éviter que les ressources informatiques ne soient consommées par des particules 'improbables'. Au cours de cette étape, un seuil est calculé [Doucet et al., 2001] pour partitionner l'ensemble des particules en fonction de leur poids, [Arulampalam et al., 2002]. Après avoir enlevé les particules qui présentent un poids inférieur au seuil considéré, les particules restantes sont dupliquées afin de maintenir un nombre constant de particules et tous les poids sont initialisés à une valeur constante (inverse du nombre total de particules).

6.3.2.2 Filtre particulaire Rao-Blackwellisé

Dans la précédente modélisation du filtre particulaire, la vitesse a été estimée directement à partir des mesures Doppler (étant à 'l'extérieur' de l'étape d'estimation du PF, cette estimation ne prenait donc pas en compte les estimations précédentes de l'étape de prédiction du PF). Cela revient à supposer l'inexistence de bruit dans les mesures Doppler. Afin d'éviter une telle hypothèse *a priori* peu réaliste, nous étendons le vecteur d'état de $(e_{r,t}, n_{r,t}, \delta t_t)^\top$ à $(e_{r,t}, n_{r,t}, \delta t_t, \dot{e}_{r,t}, \dot{n}_{r,t}, \dot{\delta} t_t, \ddot{e}_{r,t}, \ddot{n}_{r,t})^\top$, i.e. sa dimension augmente de trois à huit.

Cependant, le filtre particulaire précédent nécessiterait un nombre très important de particules pour explorer l'espace des solutions et le PF deviendrait inopérant. D'autre part, l'approche Rao-Blackwellisé a été proposée à la fois pour réduire la complexité et mieux approcher la solution dans le cas de fonctions convexes [Casella and Robert, 1996, Doucet et al., 2000]. Cette approche est basée sur l'idée de subdiviser le vecteur d'état afin de réduire la variance de la partie estimée par échantillonnage d'importance en exploitant des sous-structures linéaires [Schon et al., 2005]. Le cas classique correspond à la subdivision du vecteur d'état initial en deux sous-vecteurs, l'un étant estimé analytiquement et l'autre par échantillonnage d'importance (par exemple, via un PF). Ainsi, le nombre de particules requis pour une estimation précise reste numériquement raisonnable de par la réduction de la dimension du sous-système non-linéaire [Doucet et al., 2000, Schon et al., 2005].

Dans le cadre de notre problème, nous subdivisons le système de huit composantes décrivant les équations de l'étape de prédiction en deux sous-systèmes, un système linéaire et un système non-linéaire, comme suit : les équations pour calculer la position GPS à partir des observations PR (équation (6.1)) sont non-linéaires impliquant un système non-linéaire. D'autre part, l'estimation de la vitesse, connaissant la position du récepteur et les mesures Doppler, est réalisée par la résolution d'un système linéaire (équation (6.3)). Ainsi, nous définissons, en plus de $\mathbf{x}_{pf,t}$ défini dans la section 6.3.2.1, le vecteur d'état $\mathbf{x}_{kf,t} = (\dot{e}_{r,t}, \dot{n}_{r,t}, \dot{\delta} t_t, \ddot{e}_{r,t}, \ddot{n}_{r,t})^\top$.

La probabilité *a posteriori* du RBPF est factorisée :

$$p\left(\mathbf{x}_{kf,t}, \mathbf{x}_{pf,t} | \mathbf{z}_t\right) = p\left(\mathbf{x}_{kf,t} | \mathbf{x}_{pf,t}, \mathbf{z}_t\right) p\left(\mathbf{x}_{pf,t} | \mathbf{z}_t\right), \quad (6.24)$$

où \mathbf{z}_t appartient à l'ensemble des observations.

Le premier terme est résolu analytiquement en utilisant le filtre de Kalman (KF), et le second terme est estimé par échantillonnage de Monte Carlo en utilisant le filtre particulaire SIR. Dans le RBPF, nous pouvons garder le même nombre de particules que dans la section 6.3.2.1, tout en ayant introduit la vitesse du récepteur dans le vecteur d'état de sorte qu'elle soit également filtrée temporellement. Le modèle proposé pour le RBPF est triangulaire :

$$\begin{pmatrix} \mathbf{x}_{pf,t} \\ \mathbf{x}_{kf,t} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{3 \times 3} & \mathbf{A}_{pf,dt} \\ \mathbf{0}_{5 \times 3} & \mathbf{A}_{kf,dt} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{pf,t-1} \\ \mathbf{x}_{kf,t-1} \end{pmatrix} + \begin{pmatrix} \mathbf{v}_{pf,t-1} \\ \mathbf{v}_{kf,t-1} \end{pmatrix} \quad (6.25)$$

où $\mathbf{A}_{pf,dt}$ et $\mathbf{A}_{kf,dt}$ sont les matrices de transition définies comme suit :

$$\mathbf{A}_{pf,dt} = \begin{pmatrix} dt & 0 & 0 & \frac{dt^2}{2} & 0 \\ 0 & dt & 0 & 0 & \frac{dt^2}{2} \\ 0 & 0 & dt & 0 & 0 \end{pmatrix}, \quad (6.26)$$

$$\mathbf{A}_{kf,dt} = \begin{pmatrix} 1 & 0 & 0 & dt & 0 \\ 0 & 1 & 0 & 0 & dt \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (6.27)$$

$\mathbf{I}_{n \times n}$ est la matrice identité carrée de dimension n , $\mathbf{0}_{m \times n}$ est la matrice nulle (tous les coefficients égaux à 0) rectangulaire de dimension $m \times n$, et \mathbf{v}_t est un bruit blanc gaussien tel que ;

$$\mathbf{v}_t = \begin{pmatrix} \mathbf{v}_{pf,t-1} \\ \mathbf{v}_{kf,t-1} \end{pmatrix} \sim \mathcal{N}(0, \mathbf{Q}_t), \quad \mathbf{Q}_t = \begin{pmatrix} \mathbf{Q}_{pf} & \mathbf{0}_{3 \times 5} \\ \mathbf{0}_{5 \times 3} & \mathbf{Q}_{kf} \end{pmatrix}. \quad (6.28)$$

\mathbf{Q}_{pf} et \mathbf{Q}_{kf} sont les matrices de covariance des bruits $\mathbf{v}_{pf,t-1}$ et $\mathbf{v}_{kf,t-1}$ (pour simplifier la notation, nous avons omis la dépendance en temps pour les matrices de covariance).

La partie non-linéaire est traitée en utilisant le filtre particulaire présenté dans la section 6.3.2.1 pour estimer le vecteur d'état associé à une particule $\mathbf{x}_{pf,t}^{(i)}$ et son poids $w_t^{(i)}$. La partie linéaire est traitée à l'aide d'un filtre de Kalman étendu appliqué au vecteur d'état $\mathbf{x}_{kf,t}^{(i)}$ pour chaque particule récursivement. Le KF comporte deux étapes principales :

6.3.2.2.1 Étape de prédiction Cette étape a lieu entre l'étape de prédiction et l'étape d'estimation du filtre particulaire SIR. Nous définissons les variables intermédiaires :

$$N_t = A_{pf,dt} P_{t-1|t-1} A_{pf,dt}^\top + Q_{pf}, \quad (6.29)$$

$$L_t = A_{kf,dt} P_{t-1|t-1} A_{pf,dt}^\top N_t^{-1}, \quad (6.30)$$

$$y_t = x_{pf,t} - x_{pf,t-1}, \quad (6.31)$$

où y_t est interprété comme une erreur de mesure, L_t et N_t sont des matrices intermédiaires modélisant l'impact du système non-linéaire sur l'estimation linéaire. L'étape de prédiction s'écrit :

$$\hat{x}_{kf,t|t-1} = A_{kf,dt} \hat{x}_{kf,t-1|t-1} + L_t (y_t - A_{pf,dt} \hat{x}_{kf,t-1|t-1}), \quad (6.32)$$

$$P_{t|t-1} = A_{kf,dt} P_{t-1|t-1} A_{kf,dt}^\top + Q_{kf} - L_t N_t L_t^\top, \quad (6.33)$$

où $P_{t-1|t-1}$ est la matrice de covariance de $x_{kf,t}$. On note que si $A_{pf,dt}$ est une matrice nulle, les équations précédentes reviennent à l'étape de prédiction du filtre de Kalman. On notera aussi que, étant donné que l'étape de prédiction présentée dans la section 6.3.2.1 intervient dans l'équation (6.31), l'étape de prédiction actuelle se produit après l'étape de prédiction de la partie non-linéaire de RBPF.

Algorithm 9: Algorithme du filtre particulaire Rao Blackwellisé.

1 **Initialisation :**

2 **for** $(i) : 1$ **to** N **do**

3 Échantillonner $\hat{x}_{pf,0}^{(i)} \sim p(x_{pf,0})$ et fixer $\hat{x}_{kf,0}^{(i)}$ et P_0 .

4 Calculer les poids $\omega_0^{(i)} = p(z_0 | x_{pf,0}^{(i)})$.

5 Calculer $\hat{x}_{kf,0|0}^{(i)}$ et $P_{0|0}$ et fixer $t = 1$.

6 **end**

7 **for** $t > 0$ **et** $(i) : 1$ **to** N **do**

8 Calculer $\hat{x}_{pf,t|t-1}^{(i)}$.

9 Calculer $\hat{x}_{kf,t|t-1}^{(i)}$ et $P_{t|t-1}^{(i)}$.

10 Mettre à jour $\omega_t^{(i)}$ et le normaliser.

11 Calculer $\hat{x}_{kf,t|t}^{(i)}$ et $P_{t|t}^{(i)}$.

12 Rééchantillonnage des particules si nécessaire.

13 **end**

14 En sortie : $\{\hat{x}_{pf,t}^{(1 \dots N)}, \omega_t^{(1 \dots N)}\}$ et $\{\hat{x}_{kf,t|t}^{(1 \dots N)}, P_{t|t}^{(1 \dots N)}\}$

6.3.2.2.2 Étape d'estimation Cette étape a lieu entre l'étape d'estimation et l'étape de rééchantillonnage du filtre particulaire SIR. Elle est identique à l'étape de correction classique du filtre de Kalman étendu. L'étape d'estimation s'écrit

$$\hat{x}_{kf,t|t} = \hat{x}_{kf,t|t-1} + K_t (\hat{\rho}_t - C_t \hat{x}_{kf,t|t-1}), \quad (6.34)$$

$$P_{t|t} = P_{t|t-1} - K_t C_t P_{t|t-1}, \quad (6.35)$$

$$K_t = P_{t|t-1} C_t^\top (C_t P_{t|t-1} C_t^\top + R_t^{-1}), \quad (6.36)$$

où C_t est la matrice d'observation des mesures Doppler déduites de l'équation (6.3).

Cette correction analytique du sous-vecteur $\hat{x}_{k_f,t|t}$ est indépendante de l'estimation de $\hat{x}_{p_f,t|t}$ qui est effectuée en fonction de l'étape d'estimation présentée dans la section 6.3.2.1 (voir l'équation (6.23)). L'algorithme 9 présente le filtre particulaire Rao-Blackwellisé.

L'un des objectifs de cette étude était de vérifier l'intérêt de supprimer les données aberrantes parmi l'ensemble des mesures acquises, soit PR soit (PR, Dp). Ceci peut être réalisé en comparant les résultats de localisation obtenus avec la détection de données aberrantes associées au filtrage particulaire ou sa version Rao-Blackwellisée.

6.4 Expérimentation et résultats

Comme présenté dans le chapitre 4, afin de tester notre méthode de localisation, nous avons acquis des données dans deux environnements contraints : en canyon urbain et en forêt, caractérisés par la réception de signaux en NLOS. La figure 4.7a, dans le chapitre 4, montre la trajectoire du récepteur dans le sud de la région parisienne (France). La trajectoire est d'une longueur de 5km pour une durée totale de l'expérimentation de 11 min.

Dans l'expérience considérée à présent, l'APS-3 est utilisé à deux fins : pour établir la vérité terrain et obtenir les données brutes utilisées dans les algorithmes de localisation. Cependant, les données brutes utilisées ne sont pas post-traitées par l'APS-3.

La configuration des satellites durant l'expérimentation est illustrée sur la figure 4.6 du chapitre 4. Le nombre de satellites disponibles varie entre quatre et onze avec une valeur moyenne égale à neuf satellites.

6.4.1 Résultats de localisations

Spécifions tout d'abord les paramètres de l'algorithme proposé :

- Dans l'algorithme 1, $n_{ep} = 3$, $\lambda_1^T = (200 \ 20 \ 200 \ 20 \ 10000 \ 20 \ 0 \ 0)$ et $\lambda_2^T = (20 \ 2 \ 20 \ 20 \ 2 \ 20 \ 10000 \ 10000 \ 0 \ 0)$;
- Dans l'EKF, le SIR-PF et le RBPF, l'imprécision des mesures PR est de $\sigma_{PR} = 5$ m, et l'imprécision des mesures Dp est de $\sigma_{Dp} = 2$ Hz;
- dans le SIR-PF et le RBPF, le nombre de particules est fixé à 3000.

Les performances globales en localisation sont représentées avec des courbes de fonction de répartition des erreurs de localisation : meilleur est le résultat, plus grande est la surface sous la courbe de cette fonction de répartition. Dans cette étude, nous considérons onze méthodes de localisation. Deux d'entre elles sont les solutions des GPS eux-mêmes : soit l'UBLOX soit le GPS GARMIN. Les solutions GARMIN et UBLOX EVK-5T sont tracées comme des références, car il serait peu objectif de comparer le GPS haute précision avec les GPS bas coût. Toutefois, nous notons que la solution GARMIN semble plutôt intéressante, et même si l'algorithme GARMIN est inconnu, nous pouvons deviner qu'il utilise

TABLE 6.1 – Erreurs de localisation en centile. Les algorithmes de localisation considérés sont EKF, *Extended Kalman Filter*; PF, *Particle Filter*; RBPF, *Rao-Blackwell Particle Filter* et ORKF, *Outlier Robust Kalman Filter*.

Méthode de localisation	% Erreur < 3 m	% Erreur < 6 m	% Erreur < 9 m
UBLOX	20.9%	47.15%	64.92%
GARMIN	28.6%	72.97%	90.72%
EKF	37.26%	71.66%	80.75%
EKF + NFA (PR)	40.94%	81.82%	91.73%
EKF + NFA (PR,Dp)	37.13%	74.88%	96.49%
ORKF	40.83%	74.77%	83.22%
PF	21.1%	55.02%	75.12%
PF + NFA (PR)	44.6%	77.19%	89.85%
PF + NFA (PR,Dp)	59.95%	87.05%	94.61%
RBPF + NFA (PR)	61%	85.78%	93.38%
RBPF + NFA (PR,Dp)	61.96%	90.11%	98.28%

un pré-traitement des données. Par exemple, si on utilise le masque d'élévation du satellite en rejetant les satellites ayant une élévation inférieure à celle 15° , selon la figure 4.6 du chapitre 4, les satellites S_4 , S_{10} , S_{11} , S_{31} et S_{32} ne seront pas utilisés, ce qui correspond aux données aberrantes les plus fréquentes, comme nous le verrons plus loin.

Les autres méthodes correspondent aux différentes versions du filtre de Kalman étendu, au filtre particulaire et au filtre particulaire Rao-Blackwellisé : sans enlever les données aberrantes, en couplant le filtre avec la détection des données PR aberrantes ou avec la détection des données aberrantes dans (PR, Dp). Dans les trois filtres (EKF, PF et RBPF), la solution initiale est fournie soit par la méthode des moindres carrés soit par la sortie de l'algorithme 8 quand il y a une étape de détection de données aberrantes. À titre de comparaison, nous mettons en œuvre également une méthode robuste aux données aberrantes récente appelée ORKF (*Outlier Robuste Kalman Filter*) [Agamennoni et al., 2011]. Cette dernière, présentée au chapitre 3, est similaire à l'EKF, sauf que la covariance du bruit d'observation est estimée de façon récursive à l'intérieur de l'étape d'estimation (relâchant l'hypothèse sur la précision des mesures PR et Dp).

La figure 6.1 et le tableau 6.1 nous permettent de tirer les conclusions suivantes :

- Parmi les algorithmes implémentés, le filtre particulaire (PF) fournit des résultats plutôt décevants avec une erreur inférieure à 6 m dans seulement 55% des cas. Cette performance relativement mauvaise du PF par rapport à l'EKF par exemple, est probablement dû au fait que la vitesse ne fait pas partie du vecteur d'état. Elle n'est donc pas du tout filtrée, à l'inverse du cas de l'EKF.
- L'ORKF a de meilleures performances que la version la plus simple de PF et de l'EKF classique et des performances similaires à l'EKF + NFA(PR,Dp) en termes d'erreurs inférieures à 6 m.
- En supprimant les données aberrantes PR à l'entrée des filtres, l'EKF + NFA(PR) et le PF + NFA(PR) donnent de bien meilleurs résultats en localisation qu'en utilisant l'ensemble des données avec l'EKF, le PF ou même l'ORKF en termes d'erreurs inférieures à 6 m. D'ailleurs, si

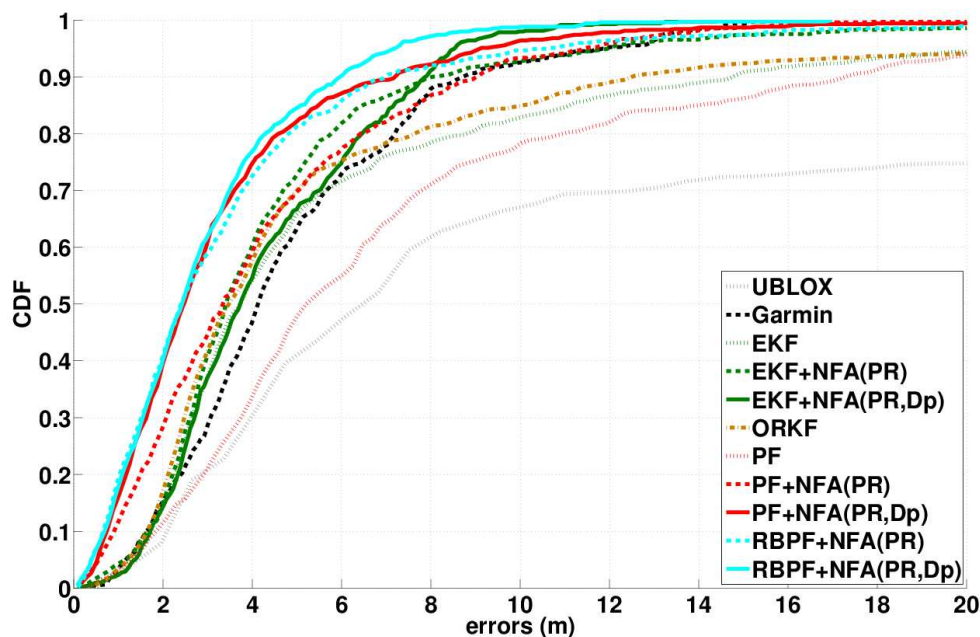


FIGURE 6.1 – Fonction de répartition des erreurs obtenues par les quatre versions de l'EKF, les cinq versions du filtre particulaire et les solutions fournies par les deux GPS pour une durée d'expérience totale de 11min et 40s.

l'EKF + NFA(PR) fait toujours mieux que le PF + NFA(PR) pour les erreurs inférieures à 6 m, l'écart a diminué, et en termes d'erreurs inférieures à 3 m, le PF + NFA(PR) surpasse l'EKF + NFA(PR).

- En supprimant également les données aberrantes D_p , le PF + NFA(PR, D_p) fournit de meilleurs résultats que les méthodes précédentes. Par exemple, son 95^{ème} centile correspond à une erreur inférieure à 9 m, tandis que l'erreur de PF + NFA(PR) pour le même centile est de 11,5m. Ceci illustre clairement l'intérêt de supprimer également les données aberrantes Doppler, d'autant qu'elles ne sont pas filtrées par l'étape d'estimation du PF. Cependant, dans le cas de l'EKF où les vitesses sont filtrées, l'effet d'éliminer les données aberrantes Doppler est moins clair : comme nous pouvons le voir sur la statistique d'erreurs inférieures à 9 m.
- En supprimant les données aberrantes PR, le RBPF + NFA(PR) a les mêmes performances en localisation que la version PF + NFA(PR, D_p). Ceci peut être expliqué par le fait que, de par le filtrage de l'estimation de la vitesse, le RBPF est relativement robuste aux données aberrantes dans les mesures Doppler. Il dépasse également l'EKF + NFA(PR) en performance.
- Enfin, en supprimant également les données aberrantes D_p , le RBPF + NFA(PR, D_p) surpasse toutes les autres méthodes. Selon le tableau 6.1, si la performance du PF + NFA(PR, D_p) et les deux RBPF sont proches pour des erreurs de moins de 3 m, un niveau de confiance plus élevé est atteint par le RBPF + NFA(PR, D_p) pour les erreurs inférieures à 6 m et 9 m.

Le tableau 6.2 nous montre les précisions globales en localisation. La précision a été évaluée par trois indicateurs : $Norm_1$, $RMSE$, la moyenne et l'écart-type des erreurs (cf. paragraphe 4.2.2 du chapitre 3). Les résultats sont cohérents avec la figure 6.1 : parmi les algorithmes mis en œuvre, si on considère toutes les données, l'EKF et l'ORKF montrent de bonnes performances, et si on accepte d'éliminer les données aberrantes (que ce soit PR ou (PR, D_p)), le RBPF surpasse les autres approches. Les meilleurs résultats sont obtenus par NFA(PR, D_p) couplé avec le RBPF, même si l'intérêt de l'éli-

TABLE 6.2 – Erreur de localisation (en m) en coordonnées (Est, Nord), erreurs $Norm_1$ et $RMSE$, et erreurs en moyenne et en écart-type : Comparaison des quatre versions de l'EKF, cinq versions du filtre particulaire et solutions fournies par les deux GPS pour une durée d'expérience totale de 11min et 40s.

Mesures d'erreurs	Algorithme de localisation	Données toutes les données	Données	
			NFA (PR) <i>inliers</i>	NFA (PR,Dp) <i>inliers</i>
$Norm_1$	UBLOX	(11.92,10.20)	-	-
	GARMIN	(3.35,2.76)	-	-
	EKF	(3.76,4.50)	(2.63,3.18)	(3.31,2.24)
	ORKF	(3.55,4.31)	-	-
	PF	(6.68,6.72)	(2.61,2.83)	(1.82,2.41)
	RBPF	-	(1.84,2.69)	(1.62,2.17)
$RMSE$	UBLOX	(20.44,18.60)	-	-
	GARMIN	(4.73,3.35)	-	-
	EKF	(5.77,7.47)	(3.43,5.00)	(3.92,3.09)
	ORKF	(5.55,7.79)	-	-
	PF	(9.09,9.49)	(3.48,3.86)	(2.95,3.51)
	RBPF	-	(3.37,3.53)	(2.51,3.20)
$(\mu_{loc}, \sigma_{loc})$	UBLOX	(16.72,22.02)	-	-
	GARMIN	(4.91,3.08)	-	-
	EKF	(6.40,6.96)	(4.59,3.96)	(4.37,2.42)
	ORKF	(6.13,7.36)	-	-
	PF	(10.43,7.99)	(4.25,3.41)	(3.37,3.11)
	RBPF	-	(3.53,3.56)	(2.96,2.25)

mination des données aberrantes peut également être remarqué dans le cas de l'EKF ou du PF. Enfin, pour quantifier l'amélioration due à la détection des données aberrantes par le NFA, nous avons exécuté le RBPF avec un masque d'élévation supprimant les satellites en dessous de 15° (comme il est fait habituellement sur la plupart des récepteurs GNSS). Les résultats sont : $Norm_1 = (2.82, 3.10)$, $RMSE = (4.68, 4.38)$ et $(\mu_{loc}, \sigma_{loc}) = (4.64, 4.42)$. Comme prévu, les résultats sont moins précis que le RBPF + NFA(PR,Dp) ou même RBPF + NFA(PR), montrant que la détection de données aberrantes est plus pertinente que l'utilisation d'un simple critère d'élévation du satellite.

Le tableau 6.3 montre l'erreur de localisation calculée sur les trois sous-parties de la trajectoire correspondant aux trois qualités de solution RTK. D'après le tableau 6.3, nous observons une 'corrélacion' entre la qualité du résultat de la localisation et la qualité de la solution fournie par le récepteur RTK : la localisation est plus précise avec le mode *RTK fixed* qu'avec le mode *RTK float*, et la partie différentielle présente des résultats en localisation plus mauvais. Il y a deux interprétations/explications à cette observation : (i) l'imprécision de la vérité terrain dans le cas *RTK float* ou différentiel introduit une erreur supplémentaire qui dégrade légèrement la précision *estimée* de la localisation ; (ii) la solution *RTK fixed* se produit principalement dans les zones dégagées (alors que la solution *RTK float* se produit plutôt dans des environnements urbains et la solution différentielle dans la partie forêt où la localisation est généralement bonne. En effet, en regardant la distribution des précisions de localisation *versus* la solution RTK pour les autres méthodes testées, nous rappelons que, comme souligné au début du

TABLE 6.3 – Erreurs de localisation de la méthode proposée (en m) en coordonnée (Est, Nord), erreurs sur la $Norm_1$ et le $RMSE$ et les erreurs en moyennes et en écart-type en fonction de la qualité de la solution RTK utilisée comme vérité terrain.

	Qualité de la solution	BBPF + NFA (PR)	BBPF + NFA (PR,Dp)
$Norm_1$	<i>RTK fixed</i>	(1.44,2.08)	(1.27,1.74)
	<i>RTK float</i>	(2.21,3.03)	(1.91,2.55)
	Différentiel	(2.16,4.16)	(2.03,2.62)
$RMSE$	<i>RTK fixed</i>	(2.56,3.14)	(1.76,2.45)
	<i>RTK float</i>	(3.19,4.35)	(2.56,3.26)
	Différentiel	(3.70,6.02)	(2.75,3.85)
$(\mu_{loc}, \sigma_{loc})$	<i>RTK fixed</i>	(2.74,2.99)	(2.38,1.85)
	<i>RTK float</i>	(4.05,3.57)	(3.44,2.31)
	Différentiel	(5.08,4.96)	(3.68,3.00)

chapitre , aussi que les résultats sont plus précis dans la partie *RTK fixed* de la trajectoire.

6.4.2 Validation des données aberrantes estimées

Dans cette section, nous cherchons à vérifier l'efficacité de l'algorithme 8 en termes de détection de données aberrantes. La partie délicate est d'avoir une 'vérité terrain' en termes d'*outliers*. Tout d'abord, notons que la définition d'une donnée aberrante est elle-même non triviale : du point de vue statistique, une valeur aberrante est une mesure *considerably dissimilar or inconsistent with the remainder of the data* [Ramaswamy et al., 2000], alors que du point de vue physique et pour l'application considérée, une donnée aberrante est une mesure affectée par des multi-trajets ou une réception en NLOS. Dans cette étude, nous adoptons la définition statistique, et nous estimons les biais comme dans [Le Marchand et al., 2009].

Parmi l'ensemble (PR,Dp), nous cherchons le sous-ensemble d'observations qui se comportent comme des *outliers* du point de vue statistique. La seule 'vérité terrain' que nous ayons est la position du récepteur qui est fournie par l'APS-3 GPS + GLONASS RTK. La construction d'une 'vérité terrain' concernant les *outliers* sur la base d'une vérité terrain concernant la position du récepteur se déroule en deux étapes : (i) tout d'abord, l'estimation des biais entre les mesures observées et celles attendues ; (ii) d'autre part, l'analyse des biais afin de les classer comme induits par des *outliers* ou par des *inliers*.

6.4.2.1 Estimation de δt pour une analyse qualitative

Pour la première étape, nous devons estimer les mesures 'attendues' par le récepteur à partir de la vérité terrain. Celle-ci permet de calculer la distance euclidienne entre le satellite S_j et la position du récepteur, d_j . Néanmoins, nous avons encore besoin d'estimer le biais d'horloge δt . Dans [Le Marchand et al., 2009], l'équation $c\delta t = \frac{1}{N} \sum_{j=1}^N (\rho_j - d_j)$ était utilisée (nous rappelons que $d_j =$

$\sqrt{(e_r - e_{S_j})^2 + (n_r - n_{S_j})^2 + (u_r - u_{S_j})^2}$). Toutefois, l'estimation par la moyenne n'est ni robuste à la présence d'*outliers*, ni au fait que les oscillateurs embarqués sur les récepteurs GPS ne soient ni stables ni précis. Ainsi, nous utilisons plutôt le M estimateur [Huber, 1964] comme une solution simple parmi les estimateurs robustes :

En supposant e le vecteur des résidus des estimations des biais temporels ($e_j = \rho_j - d_j - c\tilde{\delta}t$), $\alpha(e_j)$ étant un coefficient de pondération défini par $\alpha(e_j) = \frac{1}{\text{abs}(e_j)}$, la valeur optimale du biais temporel $c\hat{\delta}t$ et l'estimation du biais PR $\tilde{\Delta}m_j$ sont données par :

$$\begin{cases} c\hat{\delta}t &= \frac{\sum_{j=1}^N \alpha(e_j)(\rho_j - d_j)}{\sum_{j=1}^N \alpha(e_j)}, \\ \tilde{\Delta}m_j &= \rho_j - d_j - c\hat{\delta}t. \end{cases} \quad (6.37)$$

De la même manière, nous estimons les biais $\tilde{\Delta}\dot{m}_j$ sur les mesures Doppler en connaissant à la fois la vitesse et la position du récepteur GPS :

$$\begin{cases} c\hat{\delta}t &= \frac{\sum_{j=1}^N \alpha(\dot{e}_j)(\dot{\rho}_j - \dot{d}_j)}{\sum_{j=1}^N \alpha(\dot{e}_j)}, \\ \tilde{\Delta}\dot{m}_j &= \dot{\rho}_j - \dot{d}_j - c\hat{\delta}t. \end{cases} \quad (6.38)$$

avec \dot{e} le vecteur des résidus des estimations des dérivées temporelles ($\dot{e}_j = \dot{\rho}_j - \dot{d}_j - c\tilde{\delta}\dot{t}$).

Les figures 6.2 et 6.3 nous permettent de vérifier qualitativement la cohérence entre les biais importants (soit sur les mesures PR ou Doppler) et les données aberrantes détectées par le NFA. Plus précisément, ces figures montrent la variation temporelle des biais estimés pour observations PR et Doppler de chaque satellite. Les valeurs détectées comme des données aberrantes par l'algorithme NFA sont indiquées par un symbole rouge. Nous notons également que le signal estimé dans l'équation (6.37) est probablement affecté par des bruits atmosphériques et électroniques qui diffèrent d'un satellite à un autre. Cette spécificité par satellite induit différents biais, même entre les courbes cohérentes (voir les courbes S_{17} , S_{20} et S_{23} dans la figure 6.2). Dans le cas Dp, l'estimation est moins affectée par le bruit atmosphérique, de sorte que les pics de la figure 6.3 correspondant à une réception NLOS potentielle ou à un signal en multi-trajet apparaissent clairement.

6.4.2.2 Classification des biais pour une analyse quantitative

Pour évaluer quantitativement l'efficacité de la détection, nous avons étiqueté les données comme des *outliers* ou *inliers*. Un tel étiquetage n'a été fait que pour $\tilde{\Delta}m_j$ (en raison de la difficulté objective de l'étiquetage de $\tilde{\Delta}\dot{m}_j$) par un opérateur humain comme suit. Pour chaque *epoch* t de l'expérience, un biais $\tilde{\Delta}m_j$ est étiqueté comme *outlier* si, à t , il ne semble compatible ni avec le biais moyen du satellite considéré, ni avec les autres biais satellites. En pratique, une étape de seuillage par rapport à la valeur

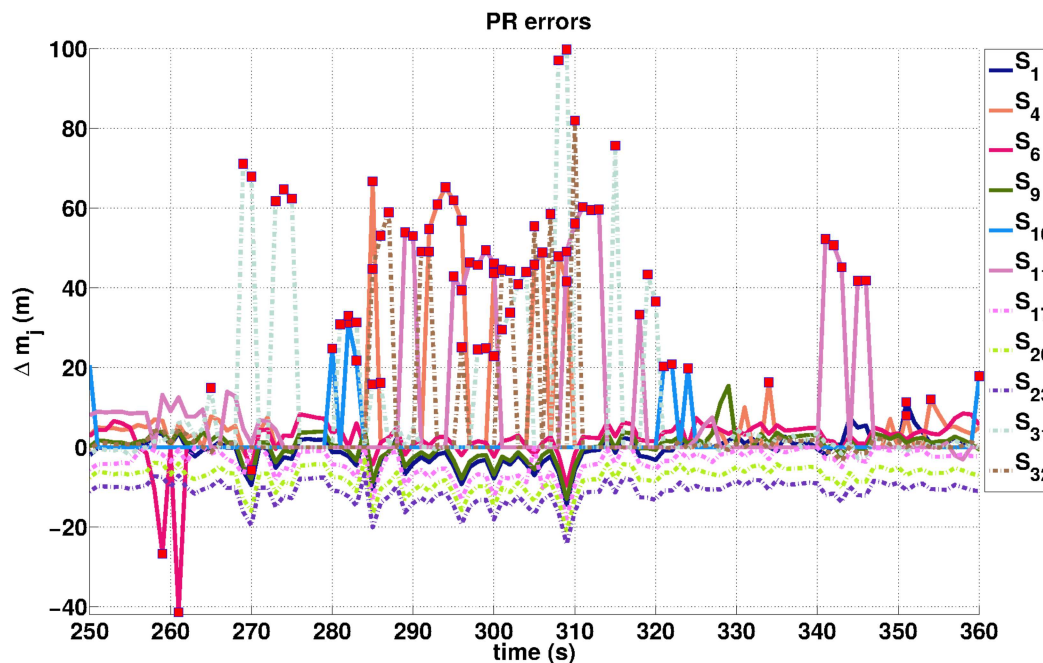


FIGURE 6.2 – Estimation de $\tilde{\Delta}m_j$ sur les mesures PR acquises par les différents satellites (de numéros compris entre 1 et 32). Les petits carrés rouges indiquent les mesures PR détectées comme aberrantes par le NFA.

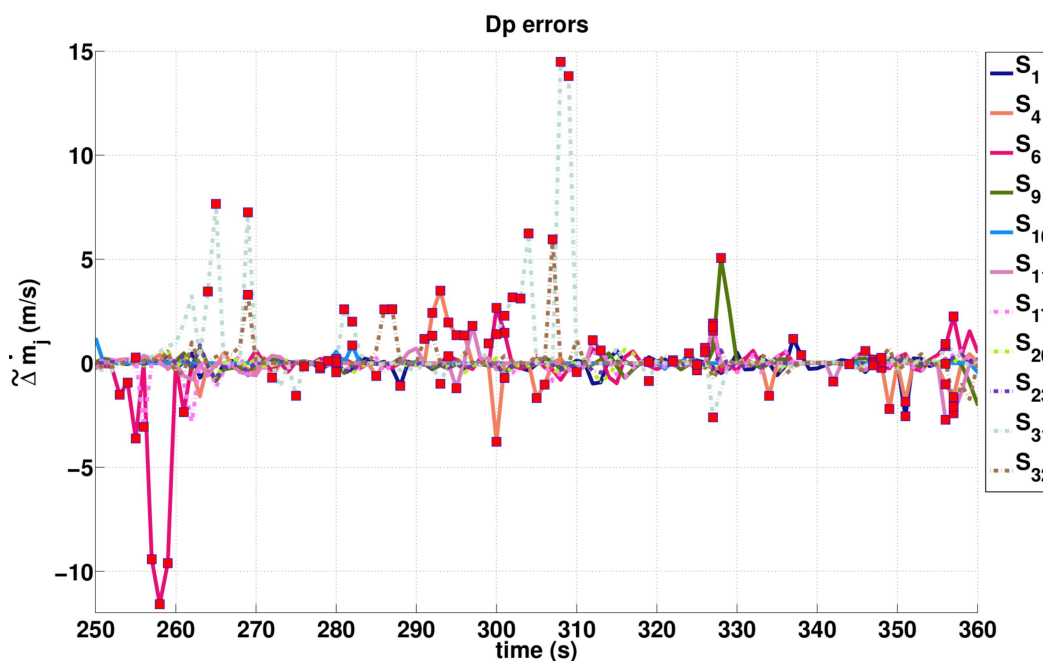


FIGURE 6.3 – Estimation de $\tilde{\Delta}m_j$ sur les mesures Doppler acquises par les différents satellites (de numéros compris entre 1 et 32). Les petits carrés rouges indiquent les mesures Doppler détectées comme aberrantes par le NFA.

médiane de tous les biais $\tilde{\Delta}m_j$ à t est appliquée d'abord (test de cohérence avec les autres satellites), et est suivie par une inspection visuelle des biais sélectionnés. Par exemple, à l'instant 351s, même si S_{23} présente une valeur assez importante pour $\tilde{\Delta}m_j$, seuls S_1 et S_4 sont étiquetés comme *outliers*. Même si

l'étiquetage précédent comprend une partie subjective, nous supposons qu'il est *statistiquement* non biaisé, et nous l'utilisons pour analyser *statistiquement* les résultats de détection d'*outliers*.

À partir du précédent étiquetage des biais et de l'ensemble des *inliers* \mathcal{D} fournit par l'algorithme 8, nous calculons le nombre de vrais positifs (VP , $PR \in \mathcal{D}$ avec un biais étiqueté *inlier*), de faux positifs (FP , $PR \in \mathcal{D}$ avec un biais étiqueté *outlier*, cf. chapitre 3 paragraphe 4.2.1), de faux négatifs (FN , $PR \notin \mathcal{D}$ avec un biais étiqueté *inlier*) et de vrais négatifs (VN , $PR \notin \mathcal{D}$ avec un biais étiqueté *outlier*). À partir de ces statistiques, les mesures d'*Accuracy* et de *Precision* sont calculées selon l'équation 4.1 (chapitre 3, paragraphe 4.2.1). L'ensemble des échantillons comporte 3498 échantillons PR correspondant à une sous-partie de l'expérience (de 7 min) où les biais $\tilde{\Delta}m_j$ sont étiquetés. Le tableau 6.4 montre les valeurs quantitatives de qualité de la détection obtenues. Les deux méthodes de couplage présentées entre le filtre particulière et la détection d'*outliers*, qu'elles soient limitées aux PR ou aux (PR, Dp) sont appelées "NFA(PR)" et "NFA(PR,Dp)", respectivement. En comparant ces deux approches, nous notons que la performance de chacune d'elles est très élevée. En outre, elles apparaissent très proches étant donné l'imprécision statistique intrinsèque au processus d'étiquetage pour la vérité terrain.

TABLE 6.4 – Performance de l'algorithme 8 pour la détection d'*outliers* parmi les mesures PR ou (PR,Dp).

	VP	FP	FN	VN	<i>Accuracy</i>	<i>Precision</i>
NFA (PR)	3131	39	49	279	97.5	98.7
NFA (PR,Dp)	3112	91	45	250	96.1	97.2

6.4.2.3 Corrélation entre les *outliers* PR et Doppler

Ayant, au moins qualitativement, évalué positivement la détection des données aberrantes, nous pouvons aussi interpréter les résultats en termes de présence de données aberrantes parmi les mesures Doppler.

En termes de statistiques globales, au cours de l'expérience, le NFA(PR) exclut 9,83% des observations PR disponibles. Le NFA(PR,Dp) rejette moins d'observations PR (8,37%), mais enlève 2,85% des observations Doppler disponibles. En termes de corrélation entre les *outliers* PR et Doppler, parmi les *outliers* Doppler, 54% sont également des *outliers* PR. Ainsi, nous pouvons en déduire que, selon ces statistiques, dans des environnements contraints, les mesures Doppler présentent trois fois moins d'*outliers* que les mesures PR, mais, néanmoins, souffrent de phénomènes de signaux multi-trajet et de réception en NLOS.

6.5 Conclusion

Dans ce chapitre, une nouvelle approche capable de faire face à un nombre important de données aberrantes (parmi les mesures de pseudo-distance et celles Doppler) a été présentée pour la localisation GNSS. Sur la base d'une modélisation *a contrario*, le critère du nombre de fausses alarmes (NFA)

nous permet de répartir les mesures de pseudo-distance et Doppler entre *inliers* et *outliers*. Les données aberrantes détectées sont retirées de l'ensemble des observations pour obtenir une estimation robuste de la position et de la vitesse du récepteur. Deux modèles basés sur le filtrage particulaire ont été considérés et comparés pour le processus de localisation. Le premier modèle (PF) ne filtre que la position du récepteur, tandis que le second modèle (RBPF) est un filtre plus complet qui gère la position du récepteur et de la vitesse en utilisant les deux observations PR et Doppler dans son étape d'estimation. Des tests ont été effectués dans le cas d'un récepteur à bord d'un véhicule circulant dans un canyon urbain et en forêt. Les résultats montrent que, en excluant les *outliers* et en filtrant le bruit des observations, la localisation est plus précise.

Dans le chapitre 2, nous avons introduit le problème de la localisation en explicitant le fait que cette dernière puisse être représentée soit de façon ponctuelle soit en termes d'ensemble(s). Dans ce chapitre et le précédent, nous nous sommes focalisés sur une représentation ponctuelle de la localisation. Cependant cette dernière ne permet pas de représenter l'imprécision de l'estimation (imprécision liée par exemple au nombre de mesures utilisées). Dans le chapitre suivant, le formalisme des fonctions de croyance va nous offrir un cadre théorique pour manipuler à la fois les incertitudes et les imprécisions des sources d'information via des entités appelées 'fonctions de masse'. Une représentation sous forme d'ensembles similaire à celle de l'analyse par intervalles ainsi que les différents outils de manipulation de ces ensembles seront utilisées pour combiner les sources. Notre contribution principale sera alors à nouveau dans la définition d'un algorithme de détection des données aberrantes adapté cette fois à la représentation ensembliste.

Approche crédibiliste appliquée à la localisation GNSS

Sommaire

7.1 Introduction	115
7.2 Fonctions de croyance	116
7.2.1 Représentation des bbas 2D	119
7.2.2 Approximation de bba	120
7.3 Approche proposée pour la détection d'outliers	122
7.3.1 Algorithme du RANSAC évidentiel	122
7.3.2 Exemple jouet pour illustrer le RANSAC évidentiel	125
7.3.3 Exemple d'application sur les bbas 2D : estimation des paramètres d'une droite	127
7.4 Application à la localisation GPS	128
7.4.1 Formulation évidentielle du problème de localisation	128
7.4.2 Estimation dans le cas dynamique	129
7.5 Résultats	131
7.5.1 Cas statique	132
7.5.2 Cas dynamique	133
7.6 Conclusion	137

7.1 Introduction

Dans les chapitres précédents, nous nous sommes intéressés aux méthodes statistiques de détections de données aberrantes pour la localisation robuste. Ces approches probabilistes sont adaptées à une représentation ponctuelle de la position du récepteur (ou plus généralement du vecteur d'état considéré) et les incertitudes sont modélisées de manière fine et sophistiquée. Néanmoins, l'imprécision se distingue mal de l'incertitude dans de telles approches. Or, cette imprécision est inhérente aux données. Citons par exemple le cas de paramètres ou de caractéristiques des capteurs mal connus.

Aussi, des méthodes alternatives qui se concentrent sur la modélisation de l'imprécision ont-elles été proposées. En particulier, l'analyse par intervalles (AI) [Jaulin et al., 2001b] prend en compte l'imprécision de toutes les sources (capteurs notamment) pour fournir une solution 'garantie', sous l'hypothèse que toutes les erreurs soient bornées dans un intervalle connu. Le calcul ensembliste a été

appliqué à la localisation, que ce soit en intérieur (un robot équipé d'odomètres et de capteurs ultra-son [Seigneur et al., 2009]) ou en extérieur (avec un véhicule intelligent équipé d'odomètres, d'un modèle numérique de terrain et d'un GPS [Drevelle and Bonnifait, 2013]). Les hypothèses sur les bornes d'erreur des capteurs ne sont elles-mêmes pas toujours correctes (notamment en cas de données aberrantes). Aussi y a-t-il un risque que l'estimateur fournisse une solution vide. La q -relaxation, présentée au chapitre 3 section 3.3.1, a pour but d'éviter une solution vide en relâchant q contraintes parmi les N initiales. Le contrôle du paramètre q est critique et dépend des hypothèses sur la validité des bornes d'erreur des mesures (elles-mêmes reliées aux hypothèses sur la présence d'*outliers*).

Récemment, Pichon et al. [Pichon et al., 2015] ont présenté l'idée d'optimiser la décision à partir d'un sous-ensemble de sources dans la théorie des fonctions de croyance (BFT). D'une part, cette approche s'appuie sur le principe de la q -relaxation, mais d'autre part, elle s'en distingue par plusieurs aspects. Le formalisme est celui des fonctions de croyance, de sorte que les observations soient représentées et manipulées par l'intermédiaire d'une fonction de masse ou bba (*basic belief assignment*). En relaxant q bbas parmi N , le résultat est calculé sans spécifier les bbas à relâcher en considérant toutes les disjonctions de $N - q$ bbas. Les travaux de [Pichon et al., 2015] généralisent l'approche de l'analyse par intervalles avec q -relaxation au sens où la théorie des fonctions de croyance va gérer à la fois les imprécisions et les incertitudes. Néanmoins, l'approche proposée a une complexité importante, ce qui rend difficile son application à un nombre important de sources.

Dans ce chapitre, nous proposons une approche inspirée de la méthode RANSAC mais en se plaçant dans le cadre des fonctions de croyance, pour détecter de façon itérative les *outliers* à partir d'une mesure de cohérence précisée ultérieurement. Cette méthode permettra, comme précédemment, de partitionner l'ensemble des observations en deux sous-ensembles, *inliers* et *outliers*. Une comparaison avec la méthode proposée par Pichon et al. sera effectuée pour montrer l'intérêt de notre approche. Enfin, une partie estimation dynamique dans le cadre évidentiel sera abordée pour prendre en compte la bba de prédiction dans le processus de localisation.

7.2 Fonctions de croyance

Nous ne présentons ici que les outils de la théorie des fonctions de croyance directement utilisés dans notre approche. Pour un lecteur ne connaissant pas cette théorie, quelques éléments supplémentaires sont donnés en annexe.

Fonction de masse :

Soit Ω le cadre de discernement, c'est-à-dire, l'ensemble des hypothèses mutuellement exclusives représentant l'espace des solutions. La fonction de masse m est définie de 2^Ω vers $[0, 1]$ telle que : $\sum_{A \in 2^\Omega} m(A) = 1$, où 2^Ω est l'ensemble des disjonctions d'hypothèses possibles du cadre Ω , ensemble de taille $2^{|\Omega|}$, et $m(A)$ représente la croyance que la solution appartienne à l'ensemble $A \in 2^\Omega$, sans inclure les sous-ensembles de A . Une telle modélisation en termes de fonction de masse est notée bba. Si $m(A) \geq 0$, A est dit élément focal de la bba m .

Une bba peut également être définie par sa fonction de plausibilité, notée Pl et définie de 2^Ω vers $[0, 1]$ telle que $\forall A \in 2^\Omega, Pl(A) = \sum_{B \in 2^\Omega, A \cap B \neq \emptyset} m(B)$. À partir de Pl , la fonction contour, notée pl , est alors définie comme la plausibilité sur les singletons de Ω .

Combinaison de bbas :

Soient m_1 et m_2 deux bbas définies dans Ω . Si m_1 et m_2 représentent une information provenant de deux sources indépendantes S_1 et S_2 respectivement, elles peuvent être combinées de façon conjonctive suivant la règle :

$$\forall A \in 2^\Omega, m_{1 \odot 2}(A) = m_1 \odot m_2(A) = \sum_{\substack{(B,C) \in 2^\Omega \times 2^\Omega, \\ B \cap C = A}} m_1(B) m_2(C). \quad (7.1)$$

Cette règle conjonctive se distingue de la règle de Dempster notamment dans la modélisation de l'espace des solutions. La règle de Dempster suppose un monde fermé où les hypothèses du cadre de discernement sont exhaustives. En cas de conflit, les masses des éléments focaux sont renormalisées de sorte que la somme des masses soit égale à 1 tout en forçant une masse nulle pour l'ensemble vide (\emptyset). La règle conjonctive considère un monde ouvert où l'ensemble vide peut devenir élément focal en cas de conflit.

Récemment, [Pichon et al., 2015] a présenté l'idée de relâcher des sources dans la combinaison de bbas. Supposons que seules r sources sont à considérer parmi les N sources disponibles (H_r^N est cette hypothèse qui représente le fait de relâcher $q = N - r$ sources). Selon cette hypothèse, pour tout N-uplet d'hypothèses, $A = \{A_1, A_2, \dots, A_N\}, \forall i \in [1, N] A_i \in \Omega$, seules r d'entre elles doivent être considérées sans connaître lesquelles. Soit alors $\Gamma_r(A)$ l'hypothèse représentant cette connaissance :

$$\Gamma_r(A) = \bigcup_{\mathcal{A} \subseteq \{A_1, \dots, A_N\}, |\mathcal{A}|=r} (\cap_{A_i \in \mathcal{A}} A_i). \quad (7.2)$$

La masse de l'hypothèse B est la somme, sur tous les $\Gamma_r(A)$ correspondant à B , du produit des masses des éléments de A [Pichon et al., 2015] :

$$m[H_r^N](B) = \sum_{\substack{A \subseteq \Omega^N; \\ \Gamma_r(A)=B}} \left[\prod_{i=1}^N m_i^\Omega(A_i) \right], \quad (7.3)$$

Une telle règle de combinaison généralise les règles de combinaisons classiques [Pichon et al., 2015], telles que la règle conjonctive pour $r = N$ et la règle disjonctive pour $r = 1$. Lorsque la bba est catégorique (ne contient qu'un seul élément focal), l'approche est similaire à la q-relaxation appliquée en analyse par intervalles.

Le tableau 7.1 présente un exemple tiré de l'article de Pichon et al. [Pichon et al., 2015], pour montrer l'application des équations 7.2 et 7.3. Trois fonctions de masse sont définies sur un cadre de dis-

TABLE 7.1 – Comparaison statistique en variant la largeur des éléments focaux.

A	m_1	m_2	m_3	$m [H_1^3]$	$m [H_2^3]$	$m [H_3^3]$
\emptyset	0	0	0	0	0	0.36
θ_1	0.5	0	0	0	0.06	0.2
$\theta_1 \cup \theta_2$	0	0.2	0	0	0.04	0.04
θ_3	0	0	0.6	0	0	0.24
$\theta_1 \cup \theta_3$	0	0	0	0	0.24	0
Ω	0.5	0.8	0.4	1	0.66	0.16

cernement comprenant trois hypothèses $\Omega = \{\theta_1, \theta_2, \theta_3\}$ et ayant chacune deux éléments focaux. Sur les trois colonnes de droite sont effectuées une combinaison conjonctive suivant trois hypothèses :

- H_1^3 : 1 mesure sur 3 est fiable ;
- H_2^3 : 2 mesures sur 3 sont fiables ;
- H_3^3 : toutes les mesures sont fiables.

Calculons par exemple la masse de θ_1 après combinaison conjonctive sous l'hypothèse que toutes les sources sont fiables :

$$\begin{aligned} \Gamma_3(\theta_1) &\in \{(\theta_1)_1 \cap (\theta_1 \cup \theta_2)_2 \cap (\Omega)_3, (\theta_1)_1 \cap (\Omega)_2 \cap (\Omega)_3\}, \\ m[H_3^3](\theta_1) &= m_1(\theta_1) m_2(\theta_1 \cup \theta_2) m_3(\Omega) + m_1(\theta_1) m_2(\Omega) m_3(\Omega). \end{aligned} \quad (7.4)$$

La notation $(\theta_1)_1$ signifie que l'hypothèse θ_1 est considérée/sélectionnée par la source 1 (autrement dit, c'est la source 1 qui promeut l'hypothèse θ_1). En revanche, sous l'hypothèse que deux des trois sources seulement sont fiables, la masse de θ_1 est estimée selon :

$$\begin{aligned} \Gamma_2(\theta_1) &= ((\theta_1)_1 \cap (\theta_1 \cup \theta_2)_2) \cup ((\theta_1 \cup \theta_2)_2 \cap (\theta_3)_3) \cup ((\theta_1)_1 \cap (\theta_3)_3), \\ m[H_2^3](\theta_1) &= m_1(\theta_1) m_2(\theta_1 \cup \theta_2) m_3(\theta_3). \end{aligned} \quad (7.5)$$

Le critère de décision proposé dans [Pichon et al., 2015] est une mesure de cohérence définie comme le maximum de la fonction contour :

$$\phi_{m^H}(m) = \max Pl [m^H]. \quad (7.6)$$

Les résultats sur les différentes hypothèses de nombre de sources fiables nous donnent : $\phi_{H_1^3} = 1$; $\phi_{H_2^3} = 1$; $\phi_{H_3^3} = 0.4$. Dans la pratique, il s'agit de ne considérer que les hypothèses de nombre de sources fiables ayant une mesure de cohérence supérieure à un seuil et, parmi ces hypothèses, de choisir celle maximisant le nombre de sources fiables. La décision est prise en considérant l'hypothèse qui optimise à la fois la contrainte de l'équation 7.6 et le nombre de sources fiables. Pour l'exemple, avec un seuil fixé à 0.8, la décision sera en faveur de l'hypothèse H_2^3 .

Mesure d'incohérence :

Classiquement, la masse de l'ensemble vide $m(\emptyset)$, calculée selon l'équation 7.1, fournit une mesure de désaccord entre sources :

$$\kappa = m_{1 \otimes 2}(\emptyset) = \sum_{\substack{\forall B, C \in 2^\Omega \times 2^\Omega \\ |B \cap C| = 0}} m_1(B) m_2(C). \quad (7.7)$$

Notons cependant que la mesure de conflit a longtemps été un sujet de recherche (e.g., [Roquel et al., 2014]) d'autant plus qu'elle peut avoir différentes origines [Lefevre et al., 2002, Destercke and Burger, 2013] : une bba erronée due à un cadre de discernement incomplet ou une observation erronée, la combinaison conjonctive d'un nombre important de sources, ou une confiance trop élevée en une source bruitée. Dans ce chapitre, la masse de l'ensemble vide $m(\emptyset)$ est utilisée comme simple mesure d'incohérence pour détecter les sources non fiables.

Décision :

Après la combinaison de différentes sources, une décision est prise généralement en utilisant soit le maximum de plausibilité parmi les singletons, soit le maximum de probabilité pignistique, défini comme suit :

$$\forall H \in \Omega, \text{Bet}P(H) = \sum_{A \in 2^\Omega, H \in A} \frac{m(A)}{|A| (1 - m(\emptyset))}. \quad (7.8)$$

7.2.1 Représentation des bbas 2D

Dans notre application, le cadre de discernement est un ensemble discret de pavés où une localisation est possible suivant les coordonnées Est et Nord. Suivant les hypothèses *a priori* sur l'espace des solutions et sur le pas de discrétisation, la dimension d'un tel cadre de discernement peut être très grande. Par exemple, si on considère une surface de 100m par 100m avec une résolution de 1m², le cardinal du cadre de discernement est $|\Omega| = 10000$. C'est un nombre élevé comparé aux applications classiques en fonctions de croyance. Pour nous adapter à la spécificité de ces bbas 2D, nous représentons chaque élément focal par un sous-ensemble de pavés, ce qui permet à la fois une représentation moins éparse et une manipulation aisée de ces derniers [André et al., 2015, Rekik et al., 2015]. En effet, grâce à la redéfinition des opérateurs basiques de combinaison (intersection, union), une telle représentation, bien que non univoque, est efficace. Notons qu'elle est également plus précise que de traiter indépendamment les intervalles à une dimension comme dans [Nassreddine et al., 2010].

À partir de la représentation précédente des bbas 2D (qui, nous le redisons, est non univoque), nous souhaitons une représentation aussi légère que possible en nombre de sous-pavés. Pour ce faire, nous devons établir un sous-pavage sans chevauchement des sous-pavés de l'élément focal considéré. Une procédure similaire appelée 'régularisation' est utilisée en analyse par intervalles. Pour éviter toute confusion avec la régularisation utilisée au chapitre précédent, nous l'appellerons ici 'simpli-

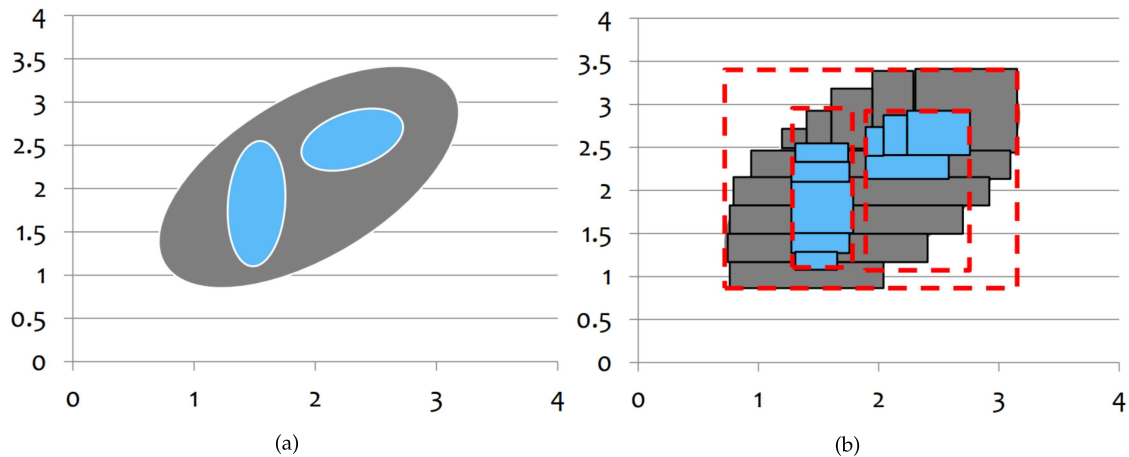


FIGURE 7.1 – Exemple d'une bba 2D.

fication'. D'un point de vue pratique, nous pouvons créer une imagerie d'une résolution ε . Chaque élément focal est composé de sous-pavés de résolution minimale. L'étape suivante permet alors de fusionner ces sous-pavés élémentaires en sous-pavés de plus grande taille de façon à réduire au maximum le nombre de sous-pavés de chaque élément focal de la bba. La figure 7.1 présente un exemple d'une bba 2D à deux éléments focaux (figure 7.1a) et la représentation discrète de ses éléments focaux en 2D (figure 7.1b). Les encadrements en traits discontinus rouges de la figure 7.1b sont une représentation sous forme d'un produit cartésien d'éléments focaux 1D sur chaque axe de façon indépendante. Nous notons que cette représentation que nous appelons 2X1D est plus imprécise qu'une représentation en 2D.

7.2.2 Approximation de bba

La simplification du paragraphe précédent n'entraîne pas de perte d'information. Malheureusement, elle n'est pas suffisante pour lutter contre un effritement de la masse, causé dans notre cas par le nombre élevé de combinaisons effectuées. En effet, la combinaison conjonctive d'un grand nombre de bbas entraîne une augmentation exponentielle du nombre d'éléments focaux de la bba résultante et une charge calculatoire importante. Le but de l'approximation est alors de réduire le nombre d'éléments focaux dans une bba tout en préservant le maximum d'information.

Des méthodes sophistiquées ont été proposées [Dubois and Prade, 1990] [Denoeux, 2001] pour approximer la bba moyennant un coût calculatoire élevé. Dans ce qui suit, nous présentons deux méthodes simples à mettre en œuvre et rapides, fondées sur l'élimination d'éléments focaux. Leurs différences se situent principalement au niveau du critère de sélection et dans la manière dont les éléments focaux sont regroupés.

7.2.2.1 Summarization

La méthode d'approximation la plus simple et rapide à utiliser consiste à regrouper en un seul élément focal les éléments de masses faibles, et transférer la somme de leur masse à leur union. Cette méthode, appelée *summarization* [Lowrance et al., 1986], est la plus populaire de part son efficacité à éliminer les petites masses résiduelles.

Plus précisément, soit $\Omega = \{\theta_1, \dots, \theta_p\}$ le cadre de discernement de la fonction de masse m . En notant $m(\theta_i) = m_i$ et en supposant que les masses des éléments focaux sont triées de façon décroissante, alors la construction de la fonction m' avec p éléments focaux, tel que $P > p > 1$, est obtenue par :

$$\begin{aligned} m'(\theta_i) &= m_i; \quad i = 1, \dots, p-1 \\ m'(\theta'_p) &= \sum_{i=p}^P m_i; \\ \theta'_p &= \bigcup_{i=p}^P \theta_i. \end{aligned} \tag{7.9}$$

Cette méthode a deux inconvénients majeurs. Le premier est la non prise en compte de la taille de l'élément focal (ou de son cardinal) dans le critère de sélection, qui est une caractéristique importante dans notre application. Le second est d'augmenter paradoxalement la plausibilité des hypothèses éliminées de l'ensemble des éléments focaux.

7.2.2.2 Agrégation itérative

Cette méthode permet une agrégation des éléments focaux deux par deux en minimisant un critère donné. Plusieurs critères ont été proposés. Par exemple, [Harmanec, 1999] propose un critère de minimisation de la différence entre la somme des valeurs de crédibilités considérant des bbas avant et après approximation. Dans la même idée, [Denoeux, 2001] propose un critère minimisant la somme des plausibilités sur les singletons.

Dans ce chapitre, nous reprenons l'idée, proposée dans [André et al., 2015], de combiner la méthode d'agrégation avec un critère de similarité entre éléments focaux. La mesure de similarité de Jousselme [Jousselme et al., 2001] permet de regrouper les éléments focaux en considérant à la fois leur masse et leur cardinal. Le critère s'écrit après simplification [André et al., 2015] :

$$D_J(A, B) = m(A)^2 \left(1 - \frac{|A|}{|A \cup B|}\right) + m(B)^2 \left(1 - \frac{|B|}{|A \cup B|}\right). \tag{7.10}$$

L'équation 7.10 favorise ainsi le regroupement d'éléments focaux de cardinal élevé. Néanmoins, les performances de cette méthode, de même que la *summarisation*, dépendent du choix du nombre d'éléments focaux de la bba approximée. Un mauvais choix de ce paramètre peut conduire à un résultat médiocre en termes d'approximation de la bba.

7.3 Approche proposée pour la détection d'*outliers*

Revenons très brièvement sur le RANSAC (présenté aux chapitres 3 et 5), pour en repreciser les deux idées fondamentales. La première est d'éviter une recherche exhaustive dans l'espace des solutions : les solutions à tester sont calculées suivant un modèle qui consiste généralement à considérer seulement les solutions exactes issues des sous-ensembles de mesures. Pour notre approche, nous avons suivi l'idée d'une exploration guidée de l'espace des solutions. Néanmoins, la manière dont est faite l'exploration diffère de celle utilisée par le RANSAC. Elle se fonde sur un ajout/suppression de sources dans/depuis l'ensemble des sources courantes considérées comme *inliers* (fiables).

La seconde idée du RANSAC est de garder la solution la plus cohérente, c'est-à-dire, qui conduit au plus grand nombre de données *inliers* déterminées comme les mesures qui présentent un niveau de bruit inférieur à un seuil donné. En suivant le même principe et celui de la q-relaxation, puisque un critère sur le plus grand nombre d'*inliers* est aussi un critère sur le moins grand nombre d'*outliers*, nous choisirons la solution correspondant au plus grand nombre de sources sélectionnées selon un seuil calculé à partir d'une mesure de cohérence ou d'incohérence.

Dans la section suivante, nous proposons un algorithme nommé 'RANSAC évidentiel' (EV-RANSAC) puisqu'au lieu de manipuler des données et des mesures, nous manipulons des bbas. Comme dans le RANSAC classique, la solution est obtenue en explorant partiellement l'espace des solutions et en maximisant le nombre de sources cohérentes (consensuelles) pour un seuil donné comme paramètre *a priori*.

7.3.1 Algorithme du RANSAC évidentiel

L'algorithme 10 présente le RANSAC évidentiel dans le cas où le conflit $m(\emptyset)$ est utilisé comme mesure d'incohérence. L'algorithme s'en déduit de façon immédiate. Il est possible de remplacer cette mesure par une autre mesure de cohérence ou d'incohérence. Par exemple, nous avons aussi implémenté la méthode en utilisant le maximum de la fonction contour comme mesure de cohérence.

7.3.1.1 Exploration de l'espace des solutions

Comme dit précédemment, nous allons nous déplacer dans l'espace des solutions suivant certains mouvements, définis via le graphe \mathcal{G} . Considérant N sources, les nœuds du graphe \mathcal{G} sont les sous-ensembles de sources au nombre de 2^N . Deux nœuds sont connectés par une arête (ou un arc) si les deux sous-ensembles de sources (représentés par les deux nœuds considérés) ne diffèrent que d'une seule source. En pratique, les nœuds sont notés par des mots binaires représentant les sous-ensembles de sources (e.g., pour $N = 4$, tous les mots de 4 bits, de sorte que les nœuds correspondant à une seule source $\in \{0001, 0010, 0100, 1000\}$, ceux correspondant à deux sources $\in \{0011, 0101, 0110, 1001, 1010, 1100\}$, etc.). Par la suite, les nœuds sont connectés si et seulement si leur distance de Hamming est égale à 1 (e.g., 0001 n'a que 4 voisins : 0000, 0011, 0101 et 1001). Nous introduisons aussi la notation $|b|$ qui est, pour chaque mot binaire b , le nombre de bits égal à 1 corres-

Algorithm 10: RANSAC évidentiel ; entrées : ensemble des bbas \mathcal{M} , cadre de discernement Ω , seuil de conflit κ_{th} , nombre maximum d'itérations l_{max} ; sorties : ensemble \mathcal{I}_S des sources étiquetées comme *inliers*.

```

1  $N \leftarrow |\mathcal{M}| ; l \leftarrow 1 ;$ 
2 Créer le graphe  $\mathcal{G}$  ayant  $2^N$  nœuds (codés en mots binaires) et des arêtes entre les nœuds à une
   distance de Hamming de 1 ;
3 Initialiser  $\mathcal{G}$  nœuds (conflit et 'père') à une valeur de  $-1$  ;
4 for  $j \leftarrow 1$  to  $\frac{N \times (N-1)}{2}$  do
5    $b \leftarrow$  code binaire du  $j^{\text{ème}}$  nœud composé de 2 sources ;
6   Fixer  $i_1$  et  $i_2$  égaux aux indices des sources dans le sous-ensemble  $b$  (e.g.,
    $\{i_1, i_2\} \leftarrow \{n \mid b \& (1 \ll n) = 1\}$ ) ;
7    $\kappa_b \leftarrow \bigoplus_{i \in \{i_1, i_2\}} m_i (\emptyset) ;$ 
8 end
9 Nœud initial  $b_0 \leftarrow \arg \min_{b, |b|=2} \kappa_b ;$ 
10 Fixer la valeur du conflit du nœud  $b_0$  à  $\kappa_{b_0} ;$ 
11 Appeler la fonction  $\text{RecursiveEvRANSAC}(\mathcal{G}, \mathcal{M}, \Omega, \kappa_{th}, b_0, l, l_{max}) ;$ 
12  $\kappa_{\hat{b}} = +\infty ; n \leftarrow N ;$ 
13 while  $\kappa_{\hat{b}} > \kappa_{th}$  et  $n > 0$  do
14   for  $j \leftarrow 1$  to  $\binom{n}{N}$  do
15      $b \leftarrow$  code binaire du  $j^{\text{ème}}$  nœud composé de  $N$  sources ;
16      $\kappa_b \leftarrow$  valeur de conflit de  $b ;$ 
17     if  $\kappa_b \in [0, \kappa_{\hat{b}})$  then
18        $\kappa_{\hat{b}} \leftarrow \kappa_b ; \hat{b} \leftarrow b ;$ 
19     end
20   end
21   if  $\kappa_{\hat{b}} \geq \kappa_{th}$  then
22      $n \leftarrow n - 1 ;$ 
23   end
24 end
25 Mettre dans  $\mathcal{I}_S$  les indices des sources en  $\hat{b}$  (e.g.,  $\mathcal{I}_S \leftarrow \{n \mid \hat{b} \& (1 \ll n) = 1\}$ ) ;
```

pendant au cardinal du sous-ensemble de sources représenté par le nœud b .

Dans le but de sauvegarder les chemins d'exploration, les nœuds de \mathcal{G} contiennent aussi deux autres valeurs : la valeur du conflit induite par la combinaison des nœuds sources avec le nœud précédent sur le chemin d'exploration (appelé 'père'), et le code binaire du père dont la combinaison courante est issue. Ces valeurs sont seulement calculées durant l'exploration, si bien que pour tout nœud non atteint, la valeur de conflit ne sera pas calculée (elle est initialisée à une valeur négative, indiquant qu'elle n'est pas valide). Notons que, puisque l'exploration de l'espace des solutions n'est que partielle, certaines de ces valeurs ne seront jamais calculées.

Le chemin d'exploration commence par les nœuds composés de deux sources présentant les valeurs de conflit les moins élevées. Ce choix est un compromis entre complexité et séparabilité des

Algorithm 11: Fonction RecursiveEvRANSAC; paramètres : graphe \mathcal{G} , ensemble des bbas \mathcal{M} , cadre de discernement Ω , κ_{th} seuil de conflit, nœud courant b , itération courante l , nombre maximum d'itérations l_{max}

```

1  $N \leftarrow |\mathcal{M}|$  et initialiser  $\kappa_{min} \leftarrow \kappa_{th}$  et  $\hat{b} \leftarrow \emptyset$ ;
2 for  $j \leftarrow 1$  to  $N$  do
3   if  $j \notin b$  then
4      $b' \leftarrow b \cup j$ ;  $\kappa_{b'} \leftarrow \bigoplus_{i \in b'} m_i(\emptyset)$ ;
5     if  $\kappa_{b'} < \kappa_{min}$  then
6        $\kappa_{min} \leftarrow \kappa_{b'}$ ;
7        $\hat{b} \leftarrow b'$ ;
8     end
9   end
10 end
11 if  $l < l_{max}$  et  $|\hat{b}| < N$  then
12   while  $\hat{b} = \emptyset$  et  $|b| > 0$  do
13      $\kappa_b \leftarrow$  valeur du conflit en  $b$ ;
14      $\hat{b} \leftarrow \arg \min_{b', |b'|=|b|, \kappa_{b'} > \kappa_b} \kappa_{b'}$ ;
15      $b \leftarrow$  nœud père de  $\hat{b}$ ;
16   end
17   RecursiveEvRANSAC( $\mathcal{G}, \mathcal{M}, \Omega, \kappa_{th}, \hat{b}, l + 1, l_{max}$ );
18 end

```

nœuds selon les valeurs de conflit. D'une part, pour les nœuds comportant une seule source, les valeurs de conflit sont nulles, et même pour les nœuds à deux sources, une valeur nulle du conflit peut être atteinte par plusieurs nœuds. D'autre part, en considérant des nœuds de k sources, le nombre de combinaisons de sources à tester pour choisir le nœud initial est de $\binom{N}{k}$, une valeur qui augmente exponentiellement pour k inférieur à $\frac{N}{2}$. À la fin de l'exploration, i.e. quand le nombre maximum d'itérations est atteint, nous sélectionnons la meilleure solution comme étant celle qui a le cardinal le plus élevé avec une valeur de conflit inférieure au seuil. Si plusieurs solutions satisfont le critère, la solution retenue est celle qui possède la valeur de conflit minimale.

7.3.1.2 Fonction récursive

L'algorithme 10 présente l'approche EV-RANSAC proposée, incluant une fonction récursive utilisée pour parcourir le graphe \mathcal{G} . Cette dernière est décrite dans l'algorithme 11 qui utilise les mêmes notations que l'algorithme 10. À partir d'un nœud passé en argument de la fonction, nous examinons toutes les combinaisons ajoutant une source à ce nœud. Pratiquement, pour toute source ayant un indice $j \in \{1, \dots, N\}$, si la source j n'appartient pas effectivement au sous-ensemble de sources b , nous calculons le conflit (ou le critère de cohérence ou d'incohérence) associé au sous-ensemble de sources $j \cup b$. Le prochain nœud choisi est alors celui qui minimise ce conflit. Notons que toutes les opérations ou tests sont simplement effectués par des opérateurs bit à bit.

TABLE 7.2 – Initialisation de 5 sources pour illustrer le principe du RANSAC évidentiel avec $\Omega = \{\theta_1, \theta_2, \theta_3\}$.

A	S_1	S_2	S_3	S_4	S_5
θ_1	0	0	0	0.7	0.3
θ_2	0.05	0	0.8	0	0
$\theta_1 \cup \theta_2$	0.75	0.8	0	0.3	0
$\theta_2 \cup \theta_3$	0	0	0.2	0	0
Ω	0.2	0.2	0	0	0.65

La récursion est stoppée si le nombre maximum d'itérations est atteint ($l_{max} \ll 2^N - \binom{1}{N} - \binom{2}{N}$) ou si le nœud qui inclut l'ensemble des sources est la solution finale. À partir du nœud courant, la récursion ajoute une nouvelle source (celle qui minimise la valeur du conflit) à condition que le conflit reste en dessous du seuil maximal. Dans le cas contraire, il recherche le nœud de même cardinal (que le sous-ensemble de sources b) n et ayant la valeur du conflit 'suivante' dans la liste ordonnée des nœuds (valeur de conflit la moins élevée parmi les nœuds ayant des valeurs plus élevées que le nœud courant b). En termes de graphe, cela revient à revenir au nœud 'père' (de cardinal de $n - 1$) et puis continuer l'exploration parmi ses voisins de cardinal n (dont les valeurs de conflit ont déjà été calculées lors du traitement du nœud père). Si un tel nœud n'existe pas (tous les nœuds 'frères' ont été testés), le nœud suivant est recherché parmi les 'frères' du 'père' et ainsi de suite.

7.3.2 Exemple jouet pour illustrer le RANSAC évidentiel

Pour illustrer le fonctionnement de l'algorithme, la figure 7.2 présente un exemple avec cinq sources où l'une d'entre elle est en conflit fort avec la combinaison des sources restantes. Le tableau 7.2 présente les fonctions de masse de chaque source sur le cadre de discernement composé de trois hypothèses $\Omega = \{\theta_1, \theta_2, \theta_3\}$. La disposition des sources illustre les différents cas de figure que peut rencontrer l'algorithme proposé. Selon l'algorithme 10, nous commençons par calculer le conflit associé à chaque paire de sources représentées par le premier niveau de combinaison du graphe de la figure 7.2. Pour l'exemple, nous nous restreignons (à cette étape) à ne combiner les sources qu'avec la source S_1 . La combinaison conjonctive des sources est représentée par les indices à l'intérieur des nœuds pour une meilleure lisibilité (e.g. 12 au lieu de S_{1n2}).

Les mesures de conflit (mesure d'incohérence pour cet exemple) sont affichées à côté de chaque nœud sauf si le conflit est nul, ce qui est le cas pour la conjonction des sources S_2 et S_3 avec la source S_1 . Ayant un conflit nul sur 2 des 4 conjonctions, nous sélectionnons aléatoirement l'une d'entre elles pour entrer dans la partie récursive de notre approche (algorithme 11). Nous continuons alors d'explorer le graphe en combinant le résultat de S_{1n2} avec les sources restantes, ce qui nous donne un conflit nul sur la combinaison S_{1n2n3} , un conflit de 0.015 pour S_{1n2n5} et de 0.035 pour S_{1n2n4} . Les valeurs de conflit pour les nœuds du même niveau, mais qui n'ont pas été explorés (tous les nœuds gris), sont présentées afin d'avoir une idée de la pertinence de la mesure de conflit comme mesure

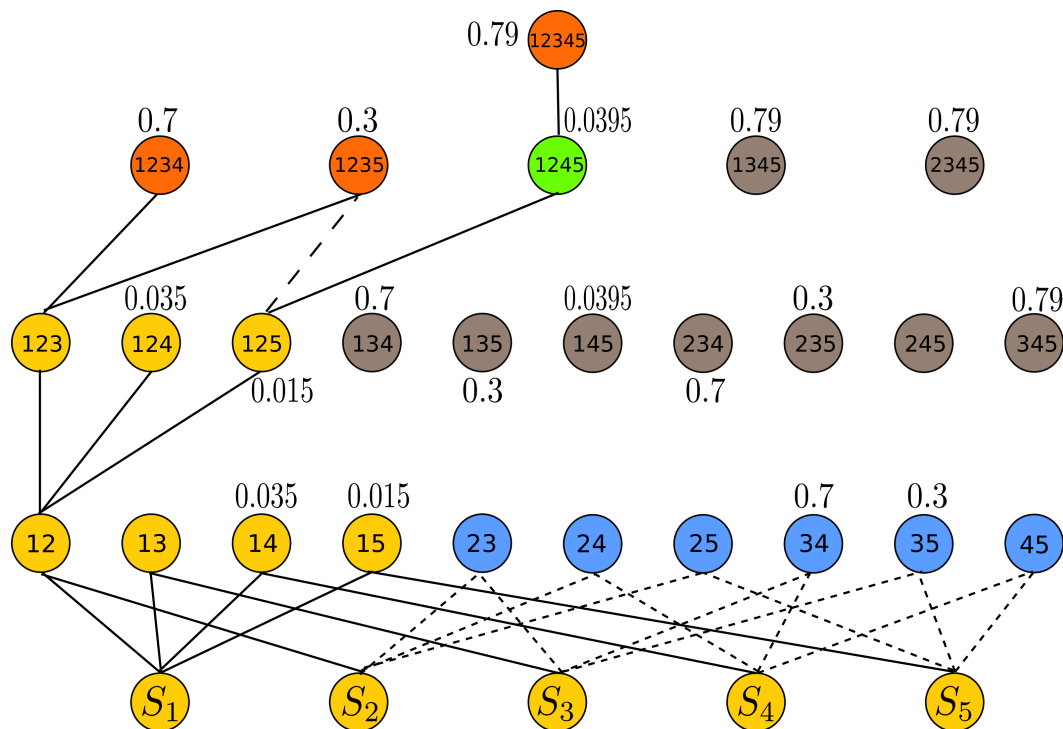


FIGURE 7.2 – Illustration de l’approche EV-RANSAC avec 5 sources dont 1 défaillante ; les 5 sources sont au premier niveau, le deuxième niveau représente toutes les combinaisons de deux sources, et ainsi de suite. Les nombres réels à côté des nœuds sont les valeurs du conflit quand elles sont calculées et non nulles.

d’incohérence. L’exploration du graphe continue en combinant $S_{1n_2n_3}$ avec les sources restantes (S_4 et S_5), ce qui nous donne dans les deux cas (nœuds en rouge) un conflit supérieur au seuil maximum fixé à $\kappa_{th} = 0.2$. L’algorithme retourne alors au nœud ‘père’ (S_{1n_2}) du nœud courant ($S_{1n_2n_3}$) pour continuer l’exploration à partir du deuxième nœud le moins conflictuel, c’est-à-dire $S_{1n_2n_5}$. Le nœud $S_{1n_2n_5}$ est combiné avec la source S_4 , la combinaison avec la source S_3 ayant déjà été effectuée à la précédente itération. La valeur du conflit obtenue est inférieure à κ_{th} (nœud vert), qui correspond à l’ensemble le plus cohérent, pour le seuil combinant le plus grand nombre de sources. Il n’est pas nécessaire de continuer la combinaison conjonctive de toutes les sources puisque le conflit maximum en combinant 4 sources parmi 5 est largement supérieur au seuil. Dans cet exemple, nous voyons alors que la source conflictuelle est S_3 et que le conflit apparaît principalement avec S_4 .

Au total, 10 nœuds ont été testés sur 25 pour arriver à une combinaison cohérente. Le chemin d’exploration du graphe aurait pu être plus long ou plus court suivant le choix du premier nœud père au niveau des premières combinaisons. Le choix du nœud S_{1n_3} aurait induit un nombre d’itérations (la source S_3 est la plus conflictuelle) plus élevé, alors que le choix du nœud S_{2n_5} (en commençant par la source S_5) aurait induit moins d’itérations. Les paramètres de seuil sur la mesure d’incohérence (ou de cohérence) ainsi que le nombre d’itérations sont critiques et doivent être choisis selon l’application.

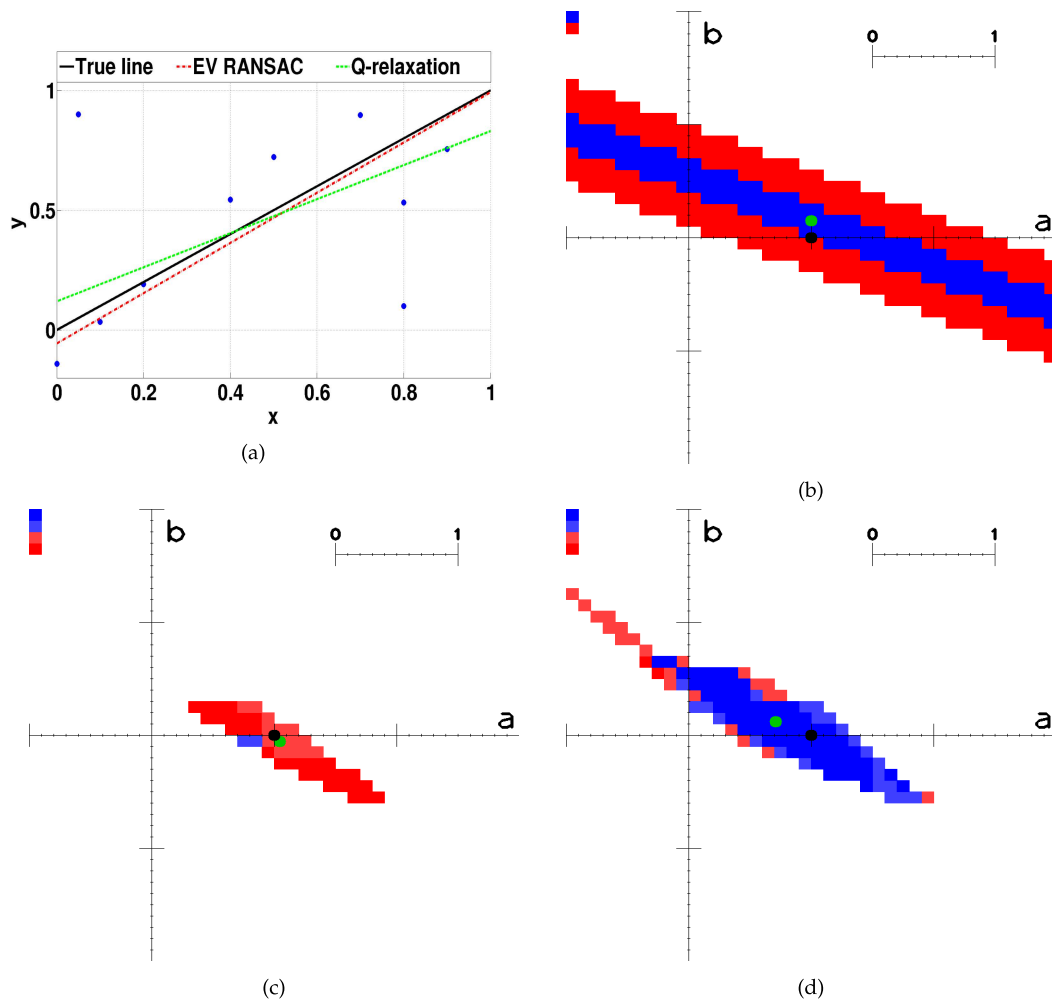


FIGURE 7.3 – Comparaison entre EV-RANSAC et la q-relaxation évidentielle dans le cas d'un exemple (droite $y = ax + b$) avec 10 points et 2 outliers : (a) dataset, (b) 4^{ème} point représenté sous forme de bba dans l'espace des paramètres (a, b) , (c) solution fournie par EV-RANSAC, (d) solution fournie par la q-relaxation évidentielle.

7.3.3 Exemple d'application sur les bbas 2D : estimation des paramètres d'une droite

Pour illustrer l'intérêt du RANSAC évidentiel, nous considérons l'exemple classique de l'estimation des paramètres d'une droite à partir d'un nuage de points comportant des *inliers* et des *outliers*. La figure 7.3 montre 10 points tirés aléatoirement : 8 d'entre eux présentent du bruit autour de la droite à estimer (pente de 1 et biais de 0) alors que les deux autres points sont aberrants. Plusieurs approches ont été proposées pour résoudre ce type de problème : qu'elles soient statistiques, comme présenté dans le chapitre 5 ou utilisant la reconnaissance de forme (e.g., transformée de Hough). Notre but est ici d'illustrer le résultat fournit par l'EV-RANSAC et sa capacité à effectuer une estimation robuste. Étant donné que deux paramètres doivent être estimés (a la pente et b l'offset de la droite), le cadre de discernement est un espace compact 2D que nous bornons comme suit : $a \in [-1, 3]$ et $b \in [-2, 2]$, et discrétisé avec un pas de discrétisation égal à 0.1. $|\Omega| = 40 \times 40$. Les sources sont les points 2D (*inliers* et *outliers*). Chacun d'eux génère une bba contenant deux éléments

focaux (comme plus tard dans la partie application). Pour l'estimation de la droite, les éléments focaux générés par les points (x_0, y_0) sont consonants et forment une bande autour du segment de droite $b = y_0 - a \times x_0$, $(a, b) \in [-1, 3] \times [-2, 2]$ (spécifiquement, la largeur de A_1 est de 0.1 et celle de A_2 est de 0.3, $m(A_1) = 0.49$ et $m(A_2) = 0.51$, le choix des largeurs des éléments focaux est ici empirique et doit être supérieurs au pas de discrétisation).

La figure 7.3 présente l'exemple traité : les données en (a) ; un exemple de bba avec deux éléments focaux consonants en (b) ; le résultat obtenu par une combinaison conjonctive des bba sélectionnées comme *inliers* par la méthode proposée EV-RANSAC en (c) et celui fourni par la q-relaxation évidentielle [Pichon et al., 2015] en (d). Nous notons que, dans cet exemple en particulier, nos résultats sont plus spécifiques (engagés) que ceux de la q-relaxation évidentielle, tout en étant également plus précis.

7.4 Application à la localisation GPS

7.4.1 Formulation évidentielle du problème de localisation

Pour nous adapter au nouveau cadre formel des fonctions de croyance, nous reformulons légèrement notre problème de localisation GPS et nos équations. Le modèle d'observation représenté dans le repère ENU (comme dans le chapitre 6), est légèrement différent de l'équation 6.1 et s'écrit :

$$\rho_j = \sqrt{(e_r - e_{S_j})^2 + (n_r - n_{S_j})^2 + (u_r - u_{S_j})^2} + c\delta t + \varepsilon_{ae} + \varepsilon_{mp}. \quad (7.11)$$

Dans l'équation 7.11, deux types de bruits sont distingués : les bruits atmosphériques et électroniques ε_{ae} et les effets de multi-trajets ε_{mp} . Ces deux termes vont servir à l'initialisation des deux éléments focaux de la bba associée à une mesure de pseudo-distance.

Dans le cadre des fonctions de croyance, il est assez difficile de manipuler des dimensions supérieures à deux pour un cadre de discernement. Pour une représentation en deux dimensions, les coordonnées Est et Nord sont les plus importantes en localisation appliquée à des mobiles terrestres. Elles sont représentées par le vecteur $\mathbf{X}_{bba} = (e_r, n_r)^\top$. Pour des applications terrestres, nous pouvons alors supposer que l'altitude (ou plus précisément la coordonnée Up) puisse être calculée *a posteriori* à partir du vecteur \mathbf{X}_{bba} en utilisant un modèle numérique de terrain. En ce qui concerne l'estimation du biais d'horloge, un filtre de Kalman est utilisé avec l'hypothèse d'un modèle linéaire gaussien et une dérive d'horloge estimée *a priori* par différence finie.

Alors, l'estimation de la pseudo-distance est :

$$\tilde{\rho}_j(\mathbf{X}_{bba}) = \left[(e_r - e_{S_j})^2 + (n_r - n_{S_j})^2 + \alpha_j \right]^{1/2} + \beta, \quad (7.12)$$

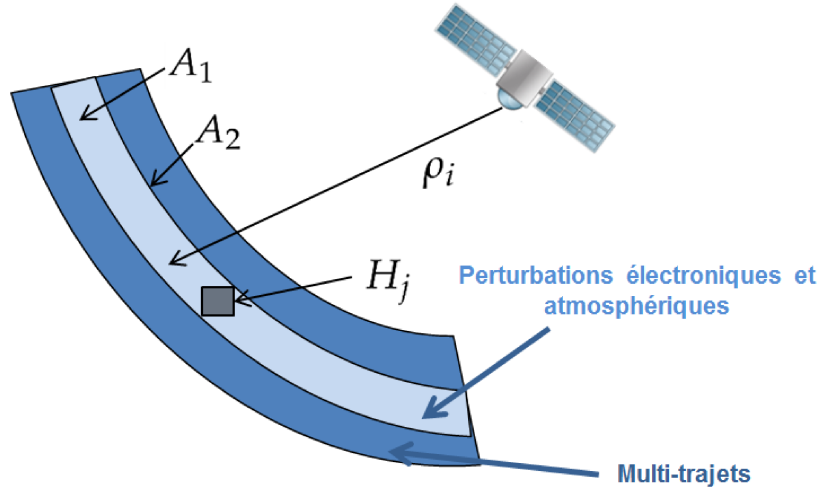


FIGURE 7.4 – Initialisation de la bba.

où $\alpha_j = (\tilde{u}_r - u_{S_j})^2$ est la différence au carré des coordonnées Up d'un satellite S_j et du récepteur, et $\beta = c\tilde{\delta}t$ l'estimation par le filtre de Kalman du biais temporel multiplié par la vitesse de la lumière.

De la même manière que dans l'exemple de la droite (voir la figure 7.3), le cadre de discernement est borné, de largeur et hauteur (w, h) , avec un pas de discrétisation R identique sur les axes Est et Nord, ce qui nous donne une taille du cadre de discernement $|\Omega| = \frac{w \times h}{R^2}$.

L'initialisation des bbas est une tâche sensible et primordiale pour toute modélisation sous forme de fonctions de croyance. Dans notre cas, les bbas sont consonnantes avec deux éléments focaux, A_1 et A_2 , qui sont définis suivant le modèle présenté par l'équation 7.11 : A_1 modélise l'imprécision des bruits atmosphériques et électroniques, et A_2 inclue en plus l'impact des signaux multi-trajets (comme illustré sur la figure 7.4). Les éléments focaux représentent alors deux niveaux d'imprécision et sont naturellement consonnantes comme dans [André et al., 2015] (l'élément focal A_1 est toujours incluse dans l'élément focal A_2).

L'initialisation des bbas est effectuée suivant l'équation 7.13 :

$$j \in [0, \Omega] : H_j \subset \begin{cases} A_1 & \text{si } \left| \tilde{\rho}_i(\mathbf{X}_{bba}(H_j)) - \rho_i \right| \leq \sigma_{noise} \\ A_2 & \text{si } \left| \tilde{\rho}_i(\mathbf{X}_{bba}(H_j)) - \rho_i \right| \leq \sigma_{noise} + \sigma_m \end{cases}, \quad (7.13)$$

avec $i \in [1, N_{PR}]$ et N_{PR} le nombre de PR (ou de satellites) à l'epoch considérée.

7.4.2 Estimation dans le cas dynamique

En filtrage temporel, nous distinguons deux étapes importantes : l'étape de prédiction, qui permet de projeter l'état du système à l'instant suivant, et l'étape d'estimation qui sert de correction ou de mise à jour de l'état prédit. Il est nécessaire, dans le cas des bbas 2D, de prendre en considération la

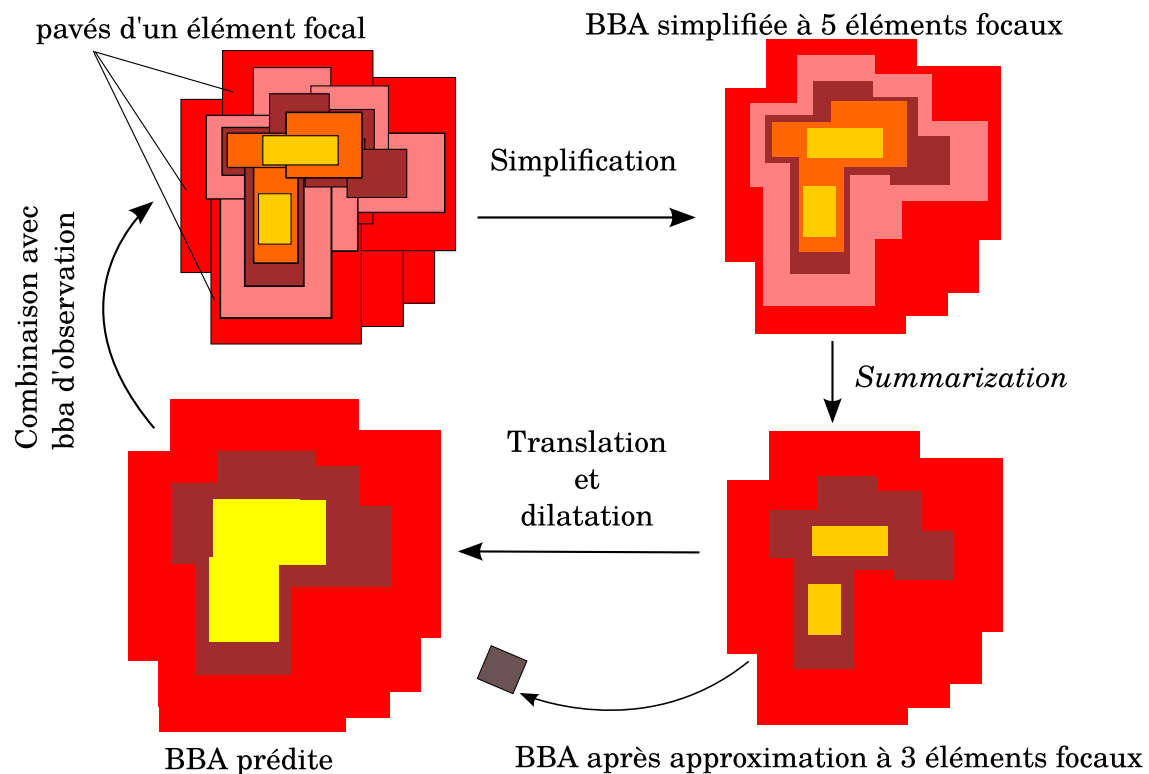


FIGURE 7.5 – Processus d'estimation de la bba 2D

bba prédite pour garder un certain niveau de précision dans l'estimation de la localisation. La figure 7.5 illustre le processus d'estimation dynamique de localisation dans le cas de bbas 2D utilisées dans ce chapitre.

7.4.2.1 Étape de prédiction

Comme tout processus de fusion, une projection de la position à $t - 1$ vers l'instant t est nécessaire pour suivre le modèle de déplacement de l'objet à localiser. L'étape de prédiction, en plus de la position, projette aussi l'imprécision de la localisation qui est liée aux approximations du modèle de déplacement, comme dans le filtrage statistique (filtre de Kalman).

Dans notre cas, la bba est déplacée suivant un modèle linéaire fonction de la vitesse. Cette dernière est obtenue par différences finies mais pourra également être calculée, dans le cas du GPS, en utilisant les mesures Doppler. Ainsi, tous les éléments focaux de la bba sont déplacés de la même manière. Par la suite, chaque élément focal est dilaté suivant un élément structurant (dans notre cas, un disque de rayon donné, approximé par un ensemble de pixels carrés selon la résolution choisie pour l'étape de simplification), pour ajouter de l'imprécision liée aux erreurs du modèle de déplacement. Cette dilatation est appliquée à chaque élément focal et n'augmente pas la complexité de la bba.

7.4.2.2 Combinaison des bbas de prédiction et d'observation

L'étape de correction ou de mise à jour consiste à fusionner deux bbas, une bba de prédiction et une bba d'observation obtenue par combinaison de sources. Cette bba d'observation est issue de la combinaison conjonctive des bbas *inliers* détectées précédemment. Il existe différentes façons de combiner les deux bbas d'observation et de prédiction, mais dans le cas d'un filtrage temporel et dans le but d'améliorer la précision d'un processus donné, deux règles nous semblent naturelles : la règle conjonctive et la révision.

La règle conjonctive, qui est commutative, est probablement la plus intuitive car elle permet d'affiner la localisation avec les informations complémentaires venant des deux bbas (pour peu qu'elles contiennent la vraie solution).

Pour ce qui est de la révision, non commutative, cette règle se fonde sur la mise à jour d'une information *a priori* avec une nouvelle donnée d'entrée, à l'instar de la probabilité conditionnelle en théorie des probabilités. Dans notre cas, la 'logique' voudrait que la prédiction soit l'information *a priori* et l'observation la nouvelle information acquise à chaque pas de temps. Soit m et m_I deux bbas sur 2^Ω , la règle de révision de m par m_I , pour toute hypothèse $C \neq \emptyset$, est définie dans [Ma et al., 2011] :

$$(m \circ_r m_I)(C) = \sum_{A \cap B = C} \sigma_r(A, B) m_I(B), \quad (7.14)$$

où

$$\sigma_r(A, B) = \begin{cases} \frac{m(A)}{Pl(B)} & \text{si } Pl(B) > 0 \\ 0 & \text{si } Pl(B) = 0 \text{ et } A \neq B \\ 1 & \text{si } Pl(B) = 0 \text{ et } A = B \end{cases} \quad (7.15)$$

Nous allons à présent appliquer l'approche proposée à la localisation GPS en utilisant comme sources d'informations les pseudo-distances modélisées sous forme de bbas consonantes.

7.5 Résultats

Une première évaluation des approches évidentielles a été menée en utilisant le deuxième jeu de données (en statique et en environnement urbain, paragraphe 4.1.3 du chapitre 3). Contrairement à la démarche entreprise dans le chapitre 5 pour tester l'approche *a contrario* où une fenêtre temporelle est considérée, nous estimons la position du récepteur indépendamment à chaque *epoch*. Le filtrage de Kalman sur le paramètre Up et sur le biais temporel est le seul à exploiter les *a priori* sur les mesures pour assurer la non-divergence des méthodes développées dans ce chapitre.

L'évaluation de la méthode du RANSAC évidentielle appliquée à l'estimation dynamique de la localisation se fera en utilisant le jeu de données 3, présenté dans le chapitre 4. Nous considérerons

TABLE 7.3 – Performances statistiques en fonction du rayon des éléments focaux.

Méthode	rayon des éléments focaux	M_1 (m)		M_2 (m)	
		68 %	95 %	68 %	95 %
EvRANSAC	(4, 20)	10.72	20.81	15.43	23.45
	(6, 15)	10.07	16.54	12.36	17.84
	(6, 20)	9.51	17.74	13.36	20.25
Pichon and al. 2015	$\sigma(4, 20)$	10.92	31.63	23.90	49.49
	(6, 15)	8.31	29.22	15.92	40.46
	(6, 20)	8.63	17.95	19.10	32.38
AI q-relaxation	$\sigma(12)$	10.63	17.71	15.99	20.36
	(15)	11.18	18.31	14.35	20.24
	(20)	11.86	23.57	18.32	25.59

alors les données en zone urbaine et en forêt, comme dans les chapitres 5 et 6.

L'évaluation des résultats est loin d'être triviale puisque nous obtenons une solution sous forme d'ensemble(s) tandis que la vérité terrain de localisation est ponctuelle. Une première possibilité (qui a été adoptée dans le cas des tests en statique) est de calculer une solution (de localisation) ponctuelle à partir de la bba finale. Une seconde possibilité (plus intéressante à notre sens et adoptée dans le cas des tests en dynamique) est de prendre en compte à la fois la croyance en un ensemble et sa géométrie. Nous proposerons alors un critère dans ce sens.

7.5.1 Cas statique

Notre approche est comparée avec la méthode proposée par Pichon et al. [Pichon et al., 2015] ainsi qu'avec l'analyse par intervalles (AI) utilisant la q-relaxation. Dans l'algorithme 10, nous initialisons le seuil sur le conflit à $\kappa_{th} = 0.2$ et le nombre d'itérations maximum à $l_{max} = \min(3N, N + 2^{N-3})$, avec N le nombre de satellites par *epoch*. En outre, ayant déterminé le sous-ensemble de bbas *inliers*, ces dernières sont combinées de façon conjonctive (voir l'équation 7.1) afin d'obtenir la bba à partir de laquelle la localisation du récepteur est estimée. Dans [Pichon et al., 2015], le seuil utilisé est une mesure de cohérence ϕ_{min} représentant le maximum de la fonction contour. Pour notre comparaison, la valeur de ce seuil est fixée à $\phi_{min} = 0.8$. Finalement, pour l'AI avec q-relaxation, le nombre de contraintes relâchées est le nombre minimum de mesures prises en compte pour qu'une solution existe ($\neq \emptyset$). Cette méthode, introduite par [Jaulin et al., 1996], est appelée GOMNE (*Guaranteed Outlier Minimal Number Estimator*). Néanmoins, elle n'assure pas que la solution obtenue soit correcte en particulier il peut rester des *outliers*. Aussi avons-nous modifié le critère d'arrêt initial pour prendre en compte d'éventuels *outliers* supplémentaires et pour éviter une solution erronée [Jaulin et al., 2002]. Le nombre de contraintes à relâcher est alors $q = q_{min} + r$, avec q_{min} obtenu par GOMNE et r une 'marge' fixée à

1. Nous notons que seules les variables Est et Nord sont estimées par l'analyse par intervalles pour une comparaison équitable avec les approches évidentielles limitées comme dit précédemment au cas 2D. Pour l'initialisation, l'élément focal le plus large de la bba de la mesure PR considérée est le pavé attribué à la mesure utilisée par l'analyse par intervalles avec q-relaxation.

Dans cette sous-partie (cas statique), les méthodes sont comparées en termes de précision de localisation. À partir de la bba finale (qui résulte de la combinaison conjonctive des bbas *inliers*), nous calculons la probabilité de distribution fournie par la fonction BetP (voir l'équation 7.8). À partir de cette distribution, deux types d'erreurs sont calculées. La première nous permet d'évaluer si la solution est biaisée. Elle s'écrit comme une espérance statistique, connaissant la distribution de probabilité : $E_1 = \sum_{i=1}^{|\Omega|} \varepsilon(H_i) \text{BetP}(H_i)$, où $\varepsilon(H_i)$ est la différence entre le centroïde du pavé élémentaire associé à l'hypothèse H_i et la vérité terrain. La deuxième mesure nous permet d'évaluer l'imprécision de la solution (en évitant que les erreurs se compensent). Elle s'écrit comme l'espérance des erreurs de $Norm_1$: $E_2 = \sum_{i=1}^{|\Omega|} |\varepsilon(H_i)| \text{BetP}(H_i)$.

Le tableau 7.3 montre les erreurs obtenues pour les différentes méthodes de localisation (RANSAC évidentiel, q-relaxation évidentielle, AI avec q-relaxation) pour différents rayons d'éléments focaux A_1 et A_2 (i.e., en supposant différents niveaux de bruit sur les mesures de pseudo-distances). Il nous permet d'évaluer la robustesse des approches à ce paramètre (niveau de bruit) qui est fixé empiriquement. Le tableau 7.3 présente les *p-values* correspondant aux 68^{ème} et 95^{ème} percentiles (valeurs choisies comme étant $1 \times \sigma$ et $2 \times \sigma$ de la distribution normale). Nous notons que l'approche proposée a de meilleures performances suivant la mesure E_2 et suivant le 95^{ème} percentile de la mesure E_1 . Ceci est dû au fait qu'en utilisant la règle de combinaison conjonctive sur les bbas *inliers*, le résultat est beaucoup plus engagé qu'avec la q-relaxation (que ce soit la méthode évidentielle ou l'AI), ce qui a déjà été remarqué dans l'exemple jouet de la droite. Nous notons que les meilleurs résultats sont obtenus avec un niveau de bruit de 6m et 15m respectivement pour chaque élément focal. Pour le 68^{ème} percentile de la mesure E_1 , la méthode de Pichon et al dépasse en performance toutes les méthodes pour le même niveau de bruit, suivie par la méthode EV-RANSAC. Toutefois, la q-relaxation évidentielle paraît moins robuste, au niveau de la représentation des bruits par une largeur dans les éléments focaux, que la méthode de localisation basée sur le RANSAC évidentiel.

7.5.2 Cas dynamique

Pour le cas dynamique, les paramètres de l'algorithme EV-RANSAC sont les suivants : nous utilisons le conflit comme mesure d'incohérence avec un seuil fixé $\kappa_{max} = 0.2$. Les autres paramètres de la modélisation évidentielle sont : largeurs de bandes de la bba de pseudo-distance sont fixées à (6, 15)m pour l'élément focal 1 et 2 de la bba, correspondant aux meilleurs résultats obtenus dans la partie statique (voir le tableau 7.4), cadre de discernement défini sur une surface de $80 \times 80\text{m}^2$ avec une résolution de $2 \times 2\text{m}^2$, soit $|\Omega| = 1600$ hypothèses singletons.

La figure 7.6 présente un exemple de combinaisons entre une observation à une *epoch* considérée et différentes bbas de prédiction issues de différents types de combinaisons. La première combinai-

TABLE 7.4 – Résultat de localisation sur un cadre de discernement de taille 80×80 avec une résolution de 2×2 .

Méthodes	$Norm_1$	RMSE	μ_{loc}	σ_{loc}
LS	(4.49, 4.89)	(6.02, 8.60)	7.3786	7.4716
NFA_1	(3.1, 4.4)	(4.0, 5.9)	5.93	3.88
observation	(3.98, 3.31)	(4.70, 4.96)	5.81	3.60
Révision 1	(4.01, 3.47)	(4.69, 5.01)	6.02	5.84
Révision 2	(4.04, 3.61)	(4.71, 5.28)	6.14	3.52
règle conjonctive	(3.85, 4.19)	(4.61, 5.77)	6.41	3.69

son représente la règle conjonctive classique (colonne de gauche). La deuxième combinaison (notée ‘Révision 1’) représente la révision de la bba d’observation sachant un *a priori* sur la bba prédiction. Dans la dernière combinaison (notée ‘Révision 2’) les rôles des bbas de prédiction et d’observation sont inversés par rapport à la ‘Révision 1’. Le code couleur, du rouge foncé vers le bleu foncé, est un tri de l’élément focal du plus grand au plus petit. Nous remarquons que la bba d’observation a ses éléments focaux plus concentrés autour de l’estimation réelle de position que les bbas combinées avec les prédictions, et ce pour toutes les règles utilisées. Nous remarquons aussi que les bbas de prédiction ont la même taille (le plus large élément focal) mais différent dans la disposition et la taille des éléments focaux inclus. Par conséquent, les résultats sont différents dans la disposition des éléments focaux dans les bbas combinées, même si l’élément focal le plus large et le plus petit sont similaires pour toutes les bbas résultats. Qualitativement, et pour cet échantillon, nous remarquons que tous les éléments focaux des bbas incluent la vérité terrain (en bleu).

Pour évaluer quantitativement l’approche proposée, le tableau 7.4 présente les statistiques de localisation des méthodes proposée, comparées à la méthode des moindres carrés (LS) donnée à titre d’illustration et à l’approche *a contrario* des chapitres précédents, pour comparaison entre une approche ponctuelle et une approche ensembliste. Étonnamment, les meilleurs résultats sont donnés par la bba d’observation (meilleure que le NFA_1) sans combinaison avec la bba de prédiction, dans presque toutes les mesures de performances utilisées. De plus, la règle conjonctive donne les moins bons résultats et ne semble pas être adaptée au filtrage temporel. Une des raisons pour lesquelles les bbas combinées n’ont pas de meilleures performances est peut être liée à la *summarization* à quatre éléments focaux qui aurait éliminé des éléments pertinents pour l’estimation finale de la solution. Néanmoins, une réduction du nombre d’éléments focaux éliminés lors de l’approximation de bba entraînera une charge combinatoire plus importante.

L’évaluation quantitative précédente est destinée à comparer des approches d’estimation ponctuelle. Nous avons dit qu’elle n’est peut-être pas pertinente dans le cas de solution ensembliste. À titre d’illustration, l’analyse par intervalles préfère se placer en termes de ‘garantie’ de la solution plutôt que de précision de cette dernière. Un critère qui tiendrait compte de l’imprécision des solutions obtenues paraît alors plus appropriée pour une comparaison avec des approches de nature ensembliste. Nous proposons le critère suivant qui nous permet de quantifier la précision de localisation

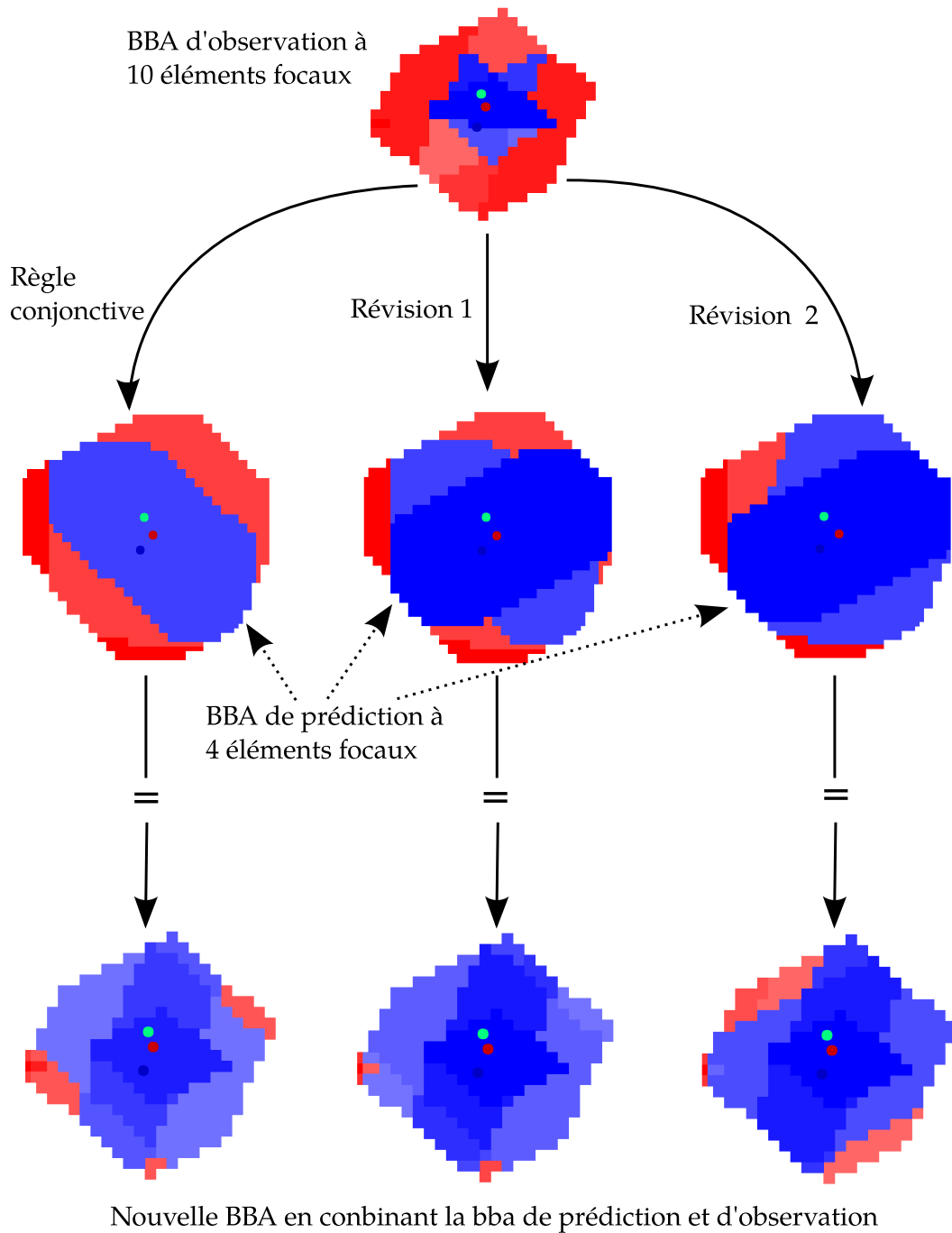


FIGURE 7.6 – Résultats de combinaison entre une bba d'observation et de prédiction en utilisant différentes règles de combinaison. Le point bleu représente la vérité terrain, le point vert l'estimation par moindres carrés et le point rouge la méthode EV-RANSAC.

versus le fait d'avoir une solution garantie :

$$E_{BBA} = \sum_{i=1}^m (|A| + \gamma(A) \lambda) \frac{m(A)}{m(\emptyset)}, \quad (7.16)$$

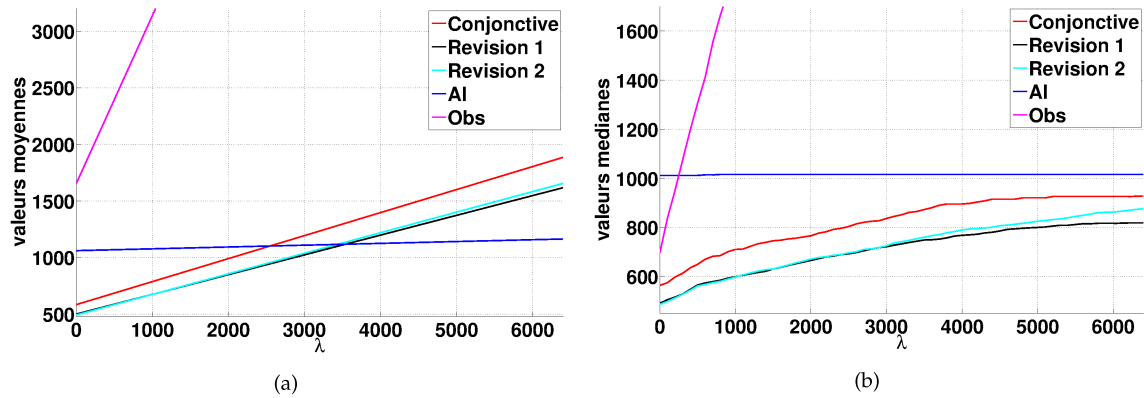


FIGURE 7.7 – Performances selon le critère proposé (équation 7.16) en fonction du coefficient de pénalité λ ; (a) valeurs moyennes (b) valeurs médianes, sur l’expérience en zone urbaine

où

$$\gamma(A) = \begin{cases} 0 & \text{si } X_{RTK} \in A \\ 1 & \text{si } X_{RTK} \notin A \end{cases}, \quad (7.17)$$

et λ est une pénalité dans le cas où l’hypothèse vérité terrain n’appartient pas à l’élément focal A . La valeur de λ peut être vue comme l’ajout d’une pénalité (plus ou moins élevée en fonction de la valeur du paramètre) aux éléments focaux qui n’incluent pas la vérité terrain (condition $\gamma(A)$). Si $\lambda = 0$, seul la précision de la bba est prise en compte indépendamment du fait qu’elle inclut ou non la vraie position du récepteur. Si $\lambda \rightarrow \infty$, le fait que chaque élément focal doit contenir la solution devient une contrainte dure (strict). Ce critère s’applique aussi pour l’AI avec un seul élément focal et $\lambda \gg 0$ (exigence de solution ‘garantie’).

Les figures 7.7a et 7.7b présentent les résultats sur les différentes méthodes ensemblistes en fonction des valeurs du coefficient de pénalité λ . Nous remarquons sur les deux figures que les valeurs moyennes et médianes de l’analyse par intervalles ne varient que très légèrement, impliquant que la plupart des ensembles estimés contiennent la vérité terrain. En revanche, l’estimation par l’approche évidentielle des bbas d’observation diverge très rapidement. Cela est dû à la présence de petit éléments focaux à fortes masses ne contenant pas la vérité terrain, ce qui augmente rapidement le critère E_{BBA} . Les différentes combinaisons de bbas de prédiction et d’observation donnent des résultats plus intéressantes que les résultats de l’AI et des bbas d’observation. Nous observons que, contrairement au cas des critères du tableau, les contributions faisant intervenir la bba de prédiction en plus de l’observation sont plus satisfaisantes. Cela met en évidence le fait que les bbas sont beaucoup plus engagées que la bba d’observation, même si le prix de cet engagement supplémentaire est un léger biais sur l’estimation ponctuelle de la localisation. La combinaison des bbas a permis d’améliorer la précision des éléments focaux et de réduire la masse de ceux qui ne sont pas compatibles avec la vérité terrain. En augmentant la valeur de λ , nous pénalisons de plus en plus la non inclusion de la solution vérité terrain dans les éléments focaux. Sur les courbes représentant la moyenne statistique des valeurs du critère sur la durée de l’expérience (7.7a), l’AI devient préférable à l’approche évidentielle pour les très fortes valeurs de λ (intersection en 3500, 2600). Cependant quand nous considérons la médiane plutôt

que la moyenne statistique, l'approche évidentielle reste meilleure quelque soit λ (7.7b). Cela signifie qu'il existe dans l'expérience un certain nombre d'*epochs* où l'analyse évidentielle échoue à inclure la vraie solution dans les éléments focaux de masse significative. En extrayant quelques uns de ces cas, nous observons que le nombre d'*outliers* détectés est trop important qu'effectivement l'estimation soit correcte. Notons cependant que ce nombre de cas est relativement restreint puisqu'il n'affecte pas significativement la médiane.

7.6 Conclusion

Dans ce chapitre, nous avons proposé une adaptation du principe de la méthode RANSAC pour sélectionner les sources pertinentes dans un cadre évidentiel. Selon son principe, la solution retenue est celle qui est la plus 'consensuelle', à savoir compatible avec le plus grand nombre d'observations considérées comme sources d'information. Cette approche est comparée à la q-relaxation classique et sa version évidentielle qui, toutes deux, recherchent la solution qui relâche le plus petit nombre de sources. En appliquant ces algorithmes aux traitements des mesures PR, nous observons que l'EV-RANSAC fournit une localisation plus précise.

Dans le cas dynamique pour rétablir une forme de filtrage temporel et de prise en compte de la prédiction dans le processus de localisation, nous avons testé différentes combinaisons possibles. Nous avons également introduit un nouveau critère qui évalue les performances d'une solution de forme ensembliste en prenant en compte à la fois son engagement (précision de l'ensemble) et sa fiabilité (garantie de la solution). Des travaux supplémentaires restent à mener pour développer le filtrage temporel dans le cadre des fonctions de croyance pour une meilleure prise en compte des informations *a priori*, tout en gardant une charge combinatoire raisonnable.

Conclusion et perspectives

Dans cette thèse, nous avons abordé le problème de la détection de données aberrantes dans les mesures GNSS pour la localisation robuste. L'application des approches développées dans cette thèse est orientée vers la localisation automobile et contrainte par une utilisation exclusive des données brutes fournies par le récepteur GPS sans capteurs supplémentaires.

Les deux premiers chapitres du manuscrit dressent l'état de l'art pour notre problème. Dans le premier, nous présentons les systèmes GNSS, les différentes avancées technologiques actuelles et à venir ainsi que les différents phénomènes qui biaisent l'estimation de la mesure de pseudo-distance et celle Doppler. Le deuxième chapitre s'intéresse aux méthodes de détection d'*outliers*, pour notre cas, les signaux multi-trajets et les réceptions NLOS. Parmi ces méthodes très variées : approche ensembliste, approches statistiques, méthodes par consensus, l'approche *a contrario* offre l'avantage d'être une méthode d'estimation sans paramètre et robuste à des taux d'*outliers* élevés. L'application de cette approche couplée avec une stratégie de type RANSAC en localisation GPS et l'une des principales contributions de cette thèse. Précisément ces dernières sont les suivantes :

La contribution principale de cette thèse est un algorithme de détection de mesures de pseudo-distances aberrantes exploitant la modélisation *a contrario*. Contrairement aux algorithmes présents dans les récepteurs GPS qui détectent au plus un *outlier* par *epoch*, l'approche présentée (au chapitre 5 de ce manuscrit) peut identifier un grand nombre de données aberrantes. À cause du manque de redondance des mesures GPS et en attendant le développement des systèmes compatibles avec le GPS, les *outliers* sont détectés en considérant une fenêtre temporelle de mesures. Pour pallier le problème du nombre important de combinaisons à tester, une stratégie fondée sur l'historique des *inliers* détectés sur l'horizon de la fenêtre temporelle considérée a été définie. Cette astuce méthodologique nous permet d'estimer une solution en temps réel et satisfaisante aux exigences de l'application. Deux mesures de significativité ont été définies, appliquées et comparées. La première repose sur un modèle de bruit, dit naïf, uniforme et la seconde sur un modèle gaussien. Durant les tests effectués, le critère issu du modèle naïf uniforme est apparu plus prudent, en termes de détection d'*outliers*, que celui issu de l'hypothèse gaussienne. Sur les données simulées où il existe un contrôle sur les données aberrantes ajoutées, le nombre de faux négatifs (*inliers* détectés à tort comme *outliers*) est plus important sous l'hypothèse d'un modèle naïf uniforme que sous une hypothèse gaussienne. En termes de précision de localisation, le manque de redondance des mesures pertinentes sous le modèle naïf uniforme induit par la suite des performances inférieures à celles obtenues en considérant le modèle gaussien. Sur les jeux de données réelles en statique et en dynamique sous différentes conditions, la comparaison avec des méthodes statistiques robustes a montré des performances très intéressantes pour le modèle naïf

gaussien. Ce travail a permis la publication d'un article de conférence [Zair et al., 2015a] et d'un article de revue [Zair et al., 2016a].

Notre seconde contribution (présentée au chapitre 6 de ce manuscrit) concerne l'introduction des mesures Doppler dans le processus de localisation. Nous proposons alors une localisation basée sur le filtre particulaire avec détection d'*outliers* conjointement dans les mesures de pseudo-distance (PR) et/ou Doppler. En effet, même si *a priori* les mesures Doppler sont moins impactées par les multi-trajets que les PR, [Le Marchand et al., 2009] caractérise les mesures PR et Doppler en environnement urbain et observe que les erreurs Doppler dépendent de la variation des erreurs PR. Nous étendons l'approche *a contrario* pour détecter d'éventuels *outliers* dans les mesures PR et/ou Doppler. Pour la localisation, puisque nous avons besoin de filtrer les bruits de mesures (PR ou Doppler) en plus d'éliminer les *outliers* sachant que notre modèle d'observation est non-linéaire, nous sommes naturellement orientés vers le filtrage particulaire. Cependant, avec l'estimation instantanée de la vitesse du récepteur, la dimension du vecteur d'état devient conséquente (8) de même que le nombre de particules nécessaire à l'exploration de l'espace des solutions. Notre première solution consistait à considérer les mesures Doppler dans l'étape de prédiction afin de retirer l'estimation de vitesse du vecteur d'état [Zair et al., 2015b]. Notre deuxième solution, plus élaborée, combine l'échantillonnage d'importance pour l'estimation liée au modèle d'observation des pseudo-distances et une résolution analytique par filtre de Kalman pour l'estimation de la vitesse liée aux mesures Doppler [Zair et al., 2016c]. Cette dernière modélisation, connue génériquement sous le nom de filtre de Rao-Blackwell, corrige l'estimation de la vitesse dans l'étape de d'estimation contrairement à la première solution proposée. L'approche *a contrario* fournit les données pertinentes à l'algorithme de localisation décrit précédemment. Hormis la contribution méthodologique précédente, ce travail a donné lieu à une contribution thématique puisque les expérimentations réalisées en milieu contraint ont permis de quantifier la présence de données aberrantes parmi les mesures Doppler et ont démontré l'intérêt de détecter et d'éliminer aussi ces *outliers* Doppler pour améliorer la localisation.

Notre troisième contribution (présentée au chapitre 7) est plus en amont. Nous avons proposé une approche crédibiliste pour la détection des mesures aberrantes dans les pseudo-distances. En analyse par intervalles, la q-relaxation est la méthode robuste utilisée pour détecter les *outliers* en relâchant des contraintes tant qu'une solution non vide n'est pas obtenue. L'idée a été étendue par [Pichon et al., 2015] au cadre des fonctions de croyance. Cette approche généralise l'AI en gérant simultanément les incertitudes et les imprécisions des observations. L'approche proposée s'inspire du RANSAC en choisissant parmi les combinaisons de sources possibles, la plus compatible selon une mesure de cohérence ou d'incohérence [Zair et al., 2016b]. Par ailleurs notre application nécessitait la définition et la manipulation de bbas 2D. Pour rester raisonnables en termes de charge combinatoire, une étape de simplification et d'approximation est alors effectuée. Les résultats obtenus sont encourageants, malgré des artefacts encore mal compris dans la partie filtrage ou la fusion des bbas de prédiction et d'estimation. Plusieurs approfondissements peuvent être apportés dans cet axe de recherche que ce soit dans la manière de gérer le filtrage temporel ou dans les mesures pour quantifier la performance de telle approches entre elles ou par rapport aux méthodes dites d'estimation ponctuelle.

Perspectives

Pour la suite de cette thèse, nous envisageons les perspectives suivantes. Les deux premières peuvent être à court terme, tandis que les suivantes sont à moyen terme.

Nous souhaitons évaluer l'intérêt d'introduire un modèle de correction atmosphérique pour atténuer l'effet de ces perturbations sur la détection des mesures impactées par des multi-trajets. La correction peut être faite en utilisant par exemple les informations fournis par un système SBAS comme le EGNOS en Europe.

Nous souhaitons également constituer un nouveau jeu de données dans des environnements plus contraint que ceux présentés dans ce manuscrit. Pour éviter les problèmes de latence liée à l'envoi des corrections de la base vers le récepteur mobile via le réseau GSM, la vérité de terrain sera établie par un post-traitement qui offre l'avantage d'augmenter la fréquence d'acquisition des mesures GNSS. La constitution de ce nouveau jeu de données pourra être l'occasion d'acquérir des mesures issues des constellations GPS et GLONASS pour augmenter la redondance des observations et permettre une détection d'*outliers* sur une seule *epoch*.

Notre troisième perspective vise à implémenter les algorithmes proposés sur des architectures parallèles. En effet, certaines parties des approches proposées sont aisément transposables pour une exécution en parallèle comme le tirage aléatoire des différentes hypothèses dans l'AC-RANSAC, l'estimation des poids des particules et leur déplacement dans les étapes d'estimation et de prédiction des filtres particulaires proposées ou l'initialisation des bbas ainsi que l'exploration du graphe des combinaisons dans l'EV-RANSAC.

Notre quatrième perspective concerne l'interprétation des données détectées en tant qu'*outliers*. L'estimation du biais multi-trajets requière une connaissance *a priori* comme une carte digitale l'altitude du récepteur au moment de l'acquisition. Nous pouvons aussi mieux détecter les mesures avec des lois de distribution modélisant les réceptions NLOS comme la loi de Rice ou la loi de Rayleigh pour les multi-trajets en LOS.

Une perspective intéressante sera d'appliquer l'approche *a contrario* développée non seulement à l'élimination des mesures aberrantes mais aussi aux choix des particules les plus cohérentes. Pour ce faire, nous interpréterons les particules comme des points de coordonnées, le vecteur d'état, et de valeur, leur poids. Typiquement pour un vecteur d'état de dimension 2, nous obtenons une image 2D des poids des particules dans laquelle nous pouvons détecter un ensemble de pixels cohérents sous l'hypothèse du modèle naïf considéré. Cette méthode permettra de choisir automatiquement les particules pertinentes durant l'étape de rééchantillonnage du filtre particulaire plutôt que de fixer un seuil *a priori*.

Une dernière perspective concerne une analyse plus approfondie du conflit généré par la combinaison conjonctive de plusieurs sources dans le cadre des fonctions de croyances. Le conflit a été utilisé comme une mesure d'incohérence pour détecter les sources *inliers*. Nous voulons par la suite approfondir notre analyse pour quantifier le degré de conflit apporté par/associé à chaque élément focal [Roquel et al., 2014]. Nous serons alors en mesure d'utiliser cette information pour guider l'ap-

proximation de bba (à l'instar de l'élimination des particules proposée plus haut).

Annexe

A.1 Nombre de fausses alarmes : exemple des dates d'anniversaire

Nous allons dans cette annexe illustrer le principe d'Helmholtz par l'exemple des dates d'anniversaire tiré de [Desolneux et al., 2007] et la simplicité des calculs à effectuer lorsque nous raisonnons en espérance du nombre d'occurrences plutôt qu'en probabilité.

Soit une population de N personnes et que leurs dates d'anniversaire vues comme des variables aléatoires indépendantes et uniformément réparties sur les 365 jours de l'année. Soit X la variable aléatoire représentant le nombre de personnes possédant la même date d'anniversaire. Alors, la probabilité qu'il y ait au moins deux personnes ($n = 2$) ayant la même date d'anniversaire est égale à 1 moins la probabilité que toutes les dates d'anniversaire soient différentes, soit :

$$P(X \geq 2) = 1 - \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{N-1}{365}\right) = 1 - \prod_{k=1}^{N-1} \left(1 - \frac{k}{365}\right). \quad (\text{A.1})$$

Prenons quelques exemples numériques : $N = 3, P(X \geq 2) = 0.0082$, $N = 25, P(X \geq 2) = 0.05687$ et $N = 50, P(X \geq 2) = 0.9704$. Plus le nombre de personnes augmente, moins il y a lieu de s'étonner que deux personnes aient une date d'anniversaire identique.

Le calcul précédent était relativement facile, mais le calcul est plus ardu si on considère plus que deux personnes avec la même date d'anniversaire. Pour $n = 3$, la probabilité $P(X \geq 3)$ vaut 1 moins la probabilité que toutes les personnes aient des dates d'anniversaire différentes moins la probabilité que seuls des couples aient la même date d'anniversaire. Ce calcul n'est pas trivial que cela car : il faut faire varier le nombre de couples (de même date anniversaire) de 1 jusqu'à $N/2$, et pour un même nombre de couples, il faut faire varier les couples ayant la même date de naissance et les dates anniversaires des $N - i$ personnes ayant des dates de naissances différentes. La probabilité est alors :

$$P(X \geq 3) = P(X \geq 2) - \underbrace{\sum_{i=1}^{N/2} \frac{\prod_{k=1}^i \binom{N-2(k-1)}{2}}{i!} \prod_{k=1}^i \left(\frac{1}{365}\right) \prod_{k=1}^{N-i} \left(1 - \frac{k}{365}\right)}_{\left(\frac{1}{365^N}\right) \sum_{i=1}^{N/2} \frac{\prod_{k=0}^{i-1} \binom{N-2k}{2}}{i!} \prod_{k=1}^{N-i} (365 - k)}. \quad (\text{A.2})$$

L'expression au-dessus du deuxième terme de l'équation A.2 représente la probabilité décrite précédemment. De la même manière, le calcul s'effectue pour des nombres $n > 3$, avec des calculs de moins en moins évidents. Maintenant, si nous raisonnons en termes d'espérance du nombre d'occurrences, nous pouvons obtenir une borne supérieure de la probabilité recherchée. L'espérance qu'au moins n personnes parmi N aient la même date d'anniversaire est $\left(\frac{1}{365}\right)^{n-1}$ multipliée par le nombre de n -uplets parmi N $\binom{N}{n}$. Pour $n = 3$, on a $P(X \geq 3) \leq \frac{1}{365^2} \frac{N(N-1)(N-2)}{6}$. En prenant le même nombre de personnes que l'exemple précédent : $N = 25$, $P(X \geq 3) = 0.0164 \leq 0.0173$ et $N = 50$, $P(X \geq 3) = 0.1264 \leq 0.1471$. Ainsi, l'espérance du nombre d'occurrences donne une bonne approximation de la probabilité avec une complexité de calcul bien inférieure.

A.2 Formulation de la jacobienne de l'algorithme 6

Nous détaillons ici le calcul de la matrice jacobienne du chapitre 3. Soit la fonction non-linéaire :

$$f_i(\mathbf{X}) = \sqrt{(x_{S_i} - X_r)^2 + (y_{S_i} - Y_r)^2 + (z_{S_i} - z_r)^2} + c\Delta t, \quad (\text{A.3})$$

Où $\mathbf{X}^\top = (X_r, Y_r, z_r, \Delta t)$ est le vecteur d'état inconnu, $X_r = x_r + \dot{x}_r dt$, $Y_r = y_r + \dot{y}_r dt$ et $\Delta t = \delta t + \dot{\delta} t dt$.

Nous posons

$$R(i) = \sqrt{(x_{S_i} - X_r)^2 + (y_{S_i} - Y_r)^2 + (z_{S_i} - z_r)^2} \text{ pour tout } i \in \{1 \dots n\}.$$

Nous définissons la matrice \mathbf{Y} , composée de l'ensemble des positions 3D des satellites, alors la jacobienne est :

$$J(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} \frac{\partial f_1(\mathbf{X})}{\partial x_r} & \frac{\partial f_1(\mathbf{X})}{\partial \dot{x}_r} & \dots & \frac{\partial f_1(\mathbf{X})}{\partial \delta t} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(\mathbf{X})}{\partial x_r} & \frac{\partial f_n(\mathbf{X})}{\partial \dot{x}_r} & \dots & \frac{\partial f_n(\mathbf{X})}{\partial \delta t} \end{pmatrix}, \quad (\text{A.4})$$

où

$$\left\{ \begin{array}{l} \frac{\partial f_i(\mathbf{X})}{\partial x_r} = -\frac{(x_{S_i} - (x_r + \dot{x}_r dt))}{R(i)}, \\ \frac{\partial f_i(\mathbf{X})}{\partial \dot{x}_r} = -dt \frac{(x_{S_i} - (x_r + \dot{x}_r dt))}{R(i)}, \\ \frac{\partial f_i(\mathbf{X})}{\partial y_r} = -\frac{(y_{S_i} - (y_r + \dot{y}_r dt))}{R(i)}, \\ \frac{\partial f_i(\mathbf{X})}{\partial \dot{y}_r} = -dt \frac{(y_{S_i} - (y_r + \dot{y}_r dt))}{R(i)}, \\ \frac{\partial f_i(\mathbf{X})}{\partial z_r} = -\frac{(z_{S_i} - z_r)}{R(i)}, \\ \frac{\partial f_i(\mathbf{X})}{\partial \delta t} = c, \\ \frac{\partial f_i(\mathbf{X})}{\partial \dot{\delta} t} = c dt. \end{array} \right.$$

A.3 Détails de l'équation 5.12 de la stratégie de sélection

L'équation 5.12 décrit le nombre de tests maximum à effectuer pour explorer tous l'espace des solutions. Les termes de cette équation résultent des hypothèses sur la stratégie de sélection pour se concentrer sur les configurations favorables. Nous écrivons l'équation 5.12 sous la forme d'un produit de trois termes :

$$N_{test} = \underbrace{\binom{N_{PR}}{1}}_{T_1} \underbrace{\binom{n_{ep} - 1}{(d-1) \bmod (n_{ep} - 1)}}_{T_2} \underbrace{\prod_{j=1}^{n_{ep}-1} \binom{\phi(j)}{\psi(j)}}_{T_3}. \quad (\text{A.5})$$

Le terme T_1 dérive de l'hypothèse que l'*epoch* actuelle contient au moins un satellite *inlier*. Nous allons donc tirer aléatoirement, à chaque tirage, un satellite (ou une mesure) dans l'*epoch* courante.

Les termes T_2 et T_3 sont issus de la deuxième hypothèse de la section 5.3.4.3 permettant de sélectionner différents satellites et de tirer aléatoirement un satellite par *epoch* sur la fenêtre temporelle choisie.

Soit l'exemple suivant où $d = 7$, $n_{ep} = 5$, $N_{PR} = 8$. (nous rappelons que d est le nombre de satellites à sélectionner, n_{ep} est la taille de la fenêtre temporelle et N_{PR} est le nombre de satellites total) :

- Après avoir sélectionné un satellite de l'*epoch* courante, nous devons choisir $d - 1 = 6$ satellites parmi les $n_{ep} - 1 = 4$ *epochs* restantes. Le terme $T_2 = \binom{4}{2} = 6$, car il existe 6 combinaisons qui permettent la sélection de 6 satellites parmi 4 *epoch*. En supposant une distribution uniforme des tirages des différentes *epochs*, la configuration la plus probable est que 2 des *epochs* contiennent 2 satellites tirés et 2 des *epochs* contiennent 1 seul satellite tiré. Les différents cas de répartition des satellites tirés pour les 4 *epochs* sont alors : $\{2, 2, 1, 1\}, \{2, 1, 2, 1\}, \{2, 1, 1, 2\}, \{1, 2, 2, 1\}, \{1, 2, 1, 2\}, \{1, 1, 2, 2\}$. Comme expliqué, ceci suppose que le cas où 3 satellites seraient sélectionnés sur une même *epoch* n'est pas pris en compte pour cet exemple.
- Dans cet exemple, le terme T_3 vaut $\binom{7}{1} \binom{6}{1} \binom{5}{2} \binom{3}{2} = 1260$. Ce terme décrit le nombre de tirages à effectuer sur les *epochs* restantes via deux fonctions intermédiaires $\psi(j)$ et $\phi(j)$. La fonction $\psi(j)$ explicite le nombre de satellites à sélectionner pour les *epochs* restantes et la fonction $\phi(j)$ représente le nombre de satellites par *epoch* parmi lesquels s'effectua le tirage. La fonction $\phi(j)$ dépend de $\psi(j)$ comme le montre l'équation 5.12. En reprenant l'exemple, nous allons commencer pour un tirage aléatoire d'un satellite parmi $N_{PR} - 1 = 7$, ensuite on tire un satellite de la deuxième *epoch* parmi $N_{PR} - 2 = 6$, puis deux satellites parmi 5 de la troisième *epoch* et enfin deux satellites parmi 3 de la dernière *epoch*.
- En imposant les contraintes de sélection expliquées précédemment, nous trouvons alors (toujours pour le cas de l'exemple) un nombre de tests $N_{test} = \underbrace{\binom{8}{1}}_{T_1} \underbrace{\binom{4}{2}}_{T_2} \underbrace{\binom{7}{1} \binom{6}{1} \binom{5}{2} \binom{3}{2}}_{T_3} =$

60480, alors que sans stratégie de sélection, le nombre de tests à effectuer théoriquement aurait été $\binom{n_{ep} \times N_{PR}}{d} = 18643560$ ce qui correspond à un rapport de 308 avec notre nombre de tests.

A.4 Formulation de la jacobienne de l'algorithme 12

L'équation (6.20) est résolue en utilisant l'algorithme 12. Ce dernier est présenté dans le cas de l'estimation du vecteur d'état ξ_2 . Le cas de l'estimation du vecteur d'état ξ_1 peut être déduite comme un cas particulier. Les données d'entrées sont les observations, l'estimation précédente du vecteur d'état $\xi_{(2,t|t-1)}$ et le vecteur de régularisation λ . Comme dans la plupart des applications pratiques, les paramètres de régularisation sont ajustés (ou appris) avec un premier jeu de données. Dans notre cas, nous n'avons pas régularisé le biais d'horloge et la dérive d'horloge, δt et $\dot{\delta t}$, de sorte que les coefficients correspondant à ces paramètres dans λ sont nuls. Cet algorithme inclut le calcul de la jacobienne qui se présente comme suit. Soient deux fonctions non linéaires :

$$f_i(\mathbf{X}) = \sqrt{(E_r - e_{S_i})^2 + (N_r - n_{S_i})^2 + (U_r - u_{S_i})^2} + c\Delta t, \quad (\text{A.6})$$

$$g_i(\mathbf{X}) = (\dot{E}_r - \dot{e}_{S_i})a_{e_i} + (\dot{N}_r - \dot{n}_{S_i})a_{n_i} + (\dot{U}_r - \dot{u}_{S_i})a_{u_i} - c\dot{\delta t}, \quad (\text{A.7})$$

où $\mathbf{X}^\top = (e_r, n_r, u_r, \delta t, \dot{e}_r, \dot{n}_r, \dot{u}_r, \dot{\delta t}, \ddot{e}_r, \ddot{n}_r)$ est le vecteur d'état inconnu. Alors :

Algorithm 12: Estimation non-linéaire avec un terme de régularisation ; entrées : Matrice \mathcal{S} , dont les lignes représentent les observations $(x_{S_i}, y_{S_i}, z_{S_i}, \dot{x}_{S_i}, \dot{y}_{S_i}, \dot{z}_{S_i}, \dot{\rho}_i)$, $\xi_{(2,t|t-1)}$, λ ; sorties $\tilde{\xi}_2$; Cet algorithme permet d'estimer $\tilde{\xi}_1$ en enlevant les mesures Doppler $\tilde{\rho}$.

```

1 Fixer le seuil de précision  $\varepsilon$  ;
2 if  $\lambda_2 = 0$  then
3   | Estimer  $\tilde{\xi}_2$  en utilisant l'algorithme de Gauss–Newton ;
4 else
5   | Initialiser  $\xi_2$  et  $\tilde{\xi}_2$  à  $\xi_{(2,t|t-1)}$  ;
6   | Définir  $\phi$ , le vecteur de  $\frac{d}{2}$  mesures PR et  $\frac{d}{2}$  mesures Doppler issues de  $\mathcal{S}$  ;
7   | while  $|\tilde{\xi}_2 - \xi_2| > \varepsilon$  do
8     |    $\xi_2 \leftarrow \tilde{\xi}_2$  ;  $\tilde{\phi} \leftarrow [\tilde{\rho} \tilde{\rho}^\top]^\top$  ;
9     |   Calculer la jacobienne  $J(\xi_2, \mathcal{S})$  ;
10    |    $\tilde{\xi}_2 \leftarrow (J^\top J + \lambda_2 I)^{-1} J^\top (\tilde{\phi} - \phi) + \xi_2$  ;
11  | end
12 end

```

$$\left\{ \begin{array}{l} E_r = e_r + \dot{e}_r dt + \ddot{e}_r \frac{dt^2}{2}, \\ N_r = n_r + \dot{n}_r dt + \ddot{n}_r \frac{dt^2}{2}, \\ U_r = u_r + \dot{u}_r dt, \\ \dot{E}_r = \dot{e}_r + \ddot{e}_r dt, \\ \dot{N}_r = \dot{n}_r + \ddot{n}_r dt, \\ \Delta t = \delta t + \dot{\delta} t dt, \\ a_{e_i} = \frac{e_{S_i} - E_r}{R_i}, \\ a_{n_i} = \frac{n_{S_i} - N_r}{R_i}, \\ a_{u_i} = \frac{u_{S_i} - U_r}{R_i} \end{array} \right.$$

où dt est la période d'échantillonnage et $R_i = \sqrt{(E_r - e_{S_i})^2 + (N_r - n_{S_i})^2 + (U_r - u_{S_i})^2}$ est la distance récepteur/satellite pour tout $S_i, i \in \{1 \dots n\}$.

La jacobienne est :

$$J = \begin{pmatrix} \frac{\partial f_1(X)}{\partial e_r} & \frac{\partial f_1(X)}{\partial n_r} & \dots & \frac{\partial f_1(X)}{\partial \ddot{n}_r} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n(X)}{\partial e_r} & \frac{\partial f_n(X)}{\partial n_r} & \dots & \frac{\partial f_n(X)}{\partial \ddot{n}_r} \\ \frac{\partial g_1(X)}{\partial e_r} & \frac{\partial g_1(X)}{\partial n_r} & \dots & \frac{\partial g_1(X)}{\partial \ddot{n}_r} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_n(X)}{\partial e_r} & \frac{\partial g_n(X)}{\partial n_r} & \dots & \frac{\partial g_n(X)}{\partial \ddot{n}_r} \end{pmatrix} \quad (\text{A.8})$$

où :

$$\left\{ \begin{array}{l} \frac{\partial f_i(X)}{\partial e_r} = \frac{E_r - e_{S_i}}{R_i}, \quad \frac{\partial f_i(X)}{\partial n_r} = \frac{N_r - n_{S_i}}{R_i} \\ \frac{\partial f_i(X)}{\partial u_r} = \frac{U_r - u_{S_i}}{R_i}, \quad \frac{\partial f_i(X)}{\partial \delta t} = c \\ \frac{\partial f_i(X)}{\partial \dot{e}_r} = dt \frac{\partial f_i(X)}{\partial e_r}, \quad \frac{\partial f_i(X)}{\partial \dot{n}_r} = dt \frac{\partial f_i(X)}{\partial n_r} \\ \frac{\partial f_i(X)}{\partial \dot{u}_r} = dt \frac{\partial f_i(X)}{\partial u_r}, \quad \frac{\partial f_i(X)}{\partial \dot{\delta} t} = c dt, \\ \frac{\partial f_i(X)}{\partial \ddot{e}_r} = \frac{dt^2}{2} \frac{\partial f_i(X)}{\partial e_r}, \quad \frac{\partial f_i(X)}{\partial \ddot{n}_r} = \frac{dt^2}{2} \frac{\partial f_i(X)}{\partial n_r} \end{array} \right.$$

et :

$$\left\{ \begin{array}{l} \frac{\partial g_i(X)}{\partial e_r} = (\dot{E}_r - \dot{e}_s) \frac{\partial a_{e_i}}{\partial e_r} + (\dot{N}_r - \dot{n}_s) \frac{\partial a_{n_i}}{\partial e_r} + (\dot{U}_r - \dot{u}_s) \frac{\partial a_{u_i}}{\partial e_r} \\ \frac{\partial g_i(X)}{\partial n_r} = (\dot{E}_r - \dot{e}_s) \frac{\partial a_{e_i}}{\partial n_r} + (\dot{N}_r - \dot{n}_s) \frac{\partial a_{n_i}}{\partial n_r} + (\dot{U}_r - \dot{u}_s) \frac{\partial a_{u_i}}{\partial n_r} \\ \frac{\partial g_i(X)}{\partial u_r} = (\dot{E}_r - \dot{e}_s) \frac{\partial a_{e_i}}{\partial u_r} + (\dot{N}_r - \dot{n}_s) \frac{\partial a_{n_i}}{\partial u_r} + (\dot{U}_r - \dot{u}_s) \frac{\partial a_{u_i}}{\partial u_r} \\ \frac{\partial g_i(X)}{\partial \delta t} = 0 \\ \frac{\partial g_i(X)}{\partial \dot{e}_r} = a_{e_i}, \quad \frac{\partial g_i(X)}{\partial \dot{n}_r} = a_{n_i} \\ \frac{\partial g_i(X)}{\partial \dot{u}_r} = a_{u_i}, \quad \frac{\partial g_i(X)}{\partial \dot{\delta} t} = -c dt \\ \frac{\partial g_i(X)}{\partial \ddot{e}_r} = a_{e_i} dt, \quad \frac{\partial g_i(X)}{\partial \ddot{n}_r} = a_{n_i} dt \end{array} \right.$$

Les dérivées du vecteur unitaire par rapport aux coordonnées du récepteur sont données par :

$$\left\{ \begin{array}{l} \frac{\partial a_{e_i}}{\partial e_r} = \frac{(E_r - e_s)^2 - R_i^2}{R_i^3} \\ \frac{\partial a_{n_i}}{\partial n_r} = \frac{(N_r - n_s)^2 - R_i^2}{R_i^3} \\ \frac{\partial a_{u_i}}{\partial u_r} = \frac{(U_r - u_s)^2 - R_i^2}{R_i^3} \\ \frac{\partial a_{e_i}}{\partial n_r} = \frac{\partial a_{n_i}}{\partial e_r} = \frac{(E_r - e_s)(N_r - n_s)}{R_i^3} \\ \frac{\partial a_{e_i}}{\partial u_r} = \frac{\partial a_{u_i}}{\partial e_r} = \frac{(E_r - e_s)(U_r - u_s)}{R_i^3} \\ \frac{\partial a_{u_i}}{\partial n_r} = \frac{\partial a_{n_i}}{\partial u_r} = \frac{(U_r - u_s)(N_r - n_s)}{R_i^3} \end{array} \right.$$

Les dérivées du vecteur unitaire \mathbf{a} par rapport à la vitesse et à l'accélération sont supposées nulles.

A.5 Éléments complémentaires de la théorie des fonctions de croyance

Cette annexe présente quelques éléments complémentaires en fonctions de croyance, utilisés dans cette thèse. Les notations sont celles de la thèse. En particulier, Ω est le cadre de discernement.

A.5.1 Type de bba

Les principaux types de bba :

- **bba normale** : une bba est dite normale si la masse de l'ensemble vide est nulle ($m(\emptyset) = 0$), dans le cas contraire, la bba est dite sous-normale.
- **bba catégorique** : une bba est dite catégorique si il n'existe qu'un seul élément focal dans la bba ($m(A) = 1$ et $m(B) = 0, \forall B \subseteq \Omega, B \neq A$).
- **bba vide** : une bba est dite vide si Ω est le seul élément focal ($m(\Omega) = 1$).
- **bba dogmatique** : une bba est dite dogmatique si la masse de Ω est nulle ($m(\Omega) = 0$).
- **bba simple** : une bba est dite simple si elle n'a que deux élément focaux, parmi lesquels Ω .
- **bba bayésienne** : une bba est dite bayésienne si ses éléments focaux sont tous des singletons.
- **bba consonante** : une bba est dite consonante si tous ses éléments focaux sont emboîtés (ordonnés au sens de l'inclusion).

A.5.2 Représentation des bbas :

Nous notons classiquement m une bba, fonction représentative des croyances sur 2^Ω . Néanmoins la définition initiale d'une bba est une représentation sous la forme de quatre fonctions en relations bijectives. Ces quatre fonctions sont les suivantes :

Fonction de masse

La fonction de masse est une fonction de 2^Ω vers $[0, 1]$ telle que :

$$\begin{aligned} \forall A \in 2^\Omega, m(A) &\in [0, 1], \\ \sum_{A \in 2^\Omega} m(A) &= 1. \end{aligned} \tag{A.9}$$

Fonction de crédibilité

La crédibilité (Bel) est une fonction de 2^Ω vers $[0, 1]$ telle que la crédibilité d'un ensemble A puisse être interprétée comme le degré de croyance accordé spécifiquement à A plutôt qu'à son complémentaire. Elle est calculée selon :

$$Bel(A) = \sum_{B \subseteq A; B \neq \emptyset} m(B), \quad \forall A \in 2^\Omega, \quad (\text{A.10})$$

Fonction de plausibilité

La plausibilité est une fonction de 2^Ω vers $[0, 1]$ telle que la plausibilité d'un ensemble A est la croyance compatible avec A . Elle est calculée selon :

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad \forall A \in 2^\Omega, \quad (\text{A.11})$$

Fonction d'implacabilité

L'implacabilité est très proche de la crédibilité, elle n'est différente que dans le cas d'un monde ouvert avec $m(\emptyset) = 0$:

$$b(A) = \sum_{B \subseteq A; B \in 2^\Omega} m(B), \quad \forall A \in 2^\Omega, \quad (\text{A.12})$$

Fonction de communalité

La fonction de communalité est définie par :

$$q(A) = \sum_{A \subseteq B; B \in 2^\Omega} m(B), \quad \forall A \in 2^\Omega, \quad (\text{A.13})$$

Le tableau A.1 résume les différentes relations entre ces représentations.

TABLE A.1 – Fonctions de croyance

	$m(A) =$	$Bel(A) =$	$Pl(A) =$	$b(A) =$	$q(A) =$
m		$\sum_{\emptyset \neq B \subseteq A} m(B)$	$\sum_{A \cap B \neq \emptyset} m(B)$	$\sum_{B \subseteq A} m(B)$	$\sum_{A \subseteq B} m(B)$
Bel	$\sum_{B \subseteq A} (-1)^{ A - B } Bel(B)$		$1 - m(\emptyset) - Bel(\bar{A})$	$Bel(A) + m(\emptyset)$	
Pl	$\sum_{B \subseteq A} (-1)^{ A - B +1} Pl(\bar{B})$			$1 - Pl(\bar{A})$	
b	$\sum_{B \subseteq A} (-1)^{ A - B } b(B)$				
q	$\sum_{A \subseteq B} (-1)^{ B - A } q(B)$	$\sum_{B \subseteq \bar{A}} (-1)^{ B } q(B)$			

A.5.3 Mesures de similarité

Il existe un grand nombre de mesures pour quantifier la similarité entre deux fonctions de croyance. Elles permettent généralement d'évaluer la perte d'information après des approximations réalisées sur des bbas [Jousselme et al., 2001], et également dans l'optimisation des paramètres. Parmi ces mesures, nous allons citer les plus utilisées :

Divergence de Kullback-Lieber [Perry and Stephanou, 1991] C'est une généralisation de la mesure probabilité de Kullback-Lieber :

$$d_{KL} = \sum_{A \subseteq \Omega} m_1(A) m_2(A) (1 - (m_1 \oplus m_2)(A)). \quad (\text{A.14})$$

Distance de type L_1 , L_2 et L_∞

La distance L_1 sur les crédibilités [Harmanec, 1999] :

$$d_H(m_1, m_2) = \sum_{A \subseteq \Omega} |Bel_1(A) - Bel_2(A)|. \quad (\text{A.15})$$

La distance L_∞ sur les probabilités pignistiques, connue aussi sous le nom de distance de Tessem [Tessem et al., 1993] :

$$d_T(m_1, m_2) = \max_{A \subseteq \Omega} |BetP_1(A) - BetP_2(A)|. \quad (\text{A.16})$$

La distance euclidienne entre les masses :

$$d_E(m_1, m_2) = \sqrt{1/2 (\langle m_1, m_1 \rangle_E + \langle m_2, m_2 \rangle_E - 2 \langle m_1, m_2 \rangle_E)}, \quad (\text{A.17})$$

où

$$\langle m_1, m_2 \rangle_E = \sum_{A \subseteq \Omega} m_1(A) m_2(A). \quad (\text{A.18})$$

L'équation A.18 n'est pas utilisée directement, on lui préfère des variantes qui redéfinissent le produit scalaire tel que :

Pseudo-distance de Fixen & Mahler [Fixsen and Mahler, 1997]

$$\langle m_1, m_2 \rangle_{FM} = \sum_{A \subseteq \Omega} \sum_{B \subseteq \Omega} m_1(A) m_2(B) \frac{|A \cap B|}{|A| |B|}. \quad (\text{A.19})$$

Distance de Josselme [[Josselme et al., 2001](#)]

$$\langle m_1, m_2 \rangle_J = \sum_{A \subseteq \Omega} \sum_{B \subseteq \Omega} m_1(A) m_2(B) \frac{|A \cap B|}{|A \cup B|}. \quad (\text{A.20})$$

Liste des symboles

$NFA_1(\mathcal{D})$	Nombre de fausses alarmes du sous-ensemble \mathcal{D} sous l'hypothèse du modèle naïf gaussien \mathcal{M}_1
$NFA_2(\mathcal{D})$	Nombre de fausses alarmes du sous-ensemble \mathcal{D} sous l'hypothèse du modèle naïf uniforme \mathcal{M}_2
N_{test}	Nombre de tests
$\mathbf{x}_r^{ENU} \mathbf{x}_{S_j}^{ENU}$	Vecteur position du récepteur/satellite dans le repère ENU
$\dot{\mathbf{x}}_r^{ENU} \dot{\mathbf{x}}_{S_j}^{ENU}$	Vecteur vitesse du récepteur/satellite dans le repère ENU
\mathbf{a}_j	Vecteur unitaire du satellite vers le récepteur
$\delta_{\mathcal{D}}^2$	Somme des carrés des résidus
\mathcal{I}	Ensemble des indices des mesures satellite sur une fenêtre temporelle \mathcal{T}
\mathcal{D}	Ensemble des <i>inliers</i>
λ	Vecteur des coefficients de régularisation
$(\delta t, \dot{\delta t})$	Biais et dérive d'horloge
$(x_{S_i}, y_{S_i}, z_{S_i})$	Position 3D du satellite S_i
(x_r, y_r, z_r)	Position 3D du récepteur
$(\dot{x}_r, \dot{y}_r, \dot{z}_r)$	Vitesse 3D du récepteur
$\dot{\rho}_j$	Dépend de la fréquence Doppler et est appelée <i>pseudorange rate</i>
ρ	Mesure de pseudo-distance
n_{ep}	Taille de la fenêtre temporelle \mathcal{T}
$\tilde{\rho}_i(t_k \mathbf{X}_t^m)$	Estimation de la pseudo-distance à l'instant t_k supposant le vecteur d'état \mathbf{X}_t^m
\mathcal{T}	Ensemble des instants considérés pour l'estimation d'état
$\mathcal{V}_{\mathcal{D}}$	Ellipsoïde d'imprécision validant les éléments de \mathcal{D}
$\mathbf{w}_t^{(i)}$	Poids d'une particule (i) à l'instant t
$\mathbf{x}_{kf,t}^{(i)}$	Vecteur d'état du filtre de Kalman à l'instant t utilisé dans le filtre RBPF
$\mathbf{x}_{pf,t}^{(i)}$	Vecteur d'état du filtre particulaire d'une particule (i) à l'instant t
f_0	Fréquence fondamentale
f_{L1}	Fréquence de la porteuse L1
f_{L2}	Fréquence de la porteuse L2

Liste des publications

Revue internationale à comité de lecture

- Zair, S., Le Hégarat-Masclé, S., & Seignez, E. (2016). **A-contrario modeling for robust localization using raw GNSS data.** *IEEE Transactions on Intelligent Transportation Systems*, 17(5), 1354-1367. [Zair et al., 2016a]
- Zair, S., Le Hégarat-Masclé, S., & Seignez, E. (2016). **Outlier Detection in GNSS Pseudo-Range/Doppler Measurements for Robust Localization.** *Sensors*, 16(4), 580. [Zair et al., 2016c]
- Zair, S., Le Hégarat-Masclé, S. (2017). **Evidential framework for robust localization using raw GNSS data.** *Engineering Applications of Artificial Intelligence*, (soumis).

Conférences internationales à comité de lecture :

- Zair, S., Le Hégarat-Masclé, S., & Seignez, E. (2016, September). **An Evidential RANSAC Algorithm and Its Application to GNSS Positioning.** *In International Conference on Belief Functions* (pp. 241-250). Springer International Publishing. [Zair et al., 2016b]
- Zair, S., Le Hégarat-Masclé, S., & Seignez, E. (2015, September). **Coupling Outlier Detection with Particle Filter for GPS-Based Localization.** *In 2015 IEEE 18th International Conference on Intelligent Transportation Systems* (pp. 2518-2524). [Zair et al., 2015b]
- Zair, S., Le Hégarat-Masclé, S., & Seignez, E. (2015, June). **A-contrario approach for outlier detection in GNSS positioning.** *In 2015 IEEE International Conference on Location and GNSS (ICL-GNSS)* (pp. 1-6). [Zair et al., 2015a]

Table des figures

2.1	Illustration du positionnement GNSS avec des corrections satellite ou base terrestre. . .	10
2.2	Mesure du temps de trajet satellite/récepteur par comparaison de PRN.	12
2.3	Effet Doppler en navigation GNSS.	14
2.4	Illustration d'un signal en multi-trajet et une réception d'un signal en non-visibilité direct (<i>NLOS</i>).	18
2.5	(a) Mauvaise et (b) bonne configuration des satellites.	19
2.6	Principe de la trilatération par satellite.	21
3.1	Illustration de la q-relaxation avec 3 balises.	31
3.2	Test du χ^2 avec deux distributions : centrée χ_k^2 et non-centrée χ_{k,δ_T}^2 avec $k = N - 4 = 6$	35
3.3	Impact du paramètre δ sur l'estimation de la droite par RANSAC.	45
3.4	Illustration du principe d'Helmholtz sur la détection d'alignement [Desolneux et al., 2003]	47
4.1	Récepteurs GNSS utilisés durant les expérimentations.	49
4.2	Plateforme expérimentale avec trois récepteurs GPS et la base.	50
4.3	Vue du ciel de l'environnement urbain de l'expérience, superposée aux trajectoires des satellites (tracées en orange).	52
4.4	Vue d'ensemble (© Google Earth) de la trajectoire. La durée totale de l'expérience est de 48mn.	53
4.5	Configuration et trajectoire des satellites en autoroute.	54
4.6	Configuration des satellites et histogramme en milieu urbain et forêt.	55
4.7	Trajectoire durant l'expérimentation en milieu urbain et en forêt.	56
5.1	Minimisation de la courbe NFA.	63
5.2	Exemple d'illustration du critère NFA basée sur le modèle \mathcal{M}_2	66
5.3	Estimation d'une droite à partir d'un ensemble de 200 échantillons	68

5.4	Comparaison entre le RANSAC et le NFA dans le cas d'une estimation d'une droite avec 80% d' <i>outliers</i>	69
5.5	Biais d'horloge en fonction du temps, cas du GPS Ublox EVK-5T.	70
5.6	Vue globale de l'algorithme.	75
5.7	Courbes $\log(NFA)$ en fonction du nombre d' <i>inliers</i>	76
5.8	Cluster de 1750 points représentant les positions estimées avec différentes méthodes.	78
5.9	Exemple de trajectoire utilisée pour estimer le vecteur de régularisation.	81
5.10	Erreur de localisation en fonction du temps dans le cas du jeu de données acquis en milieu urbain (cas statique).	83
5.11	Résultats expérimentaux en milieu urbain.	86
5.12	Résultats expérimentaux sur la zone forestière.	88
6.1	Fonction de répartition des erreurs obtenues par les quatre versions de l'EKF, les cinq versions du filtre particulaire et les solutions fournies par les deux GPS pour une durée d'expérience totale de 11min et 40s.	108
6.2	Estimation de $\tilde{\Delta}m_j$ sur les mesures PR acquises par les différents satellites (de numéros compris entre 1 et 32).	112
6.3	Estimation de $\tilde{\Delta}\dot{m}_j$ sur les mesures Doppler acquises par les différents satellites (de numéros compris entre 1 et 32).	112
7.1	Exemple d'une bba 2D.	120
7.2	Illustration de l'approche EV-RANSAC avec 5 sources dont 1 défailante.	126
7.3	Comparaison entre EV-RANSAC et la q-relaxation évidentielle dans le cas d'un exemple (droite $y = ax + b$) avec 10 points et 2 <i>outliers</i>	127
7.4	Initialisation de la bba.	129
7.5	Processus d'estimation de la bba 2D	130
7.6	Résultats de combinaison entre une bba d'observation et de prédiction en utilisant différentes règles de combinaison.	135
7.7	Performances selon le critère proposé (équation 7.16) en fonction du coefficient de pénalité λ	136

Liste des tableaux

3.1	Exemples de M-estimateurs.	40
4.1	Spécification des récepteurs GNSS utilisés durant nos expérimentations.	50
5.1	Performance des critères NFA_1 et NFA_2 pour le diagnostic des données aberrantes : cas statique. Les deux critères NFA conduisent à des taux élevés d' <i>Accuracy</i> et de <i>Precision</i>	77
5.2	Erreurs de localisation (en m) en coordonnées (Est,Nord, Up) dans le cas de données acquises en milieu dégagé avec des <i>outliers</i> simulés par l'ajout de biais (tirés aléatoirement) provenant de trois distributions différentes.	77
5.3	Erreurs de localisations (en m) sur les données dynamiques simulées en coordonnées (Est,Nord).	79
5.4	Taux d'exclusion en % pour différents biais et différentes longueurs de fenêtre temporelle.	80
5.5	Performance des critères NFA_1 et NFA_2 pour le diagnostic des données aberrantes : cas dynamique.	81
5.6	Taux d'exclusion des mesures satellites sur les périodes T_1 et T_2	82
5.7	Erreurs de localisation (en m) des données réelles selon les coordonnées (Est, Nord, Up)	83
5.8	Erreur de localisation en m en coordonnées (Est, Nord) sur le trajet en autoroute avec le jeu de données réelles.	84
5.9	Erreur de localisation en (m) pour les coordonnées (Est, Nord) sur le trajet en milieu urbain avec le jeu de données réelles.	87
5.10	Erreur de localisation (en m) pour les coordonnées (Est, Nord) sur le trajet en milieu urbain avec le jeu de données réelles.	87
5.11	Erreur de localisation (en m) pour les coordonnées (Est, Nord) sur le trajet en forêt avec le jeu de données réelles.	89
6.1	Erreurs de localisation en centile. Les algorithmes de localisation considérés sont EKF, <i>Extended Kalman Filter</i> ; PF, <i>Particle Filter</i> ; RBPF, <i>Rao-Blackwell Particle Filter</i> et ORKF, <i>Outlier Robust Kalman Filter</i>	107
6.2	Erreur de localisation (en m) en coordonnées (Est, Nord), erreurs $Norm_1$ et RMSE, et erreurs en moyenne et en écart-type.	109

6.3	Erreurs de localisation de la méthode proposée (en m) en coordonnée (Est, Nord), erreurs sur la $Norm_1$ et le $RMSE$ et les erreurs en moyennes et en écart-type en fonction de la qualité de la solution RTK utilisée comme vérité terrain.	110
6.4	Performance de l'algorithme 8 pour la détection d' <i>outliers</i> parmi les mesures PR ou (PR,Dp).	113
7.1	Comparaison statistique en variant la largeur des éléments focaux.	118
7.2	Initialisation de 5 sources pour illustrer le principe du RANSAC évidentiel avec $\Omega = \{\theta_1, \theta_2, \theta_3\}$	125
7.3	Performances statistiques en fonction du rayon des éléments focaux.	132
7.4	Résultat de localisation sur un cadre de discernement de taille 80×80 avec une résolution de 2×2	134
A.1	Fonctions de croyance	151

Bibliographie

- [Abdel-Hafez, 2010] Abdel-Hafez, M. (2010). The Autocovariance Least-Squares Technique for GPS Measurement Noise Estimation. *IEEE Transactions on Vehicular Technology*, 59(2) :574–588. (Cit  en pages 28 et 51.)
- [Agamennoni et al., 2011] Agamennoni, G., Nieto, J. I., and Nebot, E. M. (2011). An outlier-robust kalman filter. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1551–1558. IEEE. (Cit  en pages 41 et 107.)
- [Agamennoni et al., 2012] Agamennoni, G., Nieto, J. I., and Nebot, E. M. (2012). Approximate inference in state-space models with heavy-tailed noise. *Signal Processing, IEEE Transactions on*, 60(10) :5024–5037. (Cit  en page 41.)
- [Ammar et al., 2013] Ammar, M., Le H, S., et al. (2013). An a-contrario approach for object detection in video sequence. *International Journal of Pure and Applied Mathematics*, 89(2) :173–201. (Cit  en page 47.)
- [Andr  et al., 2015] Andr , C., H egar-Mascl , S. L., and Reynaud, R. (2015). Evidential framework for data fusion in a multi-sensor surveillance system. *Engineering Applications of Artificial Intelligence*, 43 :166–180. (Cit  en pages 119, 121 et 129.)
- [Arulampalam et al., 2002] Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *Signal Processing, IEEE Transactions on*, 50(2) :174–188. (Cit  en pages 41, 42, 84, 101 et 103.)
- [Baarda, 1968] Baarda, W. (1968). A testing procedure for use in geodetic networks. *Delft, Kanaalweg 4, Rijkscommissie voor Geodesie, 1968.*, 1. (Cit  en page 36.)
- [Bancroft, 1985] Bancroft, S. (1985). An algebraic solution of the GPS equations. *Aerospace and Electronic Systems, IEEE Transactions on*, (1) :56–59. (Cit  en page 21.)
- [Barnett and Lewis, 1994] Barnett, V. and Lewis, T. (1994). *Outliers in statistical data*, volume 3. Wiley New York. (Cit  en page 28.)
- [Basseville et al., 1993] Basseville, M., Nikiforov, I. V., et al. (1993). *Detection of abrupt changes : theory and application*, volume 104. Prentice Hall Englewood Cliffs. (Cit  en page 37.)
- [Brown, 1987] Brown, A. K. (1987). Receiver autonomous integrity monitoring using a 24-satellite gps constellation. In *Institute of Navigation, Technical Meeting*, pages 256–262. (Cit  en page 36.)
- [Brown, 1992] Brown, R. G. (1992). A baseline gps raim scheme and a note on the equivalence of three raim methods. *Navigation*, 39(3) :301–316. (Cit  en page 64.)
- [B taille et al., 2013] B taille, D., Peyret, F., Ortiz, M., Miquel, S., and Fontenay, L. (2013). A new modeling based on urban trenches to improve gnss positioning quality of service in cities. *IEEE Intelligent Transportation Systems Magazine*, 5(3) :59–70. (Cit  en page 29.)
- [B taille et al., 2016] B taille, D., Peyret, F., Ortiz, M., Miquel, S., and Godan, F. (2016). Improving accuracy and integrity with a probabilistic urban trench modeling. *Navigation*, 63(3) :283–294. (Cit  en page 29.)
- [Cai and Gao, 2012] Cai, C. and Gao, Y. (2012). Modeling and assessment of combined GPS/GLONASS precise point positioning. *GPS Solutions*, 17(2) :223–236. (Cit  en page 8.)

- [Cappelle et al., 2012] Cappelle, C., Najjar, M. E. E., Charpillat, F., and Pomorski, D. (2012). Virtual 3d City Model for Navigation in Urban Areas. *Journal of Intelligent & Robotic Systems*, 66(3) :377–399. (Cit  en page 28.)
- [Casella and Robert, 1996] Casella, G. and Robert, C. P. (1996). Rao-Blackwellisation of sampling schemes. *Biometrika*, 83(1) :81–94. (Cit  en page 103.)
- [Chang et al., 2009] Chang, T.-H., Wang, L.-S., and Chang, F.-R. (2009). A Solution to the Ill-Conditioned GPS Positioning Problem in an Urban Environment. *IEEE Transactions on Intelligent Transportation Systems*, 10(1) :135–145. (Cit  en pages 69 et 70.)
- [Cheng et al., 2016] Cheng, C., Tourneret, J.-Y., Pan, Q., and Calmettes, V. (2016). Detecting, estimating and correcting multipath biases affecting GNSS signals using a marginalized likelihood ratio-based method. *Signal Processing*, 118 :221–234. (Cit  en pages 37 et 96.)
- [Chiang et al., 2013] Chiang, K.-W., Duong, T. T., and Liao, J.-K. (2013). The Performance Analysis of a Real-Time Integrated INS/GPS Vehicle Navigation System with Abnormal GPS Measurement Elimination. *Sensors*, 13(8) :10599–10622. (Cit  en page 28.)
- [Choi and Medioni, 2009] Choi, J. and Medioni, G. (2009). Starsac : Stable random sample consensus for parameter estimation. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 675–682. IEEE. (Cit  en page 45.)
- [Chum and Matas, 2005] Chum, O. and Matas, J. (2005). Matching with prosac-progressive sample consensus. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 220–226. IEEE. (Cit  en page 45.)
- [Chum et al., 2003] Chum, O., Matas, J., and Kittler, J. (2003). Locally optimized ransac. In *Pattern recognition*, pages 236–243. Springer. (Cit  en page 45.)
- [Cooper and Durrant-Whyte, 1994] Cooper, S. and Durrant-Whyte, H. (1994). A Kalman filter model for GPS navigation of land vehicles. In *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems '94. 'Advanced Robotic Systems and the Real World', IROS '94*, volume 1, pages 157–163 vol.1. (Cit  en page 24.)
- [Cui and Ge, 2003] Cui, Y. and Ge, S. S. (2003). Autonomous vehicle positioning with GPS in urban canyon environments. *Robotics and Automation, IEEE Transactions on*, 19(1) :15–25. (Cit  en page 28.)
- [Denoeux, 2001] Denoeux, T. (2001). Inner and outer approximation of belief structures using a hierarchical clustering approach. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(04) :437–460. (Cit  en pages 120 et 121.)
- [Desolneux et al., 2003] Desolneux, A., Moisan, L., and More, J.-M. (2003). A grouping principle and four applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(4) :508–513. (Cit  en pages 46, 47 et 159.)
- [Desolneux et al., 2000] Desolneux, A., Moisan, L., and Morel, J.-M. (2000). Meaningful alignments. *International Journal of Computer Vision*, 40(1) :7–23. (Cit  en pages 45, 60 et 67.)
- [Desolneux et al., 2007] Desolneux, A., Moisan, L., and Morel, J.-M. (2007). *From gestalt theory to image analysis : a probabilistic approach*, volume 34. Springer Science & Business Media. (Cit  en pages 47, 60, 63, 98 et 143.)
- [Destercke and Burger, 2013] Destercke, S. and Burger, T. (2013). Toward an axiomatic definition of conflict between belief functions. *Cybernetics, IEEE Transactions on*, 43(2) :585–596. (Cit  en page 119.)

- [Doucet, 1998] Doucet, A. (1998). On sequential simulation-based methods for bayesian filtering. (Cité en page 42.)
- [Doucet et al., 2001] Doucet, A., Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice*. Springer New York, New York, NY. (Cité en pages 102 et 103.)
- [Doucet et al., 2000] Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing*, 10(3) :197–208. (Cité en pages 41 et 103.)
- [Drevelle, 2011] Drevelle, V. (2011). *Étude de méthodes ensemblistes robustes pour une localisation multi-sensorielle intégrée. Application à la navigation des véhicules en milieu urbain*. PhD thesis, Université de Technologie de Compiègne. (Cité en page 82.)
- [Drevelle and Bonnifait, 2011] Drevelle, V. and Bonnifait, P. (2011). A set-membership approach for high integrity height-aided satellite positioning. *GPS solutions*, 15(4) :357–368. (Cité en page 32.)
- [Drevelle and Bonnifait, 2013] Drevelle, V. and Bonnifait, P. (2013). Localization confidence domains via set inversion on short-term trajectory. *IEEE Transactions on Robotics*, 29(5) :1244–1256. (Cité en page 116.)
- [Dubois and Prade, 1990] Dubois, D. and Prade, H. (1990). Consonant approximations of belief functions. *International Journal of Approximate Reasoning*, 4(5-6) :419–449. (Cité en page 120.)
- [Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395. (Cité en page 43.)
- [Fixsen and Mahler, 1997] Fixsen, D. and Mahler, R. P. (1997). The modified dempster-shafer approach to classification. *IEEE Transactions on Systems, Man, and Cybernetics-Part A : Systems and Humans*, 27(1) :96–104. (Cité en page 152.)
- [Giremus and Escher, 2007] Giremus, A. and Escher, A.-C. (2007). A GLR Algorithm to Detect and Exclude up to Two Simultaneous Range Failures in a GPS/Galileo/IRS Case. *ION GNSS 2007, 20th International Technical Meeting of the Satellite Division of The Institute of Navigation*, pages pp–2911. (Cité en page 37.)
- [Giremus et al., 2007] Giremus, A., Tourneret, J.-Y., and Calmettes, V. (2007). A particle filtering approach for joint detection/estimation of multipath effects on GPS measurements. *IEEE Transactions on Signal Processing*, 55(4) :1275–1285. (Cité en pages 42, 51, 92 et 96.)
- [Gning and Bonnifait, 2006] Gning, A. and Bonnifait, P. (2006). Constraints propagation techniques on intervals for a guaranteed localization using redundant data. *Automatica*, 42(7) :1167–1175. (Cité en page 31.)
- [Grewal et al., 2007] Grewal, M. S., Weill, L. R., and Andrews, A. P. (2007). *Global Positioning Systems, Inertial Navigation, and Integration*. John Wiley & Sons. (Cité en pages 6 et 15.)
- [Groves, 2011] Groves, P. D. (2011). Shadow matching : A new gnss positioning technique for urban canyons. *Journal of Navigation*, 64(03) :417–430. (Cité en page 29.)
- [Groves et al., 2012] Groves, P. D., Jiang, Z., Wang, L., and Ziebart, M. K. (2012). Intelligent urban positioning using multi-constellation GNSS with 3d mapping and NLOS signal detection. (Cité en page 29.)

- [Groves et al., 2015] Groves, P. D., Wang, L., Adjrak, M., and Ellul, C. (2015). GNSS shadow matching : The challenges ahead. *The Institute of Navigation*. (Cit  en page 29.)
- [Grubbs, 1969] Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11(1) :1–21. (Cit  en page 28.)
- [Gupta et al., 2014] Gupta, M., Gao, J., Aggarwal, C., and Han, J. (2014). Outlier detection for temporal data. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 5(1) :1–129. (Cit  en page 27.)
- [Gustafsson et al., 2002] Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.-J. (2002). Particle filters for positioning, navigation, and tracking. *Signal Processing, IEEE Transactions on*, 50(2) :425–437. (Cit  en page 92.)
- [Gustafsson and Gustafsson, 2000] Gustafsson, F. and Gustafsson, F. (2000). *Adaptive filtering and change detection*, volume 1. Wiley New York. (Cit  en page 37.)
- [Harmanec, 1999] Harmanec, D. (1999). Faithful approximations of belief functions. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 271–278. Morgan Kaufmann Publishers Inc. (Cit  en pages 121 et 152.)
- [Hawkins, 1980] Hawkins, D. M. (1980). *Identification of outliers*, volume 11. Springer. (Cit  en page 28.)
- [Hewitson and Wang, 2006] Hewitson, S. and Wang, J. (2006). GNSS receiver autonomous integrity monitoring (RAIM) performance analysis. *GPS Solutions*, 10(3) :155–170. (Cit  en pages 36 et 64.)
- [Hewitson and Wang, 2007] Hewitson, S. and Wang, J. (2007). GNSS receiver autonomous integrity monitoring with a dynamic model. *Journal of Navigation*, 60(02) :247–263. (Cit  en page 36.)
- [Hodge and Austin, 2004] Hodge, V. J. and Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2) :85–126. (Cit  en page 27.)
- [Hsu et al., 2015] Hsu, L.-T., Gu, Y., and Kamijo, S. (2015). 3d building model-based pedestrian positioning method using GPS/GLONASS/QZSS and its reliability calculation. *GPS Solutions*, pages 1–16. (Cit  en page 29.)
- [Huber, 1964] Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1) :73–101. (Cit  en pages 39 et 111.)
- [Huber, 1987] Huber, P. J. (1987). The place of the l_1 -norm in robust estimation. *Computational Statistics & Data Analysis*, 5(4) :255–262. (Cit  en page 38.)
- [Huerta et al., 2009] Huerta, J. M., Vidal, J., Giremus, A., and Tourneret, J.-Y. (2009). Joint particle filter and ukf position tracking in severe non-line-of-sight situations. *IEEE Journal of Selected Topics in Signal Processing*, 3(5) :874–888. (Cit  en page 29.)
- [Iwase et al., 2013] Iwase, T., Suzuki, N., and Watanabe, Y. (2013). Estimation and exclusion of multipath range error for robust positioning. *GPS solutions*, 17(1) :53–62. (Cit  en page 29.)
- [Jaulin, 2009] Jaulin, L. (2009). Robust set-membership state estimation ; application to underwater robotics. *Automatica*, 45(1) :202–206. (Cit  en page 31.)
- [Jaulin et al., 2001a] Jaulin, L., Kieffer, M., Braems, I., and Walter, E. (2001a). Guaranteed non-linear estimation using constraint propagation on sets. *International Journal of Control*, 74(18) :1772–1782. (Cit  en page 31.)

- [Jaulin et al., 2001b] Jaulin, L., Kieffer, M., Didrit, O., and Walter, E. (2001b). *Applied Interval Analysis with Examples in Parameter and State Estimation, Robust Control and Robotics*. Springer London Ltd. (Cité en pages 31 et 115.)
- [Jaulin et al., 2002] Jaulin, L., Kieffer, M., Walter, E., and Meizel, D. (2002). Guaranteed robust nonlinear estimation with application to robot localization. *Systems, Man, and Cybernetics, Part C : Applications and Reviews, IEEE Transactions on*, 32(4) :374–381. (Cité en page 132.)
- [Jaulin and Walter, 1993] Jaulin, L. and Walter, E. (1993). Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4) :1053–1064. (Cité en page 31.)
- [Jaulin et al., 1996] Jaulin, L., Walter, E., and Didrit, O. (1996). Guaranteed robust nonlinear parameter bounding. In *Proc. CESA*, volume 96, pages 1156–1161. (Cité en page 132.)
- [Jousselme et al., 2001] Jousselme, A.-L., Grenier, D., and Bossé, É. (2001). A new distance between two bodies of evidence. *Information Fusion*, 2(2) :91–101. (Cité en pages 121, 152 et 153.)
- [Jwo and Wang, 2007] Jwo, D.-J. and Wang, S.-H. (2007). Adaptive fuzzy strong tracking extended Kalman filtering for GPS navigation. *Sensors Journal, IEEE*, 7(5) :778–789. (Cité en page 61.)
- [Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1) :35–45. (Cité en page 24.)
- [Kaplan and Hegarty, 2006] Kaplan, E. D. and Hegarty (2006). *Understanding GPS : principles and applications*. Artech House, Boston. (Cité en pages 15, 21, 34, 61 et 97.)
- [Khaleghi et al., 2011] Khaleghi, B., Khamis, A., Karray, F. O., and Razavi, S. N. (2011). Multisensor data fusion : A review of the state-of-the-art. *Information Fusion*. (Cité en page 26.)
- [Knight and Wang, 2009] Knight, N. L. and Wang, J. (2009). A comparison of outlier detection procedures and robust estimation methods in GPS positioning. *Journal of Navigation*, 62(04) :699–709. (Cité en pages 38, 39, 67 et 96.)
- [Krarup et al., 1980] Krarup, T., Juhl, J., and Kubik, K. (1980). Gotterdammerung over least squares adjustment. *International archives of photogrammetry*, pages 369–378. (Cité en page 38.)
- [Kuusniemi, 2005] Kuusniemi, H. (2005). *User-level reliability and quality monitoring in satellite-based personal navigation*. (Cité en page 36.)
- [Kuusniemi et al., 2004] Kuusniemi, H., Lachapelle, G., and Takala, J. H. (2004). Position and velocity reliability testing in degraded gps signal environments. *GPS solutions*, 8(4) :226–237. (Cité en page 38.)
- [Le Marchand et al., 2008] Le Marchand, O., Bonnifait, P., Bañez-Guzmán, J., Peyret, F., and Betaille, D. (2008). Performance Evaluation of Fault Detection Algorithms as Applied to Automotive Localisation. In *European Navigation Conference-GNSS 2008*. (Cité en pages 36 et 96.)
- [Le Marchand et al., 2009] Le Marchand, O., Bonnifait, P., Ibañez-Guzmán, J., Betaille, D., and Peyret, F. (2009). Characterization of GPS multipath for passenger vehicles across urban environments. *ATTI dell'Istituto Italiano di Navigazione*, (189) :77–88. (Cité en pages 51, 91, 97, 110 et 140.)
- [Lefevre et al., 2002] Lefevre, E., Colot, O., and Vannoorenberghe, P. (2002). Belief function combination and conflict management. *Information Fusion*, 3(2) :149–162. (Cité en page 119.)
- [Li et al., 2011] Li, L., Zhong, J., and Zhao, M. (2011). Doppler-Aided GNSS Position Estimation With Weighted Least Squares. *IEEE Transactions on Vehicular Technology*, 60(8) :3615–3624. (Cité en page 94.)

- [Lowrance et al., 1986] Lowrance, J. D., Garvey, T. D., and Strat, T. M. (1986). A framework for evidential-reasoning systems. In *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI- 86)*, pages 896–901. (Cit  en page 121.)
- [Ma et al., 2011] Ma, J., Liu, W., Dubois, D., and Prade, H. (2011). Bridging jeffrey’s rule, agm revision and dempster conditioning in the theory of evidence. *International Journal on Artificial Intelligence Tools*, 20(04) :691–720. (Cit  en page 131.)
- [Mao et al., 2002] Mao, X., Wada, M., and Hashimoto, H. (2002). Nonlinear filtering algorithms for GPS using pseudorange and Doppler shift measurements. In *The IEEE 5th International Conference on Intelligent Transportation Systems, 2002. Proceedings*, pages 914–919. (Cit  en pages 25 et 91.)
- [Marais et al., 2013] Marais, J., Meurie, C., Attia, D., Ruichek, Y., and Flancquart, A. (2013). Toward accurate localization in guided transport : Combining GNSS data and imaging information. *Transportation Research Part C : Emerging Technologies*. (Cit  en page 29.)
- [Meguro et al., 2009] Meguro, J., Murata, T., Takiguchi, J., Amano, Y., and Hashizume, T. (2009). GPS Multipath Mitigation for Urban Area Using Omnidirectional Infrared Camera. *IEEE Transactions on Intelligent Transportation Systems*, 10(1) :22–30. (Cit  en page 29.)
- [Meinhold and Singpurwalla, 1989] Meinhold, R. J. and Singpurwalla, N. D. (1989). Robustification of kalman filter models. *Journal of the American Statistical Association*, 84(406) :479–486. (Cit  en page 40.)
- [Miura and Kamijo, 2014] Miura, S. and Kamijo, S. (2014). GPS Error Correction by Multipath Adaptation. *International Journal of Intelligent Transportation Systems Research*, pages 1–8. (Cit  en page 29.)
- [Moisan, 2003] Moisan, L. (2003). Mod les continus, num riques et statistiques pour l’analyse d’images. *Habilitation   Diriger des Recherches, sp cialit  : Math matiques*. (Cit  en page 45.)
- [Moisan and Stival, 2004] Moisan, L. and Stival, B. (2004). A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *International Journal of Computer Vision*, 57(3) :201–218. (Cit  en page 47.)
- [Moore et al., 1966] Moore, R. E., Cloud, M. J., and Kearfott, R. B. (1966). *Introduction to interval analysis*. Siam. (Cit  en page 30.)
- [Moulon, 2014] Moulon, P. (2014). *Positionnement robuste et pr cis de r seaux d’images*. PhD thesis, Universit  Paris-Est. (Cit  en pages 45 et 46.)
- [Nassreddine et al., 2010] Nassreddine, G., Abdallah, F., and Denoux, T. (2010). State Estimation Using Interval Analysis and Belief-Function Theory : Application to Dynamic Vehicle Localization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics*, 40(5) :1205–1218. (Cit  en pages 32 et 119.)
- [Obst et al., 2012] Obst, M., Bauer, S., and Wanielik, G. (2012). Urban multipath detection and mitigation with dynamic 3d maps for reliable land vehicle localization. In *Position Location and Navigation Symposium (PLANS), 2012 IEEE/ION*, pages 685–691. IEEE. (Cit  en page 28.)
- [Obst and Wanielik, 2013] Obst, M. and Wanielik, G. (2013). Probabilistic non-line-of-sight detection in reliable urban GNSS vehicle localization based on an empirical sensor model. In *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pages 363–368. IEEE. (Cit  en page 29.)
- [Parkinson, 1996] Parkinson, B. W. (1996). *Progress in Astronautics and Aeronautics : Global Positioning System : Theory and Applications*, volume 2. Aiaa. (Cit  en page 36.)

- [Perry and Stephanou, 1991] Perry, W. L. and Stephanou, H. E. (1991). Belief function divergence as a classifier. In *Intelligent Control, 1991., Proceedings of the 1991 IEEE International Symposium on*, pages 280–285. IEEE. (Cité en page 152.)
- [Peyraud et al., 2013] Peyraud, S., Bétaille, D., Renault, S., Ortiz, M., Mougel, F., Meizel, D., and Peyret, F. (2013). About Non-Line-Of-Sight Satellite Detection and Exclusion in a 3d Map-Aided Localization Algorithm. *Sensors*, 13(1) :829–847. (Cité en page 29.)
- [Pichon et al., 2015] Pichon, F., Destercke, S., and Burger, T. (2015). A Consistency-Specificity Trade-Off to Select Source Behavior in Information Fusion. *IEEE Tran. on Cyb.*, 45(4) :598–609. (Cité en pages 116, 117, 118, 128, 132 et 140.)
- [Pitt and Shephard, 1999] Pitt, M. K. and Shephard, N. (1999). Filtering via simulation : Auxiliary particle filters. *Journal of the American statistical association*, 94(446) :590–599. (Cité en page 42.)
- [Rabaoui et al., 2012] Rabaoui, A., Viandier, N., Duflos, E., Marais, J., and Vanheeghe, P. (2012). Dirichlet Process Mixtures for Density Estimation in Dynamic Nonlinear Modeling : Application to GPS Positioning in Urban Canyons. *IEEE Transactions on Signal Processing*, 60(4) :1638–1655. (Cité en pages 42 et 92.)
- [Rabin et al., 2010] Rabin, J., Delon, J., Gousseau, Y., and Moisan, L. (2010). MAC-RANSAC : a robust algorithm for the recognition of multiple objects. In *Fifth International Symposium on 3D Data Processing, Visualization and Transmission (3DPTV 2010)*, page 051. (Cité en pages 47 et 98.)
- [Radon-Nykodim, 1997] Radon-Nykodim, R. (1997). Optimal nonlinear filtering in GPS/INS integration. *IEEE Transactions on Aerospace and Electronic Systems*, 33(3) :835. (Cité en page 61.)
- [Ramalingam et al., 2010] Ramalingam, S., Bouaziz, S., Sturm, P., and Brand, M. (2010). Skyline2gps : Localization in urban canyons using omni-skylines. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 3816–3823. IEEE. (Cité en page 29.)
- [Ramaswamy et al., 2000] Ramaswamy, S., Rastogi, R., and Shim, K. (2000). Efficient algorithms for mining outliers from large data sets. In *ACM SIGMOD Record*, volume 29, pages 427–438. ACM. (Cité en pages 28 et 110.)
- [Rekik et al., 2015] Rekik, W., Le Hégarat-Masclé, S., Reynaud, R., Kallel, A., and Ben Hamida, A. (2015). Dynamic estimation of the discernment frame in belief function theory : Application to object detection. *Information Sciences*, 306 :132–149. (Cité en page 119.)
- [Robin, 2007] Robin, A. (2007). *Détection de changements et classification sous-pixelliques en imagerie satellitaire. Application au suivi temporel des surfaces continentales*. PhD thesis, Université René Descartes-Paris V. (Cité en page 63.)
- [Robin et al., 2010] Robin, A., Moisan, L., and Le Hégarat-Masclé, S. (2010). An A-Contrario Approach for Subpixel Change Detection in Satellite Imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11) :1977–1993. (Cité en page 47.)
- [Roquel et al., 2014] Roquel, A., Le Hégarat-Masclé, S., Bloch, I., and Vincke, B. (2014). Decomposition of conflict as a distribution on hypotheses in the framework on belief functions. *International Journal of Approximate Reasoning*, 55(5) :1129–1146. (Cité en pages 119 et 141.)
- [Sabater et al., 2012] Sabater, N., Almansa, A., and Morel, J.-M. (2012). Meaningful matches in stereo-vision. *IEEE transactions on pattern analysis and machine intelligence*, 34(5) :930–942. (Cité en page 47.)
- [Samama, 2008] Samama, N. (2008). *Global Positioning : Technologies and Performance*. Wiley. (Cité en page 11.)

- [Särkkä and Nummenmaa, 2009] Särkkä, S. and Nummenmaa, A. (2009). Recursive noise adaptive kalman filtering by variational bayesian approximations. *Automatic Control, IEEE Transactions on*, 54(3) :596–600. (Cité en page 41.)
- [Schon et al., 2005] Schon, T., Gustafsson, F., and Nordlund, P.-J. (2005). Marginalized particle filters for mixed linear/nonlinear state-space models. *IEEE Transactions on Signal Processing*, 53(7) :2279–2289. (Cité en pages 92 et 103.)
- [Schroth et al., 2008] Schroth, G., Ene, A., Blanch, J., Walter, T., and Enge, P. (2008). Failure detection and exclusion via range consensus. In *Proceedings of the European navigation conference*. (Cité en pages 44, 59 et 64.)
- [Seigneur et al., 2009] Seigneur, E., Kieffer, M., Lambert, A., Walter, E., and Maurin, T. (2009). Real-time Bounded-error State Estimation for Vehicle Tracking. *The International Journal of Robotics Research*, 28(1) :34–48. (Cité en pages 31 et 116.)
- [Skog and Handel, 2009] Skog, I. and Handel, P. (2009). In-car positioning and navigation technologies :A survey. *Intelligent Transportation Systems, IEEE Transactions on*, 10(1) :4–21. (Cité en page 28.)
- [Sliwka et al., 2011] Sliwka, J., Le Bars, F., Reynet, O., and Jaulin, L. (2011). Using interval methods in the context of robust localization of underwater robots. In *Fuzzy Information Processing Society (NAFIPS), 2011 Annual Meeting of the North American*, pages 1–6. (Cité en page 32.)
- [Sorenson and Alspach, 1971] Sorenson, H. W. and Alspach, D. L. (1971). Recursive bayesian estimation using gaussian sums. *Automatica*, 7(4) :465–479. (Cité en page 40.)
- [Sunderhauf et al., 2012] Sunderhauf, N., Obst, M., Wanielik, G., and Protzel, P. (2012). Multipath mitigation in GNSS-based localization using robust optimization. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 784–789. IEEE. (Cité en page 72.)
- [Tesseem et al., 1993] Tessem, B. et al. (1993). Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61(2) :315–329. (Cité en page 152.)
- [Teunissen, 1990] Teunissen, P. (1990). Quality control in integrated navigation systems. In *Position Location and Navigation Symposium, 1990. Record. The 1990's-A Decade of Excellence in the Navigation Sciences. IEEE PLANS'90., IEEE*, pages 158–165. IEEE. (Cité en page 36.)
- [Ting et al., 2007a] Ting, J.-A., Theodorou, E., and Schaal, S. (2007a). A kalman filter for robust outlier detection. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 1514–1519. IEEE. (Cité en page 41.)
- [Ting et al., 2007b] Ting, J.-A., Theodorou, E., and Schaal, S. (2007b). Learning an outlier-robust kalman filter. In *Machine Learning : ECML 2007*, pages 748–756. Springer. (Cité en page 41.)
- [Tzoreff and Bobrovsky, 2012] Tzoreff, E. and Bobrovsky, B.-Z. (2012). A Novel Approach for Modeling Land Vehicle Kinematics to Improve GPS Performance Under Urban Environment Conditions. *IEEE Transactions on Intelligent Transportation Systems*, 13(1) :344–353. (Cité en page 69.)
- [Valero and Moreno, 2005] Valero, J. L. B. and Moreno, S. B. (2005). Robust estimation in geodetic networks. *Física de la Tierra*, 17 :7–22. (Cité en page 39.)
- [Walter and Enge, 1995] Walter, T. and Enge, P. (1995). Weighted raim for precision approach. In *PROCEEDINGS OF ION GPS*, volume 8, pages 1995–2004. Institute of Navigation. (Cité en page 64.)
- [Wang et al., 2013] Wang, L., Groves, P. D., and Ziebart, M. K. (2013). GNSS Shadow Matching : Improving Urban Positioning Accuracy Using a 3d City Model with Optimized Visibility Scoring Scheme. *Navigation*, 60(3) :195–207. (Cité en page 29.)

- [Wieser and Brunner, 2002] Wieser, A. and Brunner, F. K. (2002). Short static gps sessions : robust estimation results. *GPS solutions*, 5(3) :70–79. (Cité en page 38.)
- [Willsky and Jones, 1976] Willsky, A. S. and Jones, H. L. (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *Automatic Control, IEEE Transactions on*, 21(1) :108–112. (Cité en page 37.)
- [Won et al., 2014] Won, D. H., Lee, E., Heo, M., Sung, S., Lee, J., and Lee, Y. J. (2014). GNSS integration with vision-based navigation for low GNSS visibility conditions. *GPS Solutions*, 18(2) :177–187. (Cité en page 28.)
- [Xie and Petovello, 2015] Xie, P. and Petovello, M. (2015). Measuring GNSS Multipath Distributions in Urban Canyon Environments. *IEEE Transactions on Instrumentation and Measurement*, 64(2) :366–377. (Cité en page 91.)
- [Yang et al., 2013] Yang, L., Knight, N. L., Li, Y., and Rizos, C. (2013). Optimal Fault Detection and Exclusion Applied in GNSS Positioning. *The Journal of Navigation*, 66(05) :683–700. (Cité en page 36.)
- [Yarlagadda et al., 2000] Yarlagadda, R., Ali, I., Al-Dhahir, N., and Hershey, J. (2000). GPS GDOP metric. *IEE Proceedings : Radar, Sonar and Navigation*, 147(5) :259–264. (Cité en page 85.)
- [Zair et al., 2015a] Zair, S., Le Hégarat-Masclé, S., and Seignez, E. (2015a). A-contrario approach for outlier detection in GNSS positioning. In *IEEE International Conference on Localization and GNSS (ICL-GNSS)*, pages 1–6. (Cité en pages 140 et 157.)
- [Zair et al., 2015b] Zair, S., Le Hégarat-Masclé, S., and Seignez, E. (2015b). Coupling outlier detection with particle filter for GPS-based localization. In *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*, pages 2518–2524. (Cité en pages 140 et 157.)
- [Zair et al., 2016a] Zair, S., Le Hégarat-Masclé, S., and Seignez, E. (2016a). A-contrario modeling for robust localization using raw GNSS data. *IEEE Transactions on Intelligent Transportation Systems*. (Cité en pages 98, 140 et 157.)
- [Zair et al., 2016b] Zair, S., Le Hégarat-Masclé, S., and Seignez, E. (2016b). An evidential ransac algorithm and its application to gnss positioning. In *International Conference on Belief Functions*, pages 241–250. Springer. (Cité en pages 140 et 157.)
- [Zair et al., 2016c] Zair, S., Le Hégarat-Masclé, S., and Seignez, E. (2016c). Outlier detection in GNSS pseudo-range/Doppler measurements for robust localization. *Sensors*, 16(4) :580. (Cité en pages 140 et 157.)
- [Zhang et al., 2006] Zhang, J., Zhang, K., Grenfell, R., and Deakin, R. (2006). GPS satellite velocity and acceleration determination using the broadcast ephemeris. *Journal of Navigation*, 59(02) :293–305. (Cité en page 93.)
- [Zhou and Li, 2015] Zhou, Z. and Li, B. (2015). GNSS windowing navigation with adaptively constructed dynamic model. *GPS Solutions*, 19(1) :37–48. (Cité en pages 74 et 94.)

Titre : Détection de données aberrantes appliquée à la localisation GPS.

Mots clés : détection de données aberrantes, localisation, GPS, estimation robuste, filtrage statistique.

Résumé : Dans cette thèse, nous nous intéressons au problème de détection de mesures GPS erronées. En effet, en zones urbaines, les acquisitions sont fortement dégradées par des phénomènes de multi-trajets ou de multiples réflexions des signaux avant d'arriver à l'antenne réceptrice. En forêt, de multiples obstacles bloquent les signaux satellites, ce qui diminue la redondance des mesures. Alors que les algorithmes présents dans les récepteurs GPS détectent au maximum une mesure erronée par *epoch*, avec une combinaison de différents systèmes de navigation, l'hypothèse d'une seule erreur à la fois n'est plus tenable et la détection et gestion des données erronées (défaillantes, aberrantes ou *outliers* selon les différentes terminologies) représente un enjeu majeur dans les applications de navigation autonome et de localisation robuste et devient un nouveau défi technologique.

La contribution principale de cette thèse est un algorithme de détection de mesures de pseudo-distances aberrantes exploitant la modélisation *a contrario*. Deux critères fondés sur l'espérance du

nombre de fausses alarmes (NFA) sont utilisés pour mesurer la cohérence d'un ensemble de mesures sous l'hypothèse d'un modèle de bruit. Notre seconde contribution concerne l'introduction des mesures Doppler dans le processus de localisation. Nous étendons la détection d'*outliers* conjointement dans les mesures de pseudo-distance aux mesures Doppler et proposons une localisation par couplage avec le filtre particulière soit SIR soit de Rao-Blackwell qui permet d'estimer analytiquement la vitesse. Notre troisième contribution est une approche crédibiliste pour la détection des mesures aberrantes dans les pseudo-distances. S'inspirant du RANSAC, nous choisissons, parmi les combinaisons d'observations possibles, la plus compatible selon une mesure de cohérence ou d'incohérence. Une étape de filtrage évidentiel permet de tenir compte de la solution précédente. Les approches proposées donnent de meilleures performances que les méthodes usuelles et démontrent l'intérêt de retirer les données aberrantes du processus de localisation.

Title: Outliers detection applied to GPS localization.

Keywords: Outlier detection, localization, GPS, robust estimation, statistical filtering.

Abstract: In this work, we focus on the problem of detection of erroneous GPS measurements. Indeed, in urban areas, acquisitions are highly degraded by multipath phenomena or signal multiple reflections before reaching the receiver antenna. In forest areas, the satellite occlusion reduces the measurements redundancy. While the algorithms embedded in GPS receivers detect at most one erroneous measurement per epoch, the hypothesis of a single error at a time is no longer realistic when we combine data from different navigation systems. The detection and management of erroneous data (faulty, aberrant or outliers depending on the different terminologies) has become a major issue in the autonomous navigation applications and robust localization and raises a new technological challenge. The main contribution of this work is an outlier detection algorithm for GNSS localization with an *a contrario* modeling. Two criteria based on number of false alarms (NFA) are used to measure

the consistency of a set of measurements under the noise model assumption. Our second contribution is the introduction of Doppler measurements in the localization process. We extend the outlier detection to both pseudo-ranges and Doppler measurements, and we propose a coupling with either the particle filter SIR or the Rao-Blackwellized particle filter that allows us to estimate analytically the velocity. Our third contribution is an evidential approach for the detection of outliers in the pseudo-ranges. Inspired by the RANSAC, we choose among possible combinations of observations, the most compatible one according to a measure of consistency or inconsistency. An evidential filtering step is performed that takes into account the previous solution. The proposed approaches achieve better performance than standard methods and demonstrate the interest of removing the outliers from the localization process.

