



HAL
open science

Apport de la prise en compte de la dépendance spatiotemporelle des séries temporelles de positions GNSS à l'estimation d'un système de référence

Clément Benoist

► **To cite this version:**

Clément Benoist. Apport de la prise en compte de la dépendance spatiotemporelle des séries temporelles de positions GNSS à l'estimation d'un système de référence. Astrophysique [astro-ph]. Université Paris sciences et lettres, 2018. Français. NNT : 2018PSLEO011 . tel-02117734

HAL Id: tel-02117734

<https://theses.hal.science/tel-02117734>

Submitted on 2 May 2019

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

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



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Observatoire de Paris

**Apport de la prise en compte de la dépendance
spatiotemporelle des séries temporelles de positions
GNSS à l'estimation d'un système de référence**

Soutenue par

Clément BENOIST

Le 28 septembre 2018

Ecole doctorale n° 127

**École doctorale d'astronomie
et d'astrophysique d'Île-de-
France**

Spécialités

Géodésie, statistique

Composition du jury :

Zuheir, ALTAMIMI Professeur, IGN/LAREG	<i>Directeur de thèse</i>
Liliane, BEL Professeur, AgroParisTech	<i>Examinatrice</i>
Christian, BIZOUARD Astronome, Observatoire de Paris	<i>Président</i>
Éric, CALAIS Professeur, Ecole Normale Supérieure	<i>Examineur</i>
Frédéric, MASSON Professeur, EOST	<i>Rapporteur</i>
Jean-Matthieu, NOCQUET Professeur Géosciences Azur	<i>Rapporteur</i>
Paul, REBISCHUNG Chargé de recherche, IGN/LAREG	<i>Encadrant</i>
Alvaro, SANTAMARIA-GOMEZ Professeur, Université de la Rochelle	<i>Examineur</i>



ÉCOLE DOCTORALE D'ASTRONOMIE ET
D'ASTROPHYSIQUE D'ÎLE-DE-FRANCE

THÈSE DE DOCTORAT

SOUTENUE PAR CLÉMENT BENOIST

APPORT DE LA PRISE EN COMPTE DE
LA DÉPENDANCE SPATIOTEMPORELLE
DES SÉRIES TEMPORELLES DE
POSITIONS GNSS À L'ESTIMATION
D'UN SYSTÈME DE RÉFÉRENCE

Soutenance le vendredi 28 septembre 2018 devant le jury composé
de :

Zuheir ALTAMIMI	<i>Directeur de thèse</i>
Liliane BEL	<i>Examinatrice</i>
Christian BIZOUARD	<i>Examineur</i>
Éric CALAIS	<i>Examineur</i>
Frédéric MASSON	<i>Rapporteur</i>
Jean-Mathieu NOCQUET	<i>Rapporteur</i>
Paul REBISCHUNG	<i>Co-encadrant</i>
Alvaro SANTAMARÍA-GÓMEZ	<i>Examineur</i>

La thèse a été encadré complémentairement par :

Xavier COLLILIEUX

Laboratoire de recherche en Géodésie (IGN)
Bâtiment Lamarck, 5 rue Thomas Mann, Case courrier 7071, 75 205 PARIS
CEDEX 13

Table des matières

Résumé	7
Abstract	9
I Généralités	15
1 Statistiques	17
1.1 Régression linéaire et moindres carrés	17
1.1.1 Moindres carrés ordinaires	18
1.1.2 Moindres carrés généralisés	18
1.1.3 Traitement du cas non linéaire	19
1.2 Maximum de vraisemblance	19
1.3 Séries chronologiques	20
1.3.1 Processus stationnaire	20
1.3.2 ARMA	21
1.3.3 Un exemple de processus à mémoire longue : les ARFIMA	22
1.3.4 Densité spectrale de puissance	22
1.3.5 Filtre de Kalman	23
1.3.6 Choix des valeurs a priori	25
1.4 Statistiques spatiales	25
1.4.1 Stationnarité et isotropie	25
1.4.2 Fonction de covariance semi-définie positive	26
1.4.3 Exemples de fonctions de covariance	26
1.4.4 Variogramme	27
2 Systèmes de référence et ITRF	29
2.1 Intérêts de l'ITRF	30
2.2 Systèmes de référence terrestre	31
2.2.1 Définition	31
2.2.2 Relation entre les systèmes de référence	31
2.3 International Terrestrial Reference System (ITRS)	32
2.4 Repère de référence terrestre	32
2.4.1 Définition	32
2.4.2 International Terrestrial Reference Frame (ITRF)	32
2.5 Processus de calcul de l'ITRF	35
2.5.1 Présentation des quatre techniques de géodésie spatiale .	35
2.5.2 Vision globale du processus d'obtention de l'ITRF	37

2.5.3	Obtention des solutions de géodésie spatiale	38
2.5.4	Combinaison intra-technique	39
2.5.5	Cumul long-terme (pour une technique)	41
2.5.6	Combinaison inter-technique	43
2.6	GNSS et ITRF	43
2.6.1	Présentation des systèmes GNSS actuels	43
2.6.2	Erreurs instrumentales des GNSS	44
2.6.3	Erreurs dues à des défauts de modélisations	45
3	État de l'art de l'analyse statistique des séries temporelles de positions GNSS	47
3.1	État de l'art sur la fonctionnelle	47
3.1.1	Signaux long-terme	47
3.1.2	Signaux saisonniers	49
3.1.3	Signaux draconitiques	50
3.2	État de l'art sur la partie stochastique	51
3.2.1	Nécessité d'un bruit corrélé	51
3.2.2	Aspect de la complexité de calcul des algorithmes	52
3.3	Dépendance spatiale	53
3.4	Filtre de Kalman pour la détermination de repères de référence	54
3.5	Problématique	55
3.5.1	Dépendance spatiale des séries de positions GNSS	55
3.5.2	Algorithmique : trouver une méthodologie rapide	55
3.5.3	Apport de la dépendance spatiale	56
II	Analyse statistique spatiotemporelle des séries temporelles de positions GNSS	57
4	Choix de la méthodologie statistique	59
4.1	Choix des données	59
4.1.1	Sélection pour l'étude des corrélations spatiales	60
4.1.2	Cas d'étude : 10 séries de coordonnées en Europe	63
4.2	Modèle cinématique : fonctionnelle	63
4.3	Première étude de la dépendance spatiale : dépendance des séries temporelles du bruit total	67
4.4	Choix du modèle temporel	67
4.4.1	Bruit corrélé	67
4.4.2	Bruit blanc variable	71
4.4.3	Bruit total	74
4.5	Modèle spatio-temporel proposé	75
4.5.1	Forme du modèle espace-état	75
4.6	Définition des différents modèles	79
4.7	Choix d'un nouveau système de coordonnées	83
4.8	Choix de modèle : validation croisée	85
4.9	Vérification de l'amélioration de l'estimation des vitesses	87

<i>TABLE DES MATIÈRES</i>	5
5 Résultats et discussion	91
5.1 Modèle ACS-VARIO : établissement du modèle à partir des vario-grammes	91
5.2 Validation croisée	95
5.2.1 Europe	95
5.2.2 Caraïbes	97
5.2.3 Côte est des États-Unis	99
5.3 Amélioration de l'estimation des vitesses	106
Conclusions et perspectives	109
III Annexes	111
A Posters et présentations	113

Résumé

Tout positionnement global précis nécessite un repère de référence tel le repère international de référence terrestre (ITRF). La détermination de l'ITRF s'appuie sur des séries temporelles de positions d'instruments géodésiques, en particulier des stations GNSS permanentes. Les séries temporelles de positions de stations GNSS sont corrélées temporellement et spatialement. De nombreuses études ont caractérisé la dépendance temporelle de ces séries et son impact sur la détermination de repères de référence. En revanche, les corrélations spatiales (entre stations proches) des séries GNSS n'ont jusqu'à présent jamais été prises en compte dans le calcul de repères de référence. L'objectif de cette thèse est donc de proposer une méthodologie pour la prise en compte de ces corrélations spatiales et d'évaluer son apport.

Les dépendances spatiales entre les séries de 195 stations GNSS sont tout d'abord évaluées à l'aide de variogrammes empiriques confirmant l'existence de corrélations jusqu'à des distances d'environ 5000 km. Des modèles de covariance exponentielle ne dépendant que de la distance inter-stations sont ajustés sur ces variogrammes empiriques.

Une méthodologie basée sur un filtre de Kalman est ensuite développée pour prendre en compte les dépendances spatiales des séries GNSS dans le calcul d'un repère de référence. Trois modèles de dépendance spatiale sont proposés : un modèle ne tenant pas compte de la dépendance spatiale (cas actuel du calcul de l'ITRF), un modèle basé sur les covariances empiriques entre séries de différentes stations, et un modèle basé sur les fonctions de covariance exponentielle mentionnées ci-dessus.

Ces différents modèles sont appliqués à trois jeux tests d'une dizaine de stations chacun situés en Europe, aux Caraïbes et sur la côte est des États-Unis. Les trois modèles sont évalués à l'aune d'un critère de validation croisée, c'est-à-dire sur leur capacité à prédire les positions des stations en l'absence de données. Les résultats sur les jeux tests d'Europe et des États-Unis montrent une amélioration considérable de cette capacité prédictive lorsque la dépendance spatiale des séries est prise en compte. Cette amélioration est maximale lorsque le modèle de covariance exponentielle est utilisé. L'amélioration est nettement moindre, mais toujours présente sur le jeu test des Caraïbes.

Les trois modèles sont également évalués sur leur capacité à déterminer des vitesses de déplacement exactes à partir de séries temporelles de positions courtes. L'impact de la prise en compte de la dépendance spatiale des séries sur l'exactitude des vitesses estimées est significatif. Comme précédemment, l'amélioration est maximale lorsque le modèle de covariance exponentielle est utilisé.

Cette thèse démontre ainsi l'intérêt de la prise en compte des dépendances spatiales entre séries GNSS pour la détermination de repères de référence. La

méthodologie développée pourra être utilisée pour le calcul de futures versions de l'ITRF.

Abstract

Any global and precise positioning requires a reference frame such as the International Terrestrial Reference Frame (ITRF). The determination of the ITRF relies on the position time series of various geodetic instruments, including in particular permanent GNSS stations. GNSS station position time series are known to be temporally and spatially correlated. Many authors have studied the temporal dependency of GNSS time series and its impact on the determination of terrestrial reference frames. On the other hand, the spatial correlations (i.e., between nearby stations) of GNSS time series have so far never been taken into account in the computation of terrestrial reference frames. The objective of this thesis is therefore to develop a methodology to account for the spatial correlations of GNSS time series, and evaluate its benefits.

The spatial dependencies between the position time series of 195 GNSS stations are first evaluated by means of empirical variograms, which confirm the existence of correlations up to distances of about 5000 km. Exponential covariance models, depending only on the distance between stations, are adjusted to these empirical variograms.

A methodology based on a Kalman filter is then developed to take into account the spatial dependencies of GNSS time series in the computation of a terrestrial reference frame. Three models of spatial dependency are proposed : a model which does not account for the spatial dependency between GNSS time series (current case of the ITRF computation), a model based on the empirical covariances between the time series of different stations, and a model based on the exponential covariance functions mentioned above.

These different models are applied to three test cases of ten stations each, located in Europe, in the Caribbean, and along the east coast of the US. The three models are evaluated with regard to a cross-validation criterion, i.e., on their capacity to predict station positions in the absence of observations. The results obtained with the Europe and US test cases demonstrate a significant improvement of this predictive capacity when the spatial dependency of the series is taken into account. This improvement is highest when the exponential covariance model is used. The improvement is much lower, but still present with the Caribbean test case.

The three models are also evaluated with regard to their capacity to determine accurate station velocities from short position time series. The impact of accounting for the spatial dependency between series on the accuracy of the estimated velocities is again significant. Like previously, the improvement is highest when the exponential covariance model is used.

This thesis thus demonstrates the interest of accounting for the spatial dependency of GNSS station position time series in the determination of terrestrial

reference frames. The developed methodology could be used in the computation of future ITRF versions.

Introduction

Déterminer les coordonnées d'un objet ou encore sa vitesse de déplacement nécessitent l'adoption d'un système de référence. Ce système nécessite la description du type de coordonnées, par exemple cartésienne ou géographique, mais aussi la position et l'orientation du système d'axes par rapport à la Terre. La mesure directe de distances par rapport à ces axes n'étant pas possible, on utilise des points intermédiaires, dits points de référence, ou encore des satellites de coordonnées connues. On appelle cet ensemble de coordonnées un repère de référence. Cependant, la Terre est un solide déformable. En effet, plusieurs phénomènes contribuent à ces déformations comme les mouvements tectoniques ou encore les marées terrestres (dues à l'attraction de la Lune et du Soleil exercée sur la Terre).

Les repères de référence terrestre modernes tels le repère international de référence terrestre (ITRF) modélisent des coordonnées qui varient au cours du temps du fait du caractère déformable de la Terre. Ces coordonnées sont décrites par des fonctions affines par morceaux (avec des discontinuités de position ou de vitesse) avec pour certains points en zone de séismes des fonctions plus complexes. L'ITRF résume ces fonctions par la donnée de positions à un instant de référence conventionnel et de vitesses pour chacune des stations de mesures du réseau. Les coordonnées ITRF sont actualisées régulièrement, des nouvelles éditions de l'ITRF sont donc régulièrement publiées, la dernière en date se nomme l'ITRF2014. Chaque nouvelle édition permet de prendre en compte de nouveaux modèles de réduction des données géodésiques, d'intégrer de nouvelles données et nouvelles stations de mesure et surtout d'intégrer les déformations non prédictibles de la Terre comme les séismes et la déformation liée à l'évolution de la fonte des glaces actuelle.

En plus de son usage dans les sciences géographiques et en particulier la cartographie, l'ITRF est utilisé dans de nombreuses applications scientifiques comme l'étude du mouvement des plaques tectoniques et l'évaluation de la montée du niveau des mers dans un repère absolu. Ces usages imposent une grande précision évaluée à 1 mm pour les positions et 0.1 mm/an pour les vitesses par la communauté scientifique [Rothacher et al., 2008]. Cette précision n'est toutefois pas encore atteinte aujourd'hui [Altamimi et al., 2016, Collilieux et al., 2014].

Le calcul de l'ITRF s'appuie principalement sur des données acquises par les techniques de géodésie spatiale qui concourent au positionnement des stations : le système de détermination d'orbite précise de satellite (DORIS), la télémétrie laser sur satellite (SLR), l'interférométrie à très longue base (VLBI) et les systèmes de positionnements globaux par satellites (GNSS). Chaque technique présente des avantages et des inconvénients, en termes de précision des coordonnées obtenues, d'exactitude de localisation du centre des masses de la Terre

ou encore de distribution des stations de mesure. Chaque lot de mesures issues d'une technique, qui est intégré dans le calcul de l'ITRF, doit être pondéré de façon optimale, l'algorithme utilisé étant issu des travaux de [Sillard, 1999].

Le calcul de l'ITRF comporte quatre étapes :

- le traitement des observations issues des différentes techniques de géodésie spatiale, traitement effectué de façon indépendante pour chaque technique par différents centres de calculs internationaux
- la combinaison dite intra-technique qui fournit des coordonnées uniques pour chaque station de chaque technique et permet un contrôle qualité en cas de détermination multiple (l'International GNSS Service – IGS – fournit des coordonnées et les matrices de variance-covariance associées à cette étape),
- le cumul long-terme, qui, pour chaque technique, permet d'estimer des positions à un instant de référence et des vitesses de déplacement, à partir de séries temporelles de positions instantanées,
- la combinaison inter-technique pour déterminer les positions et vitesses des stations des quatre techniques dans un référentiel unique.

Les mesures GNSS, dont le GPS est la déclinaison américaine, occupent une place importante dans le processus d'estimation de l'ITRF du fait de la bonne répartition des stations permanentes, de leur faible coût et de leur grande précision. Les données actuellement intégrées à ce calcul sont fournies par l'International GNSS Service (IGS). Cet organisme exploite en continu les données GNSS acquises par un réseau de stations GNSS permanentes pour fournir des produits géodésiques parmi lesquels les positions journalières de ces stations. Ce travail s'intéresse plus particulièrement à l'intégration de ces données dans le processus de calcul de l'ITRF et à l'étape de cumul long-terme. Ce cumul suppose actuellement que les coordonnées journalières IGS sont affectées d'un bruit de mesure non corrélé dans le temps, mais corrélé spatialement (information fournie par une matrice de variance-covariance). Les coordonnées sont modélisées par des fonctions affines par morceaux et des signaux périodiques [Altamimi et al., 2016].

De nombreuses études ont analysé le contenu des séries temporelles de positions GNSS. Les séries temporelles de positions, comme dans l'ITRF, sont généralement décomposées en une partie déterministe et une partie stochastique. La partie déterministe est généralement estimée par moindres carrés ; le reste est appelé bruit ou résidus. La tendance linéaire peut inclure la modélisation de discontinuités de position et, éventuellement, de vitesse dues principalement aux tremblements de terre et aux changements de matériel de mesure. Le mouvement périodique, lié en partie à la réponse de la Terre aux transferts de masse à sa surface, a été étudié par [Collilieux, 2008] en vue de sa modélisation dans le calcul de l'ITRF. Il montre que des erreurs systématiques propres aux techniques de mesure sont aussi susceptibles d'affecter les coordonnées à ces fréquences. [Lercier, 2014] a complété cette bibliothèque de fonctions par des fonctions logarithmes et exponentielles afin de modéliser les coordonnées des stations affectées par un mouvement long-terme aisé se produisant consécutivement à un séisme, mouvement appelé déformation post-sismique. Dans le processus de calcul de l'ITRF2014, ces déformations ont été modélisées à l'aide des données GNSS seules puis corrigées avant le cumul long-terme des coordonnées de chacune des techniques. Cette modélisation n'est pas remise en cause dans ce travail qui porte sur la modélisation statistique du bruit.

En effet, différentes études montrent que le bruit dans les séries temporelles sont corrélés temporellement [Zhang et al., 1997, Mao et al., 1999, Williams, 2003], corrélation actuellement ignorée dans le calcul de l'ITRF ; il est admis que le meilleur modèle temporel de bruit est la somme d'un bruit blanc et d'un bruit de scintillation [Santamaría-Gómez et al., 2011] que nous définirons prochainement (sous-section 1.3.3). Or la modélisation de ces corrélations temporelles a un impact sur les vitesses estimées et surtout sur leurs incertitudes. La dépendance spatiale des séries a été moins étudiée mais mise en évidence dans différentes études. Elle a notamment été prise en compte assez tôt par la méthode dite de filtrage spatiale [Wdowinski et al., 1997]. Elle consiste à retirer de la série temporelle de positions d'une station la moyenne des résidus de séries proches, afin de filtrer le signal spatialement corrélé. Cette méthode a toutefois l'inconvénient de ne pas caractériser cette dépendance spatiale. Celle-ci a été analysée à l'aide d'un corrélogramme par [Amiri-Simkooei et al., 2017] (le corrélogramme est une représentation qui trace la corrélation en fonction de la distance). Celui-ci a été calculé à partir des coordonnées GNSS fournies par le Jet Propulsion Laboratory (JPL) mais aucun modèle de covariance utilisable n'a été fourni ; cette étude reste à confirmer par l'étude des données calculées par l'IGS et supposées de meilleure qualité (issues d'une combinaison intra-technique).

La prise en compte de la dépendance spatio-temporelle (spatiale et temporelle) lors de la modélisation des séries de positions GNSS pourrait avoir plusieurs bénéfices :

- améliorer l'estimation des vitesses et de leurs incertitudes associées,
- améliorer la prédiction des positions dans le futur,
- améliorer la prédiction des valeurs manquantes ou aberrantes [Collilieux and Bernard, 2012],
- permettre une meilleure détection des discontinuités de position ou/et de vitesse [Gazeaux et al., 2015].

L'objectif de ce travail est de caractériser la dépendance spatiale des positions de stations GNSS (en particulier les données IGS) et de proposer une méthode de cumul opérationnelle intégrant une modélisation spatio-temporelle du bruit GNSS : nous avons donc choisi d'utiliser comme méthode centrale un filtre de Kalman, qui ne nécessite pas l'inversion de grandes matrices contrairement à la méthode des moindres carrés.

La première partie de ce manuscrit présente les généralités que j'ai jugées utiles à la bonne compréhension de ma thèse. Le premier chapitre introduit les principales méthodologies statistiques utilisées dans cette thèse. Le second chapitre décrit l'intérêt, le mode de calcul de l'ITRF ainsi que la place des GNSS dans l'ITRF. Enfin, un état de l'art sur l'analyse statistique des séries de positions GNSS sera présenté.

La seconde partie proposera mon apport personnel. Nous présenterons les données de positions GNSS utilisées et sélectionnerons les meilleures données pour cette thèse sur des critères de qualité que nous expliciterons en vue de caractériser le bruit spatiotemporel. Nous effectuerons une étude préliminaire montrant l'existence d'une dépendance spatiale sur ces données. Nous exposerons la méthodologie retenue et proposerons plusieurs modèles de dépendance spatiale. Nous verrons quel est le meilleur modèle en vue de prédire les positions dans le futur grâce à une technique appelée validation croisée. Nous montrerons finalement que la prise en compte de la dépendance spatiale améliore l'estima-

tion des vitesses de stations GNSS lorsque celles-ci possèdent des séries courtes.

Première partie

Généralités

Chapitre 1

Statistiques

Les statistiques ont pour but de décrire certaines données qui paraissent inexplicables ou sont trop nombreuses pour être facilement appréhendées. Elles peuvent aussi avoir pour but de prédire une entité (fonction, réels, entiers, etc.). Pour ces objectifs, elles s'appuient sur les lois des probabilités. L'étude des séries temporelles de positions GNSS s'appuie sur les moindres carrés, le maximum de vraisemblance, les séries chronologiques. L'apport statistique principal de la thèse sera d'introduire les modèles de statistiques spatiales dans la modélisation du bruit corrélé affectant les séries de coordonnées des stations GNSS permanentes.

1.1 Régression linéaire et moindres carrés

Cette section s'inspire de [Guyader, 2011, Tufféry, 2012]. La régression linéaire vise à expliquer une variable continue unidimensionnelle Y par une variable continue X multidimensionnelle (une variable continue par opposition à une variable discrète prend ses valeurs dans un ensemble continu, par exemple, \mathbb{R}). On considère y_1, \dots, y_n les réalisations de la variable aléatoire Y , qui correspondent aux observations. À la réalisation y_i sont associées les réalisations $x_{i,1}, \dots, x_{i,p}$.

On considère le modèle linéaire :

$$y_i = \sum_{j=1}^p x_{i,j} \beta_j + \varepsilon_i.$$

La variable ε_i est appelée erreur et est supposée aléatoire, cela correspond aux écarts (positifs ou négatifs) au modèle linéaire, $(x_{i,j})_{1 \leq i \leq n, 1 \leq j \leq p}$ est la matrice des régresseurs (ou matrice du plan d'expérience) que nous noterons \underline{A} , β_j sont inconnus à estimer ($\vec{\beta}$ est le vecteur des régresseurs) et les y_i sont les observations. On peut écrire ce modèle sous forme matricielle :

$$\vec{Y} = \underline{A} \vec{\beta} + \vec{\varepsilon}$$

où $\vec{\beta} = (\beta_1, \dots, \beta_p)^T$, $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ et $\vec{Y} = (y_1, \dots, y_n)^T$.

On définit une fonction de perte $L : \mathbb{R} \rightarrow \mathbb{R}^+; v \mapsto L(v)$. On prend en général :

- $L(v) = |v|$
- $L(v) = v^2$

On effectue la minimisation :

$$\arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n L \left(y_i - \sum_{j=1}^p x_{i,j} \beta_j \right) = \hat{\beta}.$$

Remarque L'ajout d'un terme de pénalisation dans la fonction de perte L peut donner les estimateurs LASSO et RIDGE, le LASSO ayant l'avantage d'être parcimonieux (le vecteur $\hat{\beta}$ a beaucoup de coefficients nuls).

1.1.1 Moindres carrés ordinaires

Dans les moindres carrés ordinaires, on utilise la fonction de perte $L : v \mapsto v^2$. Nous supposons, de plus, que \underline{A} est une matrice de rang plein. On cherche à minimiser par rapport à $\vec{\beta}$ la quantité suivante :

$$\sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{i,j} \beta_j \right)^2 = \|\vec{Y} - \underline{A}\vec{\beta}\|_2^2.$$

La minimisation s'effectue par l'annulation de la différentielle (au sens de Fréchet). On considère $g(\vec{\beta}) = \|\vec{Y} - \underline{A}\vec{\beta}\|_2^2$. On considère $\|\cdot\|_2$ comme une norme euclidienne avec comme produit scalaire sous-jacent $\langle A, B \rangle = A^T B$.

On applique la définition de la différentielle. On considère $\vec{h} \in \mathbb{R}^n$.

$$\begin{aligned} & \|\vec{Y} - \underline{A}(\vec{\beta} + \vec{h})\|_2^2 - \|\vec{Y} - \underline{A}\vec{\beta}\|_2^2 \\ &= \|(\vec{Y} - \underline{A}\vec{\beta} - \frac{1}{2}\underline{A}\vec{h}) - \frac{1}{2}\underline{A}\vec{h}\|_2^2 - \|(\vec{Y} - \underline{A}\vec{\beta} - \frac{1}{2}\underline{A}\vec{h}) + \frac{1}{2}\underline{A}\vec{h}\|_2^2 \\ &= 4\langle \vec{Y} - \underline{A}\vec{\beta} - \frac{1}{2}\underline{A}\vec{h}, \frac{1}{2}\underline{A}\vec{h} \rangle \end{aligned}$$

Ainsi la différentielle de g en $\vec{\beta}$ vaut-elle $\vec{h} \mapsto 2(\vec{Y} - \underline{A}\vec{\beta})^T \underline{A}\vec{h}$. Elle s'annule lorsque $\vec{\beta} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \vec{Y}$. On en déduit l'estimateur des moindres carrés ordinaires

$$\hat{\beta} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \vec{Y}. \quad (1.1)$$

1.1.2 Moindres carrés généralisés

Dans les moindres carrés généralisés, on suppose que $\vec{\varepsilon} \sim \mathcal{N}(0, \underline{\Sigma})$. On minimise :

$$\|\underline{\Sigma}^{-1/2}(\vec{Y} - \underline{A}\vec{\beta})\|_2.$$

ce qui a un sens, car $\underline{\Sigma}$ est une matrice symétrique définie positive. On définit le produit scalaire : $\langle \vec{s}, \vec{r} \rangle_{\underline{\Sigma}} = \vec{s}^T \underline{\Sigma}^{-1} \vec{r}$ et sa norme associée $\|\vec{s}\|_{\underline{\Sigma}} = \sqrt{\vec{s}^T \underline{\Sigma}^{-1} \vec{s}}$.

$$\begin{aligned} \|\underline{\Sigma}^{-1/2}(\vec{Y} - \underline{A}(\vec{\beta} + \vec{h}))\|_2^2 - \|\underline{\Sigma}^{-1/2}(\vec{Y} - \underline{A}\vec{\beta})\|_2^2 &= \|\vec{Y} - \underline{A}(\vec{\beta} + \vec{h})\|_{\underline{\Sigma}}^2 - \|\vec{Y} - \underline{A}\vec{\beta}\|_{\underline{\Sigma}}^2 = \\ \|\vec{Y} - \underline{A}(\vec{\beta} + \frac{1}{2}\vec{h}) + \frac{1}{2}\underline{A}\vec{h}\|_{\underline{\Sigma}}^2 - \|\vec{Y} - \underline{A}(\vec{\beta} + \frac{1}{2}\vec{h}) - \frac{1}{2}\underline{A}\vec{h}\|_{\underline{\Sigma}}^2 &= \\ 4\langle \vec{Y} - \underline{A}(\vec{\beta} + \frac{1}{2}\vec{h}), \frac{1}{2}\underline{A}\vec{h} \rangle_{\underline{\Sigma}} \end{aligned}$$

La différentielle de $g : \vec{\beta} \mapsto \|\underline{\Sigma}^{-1/2}(\vec{Y} - \underline{A}\vec{\beta})\|_2^2$ est donc $2(\vec{Y} - \underline{A}\vec{\beta})^T \underline{\Sigma}^{-1} \underline{A}\vec{h}$ et s'annule pour $\vec{\beta} = (\underline{A}^T \underline{\Sigma}^{-1} \underline{A})^{-1} \underline{A}^T \underline{\Sigma}^{-1} \vec{Y}$, qui correspond à l'estimateur des moindres carrés généralisés.

Lorsque $\underline{\Sigma}$ est pleine, on parle de moindres carrés pondérés.

1.1.3 Traitement du cas non linéaire

Il peut arriver que l'on veuille linéariser une fonction non linéaire par rapport à $\vec{\beta}$. On considère donc le modèle :

$$y_i = f_{\vec{\beta}}(x_i) + \varepsilon_i$$

où les ε_i suivent des lois normales. Nous voulons fournir un estimateur de $\vec{\beta}$. Nous allons définir une suite $(\vec{\beta}^k)_{k=0}^{\infty}$ par la donnée de $\vec{\beta}^0$ et d'un algorithme définissant la récurrence.

Soit $\vec{\beta}^k$. Nous définissons $\vec{\beta}^{k+1}$ par $\vec{\beta}^{k+1} = \vec{\beta}^k + \Delta\beta^k$. Nous commençons par opérer un développement limité d'ordre 1 :

$$f_{\vec{\beta}^{k+1}}(x_i) \approx f_{\vec{\beta}^k}(x_i) + \sum_j (\beta_j^{k+1} - \beta_j^k) \left. \frac{\partial f_{\vec{\beta}^k}}{\partial \beta_j^k} \right|_{\vec{\beta}=\vec{\beta}^k, x=x_i}$$

Avec la notation $\Delta y_i = y_i - f_{\vec{\beta}^k}(x_i)$ et $\underline{J}^{(k)}$ la matrice jacobienne en $\vec{\beta}^k$ de la fonction $\vec{\beta} \mapsto (f_{\vec{\beta}}(x_1), \dots, f_{\vec{\beta}}(x_n))$, on aboutit à l'équation :

$$y_i - f_{\vec{\beta}^k}(x_i) = \Delta y_i - \sum_j J_{i,j}^{(k)} \Delta \beta_j$$

Nous devons donc résoudre matriciellement :

$$(\underline{J}^{(k)T} \underline{J}^{(k)}) \Delta \vec{\beta} = \underline{J}^T \Delta \vec{y}.$$

Nous aboutissons donc à l'équation de récurrence :

$$\vec{\beta}^{k+1} = \vec{\beta}^k + (\underline{J}^{(k)T} \underline{J}^{(k)})^{-1} \underline{J}^T \Delta \vec{y}$$

1.2 Maximum de vraisemblance

Le maximum de vraisemblance est un procédé d'estimation de paramètres à l'aide de la maximisation d'une quantité. Nous verrons qu'il sert à estimer les paramètres de bruit dans la méthodologie développée dans ma thèse au chapitre 4 ; on utilise pour cela les logiciels CATS [Williams, 2008] et HECTOR [Bos et al., 2013a].

On considère un n -échantillon x_1, \dots, x_n issu d'une même loi de probabilité D_{θ} , les variables aléatoires étant continues. Nous considérons leur densité $f_{\vec{\theta}}(\vec{x})$; $\vec{\theta}$ est inconnue et à estimer. On définit la vraisemblance :

$$L(\vec{\theta}; x_1, \dots, x_n) = f_{\vec{\theta}}(\vec{x}).$$

Nous pouvons aussi définir la log-vraisemblance qui est le logarithme de la vraisemblance :

$$\ell(\vec{\theta}; x_1, \dots, x_n) = \ln(f_{\vec{\theta}}(\vec{x})).$$

Par exemple, pour une régression linéaire de dimension n , nous considérons que les observations suivent une loi normale centrée sur $\underline{A}\vec{\theta}$ ($x \sim \mathcal{N}(\underline{A}\vec{\theta}, \underline{\Sigma})$), ce qui aboutit à la fonction log-vraisemblance suivante :

$$\ell(\vec{\theta}, \vec{x}) = -\frac{1}{2} \left(n \ln(2\pi) + \ln \det \underline{\Sigma} + (\vec{x} - \underline{A}\vec{\theta})^T \underline{\Sigma}^{-1} (\vec{x} - \underline{A}\vec{\theta}) \right)$$

où $\vec{\theta}$ est le vecteur des régresseurs et \underline{A} la matrice des régresseurs. Dans cet exemple, $\underline{\Sigma}$ peut dépendre de $\vec{\theta}$ ce qui sera utile dans de nombreux travaux d'analyse statistique des séries temporelles GNSS présentés dans le chapitre 3. On recherche via un algorithme d'optimisation [Fletcher and Reeves, 1964, Nelder and Mead, 1965, Bélisle, 1992] le vecteur des paramètres $\vec{\theta}$ qui maximise cette fonction log-vraisemblance.

1.3 Séries chronologiques

Pour cette section, je me suis inspiré du polycopié de [Prenat et al., 2010].

Les séries chronologiques univariées ou séries temporelles univariées sont des processus aléatoires $(X_t)_{t \in T}$ à valeurs dans \mathbb{R} où T est une partie discrète de \mathbb{R} . Une série chronologique multivariée est un processus aléatoire $(X_{\vec{t}})_{\vec{t} \in T}$ à valeurs dans \mathbb{R}^k où T est une partie discrète de \mathbb{R} et $k \in \mathbb{N}$. Par exemple, une suite de positions tridimensionnelles dans l'espace indicée par le temps constitue une série chronologique multivariée comme par exemple sur la figure 3.1. Ce processus multivarié peut contenir des valeurs manquantes. Dans cette partie, nous décrivons uniquement des processus stationnaires sur une partie de \mathbb{Z} car les séries de coordonnées GNSS sont régulièrement espacées aux valeurs manquantes près.

Un exemple classique de série chronologique est le bruit blanc qui correspond à une suite de variables aléatoires indépendantes admettant une variance ; le plus souvent, ces variables aléatoires sont gaussiennes centrées.

1.3.1 Processus stationnaire

La stationnarité des processus est une propriété centrale dans l'étude des séries temporelles, les séries stationnaires étant réputées plus faciles à étudier que celles qui ne le sont pas.

Un processus $(X_t)_{t \in \mathbb{Z}}$ est dit stationnaire au sens large si :

- le processus est L^2 i.e admet une variance : $\forall t \in \mathbb{Z} \quad \mathbb{E}X_t^2 < \infty$,
- sa moyenne ou son espérance $t \mapsto \mathbb{E}X_t$ est constante sur \mathbb{Z} ,
- $\forall r, s, \tau \in \mathbb{Z} \quad \text{Cov}(X_r, X_s) = \text{Cov}(X_{r+\tau}, X_{s+\tau})$.

Un processus $(X_t)_{t \in \mathbb{Z}}$ est stationnaire au sens strict si, pour tout $k \in \mathbb{N}^*$, pour tous t_1, \dots, t_k et pour tout $\tau \in \mathbb{Z}$, la distribution conjointe de X_{t_1}, \dots, X_{t_k} et celle de $X_{t_1+\tau}, \dots, X_{t_k+\tau}$ sont les mêmes.

Lorsque le processus est L^2 , la stationnarité stricte implique la stationnarité large.

Pour un processus stationnaire au sens large et contrairement à la plupart des processus non stationnaires, il est possible de définir la fonction de covariance :

$$\gamma(\tau) = \text{Cov}(X_s, X_{s+\tau})$$

où $\tau \in \mathbb{Z}$. La fonction $\gamma(\cdot)$ est symétrique : $\gamma(\tau) = \gamma(-\tau)$.

On définit aussi la fonction de corrélation comme une normalisation de la fonction de covariance :

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}.$$

Le coefficient de normalisation est fixé de telle sorte que $\rho(0) = 1$. De même que la fonction de covariance, la fonction de corrélation est symétrique.

Une série chronologique est composée d'une partie déterministe et d'une partie aléatoire. La partie aléatoire est obtenue en soustrayant à la série chronologique totale la partie déterministe obtenue en général par régression. La partie aléatoire peut être due à une variabilité intrinsèque du phénomène ou/et à des erreurs de mesures.

1.3.2 ARMA

Une famille classique de processus stochastiques est la famille des ARMA (autoregressive moving average) qui sont des processus stochastiques centraux dans les statistiques temporelles notamment dans le cas du filtrage. On dit que $(X_t)_{t \in \mathbb{Z}}$ est un $ARMA(p, q)$ de moyenne nulle si :

- (X_t) est stationnaire de moyenne nulle,
- il existe p réels $\alpha_1, \dots, \alpha_p$, q réels β_1, \dots, β_q et des variables aléatoires normales indépendantes, identiquement distribuées $(\varepsilon_t)_{t=-\infty}^{\infty}$ tels que $X_t - \alpha_1 X_{t-1} - \dots - \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$.

$(X_t)_{t \in \mathbb{Z}}$ est un ARMA de moyenne μ si et seulement si $(X_t - \mu)_{t \in \mathbb{Z}}$ est un $ARMA(p, q)$ de moyenne nulle.

Nous pouvons définir l'opérateur « backward » : $BX_t = X_{t-1}$, et, plus généralement, $B^k X_t = X_{t-k}$. On peut considérer des polynômes de l'opérateur B . On peut définir les polynômes :

$$A(z) = 1 - \alpha_1 z - \dots - \alpha_p z^p$$

et

$$C(z) = 1 + \beta_1 z + \dots + \beta_q z^q.$$

On peut donc écrire : $A(B)X_t = C(B)\varepsilon_t$.

Un $ARMA(p, 0)$ est appelé un $AR(p)$ et un $ARMA(0, q)$ est un $MA(q)$.

Les ARMA ont un intérêt d'approximation considérable car pour tout processus stationnaire dont la fonction de covariance tend vers 0 à l'infini et pour tout $k \in \mathbb{N}$, il est possible de trouver un processus ARMA dont la fonction de covariance coïncide avec celle de ce processus jusqu'à l'ordre k .

1.3.3 Un exemple de processus à mémoire longue : les ARFIMA

Une famille plus générale de processus stochastiques que celle des ARMA est la famille des ARFIMA.

Un $ARFIMA(0, d, 0)$ est un bruit blanc (décorrélé) intégré à un ordre fractionnaire. Pour $d \in]-0.5, 0.5[$, nous définissons l' $ARFIMA(0, d, 0)$ par le processus X_t tel que $(1 - B)^d X_t$ est un bruit blanc où

$$(1 - B)^d = \sum_{k=0}^{\infty} (-1)^k \binom{d}{k} B^k = 1 - dB + \binom{d}{2} B^2 - \binom{d}{3} B^3 + \dots$$

L' $ARFIMA(0, d, 0)$, lorsque d tend vers $\frac{1}{2}$, tend à avoir une densité spectrale proportionnelle à $1/f$ ce qui correspond à un bruit rose ou bruit de scintillation [Hosking, 1981].

1.3.4 Densité spectrale de puissance

La densité spectrale sert à appréhender les composantes périodiques d'une série chronologique stationnaire. La notion de densité spectrale de puissance s'appuie sur celle de transformée de Fourier de la série. Nous supposons que la fonction de covariance vérifie cette propriété de sommabilité :

$$\sum_{k=-\infty}^{\infty} |\gamma_X(k)| < \infty.$$

On dit que la densité spectrale de puissance est la transformée de Fourier de sa fonction de covariance :

$$P_X(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \gamma_X(l) e^{il\omega}$$

On peut déduire la covariance de la densité spectrale de puissance :

$$\gamma_X(l) = \int_{-\pi}^{\pi} P_X(\omega) e^{il\omega} d\omega.$$

Une estimation non paramétrique de la densité spectrale de puissance est le périodogramme que l'on peut obtenir en utilisant une série chronologique finie régulièrement espacée $(X_t)_{t \in \llbracket 1, N \rrbracket}$:

$$\hat{S}_{per}(\omega) = \frac{1}{N} \left| \sum_{j=1}^N X_j e^{ij\omega} \right|^2.$$

Notons que le périodogramme est une estimation biaisée de la densité spectrale de puissance.

Il existe une autre vision du périodogramme qui est défini à partir de la transformée de Fourier. On considère donc un vecteur (x_1, \dots, x_n) de \mathbb{C}^n . On peut choisir de manière unique des coefficients $(c_k)_{k=1}^n$ de telle sorte que :

$$x_t = \sum_{k=\lfloor \frac{n-1}{2} \rfloor}^{\lfloor \frac{n}{2} \rfloor} c_k e^{\frac{ikt}{2\pi n}},$$

où $\lfloor \cdot \rfloor$ est la partie entière inférieure.

On a l'égalité :

$$\hat{S}_{per}\left(\frac{2\pi k}{n}\right) = \|c_k\|_2^2.$$

Cette vision a néanmoins l'inconvénient de n'offrir que des valeurs discrètes.

[Scargle, 1982] permet d'obtenir des périodogrammes de séries temporelles espacées irrégulièrement (avec, par exemple, des valeurs manquantes) appelés périodogrammes de Lomb-Scargle grâce à la formule :

$$P_X(\omega) = \frac{1}{2} \left(\frac{(\sum_j X_j \cos(\omega(t_j - \tau)))^2}{\sum_j \cos^2(\omega(t_j - \tau))} + \frac{(\sum_j X_j \sin(\omega(t_j - \tau)))^2}{\sum_j \sin^2(\omega(t_j - \tau))} \right) \quad (1.2)$$

où τ est défini de telle sorte que

$$\tan(2\omega\tau) = \frac{\sum_j \sin(2\omega t_j)}{\sum_j \cos(2\omega t_j)}.$$

Notons que, dans le cas de données régulièrement espacées, le périodogramme de Lomb-Scargle et le périodogramme classique sont les mêmes.

1.3.5 Filtre de Kalman

Le filtre de Kalman est très utilisé en automatique et traitement du signal. Son nom provient de Rodolf Kalman, bien que sa paternité soit soumise à controverse. Le filtre de Kalman permet d'estimer des états cachés à partir d'observations. Le filtre de Kalman a un caractère séquentiel, contrairement aux moindres carrés, ce qui permet une efficacité algorithmique supérieure.

Meilleur prédicteur linéaire

Nous considérons des variables aléatoires $L^2 : (X_1, \dots, X_n, V)$. Nous voulons donner un sens à la notion d'estimateur optimal de V sachant X_1, \dots, X_n . On cherchera à prédire V comme une combinaison linéaire des X_i plus une constante, la distance L^2 servira de critère. Nous cherchons donc $\lambda_0, \lambda_1, \dots, \lambda_n$ tels que :

$$\hat{V} = \lambda_0 + \sum_{i=1}^n \lambda_i X_i.$$

Ces λ_i sont choisis de telle sorte que $\mathbb{E}[(V - \hat{V})^2]$ soit minimal. Notons V^* cette solution. On dit que V^* est le meilleur prédicteur linéaire.

Modèle linéaire

Nous considérons un vecteur d'observation Y_t de dimension p et un vecteur des états X_t de dimension q . Les observations sont connues et les états doivent être estimés. On considère que les états sont reliés linéairement à un bruit près aux observations :

$$Y_t = H_t X_t + B_t \quad (1.3)$$

où H_t est une matrice déterministe de dimension $p \times q$, B_t est un vecteur gaussien de dimension p et de matrice de variance-covariance M_t appelée matrice de variance-covariance du bruit d'observation.

De manière similaire, les états sont reliés avec une relation linéaire entre l'état suivant et l'état précédent :

$$X_{t+1} = L_t X_t + A_t \quad (1.4)$$

où L_t est une matrice déterministe de dimension $q \times q$, A_t est un vecteur gaussien de dimension q et de matrice de variance-covariance T_t appelée matrice de variance-covariance du bruit de transition.

Filtre de Kalman

L'objectif du filtre de Kalman [Kalman, 1960] est d'estimer les X_k en fonction d'un ensemble d'observations Y_t . Notons $\hat{X}_{t|k}$ le meilleur prédicteur linéaire de X_t sachant Y_0, \dots, Y_k .

La prédiction, le filtrage et le lissage visent tous à estimer X_t , mais notre connaissance des Y_t diffère. La prédiction vise à estimer X_t sachant Y_0, \dots, Y_{t-1} , le filtrage vise à estimer X_t sachant Y_0, \dots, Y_t et le lissage vise à estimer X_t sachant Y_0, \dots, Y_n avec $n > t$.

Nous noterons $P_{t|t}$ (respectivement $P_{t|t-1}$) la matrice de variance-covariance associée à $\hat{X}_{t|t}$ (respectivement $\hat{X}_{t|t-1}$).

D'abord, nous devons initialiser le vecteur des états à l'instant 0 $\hat{X}_{0|0}$ et la matrice de variance-covariance associée $P_{0|0}$. En l'absence d'information a priori qui pourrait déterminer la matrice $P_{0|0}$, nous pouvons prendre $P_{0|0}$ très grand ; alors, l'influence de $\hat{X}_{0|0}$ sera très faible.

Les estimations s'effectuent par les itérations suivantes. L'état est prédit ainsi que sa matrice de variance-covariance :

$$\begin{aligned} \hat{X}_{t|t-1} &= L_{t-1} \hat{X}_{t-1|t-1} \\ \hat{P}_{t|t-1} &= L_{t-1} \hat{P}_{t-1|t-1} L_{t-1}^T + T_{t-1} \end{aligned}$$

On prédit ensuite le vecteur des observations :

$$\hat{Y}_{t|t-1} = H_t \hat{X}_{t|t-1}$$

Nous calculons ensuite l'innovation ν_t ainsi que sa covariance N_t :

$$\begin{aligned} \nu_t &= Y_t - \hat{Y}_{t|t-1} \\ N_t &= H_t \hat{P}_{t|t-1} H_t^T + B_t \end{aligned}$$

On calcule ensuite le gain :

$$G_t = \hat{P}_{t|t-1} H_t^T N_t^{-1}.$$

On peut enfin estimer l'état courant :

$$\begin{aligned} \hat{X}_{t|t} &= \hat{X}_{t|t-1} + G_t \nu_t \\ \hat{P}_{t|t} &= \hat{P}_{t|t-1} - G_t N_t G_t^T \end{aligned}$$

Les ARMA sont très utilisés dans le filtre de Kalman dans la mesure où ils peuvent approcher tout processus stationnaire en un cas (cf. sous-section 1.3.2). Tous les processus stochastiques ne s'implémentent cependant pas dans un filtre de Kalman, par exemple le bruit de scintillation.

Lisseur de Rauch-Tung-Striebel (RTS)

Ce lisseur [Rauch et al., 1965] est un algorithme qui s'effectue après le filtre de Kalman et qui l'améliore. Contrairement au filtre de Kalman, il estime chaque état conditionnellement à l'ensemble des observations considérées. Du fait qu'il est intimement lié au filtre de Kalman, il porte le nom de « lisseur de Kalman ». En effet, il réutilise les états estimés par filtre de Kalman $\hat{X}_{t|t-1}$ et $\hat{X}_{t|t}$ et les matrices de variance-covariance $P_{t|t-1}$ et $P_{t|t}$.

Après avoir fait un filtrage, on applique une formule de récurrence à l'envers :

$$\tilde{K}_t = \hat{P}_{t|t} L_{t+1}^T \hat{P}_{t+1|t}$$

Les états lissés sont calculés par récurrence :

$$\hat{X}_{t|n} = \hat{X}_{t|t} + \tilde{K}_t (\hat{X}_{t+1|n} - \hat{X}_{t+1|t})$$

$$\hat{P}_{t|n} = \hat{P}_{t|t} + \tilde{K}_t (\hat{P}_{t+1|n} - \hat{P}_{t+1|t})$$

Il existe d'autres lisseurs similaires comme le lisseur de Bryson-Frazier modifié qui a l'avantage de ne pas nécessiter l'inversion de matrices ce qui permet d'améliorer l'efficacité de calcul surtout en grande dimension.

1.3.6 Choix des valeurs a priori

Il est nécessaire d'initialiser les valeurs et les variances des états via l'attribution d'une valeur a priori $\hat{X}_{0|0}$ et d'une variance a priori $P_{0|0}$. La valeur a priori représente la prédiction de l'expert et la variance a priori représente la confiance qu'on accorde à l'expert. S'il n'y a aucune idée de l'incertitude sur la valeur a priori, on utilise des variances a priori grandes. [Herring et al., 1990] propose une formule justifiant cela et suggère, de la même façon de considérer des variances a priori 1000 fois supérieures aux variances réelles.

1.4 Statistiques spatiales

Cette section sur les statistiques spatiales tentera d'approcher comment nous pouvons modéliser rigoureusement les covariances de champs spatiaux.

1.4.1 Stationnarité et isotropie

On considère une variable aléatoire qui correspond à une fonction de $D \subset \mathbb{R}^d \rightarrow \mathbb{R}$, cette variable aléatoire est appelée champ. La difficulté des statistiques spatiales est qu'une seule réalisation de la variable aléatoire est considérée : par exemple, il n'y a qu'une seule réalisation de la variable aléatoire qui indique la position à une date donnée et pour une station donnée, il n'est pas possible de recommencer l'expérience. Il est donc nécessaire de considérer des hypothèses sur les variables aléatoires. Nous faisons l'hypothèse de la stationnarité d'ordre 2. Un champ aléatoire Z est stationnaire d'ordre 2 si, pour tout $x \in D$, la variance de $Z(x)$ existe et est finie : Z possède alors forcément une espérance. Dans ce cas, nous définissons $C(\cdot, \cdot)$ la fonction de variance-covariance :

$$C(s_1, s_2) = \text{Cov}(Z(s_1), Z(s_2)).$$

Dans le cas d'un champ d'ordre 2, on dit que Z est stationnaire si et seulement si son espérance $\mathbb{E}[Z(x)]$ est constante sur D et $C(s_1, s_2)$ ne dépend que de $h = s_1 - s_2$. On définit une nouvelle fonction de variance covariance :

$$C(h) = \text{Cov}(Z(s), Z(s+h))$$

où $h \in \mathbb{R}^d$. Toujours dans le cas de champ d'ordre 2, on parle de champ isotrope lorsque son espérance $\mathbb{E}[Z(x)]$ est constante sur D et la fonction $C(s_1, s_2)$ ne dépend que de la distance $r = \|s_1 - s_2\|$, on définit de même une nouvelle fonction de covariance :

$$C(r) = \text{Cov}(Z(s), Z(s+h))$$

où $r = \|h\|$.

1.4.2 Fonction de covariance semi-définie positive

En utilisant la positivité de la variance de $\sum_{i=1}^n \nu_i Z(s_i)$, on montre le caractère semi-défini positif de la fonction de covariance c'est-à-dire, pour tout naturel non nul n , tous $s_1, \dots, s_n \in \mathbb{R}^d$ et tous $\nu_1, \dots, \nu_n \in \mathbb{R}^d$, on a :

$$\sum_{i=1}^n \sum_{j=1}^n \nu_i \nu_j C(s_i, s_j) \geq 0. \quad (1.5)$$

Par abus de langage, on dit parfois qu'une telle covariance est définie positive. Nous avons montré que la condition (1.5) est une condition nécessaire. Dans le cas gaussien, c'est également une condition suffisante.

Cette condition peut s'écrire de manière matricielle. Nous considérons la matrice Γ carrée de taille n de terme général $\gamma_{i,j} = C(s_i, s_j)$. L'équation (1.5) peut s'écrire sous forme matricielle :

$$\vec{\nu}^T \Gamma \vec{\nu} \geq 0,$$

pour tout vecteur $\vec{\nu} = (\nu_1, \dots, \nu_n)^T$.

Si C_1 et C_2 sont des fonctions semi-définies positives, alors, pour tous $a_1, a_2 \geq 0$, la fonction $a_1 C_1 + a_2 C_2$ est semi-définie positive et le produit de fonctions $C_1 C_2$ est une fonction définie semi-positive.

Si $(C_i)_{i=0}^\infty$ est une suite de fonctions semi-définies positives et si, pour tous x et y , $\lim_{n \rightarrow \infty} C_n(x, y)$ existe et vaut $C(x, y)$ alors $C(\cdot, \cdot)$ est une fonction semi-définie positive.

Il est possible de trouver des covariances spatiotemporelles qui dépendent du décalage temporel τ et de la distance h . Il est possible de former une covariance spatiotemporelle sous une forme séparable $C_{ST}(h, \tau) = C_S(h)C_T(\tau)$ où C_S correspond à une fonction de covariance spatiale et C_T correspond à une fonction de covariance temporelle.

1.4.3 Exemples de fonctions de covariance

Nous nous plaçons sous hypothèse d'isotropie. Nous allons donner quelques exemples de fonctions de covariance semi-définies positives.

La fonction de covariance la plus simple est la fonction de Dirac qui correspond à l'effet pépité c'est-à-dire une absence de corrélation spatiale. Elle peut

correspondre à une variabilité intrinsèque du phénomène ou à une erreur de mesure. Elle peut également être due à une micro-variabilité spatiale c'est-à-dire à une variabilité spatiale dont l'échelle est inférieure à l'échelle étudiée.

La covariance exponentielle est de la forme $C_0 \exp(-|h|/a)$ où h est la distance entre les deux points considérés. La covariance gaussienne est de la forme $C_0 \exp(-(h/a)^2)$. Une famille très générale est la classe de Matérn de la forme :

$$C(h) = \sigma^2 \frac{2^{1-\lambda}}{(\lambda-1)!} \left(\sqrt{2\lambda} \frac{h}{\rho} \right)^\lambda K_\lambda \left(\sqrt{2\lambda} \frac{h}{\rho} \right)$$

où K_λ est la fonction de Bessel modifiée de deuxième espèce et où λ est un paramètre de régularité strictement positif. De même, ρ doit être strictement positif. Elle correspond à la covariance exponentielle pour $\lambda = \frac{1}{2}$. Sa limite lorsque λ tend vers ∞ est la fonction de covariance gaussienne.

1.4.4 Variogramme

Le terme « variogramme » que nous utiliserons ci-dessous correspond à la notion de semi-variogramme. L'emploi de ce terme est un abus de langage couramment utilisé.

Définitions et quelques propriétés

Le variogramme d'un champ aléatoire stationnaire d'ordre 2 $Z(\cdot)$ défini sur \mathbb{R}^d est défini par la fonction qui à h positif ou nul associe :

$$\gamma(h) = \frac{1}{2} \mathbb{E} [(Z(x+h) - Z(x))^2] = \frac{1}{2} \mathbb{V} (Z(x+h) - Z(x)).$$

Cette notion peut être généralisée aux hypothèses de stationnarité intrinsèque ou de stationnarité des incréments. Nous considérons qu'un champ aléatoire Z sur \mathbb{R}^d vérifie les hypothèses de stationnarité intrinsèque ou stationnarité des incréments si et seulement s'il existe un vecteur β et une fonction $\gamma : \mathbb{R}^d \rightarrow \mathbb{R}^+$ tels que :

$$\mathbb{E}[Z(x) - Z(x+h)] = \langle \beta, h \rangle$$

et

$$\frac{1}{2} \mathbb{V}(Z(x+h) - Z(x)) = \gamma(h).$$

La fonction $\gamma(\cdot)$ définit alors le variogramme. Notons que la stationnarité d'ordre 2 implique la stationnarité des incréments. La stationnarité d'ordre 2 implique l'existence d'une espérance et d'une variance, mais ce n'est pas le cas pour la stationnarité des incréments. À toute fonction de covariance $C(\cdot)$, on peut associer un variogramme grâce à la relation $\gamma(h) = C(0) - C(h)$. L'obtention d'une covariance à partir d'un variogramme n'est pas toujours possible, ce n'est possible que dans le cas des processus stationnaires d'ordre 2.

Une condition suffisante pour qu'un champ aléatoire Z vérifiant l'hypothèse de stationnarité des incréments soit stationnaire d'ordre 2 est que le variogramme possède une limite $\gamma_\infty \in \mathbb{R}$ en l'infini :

$$\lim_{\|h\| \rightarrow \infty} \gamma(h) = \gamma_\infty.$$

Estimation du variogramme

Dans cette sous-section, nous nous limiterons au cas d'isotropie. L'estimation du variogramme s'effectue au moyen d'un variogramme empirique. Le variogramme empirique est une représentation graphique du variogramme à l'image de l'histogramme pour la fonction de densité.

Concrètement, nous avons des données $((s_1, Z(s_1)), \dots, (s_n, Z(s_n)))$. Par souci de simplicité, nous noterons $Z_i = Z(s_i)$. Notons $r_{i,j} = \|s_i - s_j\|$ la distance séparant s_i et s_j . Nous considérons des classes de distance $D_\ell = [r_\ell, r_{\ell+1}[$ où $r_0 = 0$. Nous considérons $r_\ell^* = \frac{r_\ell + r_{\ell+1}}{2}$. Notons :

$$\gamma_\ell^* = \frac{1}{2} \text{mean}_{\|s_i - s_j\| \in D_\ell} \{(Z_i - Z_j)^2\}.$$

Le variogramme expérimental (ou empirique) est l'ensemble des points $(r_\ell^*, \gamma_\ell^*)$. Il est ensuite choisi une famille de variogrammes admissibles (par exemple, l'ensemble des variogrammes exponentiels). On estime parmi cette famille de variogrammes le variogramme le plus proche du variogramme expérimental par exemple par maximum de vraisemblance.

Krigeage Même si cette technique n'est pas utilisée dans la thèse, les variogrammes sont souvent associés au krigeage, qui permet une interpolation spatiale (spatiale au sens d'espace mathématique) ; l'estimateur du krigeage correspond au meilleur estimateur linéaire non biaisé. Il est d'abord utilisé dans le domaine minier par Danie K. Krige, qui donne son nom à la méthode. Dans les années 1960, Georges Matheron fournit une formulation mathématique plus formelle de la méthode. Cette méthode est actuellement utilisée dans de très nombreux domaines (météorologie, gravimétrie...).

Chapitre 2

Systèmes de référence et ITRF

Tout positionnement sur un territoire nécessite un système de référence [Boucher and Altamimi, 2001]. Dans l'histoire des systèmes de référence, les États développent d'abord des systèmes de référence locaux constitués d'un repère de coordonnées géographiques bidimensionnelles et d'un système de référence de référence vertical associé. La triangulation de Cassini, triangulation le long de la méridienne de Paris qui traverse l'Observatoire de Paris, est une des premières ébauches de systèmes de référence bidimensionnels. La précédente triangulation se révélera insuffisante; ainsi la triangulation des Ingénieurs Géographes sera entreprise de 1817 à 1853. La nouvelle triangulation française entreprise de la fin du XIX^e siècle à 1991 permet la matérialisation de plus de 70000 points géodésiques répartis sur l'ensemble de la France [Levallois, 1989]. L'apparition des premiers ordinateurs permet un essor des repères de référence après la Seconde Guerre mondiale. Les Européens décident de lancer l'European Datum 1950 (ED50). En 1960, le premier grand repère de référence mondial, le World Geodetic System 1960 (WGS60), est estimé à partir de données de gravimétrie, d'astro-géodésie et des campagnes HIRAN et SHORAN. Dans les années 1960 se développent le SLR (satellite laser ranging), technique qui utilise le temps d'aller-retour d'impulsions lumineuses provenant d'un émetteur laser vers un satellite muni d'un réflecteur. D'autres techniques qui seront intégrées parmi les techniques de géodésie spatiale se développeront ensuite : VLBI (very long baseline interferometry), DORIS (détermination d'orbite et radiopositionnement par satellite), GPS (global positioning system). Ainsi, les repères de géodésie issus de la géodésie spatiale sont exprimés de manière tridimensionnelle contrairement aux repères classiques. Le WGS84, un successeur de WGS60, utilisera pour la première fois le GPS. Sont ensuite développées trois versions du repère de référence BTS (BIH¹ Terrestrial System) à partir de données VLBI, SLR, LLR (lunar laser ranging), DORIS, Doppler/TRANSIT (un embryon de GPS). Suivront les éditions successives de l'ITRF (international terrestrial reference frame) à partir de 1988 dont chaque version est une amélioration de la précédente. L'ITRF est constitué de coordonnées à un instant de référence et de vitesses pour un ensemble de stations. La dernière version de

1. Bureau international de l'heure

l'ITRF fournit des incertitudes généralement inférieures à 3 mm sur les positions et à 0.2 mm/an sur les vitesses.

Après avoir défini les notions de système et de repère de référence terrestre, nous considérerons plus en détail l'ITRS (**i**nternational **t**errestrial **r**eference system) et l'ITRF et nous décrirons le processus de calcul de l'ITRF et notamment celui de l'ITRF2014. Ma thèse vise l'amélioration de l'étape de cumul décrite ci-dessous, étape qui fait partie de ce processus. Comme ma thèse utilise les données GNSS, je décrirai ensuite la technique GNSS et mentionnerai les sources d'erreurs que l'on sait affecter les séries de positions GNSS.

2.1 Intérêts de l'ITRF

L'ITRF permet un positionnement global précis sur la surface de la Terre ou à proximité. Il est donc utilisé pour des applications civiles (agriculture de précision, navigation...) et scientifiques (mesure du mouvement des plaques tectoniques, de déformations du sol, d'orbites précises de satellites, de la montée du niveau des mers...).

Nous allons discuter de la montée du niveau des mers, car elle constitue l'application la plus exigeante en termes de précision.

Montée du niveau des mers

Une application particulièrement utile et importante d'un repère de référence global est l'évaluation précise de la montée du niveau des mers. Obtenir une bonne précision sur la montée du niveau des mers actuel nécessite une évaluation précise de la contribution de l'erreur due au système de référence [Collilieux and Wöppelmann, 2011, Morel and Willis, 2005]; le choix d'un mauvais repère de référence engendrerait une incertitude incompressible allant jusqu'à plusieurs millimètres par an sur l'estimation de la montée du niveau des mers dans certaines régions du globe. Une illustration de l'importance de l'ITRF est l'adoption de l'ITRF2005 en lieu et place d'un système cohérent avec l'ITRF2000 qui a engendré une modification de l'estimation de la montée du niveau des mers [Beckley et al., 2007].

La montée du niveau des mers, même de quelques millimètres par an, a des conséquences dramatiques. [Bellard et al., 2012] montre que la montée du niveau des mers, avec d'autres phénomènes du changement climatique, a des conséquences considérables sur la biodiversité, y compris au niveau génétique et physiologique. Elle a des conséquences au niveau des littoraux. Elle a aussi des conséquences sur les espèces animales comme les tortues de mer aux Caraïbes [Fish et al., 2005] (figure 2.1). L'île de Bhola au Bangladesh a été victime en 2005 d'inondations créant un demi-million de sans-abris. Elle peut déstabiliser par des déplacements de population un pays aussi puissant que les États-Unis, ses territoires littoraux et non littoraux [Hauer, 2017]. La France est un grand territoire maritime du fait de ses îles territoires outre-mer avec 1267 îles. 12% des îles françaises sont menacées de submersion. En Nouvelle-Calédonie, des espèces végétales sont menacées [Bellard et al., 2013].

Pour cette application, il convient de diminuer le bilan d'erreur de l'ITRF évalué à 0.5 mm/an en origine par composante [Collilieux et al., 2014] et d'atteindre 0.1 millimètre par an [Rothacher et al., 2008].



FIGURE 2.1 – Tortue marine menacée par la montée du niveau des mers

2.2 Systèmes de référence terrestre

2.2.1 Définition

Un système de référence terrestre correspond à une définition mathématique abstraite [Petit and Luzum, 2010]. Il est constitué :

- d'une origine, en général, proche du centre de la Terre,
- d'une échelle, en général, proche du mètre SI,
- d'un trièdre orthonormal direct en corotation avec la Terre,
- l'axe des cotes est proche de l'axe de rotation de la Terre,
- les axes des abscisses et des ordonnées sont proches du plan équatorial.

2.2.2 Relation entre les systèmes de référence

En considérant deux systèmes de référence S_1 et S_2 et un point quelconque M de coordonnées \vec{X}_1 dans S_1 et \vec{X}_2 dans S_2 , on peut passer d'une coordonnée à l'autre par la formule :

$$\vec{X}_2 = \vec{T} + (1 + \lambda)\underline{R}\vec{X}_1 \quad (2.1)$$

où :

- \vec{T} est un vecteur de translation entre S_1 et S_2 ,
- $1 + \lambda$ correspond au rapport d'homothétie pour passer de S_1 à S_2 ,
- \underline{R} est la matrice de rotation entre les deux systèmes.

La formule (2.1) est générale entre toute paire de systèmes de référence. Il est fréquemment utilisé des approximations par linéarisation (il est considéré

que λ et les angles de rotation de \underline{R} sont petits) :

$$\vec{X}_2 = \vec{X}_1 + \vec{T} + \lambda \vec{X}_1 + \underline{\tilde{R}} \vec{X}_1, \quad (2.2)$$

où $\underline{\tilde{R}} = \underline{R} - Id$.

Il peut être considéré une relation d'évolution temporelle entre deux systèmes de référence après suppression des termes négligeables :

$$\dot{\vec{X}}_2 = \dot{\vec{X}}_1 + \dot{\vec{T}} + \lambda \dot{\vec{X}}_1 + \underline{\dot{R}} \vec{X}_1,$$

où pour une quantité y dépendante du temps, \dot{y} est la dérivée de y par rapport au temps.

2.3 International Terrestrial Reference System (ITRS)

L'ITRS est un système de référence terrestre, il est défini par [McCarthy, 1996] :

- son origine, qui est le centre des masses de la Terre entière avec les océans et l'atmosphère,
- son unité, qui est le mètre SI,
- son orientation, qui est celle du Bureau international de l'heure (BIH) à l'époque 1984.0,
- l'évolution de son orientation, imposée par la condition dite de non-rotation par rapport à la croûte terrestre.

2.4 Repère de référence terrestre

2.4.1 Définition

Une réalisation d'un système de référence définit un repère de référence qui est constitué par un ensemble de coordonnées conventionnelles de stations au cours du temps pouvant inclure des vitesses de déplacement et éventuellement une incertitude sous forme de matrice de variance-covariance. Ces jeux de coordonnées de stations associées à des points physiques sont déterminés à l'aide d'instruments géodésiques rattachés à la croûte.

2.4.2 International Terrestrial Reference Frame (ITRF)

L'ITRF est une réalisation de l'ITRS. La dernière version de l'ITRF à ce jour s'appelle l'ITRF2014 [Altamimi et al., 2016], c'est la treizième version de l'ITRF. Il succède à une longue série de repères de référence s'améliorant au fur et à mesure des progrès effectués [Altamimi et al., 2002, Altamimi et al., 2007, Altamimi et al., 2011] ; par exemple, des modèles paramétriques décrivant les mouvements de relaxation post-sismique de certaines stations ont été fournis pour la première fois avec l'ITRF2014 (nous reviendrons sur cette notion à la sous-section 3.1.1) ; le calcul de l'ITRF2014 a également tenu compte de la présence de signaux saisonniers dans les séries temporelles de positions de

stations, bien que les signaux estimés n'ont pas été inclus dans les coordonnées de référence fournies aux utilisateurs.

La table 2.2 présente les coordonnées ITRF2014 de quelques stations GNSS. Elle contient les positions à une époque de référence (ici 2010.0) et les vitesses, ainsi que leurs incertitudes sous forme d'erreurs formelles (**Sigmas**). Pour tenir compte des discontinuités présentes dans les séries temporelles de certaines stations, il peut être utile de définir plusieurs positions de référence sur des plages temporelles appelées SOLN débutant à DATA_START et finissant à DATA_END : les SOLN sont délimités par les discontinuités de position et de vitesse.

DOMES NB.	SITE NAME	TECH. ID.	X/Vx	Y/Vy	Z/Vz	Sigmas			SOLN	DATA_START	DATA_END
			-----m/m/y-----								
10001S006	Paris	GNSS OPMT	4202777.3053	171368.0862	4778660.2517	0.0006	0.0006	0.0005			
10001S006			-.01305	0.01758	0.01031	.00003	.00004	.00004			
10002M006	Grasse (OCA)	GNSS GRAS	4581690.8267	556114.9242	4389360.8453	0.0006	0.0006	0.0006	1	00:000:00000	96:277:00000
10002M006			-.01377	0.01886	0.01142	.00003	.00004	.00003			
10002M006	Grasse (OCA)	GNSS GRAS	4581690.8289	556114.9237	4389360.8441	0.0006	0.0006	0.0005	2	96:277:00000	03:113:00000
10002M006			-.01377	0.01886	0.01142	.00003	.00004	.00003			
10002M006	Grasse (OCA)	GNSS GRAS	4581690.8331	556114.9264	4389360.8486	0.0006	0.0006	0.0005	3	03:113:00000	04:295:32400
10002M006			-.01377	0.01887	0.01142	.00003	.00004	.00003			
10002M006	Grasse (OCA)	GNSS GRAS	4581690.8306	556114.9301	4389360.8511	0.0006	0.0006	0.0005	4	04:295:32400	00:000:00000
10002M006			-.01377	0.01887	0.01142	.00003	.00004	.00003			

FIGURE 2.2 – Début d'une table de l'ITRF2014 fournissant les positions à l'époque 2010 et les vitesses.

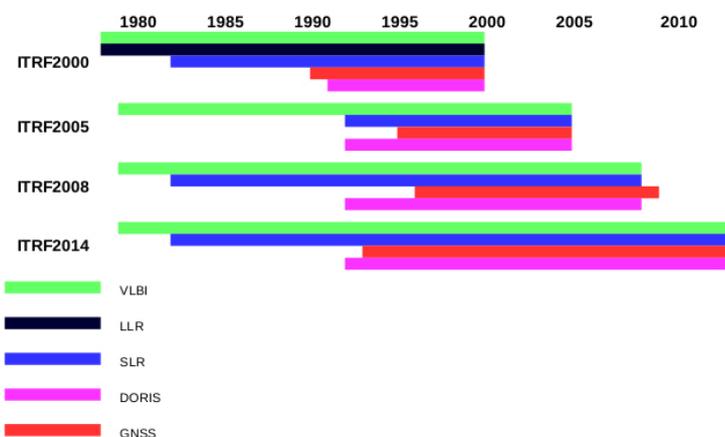


FIGURE 2.3 – Plages temporelles des données des différentes techniques de géodésie spatiale utilisées dans les calculs des versions de l'ITRF depuis l'ITRF2000. Certaines stations peuvent avoir des plages de données plus courtes [Collilieux et al., 2017].

L'ITRF est obtenu par une série de mesures de quatre techniques de géodésie spatiale :

- l'interférométrie à très longue base (VLBI)
- la télémétrie laser sur satellites (SLR)
- les systèmes de positionnement par satellites (GNSS)
- la détermination d'orbite et radiopositionnement intégré par satellite (DORIS)

La figure 2.3 présente les données des différentes techniques de géodésie spatiale utilisées dans le calcul des différentes versions de l'ITRF. Notons que le volume de données disponibles augmente pour toutes les techniques de géodésie spatiale au fil des versions successives de l'ITRF.

Nous détaillerons ces quatre techniques dans la sous-section 2.5.3.

2.5 Processus de calcul de l'ITRF

2.5.1 Présentation des quatre techniques de géodésie spatiale

En géodésie, on distingue les observations qui sont fournies directement par les appareils de mesure des solutions, qui sont issues d'un calcul et d'une modélisation à partir des observations.

Le VLBI est une technique d'interférométrie utilisant les signaux reçus de sources astronomiques telles que les quasars en mesurant la différence de temps d'arrivée en différents sites. La technique permet le positionnement des radiotélescopes avec une précision millimétrique, mais permet également la détermination des coordonnées des radiosources du repère international de référence céleste (ICRF). Des données sont disponibles depuis le début des années 1980. C'est la seule technique qui permet la mesure de l'orientation de la Terre dans

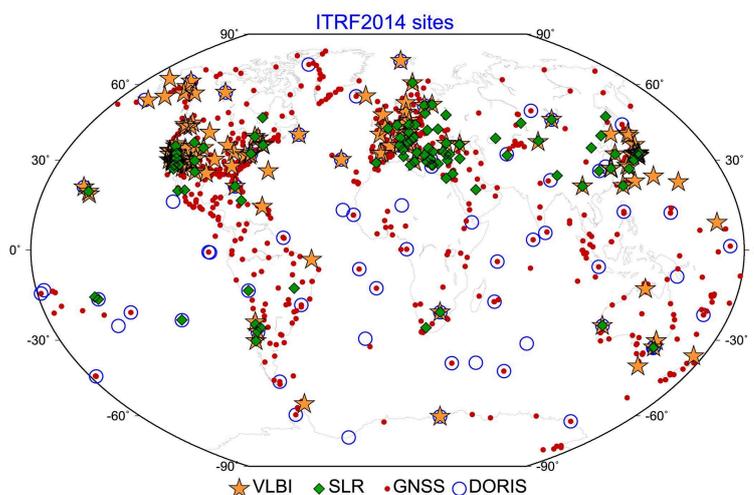


FIGURE 2.4 – Cartes des stations utilisées pour les quatre techniques utilisées par l'ITRF2014 [Altamimi et al., 2016]

l'ICRF de manière précise. L'ITRF utilise des solutions journalières fournies par l'International VLBI Service (IVS) [Schuh and Behrend, 2012]. Le VLBI contribue à définir l'échelle de l'ITRF. La figure 2.4 présente le réseau de stations VLBI avec des étoiles jaunes. La volatilité (ou répétabilité) des séries temporelles de positions VLBI vaut encore 4 mm pour les composantes horizontales et de l'ordre du centimètre pour la composante verticale.

Le SLR (respectivement le LLR) détermine la distance entre un point d'un satellite (respectivement de la Lune) et la station à partir de la mesure du temps de trajet aller-retour d'une impulsion laser. L'ITRF2014 utilise des solutions bimensuelles jusqu'en 1993, hebdomadaires au-delà, fournies par l'International Laser Ranging Service (ILRS) [Pearlman et al., 2002] dont certaines remontent au début des années 1980. Les stations SLR sont présentées avec des losanges verts sur la figure 2.4. Cette technique a une importance particulière, car elle est la seule à définir l'origine de l'ITRF2014, mais contribue aussi à définir l'échelle de l'ITRF2014. La volatilité des séries temporelles de positions SLR est de l'ordre du centimètre, quelle que soit la composante.

DORIS est un système de mesure où des balises au sol envoient un signal ce qui permet d'avoir des informations sur la vitesse du satellite récepteur via l'effet Doppler et d'en déduire la position de la balise. L'ITRF2014 utilise des solutions hebdomadaires fournies par l'International DORIS Service (IDS) [Willis et al., 2010]; les plus anciennes remontent à 1993. Les stations DORIS sont représentées par des cercles bleus sur la figure 2.4. La volatilité des séries temporelles de positions DORIS est un peu supérieure au centimètre, quelle que soit la composante.

Dans tous les GNSS, des satellites sous forme de constellations émettent des signaux en direction de la Terre recueillis par des récepteurs permanents. La figure 2.5 présente une station permanente installée à Grasse (France), qui recueille des données depuis 1995. L'ITRF2014 utilise des solutions journalières fournies par l'IGS (International GNSS Service) [Dow et al., 2009]. Ces solutions incluent 1848 stations permanentes GNSS à travers le monde. La carte du réseau



FIGURE 2.5 – Photo de la station permanente GRAS (Grasse, France)

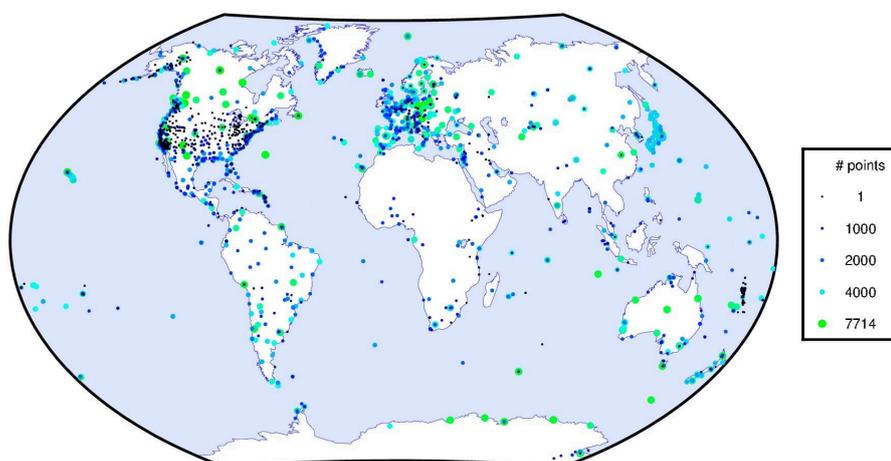


FIGURE 2.6 – Carte des stations GNSS incluses dans la contribution de l'IGS à l'ITRF2014 avec la longueur des séries de positions journalières correspondantes. La figure est extraite de [Rebischung et al., 2016]

de stations est présentée sur la figure 2.6. Nous constatons que la plus forte densité de stations se trouve aux États-Unis, au Japon et en Europe. Nous constatons que les stations dont les séries temporelles de positions sont les plus longues se trouvent principalement en Europe et au Japon.

2.5.2 Vision globale du processus d'obtention de l'ITRF

Le calcul de l'ITRF2014 suit un processus très précis :

- pour chacune des techniques, les centres d'analyse traitent les observations de la technique et fournissent des solutions, notamment les positions des stations (de manière rapide, une solution consiste en la donnée de paramètres de linéarisation, de paramètres de régression estimés par moindres carrés, de l'incertitude sur ceux-ci sous forme généralement d'une matrice de variance-covariance et éventuellement des contraintes

- appliquées).
- la **combinaison intra-technique** utilise les solutions de tous les centres d’analyse et fournit une solution moyenne par technique²,
 - le **cumul intra-technique** permet d’obtenir une solution long terme par technique,
 - la **combinaison multi-technique** utilisant les rattachements locaux permet d’obtenir la solution définitive de l’ITRF.

2.5.3 Obtention des solutions de géodésie spatiale

Le travail présenté dans cette partie est effectué par les centres d’analyse, institutions qui effectuent de manière routinière des calculs de géodésie spatiale. Ils fournissent des coordonnées de stations et leur matrice de variance-covariance associée à partir des observations de chaque technique indépendamment. Pour une technique donnée, il peut y avoir plusieurs centres d’analyse. Chaque centre d’analyse a une certaine liberté pour effectuer ses calculs ; néanmoins, des normes minimales peuvent être prévues [Luceri and Pavlis, 2016].

Considérons un ensemble d’observations \vec{v} pour une technique donnée. Grâce à une estimation par moindres carrés, on cherche un ensemble de paramètres $\vec{\theta}$. Ce vecteur $\vec{\theta}$ varie selon la technique et il contient [Rebischung et al., 2014] :

- les coordonnées des stations,
- les paramètres décrivant les orbites des satellites sauf pour le VLBI,
- les coordonnées des quasars observés pour le VLBI,
- une partie ou l’ensemble des paramètres d’orientation de la Terre,
- des paramètres de correction pour la propagation des signaux,
- des biais spécifiques à la technique utilisée.

Quitte à utiliser des linéarisations, on obtient, via les moindres carrés, l’équation normale :

$$\vec{v} = \underline{N}(\vec{\theta} - \vec{\theta}_0)$$

où \underline{N} est la matrice normale et $\vec{\theta}_0$ est un paramètre a priori pour $\vec{\theta}$ ($\vec{\theta}_0$ correspond à un point de linéarisation). Lorsque l’on souhaite ne garder qu’une partie des paramètres, on peut adapter cette équation pour obtenir une équation réduite. Le centre de produit ITRS et les centres de combinaison de l’IERS n’ont besoin que des coordonnées des stations et des paramètres d’orientation de la Terre.

Pour toutes les techniques, la matrice \underline{N} est non inversible, car aucune des quatre techniques ne peut déterminer simultanément les coordonnées des stations et les paramètres d’orientation de la Terre.

En outre, la technique VLBI est insensible à l’origine du repère terrestre [Ma et al., 2013, Rebischung et al., 2014]. En effet, les sources astronomiques utilisées dans le VLBI étant très éloignées, aucune translation de coordonnées ne fait varier les différences de temps d’arrivée des signaux des sources astronomiques au niveau des stations.

D’autre part, même si l’on fixe l’orientation du réseau de stations, les trois autres techniques (GNSS, SLR, DORIS) sont insensibles à $DUT1$, l’angle orientant la position de l’axe des X de l’ITRF sur le plan de l’équateur du repère céleste intermédiaire dans sa rotation diurne. Ceci s’explique par le fait que toute variation de $DUT1$ peut être compensée par une rotation de la constellation de satellites de telle sorte que les observations restent inchangées.

2. Cela correspond à une agrégation d’estimateur.

Néanmoins, les centres d'analyse doivent fournir une solution $\hat{\theta}$ déterminée; ils utilisent donc des contraintes. Il existe différents types de contraintes (contraintes minimales, lâches...). Ce sont ces contraintes qui définissent dans quel repère de référence sont exprimées les coordonnées. Néanmoins, toutes les contraintes utilisées par les centres d'analyse aboutissent à une équation normale de la forme :

$$(\underline{N} + \underline{N}_c)(\vec{\theta} - \vec{\theta}_0) = \vec{v}.$$

Les centres d'analyse choisissent \underline{N}_c de telle sorte que $\underline{N} + \underline{N}_c$ soit inversible. On en déduit la solution :

$$\hat{\theta} = \vec{\theta}_0 + (\underline{N} + \underline{N}_c)^{-1}\vec{v}.$$

Il en découle la matrice de variance-covariance associée à l'estimateur $\hat{\theta}$:

$$\hat{Q} = \sigma_0^2(\underline{N} + \underline{N}_c)^{-1},$$

où σ_0^2 est le facteur unitaire de variance.

Les centres d'analyse fournissent, journalièrement ou hebdomadairement selon la technique, soit les équations normales non contraintes sous la forme de la donnée de $\vec{\theta}_0$, \underline{N} et \vec{v} soit les solutions contraintes ($\vec{\theta}_0$, $\hat{\theta}$, \underline{N}_c et \underline{Q}).

Ces résultats sont présentés sous le format SINEX [IERS, 2006] pour la transmission de fichiers.

2.5.4 Combinaison intra-technique

Les centres d'analyse fournissent différentes solutions pour une technique donnée à une même date. Les centres de combinaisons utilisent des estimations $(\hat{\theta}_i, \hat{Q})_{i=1}^s$ de s centres d'analyse pour en déduire une estimation finale $(\hat{\theta}, \hat{Q})$. Par exemple, le centre de combinaison de l'IGS se situe à l'IGN (institut national de l'information géographique et forestière) depuis 2010. Il existe deux approches pour effectuer des combinaisons intra-techniques : le cumul d'équations normales ou la combinaison au niveau des solutions. La seconde solution est présentée ici. Les paramètres sont estimés par moindres carrés fondés sur les équations suivantes [Altamimi et al., 2011] :

$$X_i = X_c + t_i^X + d_i X_c - r_i^Z Y_c + r_i^Y Z_c,$$

$$Y_i = Y_c + t_i^Y + d_i Y_c + r_i^Z X_c - r_i^X Z_c,$$

$$Z_i = Z_c + t_i^Z + d_i Z_c - r_i^Y X_c + r_i^X Y_c,$$

$$x_p^i = x_p^c - r_i^Y$$

$$y_p^i = y_p^c + r_i^X$$

$$DUT1_i = DUT1_c - r_i^Z / f$$

$$\dot{x}_p^i = \dot{x}_p^c$$

$$\dot{y}_p^i = \dot{y}_p^c$$

$$LOD_i = LOD_c$$

où

- X_i, Y_i et Z_i sont les coordonnées de la station fournies par la solution du centre d'analyse i ;
- X_c, Y_c et Z_c sont les coordonnées combinées de la station qui sont une partie des sorties du processus de combinaison ;
- $x_p^i, y_p^i, \dot{x}_p^i, \dot{y}_p^i, DUT1_i, LOD_i$ (respectivement, $x_p^c, y_p^c, \dot{x}_p^c, \dot{y}_p^c, DUT1_c, LOD_c$) sont les paramètres d'orientation de la Terre fournis par la solution du centre d'analyse i (respectivement issus de la combinaison) ;
- $t_i^X, t_i^Y, t_i^Z, d_i, r_i^X, r_i^Y, r_i^Z$ sont les paramètres de transformation entre le repère fourni par le centre d'analyse i et le repère de sortie combiné ;
- $f = 1.00273781191135448$ est le taux d'avancement conventionnel de l'angle de rotation de la Terre par rapport à UT1.

On peut mettre ces équations sous une forme matricielle :

$$\begin{pmatrix} \vec{x}_1 \\ \vdots \\ \vec{x}_s \end{pmatrix} = \begin{pmatrix} \underline{J}_1 & \underline{A}_1 & & \\ & \vdots & \ddots & \\ & & & \underline{A}_s \end{pmatrix} \begin{pmatrix} \vec{x}_c \\ \vec{\theta}_1 \\ \vdots \\ \vec{\theta}_s \end{pmatrix}$$

où :

- $\underline{J}_i = \partial \vec{x}_c / \partial \vec{x}_i$ est une matrice composée de zéros et des uns qui permet d'associer tout paramètre issu des solutions fournies par les centres d'analyse au paramètre combiné correspondant ;
- $\vec{\theta}_i$ est le vecteur des sept paramètres de transformation entre la solution i et la solution combinée ;
- $\underline{A}_i = \partial \vec{x}_i / \partial \vec{\theta}_i$.

Nous allons donner une forme détaillée à \underline{A}_i . Pour cela, nous considérons que $\vec{x}_i = (X_1, Y_1, Z_1, \dots, X_k, Y_k, Z_k, x_p, y_p, DUT1, \dot{x}_p, \dot{y}_p, LOD)$ et $\theta_i = (t_i^X, t_i^Y, t_i^Z, d_i, r_i^X, r_i^Y, r_i^Z)$ alors on peut écrire \underline{A}_i :

$$\begin{pmatrix} 1 & 0 & 0 & X_1 & 0 & Z_1 & -Y_1 \\ 0 & 1 & 0 & Y_1 & -Z_1 & 0 & X_1 \\ 0 & 0 & 1 & Z_1 & Y_1 & -X_1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & X_k & 0 & Z_k & -Y_k \\ 0 & 1 & 0 & Y_k & -Z_k & 0 & X_k \\ 0 & 0 & 1 & Z_k & Y_k & -X_k & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1/f \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (2.3)$$

Dans la résolution des moindres carrés, les vecteurs $(\vec{x}_i)_{i=1}^s$ sont supposés statistiquement indépendants, bien que ce ne soit pas vrai. Ainsi la matrice de covariance est-elle diagonale par bloc $\lambda_i Q_i$. Notons qu'on utilise un système itératif pour détecter et supprimer les observations fausses et déterminer des facteurs de variance optimaux $\hat{\lambda}_i$.

2.5.5 Cumul long-terme (pour une technique)

L'objectif de cette étape est de cumuler les séries temporelles issues de la combinaison intra-technique pour fournir des positions et des vitesses pour chaque technique. Cette étape est effectuée par le centre de produit ITRS.

Pour un certain nombre d'instants t_i , nous avons les solutions de la combinaison intra-technique sous la forme de couples $(\vec{x}_i, \underline{Q}_i)_{i=1}^s$. On considère une époque de référence t^{ref} . Sont utilisées les équations [Altamimi et al., 2007] :

$$\begin{aligned} X_i &= X_c + (t_i - t^{ref})\dot{X}_c + t_i^X + d_i X_c - r_i^Z Y_c + r_i^Y Z_c, \\ Y_i &= Y_c + (t_i - t^{ref})\dot{Y}_c + t_i^Y + d_i Y_c + r_i^Z X_c - r_i^X Z_c, \\ Z_i &= Z_c + (t_i - t^{ref})\dot{Z}_c + t_i^Z + d_i Z_c - r_i^Y X_c + r_i^X Y_c. \end{aligned}$$

où :

- X_i, Y_i, Z_i sont les positions de stations issues de la combinaison intra-technique à l'instant t_i ;
- X_c, Y_c et Z_c correspondent à des positions à t^{ref} qui seront issues du cumul ;
- \dot{X}_c, \dot{Y}_c et \dot{Z}_c correspondent à des vitesses issues du cumul ;
- $t_i^X, t_i^Y, t_i^Z, d_i, r_i^X, r_i^Y$ et r_i^Z correspondent aux sept paramètres de transformation entre le repère de la solution i fournie par la combinaison intra-technique à l'instant t_i et le repère issu du cumul.

Afin de ne pas alourdir les notations, nous avons omis les termes saisonniers (annuels et semi-annuels) ainsi que les paramètres d'orientation de la Terre et les discontinuités ; ils sont néanmoins présents dans le modèle utilisé dans l'ITRF2014.

Les discontinuités de position dans les séries temporelles de positions GNSS sont dues aux changements d'équipement et aux tremblements de terre. Les discontinuités de vitesse peuvent être utilisées pour la modélisation de phénomènes non linéaires tels la fonte des glaces actuelle.

Après un tremblement de terre, la structure externe de la Terre se réajuste de manière viscoélastique ; on parle alors de mouvements post-sismiques. Une carte des stations dont les coordonnées ITRF2014 incluent une modélisation des déformations post-sismiques est présentée sur la figure 2.7 (carte extraite de [Altamimi et al., 2016]) ; ces stations sont toutes dans des zones de forte sismicité. Ces déplacements post-sismiques ont été modélisés au préalable des cumuls long terme effectués pour l'ITRF2014 par des fonctions logarithmiques ou exponentielles (cf. sous-section 3.1.1). Ces modèles ont été ensuite retirés des séries temporelles de positions de stations en entrée des cumuls par commodité et pour garder un modèle de régression linéaire.

Les équations ci-dessus peuvent se mettre sous la forme matricielle :

$$\begin{pmatrix} \vec{x}_1 \\ \vdots \\ \vec{x}_s \end{pmatrix} = \begin{pmatrix} \underline{J}_1 & (t_i - t^{ref})\underline{J}_1 & \underline{A}_1 & & \\ \vdots & \vdots & & \ddots & \\ \underline{J}_s & (t_i - t^{ref})\underline{J}_s & & & \underline{A}_s \end{pmatrix} \begin{pmatrix} \vec{x}_c \\ \vec{x}_c \\ \vec{\theta}_1 \\ \vdots \\ \vec{\theta}_s \end{pmatrix}$$

où :

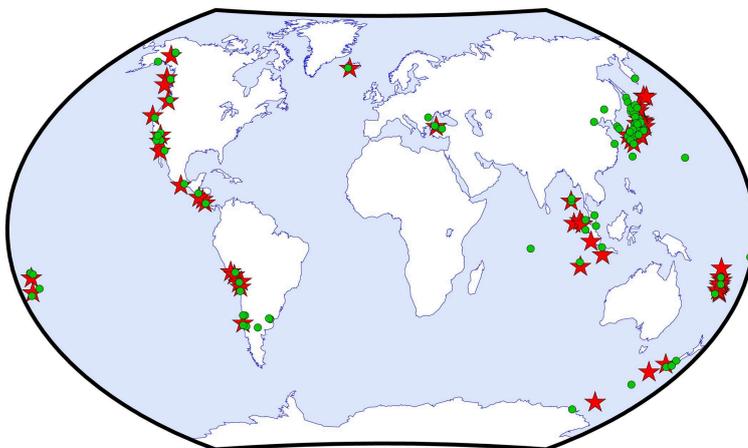


FIGURE 2.7 – Les étoiles rouges représentent les épices des séismes et les ronds verts les stations pour lesquelles a été intégré un modèle de déformation post-sismique. Cette carte est extraite de [Altamimi et al., 2016].

- $\vec{x}_1, \dots, \vec{x}_s$ sont les positions à différents instants ;
- \vec{x}_c et \vec{x}_c sont les positions à t^{ref} et les vitesses que donnera le cumul ;
- $\underline{J}_i = \partial \vec{x}_i / \partial \vec{x}_c$ est une matrice constituée de zéros et de uns qui permet d'associer à tout élément de \vec{x}_i l'élément correspondant de \vec{x}_c ;
- $\vec{\theta}_i$ correspondent aux sept paramètres de transformation entre le repère i issu de la combinaison intra-technique à l'instant t_i et le repère cumulé ;
- $\underline{A}_i = \partial \vec{x}_i / \partial \vec{\theta}_i$ (elle a une forme analogue à l'équation (2.3) si l'on supprime les lignes correspondant aux paramètres d'orientation de la Terre).

On utilise une matrice de variance-covariance diagonale par bloc de blocs diagonaux $\underline{Q}_1, \dots, \underline{Q}_s$. Il n'est ajouté aucun facteur d'échelle. Cette équation matricielle peut être résolue par moindres carrés.

Notons qu'en fonction des paramètres de transformation considérés entre les repères combinés et le repère cumulé, il peut y avoir plusieurs singularités. Il peut exister jusqu'à 14 singularités qui correspondent à la définition du repère de référence de sortie.

Face à ces singularités, il y a deux solutions possibles, par exemple pour les translations [Altamimi et al., 2007] :

- on peut imposer des conditions de non-translation et une condition similaire pour la dérivée des paramètres de translation entre le repère cumulé et un repère externe (par exemple, dans le cas de l'ITRF, la version précédente de l'ITRF) ;
- on peut imposer des contraintes internes c'est-à-dire imposer une translation nulle et une dérive nulle entre les repères individuels et le repère cumulé :

$$\sum_{i=1}^s t_i^X = 0$$

$$\sum_{i=1}^s (t_i - t^{ref}) t_i^X = 0$$

2.5.6 Combinaison inter-technique

La combinaison inter-technique vise à unifier les quatre repères long terme issus des cumuls de chaque technique qui sont fournis dans des repères de référence propres à chaque technique. Les vitesses et positions estimées grâce aux cumuls intratechnique sont combinées pour obtenir des positions et vitesses définitives pour l'ITRF dans un repère unique bien défini. Entre les repères intratechniques et le repère définitif de l'ITRF, des paramètres de transformations constitués de 3 paramètres de translation, un facteur d'échelle et 3 angles de rotation sont calculés. Il en est de même pour lier les vitesses, nous arrivons donc à un total de 14 paramètres par technique.

Les sites co-localisés, qui abritent plusieurs techniques à un même endroit, permettent d'avoir de l'information supplémentaire qu'on ajoute à la combinaison inter-technique : notamment, on suppose que deux stations colocalisées ont la même vitesse. On dispose également de la position relative des instruments déterminés durant des campagnes de mesures topométriques appelées rattachements. Une limite actuelle de l'ITRF est le nombre trop faible de sites colocalisés [Altamimi et al., 2011], bien que, entre l'ITRF2008 et l'ITRF2014, le nombre de sites colocalisés ait augmenté. De plus, l'ITRF est confronté à la discordance entre les rattachements issus de techniques topométriques et les mesures issues de la géodésie spatiale.

2.6 GNSS et ITRF

Nous traiterons les séries temporelles de positions GNSS, ainsi m'a-t-il paru important de présenter plus en détail (sous-section 2.6.1) et de lister les sources d'erreurs que l'on sait affecter les séries de positions GNSS (sous-sections 2.6.2 et 2.6.3)

2.6.1 Présentation des systèmes GNSS actuels

Beaucoup d'informations reportées ici sur les GNSS sont tirées de [Bossler, 2017].

Il existe plusieurs GNSS ; le plus connu est le GPS (global positioning system) d'origine américaine. Il existe d'autres systèmes de positionnement global par satellites, souvent déployés par volonté d'indépendance par rapport aux États-Unis³ ; la Russie déploie GLONASS, la Chine déploie Compass ou Beidou-2, l'Europe déploie Galileo. Les systèmes précités permettent le positionnement précis en tout point de la Terre. Il existe aussi des systèmes de GNSS à dimension régionale comme QZSS (quasi-zenith satellite system) pour le Japon. Notons que Beidou-1, le GNSS chinois, a étendu sa couverture initialement régionale à une couverture globale pour devenir Beidou-2. Ces systèmes sont passifs c'est-à-dire que l'utilisateur ne fait que recevoir des données sans jamais en transmettre. Les satellites de chaque système se trouvent sur des orbites différentes et les observables peuvent varier d'un système à l'autre si bien que certaines erreurs systématiques apparaissent en fonction de la constellation utilisée, comme nous

3. Cela est totalement justifié : en 1999, lors du conflit de Kargil qui opposa l'Inde au Pakistan, les États-Unis ont empêché l'accès à la technologie GPS aux Indiens.

le verrons ci-dessous. GPS et GLONASS sont décrits plus particulièrement, car le calcul de l'ITRF2014 inclut des données de ces deux constellations.

Le GPS

Le GPS est un système de navigation par satellites américain ; son entretien est réalisé par l'US Air Force. Ce projet a été lancé en 1973 initialement à finalité militaire ; les premières applications civiles commenceront dans les années 1980 avec l'accord des États-Unis. Il est complètement opérationnel depuis 1995.

Les satellites émettent au moins sur deux bandes de fréquences (L1 et L2) et pour les satellites les plus récents sur la bande de fréquence L5⁴. La combinaison des fréquences L1 et L2 permet d'éliminer l'erreur due à l'ionosphère au premier ordre [Hofmann-Wellenhof et al., 2000]. Sur ces bandes de fréquence sont modulés en phase différents types de signaux. Le premier signal civil est le code C/A (coarse/acquisition code) dont la modulation s'effectue uniquement sur L1. Le deuxième code est le code P (précision) modulé sur les bandes de fréquence L1 et L2 en quadrature de phase ; il est crypté (néanmoins, des méthodes permettent de s'en affranchir partiellement [Abraham, 2009]). Sur les dernières générations de satellites est modulé le code L2C sur la bande de fréquence L2, ouvert à tous. Le code M (code militaire) est modulé sur L1 et L2 et est destiné aux utilisateurs autorisés par le gouvernement américain. Le positionnement GPS se fonde soit sur la mesure de pseudo-distances qui sont associées à l'intervalle entre l'émission du signal par le satellite et l'arrivée du signal sur le récepteur soit sur la mesure du déphasage entre les signaux reçus et générés par le satellite.

GLONASS

GLONASS a été initié par l'URSS en 1976. En 2010, il couvre complètement la Russie et, en octobre 2011, GLONASS est complètement opérationnel. Comme pour le GPS, le système de positionnement GLONASS possède des signaux accessibles à tous et des signaux cryptés réservés aux utilisateurs autorisés par le gouvernement russe. Les orbites ont été conçues pour permettre un bon positionnement aux hautes latitudes septentrionales ou australes.

Il existe d'autres systèmes de positionnement GNSS comme Galileo ou Beidou, mais ceux-ci ne sont pas encore utilisés dans les solutions de l'ITRF.

2.6.2 Erreurs instrumentales des GNSS

Il existe de multiples sources d'erreurs instrumentales des GNSS.

Le multitrajét est une erreur due au fait que l'onde GNSS peut emprunter plusieurs trajets depuis le satellite jusqu'au récepteur via des réflexions au niveau de l'environnement de l'antenne (par exemple sur le sol). Le récepteur va observer la résultante des deux signaux direct et réfléchi ce qui engendre des biais dans les pseudo-distances et les mesures de phase. [Elósegui et al., 1995] et [King and Watson, 2010a] montrent que les biais induits par les multitrajets sur les positions de stations peuvent atteindre 0.5 mm/an en vitesse et plusieurs millimètres en position.

4. La bande de fréquence L5 vise une fiabilité du signal GPS dans des applications où une grande fiabilité est requise (par exemple, dans la navigation aérienne).

Le point dont les coordonnées sont déterminées et le point de réception des signaux GNSS appelé centre de phase de l'antenne ne sont pas confondus. Ils diffèrent selon la fréquence et la constellation. Pour déterminer les coordonnées précises grâce à la mesure de phase, il est donc nécessaire de connaître la position du centre de phase par rapport au point de référence de l'antenne dans toutes les directions et élévations. L'IGS maintient un catalogue de cartes de centre de phase. Toutefois, toutes les cartes ne sont pas connues précisément et des valeurs moyennes par type d'antenne sont utilisées. L'impact des erreurs de calibrations d'antennes sur les positions de stations atteint en moyenne 2 mm et 5 mm respectivement en horizontal et en vertical [Baire et al., 2014]. Certaines erreurs de calibrations d'antennes peuvent engendrer des erreurs en vertical et en horizontal allant jusqu'à 10 mm.

2.6.3 Erreurs dues à des défauts de modélisations

Erreurs de modélisation d'orbites

Certaines erreurs ont été constatées dans les produits IGS à des périodes spécifiques dites draconitiques (cf. sous-section 3.1.3). Une année draconitique GPS est le temps mis par le Soleil (observé depuis la Terre) pour effectuer une révolution par rapport aux nœuds orbitaux de la constellation GPS. Elle vaut 351.5 jours [Schmid et al., 2007]. Les erreurs draconitiques constatées dans les produits IGS, en particulier dans les séries temporelles de positions de stations, sont principalement dues à des défauts de modélisation des orbites des satellites GPS.

Erreurs de modélisation de marées

Les signaux relatifs aux marées sont modélisés dans les séries temporelles de positions GNSS et en sont retirés. Cependant, [Penna and Stewart, 2003] ont montré que l'aliasing d'erreurs dans les modèles de marée peut se traduire par des signaux parasites dans les séries temporelles de positions de stations GPS à des fréquences annuelles, semi-annuelles et à 14 jours, ainsi qu'à d'autres fréquences. Dans les séries IGS, on retrouve la fréquence à 14 jours. [Stewart et al., 2005] a explicité de manière mathématique ce phénomène. En analysant les signaux échantillonnés à 5 minutes, [King et al., 2008] montre la présence de séries de positions à hautes fréquences dus à des défauts de modélisation de certaines ondes de marée se traduisant dans des séries de positions journalières par des périodes annuelles et semi-annuelles.

Erreurs de modélisation de la troposphère

Les erreurs de modélisation de la troposphère conduisent à des erreurs dans les positions fournies grâce aux GNSS. [Tregoning and Herring, 2006] ont montré que les erreurs troposphériques causent des artefacts sous forme de signaux annuels et des biais dans les positions verticales estimées des stations permanentes GNSS. Néanmoins, ces erreurs ne dépassent pas 1 mm [Steigenberger et al., 2009].

Chapitre 3

État de l’art de l’analyse statistique des séries temporelles de positions GNSS

L’IGS fournit des solutions journalières de positions de stations desquelles on peut extraire une série temporelle de positions pour chaque station. Les coordonnées de stations sont analysées selon les composantes Est (E), Nord (N) et Verticale (U : Up) dans un repère topocentrique local c’est-à-dire orienté selon les directions précitées et dont l’origine est à proximité de la station. Un exemple de série temporelle est proposé sur la figure 3.1 ; il s’agit de la station OPMT (Paris, France). Les composantes horizontales sont constituées principalement d’une tendance linéaire bruitée. À partir de 2012, pour la composante Nord, on constate un terme harmonique saisonnier. Pour la composante verticale, la tendance linéaire est moins forte, mais un signal saisonnier est visible. Cette dichotomie entre les composantes horizontales et verticale confirme la pertinence du choix des coordonnées est, nord et verticale.

Généralement, un modèle linéaire est adopté pour décrire les séries temporelles de positions de stations GNSS ; il est constitué d’une fonctionnelle et d’un bruit :

$$X_i = f_\theta(t_i) + \varepsilon_i$$

où f_θ , appelée la fonctionnelle, est généralement linéaire en θ , par exemple : $f_\theta(t) = x^\circ + \dot{x}(t - t_i)$ avec $\theta = (x^\circ, \dot{x})$. Les quantités ε_i représentent le bruit et ne sont pas forcément décorrélées.

3.1 État de l’art sur la fonctionnelle

3.1.1 Signaux long-terme

La fonctionnelle utilisée par l’ITRF2014 [Altamimi et al., 2016] est une fonction linéaire par morceaux avec, dans certains cas, des logarithmes et des expo-

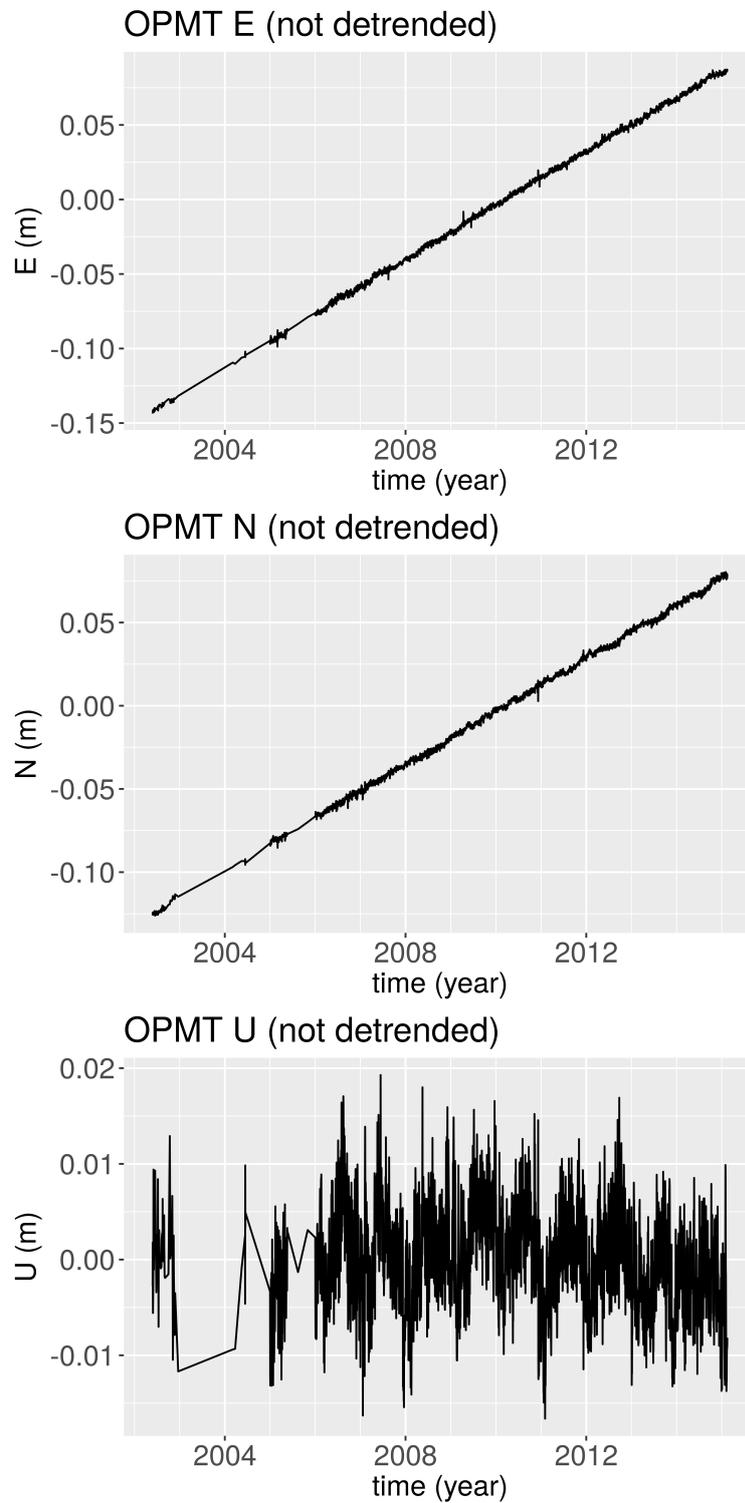


FIGURE 3.1 – Séries temporelles de position de la station OPMT (Paris, France).

mentielles. Cette sous-section présentera l'état de l'art justifiant cette modélisation.

La croûte terrestre est soumise à la tectonique des plaques qui produit un mouvement linéaire loin des zones de déformations aux échelles de temps considérées [Fowler, 2004].

Après un tremblement de terre, on constate souvent une discontinuité de position dans les séries de positions de stations proches de l'épicentre (les discontinuités peuvent aussi être dues à des changements d'équipement ou sont d'origine inconnue). Il peut ensuite se produire un mouvement post-sismique qui correspond à un réajustement viscoélastique de la partie superficielle de la Terre (croûte et manteau supérieur) dans la région du séisme, ce qui est appelé communément relaxation post-sismique. La modélisation du mouvement de déformation post-sismique dans l'ITRF2014 (ou attribué comme tel) est une avancée majeure dans la modélisation des signaux long-terme ; elle est le fruit d'un long processus de réflexion sur la question [Lercier, 2014].

[Savage and Prescott, 1978] propose un modèle du mouvement de déformation post-sismique fondé sur un mécanisme de relaxation exponentielle :

$$D(t) = A(1 - e^{(t-t_s)/\tau_0}).$$

où A est l'amplitude de la déformation, τ_0 est le temps caractéristique de la relaxation et t_s est l'instant du séisme.

[Marone et al., 1991] propose un autre modèle du mouvement de déformation post-sismique :

$$D(t) = A \ln \left(1 + \frac{t - t_s}{\tau_0} \right).$$

Certaines approches intègrent les deux modèles simultanément [Heinkelmann et al., 2008].

[Lercier, 2014] a appliqué les deux modèles aux séries temporelles de 22 stations soumises à des tremblements de terre. Elle a effectué une sélection de modèles via des critères de maximum de vraisemblance pénalisée (AIC, BIC). Elle montre l'intérêt de l'utilisation de tels modèles. Cette étude a fortement inspiré l'ITRF2014 [Altamimi et al., 2016] qui inclut désormais des modèles de déformations post-sismiques.

Le rebond post-glaciaire correspond à un réajustement visqueux du manteau à la suite de la déglaciation consécutive à la dernière période de glaciation. Il ne produit pas un mouvement linéaire, mais son échelle de temps est l'ordre de dizaines de milliers d'années ; il est raisonnable de le considérer comme linéaire sur quelques dizaines d'années [Peltier et al., 2015, Lambeck et al., 2014]. Il constitue la principale source de mouvement vertical jusqu'à l'ITRF2008 [Altamimi et al., 2011]. En revanche, la fonte des glaces actuelle due au réchauffement climatique, qui a été modélisée par [Shepherd et al., 2012], révèle un mouvement non linéaire dans les séries de positions GNSS [Khan et al., 2010], qui a été pris en compte dans l'ITRF2014 par l'introduction de discontinuités de vitesses.

3.1.2 Signaux saisonniers

Des signaux saisonniers sont visibles dans les séries temporelles de positions ; ils sont particulièrement importants sur la composante verticale. [Dong et al., 2002] a analysé des séries temporelles GNSS journalières sur une période de 4.5

ans. Il repère un certain nombre de phénomènes susceptibles d'expliquer les mouvements périodiques et tente de les modéliser à partir de modèles préexistants. Les auteurs ont étudié la marée polaire, ce qui n'a plus de sens actuellement, car elle a été entre temps corrigée dans l'ITRF. Les auteurs mentionnent les marées océaniques, dont certaines ondes ont des fréquences d'aliasing annuelle et semi-annuelle via un échantillonnage journalier. Ces erreurs de modélisation de marées peuvent atteindre 5 mm en vertical pour les périodes annuelle et semi-annuelle, avec des valeurs médianes de 0.8 mm pour la période annuelle et 0.6 mm pour la période semi-annuelle [King et al., 2008]. Ils considèrent aussi la surcharge atmosphérique ; les valeurs maximales d'amplitudes annuelles atteignent 4 mm, les fortes amplitudes annuelles se situent souvent sur la plaque eurasiatique et dans la péninsule arabique, bien qu'en général, ces amplitudes ne dépassent pas 0.5 mm. Les surcharges océaniques ne sont pas dues uniquement aux marées ; par exemple, l'hydrodynamique ou la force de surface exercée par le vent peuvent jouer un rôle. Leurs amplitudes verticales annuelles maximales sont de l'ordre de 2-3 mm sur les côtes (maritimes), cette amplitude est généralement inférieure 1 mm sur les continents. Les auteurs ont considéré les surcharges dues à la neige et à l'humidité des sols. Les plus fortes amplitudes annuelles des mouvements dus à l'humidité du sol atteignent 7.1 mm, 5.1 mm et 5.3 mm pour les stations BRAZ (Brasilia, Brésil), NSSP (Yerevan, Arménie) et QUIN (Quincy, États-Unis). À l'intérieur des continents, l'amplitude annuelle est généralement de l'ordre de 2 à 3 millimètres. Quant aux surcharges liées à la neige, elles induisent les signaux annuels de l'ordre de 3 à 5 mm dans les régions polaires. Les auteurs montrent que les phénomènes précités représentent 40 % de la puissance du signal annuel observé.

Les auteurs ont cherché d'autres origines possibles des signaux saisonniers, ils n'ont pas réussi, malgré leur recherche à l'époque, à quantifier les erreurs d'orbite des satellites, les erreurs dues à la troposphère et celles dues aux mauvaises modélisations du centre de phase. Les effets de la dilatation thermique des roches induisent une amplitude annuelle de 0.5 mm. Les auteurs notent enfin des variabilités sur les signaux annuels verticaux en fonction du logiciel utilisé, qui est généralement de 2-3 mm, mais peut atteindre 5 à 7 mm.

La longueur de la série a un effet sur l'estimateur de la vitesse de la station. Pour étudier cette question, [Blewitt and Lavallée, 2002] analysent 23 séries de positions GNSS de 3,5 années. Ils considèrent qu'une bonne approximation de la vraie vitesse peut être obtenue pour des séries temporelles dont la longueur dépasse 3,5 années, ce qui se fonde sur des considérations théoriques développées par les auteurs. Ils tronquent la série pour ne garder d'une série qu'une durée définie, dont ils calculent l'estimateur de la vitesse. Ils montrent ainsi que les séries courtes ont des estimateurs de vitesses biaisés par le signal annuel, de manière d'autant plus importante que la série de positions restante est courte.

Le calcul de l'ITRF2014 [Altamimi et al., 2016] inclut l'estimation de signaux annuels et semi-annuels mais ces signaux ne font pas partie des coordonnées de référence fournies aux utilisateurs, en raison de désaccords sur les signaux saisonniers estimés sur les sites colocalisées.

3.1.3 Signaux draconitiques

La période draconitique est la période prise par le Soleil (observé depuis la Terre) pour effectuer une révolution par rapport aux nœuds orbitaux de la

constellation GPS [Schmid et al., 2007]. Des signaux périodiques aux harmoniques de l'année draconitique sont présents dans les séries temporelles de résidus GNSS, mais pas dans les séries temporelles de résidus SLR ou VLBI [Ray et al., 2008, Collilieux et al., 2007]. L'origine des signaux draconitiques n'est pas, de manière définitive, connue, bien qu'elle semble liée majoritairement à des erreurs de modélisation des orbites GPS. Plusieurs origines cumulatives sont proposées : les erreurs d'orbites, les multi-trajets [Ray et al., 2008]. Des progrès récents dans la modélisation des orbites GPS ont permis de réduire l'amplitude des signaux draconitiques dans les séries temporelles de positions [Rodriguez-Solano et al., 2012]. Ceux-ci sont cependant toujours nettement présents dans les séries temporelles utilisées pour calculer l'ITRF2014 [Rebischung et al., 2015].

Ces signaux draconitiques ont été pris en compte via l'estimation de sinusoides aux harmoniques de l'année draconitique GPS dans certaines études [Bogusz and Klos, 2016].

3.2 État de l'art sur la partie stochastique

3.2.1 Nécessité d'un bruit corrélé

La première étude pour appréhender l'importance de la prise en compte du bruit corrélé dans les séries temporelles de positions GNSS a été réalisée par [Zhang et al., 1997]. Ils utilisent les séries temporelles journalières de positions de 10 stations GPS pendant 19 mois entre deux tremblements de terre et montrent que le bruit est corrélé. Ils utilisent le maximum de vraisemblance ainsi que l'analyse spectrale pour estimer les paramètres de bruit. Ils montrent qu'un modèle de bruit blanc ne suffit pas à décrire le bruit, il faut rajouter un bruit de loi de puissance, dont la densité spectrale de puissance est proportionnelle à $1/f^\alpha$ (α est appelé l'indice spectral de bruit de loi-puissance). Ils remarquent que les incertitudes sur la vitesse calculée à partir des séries temporelles de positions GNSS sont plus élevées d'un facteur 2 à 6 lorsqu'un bruit corrélé est considéré.

Au fil du temps, les études traitent des séries temporelles de plus en plus longues. [Mao et al., 1999] effectuent des calculs similaires à partir de 3 ans de données sur 23 stations sélectionnées parmi un ensemble de 120 stations et réparties à travers le monde avec une forte concentration au niveau de la Californie. La publication montre un avantage du maximum de vraisemblance comparé aux méthodes spectrales et montre que le meilleur modèle de bruit est le bruit blanc additionné avec du bruit de scintillation (qui est un bruit de loi de puissance dont l'indice spectral vaut 1). Ils observent une différence spatiale entre les amplitudes de bruit, en particulier pour la direction verticale, ce qui peut s'expliquer par des phénomènes météorologiques qui rendent plus difficile la modélisation des délais troposphériques ou par l'effet de l'humidité qui peut engendrer des erreurs.

Une étude théorique intéressante est [Williams, 2003] qui fournit la matrice de covariance de tout bruit de loi de puissance ainsi qu'une expression de l'incertitude sur la vitesse selon le modèle de bruit. En particulier, si les mesures sont également espacées, la formule d'incertitude sur la vitesse est, pour le modèle de bruit blanc :

$$\sigma_{\dot{x}}^2 \sim_{n \rightarrow \infty} \frac{12\sigma_w^2}{nT^2},$$

où σ_w est l'amplitude du bruit blanc, n est le nombre de mesures et T est l'amplitude temporelle de la série (différence entre la première et la dernière valeur temporelle). Lorsque le bruit est une marche aléatoire, la formule devient :

$$\sigma_x^2 = \sigma_{RW}^2/T$$

où σ_{RW} est l'amplitude de la marche aléatoire. Pour le bruit de scintillation, on a :

$$\sigma_x^2 \approx \frac{9\sigma_{BS}^2}{16\Delta T^2(n^2 - 1)}.$$

L'article donne une approximation de cette quantité pour un modèle de bruit de bruit de loi de puissance d'indice spectral quelconque.

Plusieurs modèles de bruits temporels de séries de positions GNSS ont été comparés via un maximum de vraisemblance pénalisée par [Santamaría-Gómez et al., 2011], ce qui permet de tenir compte du surapprentissage. Il a étudié plusieurs modèles de bruit blanc :

- le bruit blanc homogène dont la variance σ_w^2 est constante au cours du temps,
- le bruit blanc variable dont l'écart-type est proportionnel aux écarts-types fournis avec les données de positions [Williams, 2008],
- le bruit blanc dépendant en temps dont la variance décroît exponentiellement en fonction du temps.

Ils montrent que le meilleur modèle de bruit blanc est le bruit blanc variable. Il n'est pas aisé de comparer les modèles de bruits corrélés, car le processus peut connaître un surapprentissage c'est-à-dire un surajustement du modèle sous l'effet d'un trop grand nombre de paramètres ; [Santamaría-Gómez et al., 2011] tendent cependant à montrer que le bruit de scintillation est, pour trois quarts des stations étudiées, le meilleur modèle de bruit parmi, entre autres, le bruit de loi-puissance (dont le spectre est proportionnel à $1/f^\alpha$), la marche aléatoire, le bruit de scintillation avec une marche aléatoire, le processus de Gauss-Markov et le processus de Gauss-Markov généralisé.

Il est à noter que la nature du bruit présent dans les séries de positions GNSS varie avec le traitement des données. [Amiri-Simkooei et al., 2017] montrent par exemple l'émergence d'une marche aléatoire en comparant deux retraitements.

L'origine du bruit de scintillation est difficile à comprendre. Un mécanisme possible est la combinaison des erreurs dues aux multitrajets et des variations temporelles de la constellation GNSS [King and Watson, 2010b], bien que ce mécanisme ne suffise pas à expliquer le niveau de bruit de scintillation observé ni ses corrélations spatiales. [Rebischung et al., 2017] semblent montrer qu'une partie du bruit de scintillation peut provenir des effets de surcharge.

3.2.2 Aspect de la complexité de calcul des algorithmes

La difficulté de l'estimation des paramètres de bruit par maximum de vraisemblance est sa complexité (vitesse) en $O(n^3)$. [Williams, 2008] propose une optimisation du maximum de vraisemblance qui réduit le nombre de paramètres à optimiser d'une unité ce qui accélère les calculs. [Bos et al., 2008] proposent une optimisation algorithmique faisant passer la complexité du maximum de vraisemblance de cubique à quadratique, cette méthode suppose l'absence de valeurs

manquantes dans les données ; [Bos et al., 2013b] gardent le même ordre de grandeur de complexité, mais s'affranchissent de la contrainte d'absence de valeurs manquantes sous conditions, conditions partiellement levées grâce à [Langbein, 2017].

Ces différentes optimisations sont partiellement reprises dans les logiciels CATS [Williams, 2008] et HECTOR [Bos et al., 2013a].

3.3 Dépendance spatiale

Le premier article à étudier la dépendance spatiale dans les séries temporelles de positions est [Wdowinski et al., 1997]. Ils étudient des séries temporelles journalières de positions GPS de 100 jours centrées sur le tremblement de Terre du 28 juin 1992 à Landers en Californie. Ces séries sont issues de 6 stations dont deux sont colocalisées. Elles contiennent des signaux caractéristiques du séisme. Les séries sont séparées en une tendance long-terme dépendant de la station et un résidu dépendant aussi de la station. Ils observent que les séries de résidus sont corrélées d'une station à l'autre et retirent la moyenne des résidus à chacune des séries ce qui permet d'obtenir des séries filtrées moins bruitées.

L'analyse en composantes principales (ACP) est utilisée par [Dong et al., 2006]. Ils ont sélectionné 152 stations dans la Californie du Sud sur des critères de durée et de faible pourcentage de valeurs manquantes. Les données sont interpolées. Les données, dont on a enlevé une tendance linéaire, font l'objet d'une analyse en composante principale et une expansion de Karhunen-Loève ; ces deux algorithmes permettent d'obtenir des données résiduelles qui correspondent aux données sans tendance dont on a retiré un mode commun spatial calculé selon les deux algorithmes. Les deux méthodes aboutissent à des résultats similaires. Ces algorithmes ont permis de détecter des effets locaux pour trois stations. Les deux algorithmes sont d'après les RMS très légèrement meilleurs que l'algorithme de [Wdowinski et al., 1997].

Une étude de la dépendance spatiale des séries de positions GNSS a été réalisée par [Amiri-Simkooei et al., 2017]. Cette étude utilise les positions des première et seconde campagnes de retraitement des données du centre d'analyse JPL appelées **repro** et **repro2**. Ils utilisent deux jeux de données fournis par le centre d'analyse JPL : le premier contient uniquement des séries temporelles journalières de positions **repro** de 66 stations GNSS, le second contient des séries temporelles de positions à la fois **repro** et **repro2** de 89 stations. Ils calculent des corrélogrammes en est, nord et verticale. Les corrélogrammes sont quasi-nuls pour les composantes croisées, alors qu'il existe une dépendance spatiale jusqu'à une distance angulaire de 30° sur les trois composantes Est-Est, Nord-Nord et Verticale-Verticale. Ils remarquent un niveau de corrélation verticale qui est plus faible que celui des corrélations horizontales. Pour les composantes Nord et verticale, la corrélation spatiale est plus élevée pour **repro** que **repro2**. Néanmoins, ce travail a des limites, les corrélogrammes sont simplement lissés pour dégager une tendance, ils n'ajustent pas dessus des variogrammes théoriques (tels que les variogrammes exponentiels ou gaussiens). Une telle construction ne permet pas d'effectuer de prédiction dans le futur ou d'imputer des valeurs manquantes.

Une source importante de corrélation spatiale : les surcharges Les surcharges (atmosphériques, océaniques, etc.) désignent la force exercée sur la surface terrestre par les transferts de masse de l'atmosphère, l'océan, etc. Ces forces engendrent des mouvements de la croûte, mouvements plus importants sur la composante verticale qu'horizontale. Des modèles étudient les surcharges hydrologiques [Reichle et al., 2011], océaniques [Menemenlis et al., 2008] ou atmosphériques [Dee et al., 2011]. [Xu et al., 2017] montrent que la prise en compte des surcharges atmosphérique, océanique, due à la neige et à l'hydrologie continentale réduit le signal annuel de 43 % en vertical et d'environ 20 % en horizontal. L'échelle spatiale des mouvements de surcharge fait qu'ils sont corrélés sur de grandes distances. Ils peuvent donc expliquer une partie de la dépendance spatiale des signaux saisonniers observés dans les séries de positions GNSS. De plus, si une partie du bruit de scintillation observé dans les séries provient du mouvement de surcharge [Rebischung et al., 2017], alors ceux-ci peuvent expliquer en partie la dépendance spatiale du bruit.

3.4 Filtre de Kalman pour la détermination de repères de référence

Le calcul de l'ITRF est basé sur l'ajustement par moindres carrés, néanmoins le filtre de Kalman a été utilisé très tôt pour déterminer des repères de référence. Par rapport aux moindres carrés, le filtre de Kalman a l'avantage de traiter les données de manière séquentielle ce qui réduit la taille des matrices à inverser.

Une première utilisation du filtre de Kalman pour calculer les systèmes de référence a été décrite dans [Dong et al., 1998]. Cette étude propose une méthodologie d'élaboration d'un repère de référence similaire à celle de l'ITRF. L'article utilise des données terrestres (EDM : electronic distance measurement) et de géodésie spatiale (GPS) au niveau du sud de la Californie. Les données EDM s'étendent de 1972 à 1992 pour un réseau de 119 stations et celles (journalières) du GPS de 1986 à 1995 avec un réseau de 114 stations¹; elles ont été acquises dans une région des Transverse Ranges à la vallée de Mexicali. Ils proposent de déduire des quasi-observations à partir des mesures de chaque technique. La combinaison des quasi-observations est effectuée à partir d'un filtre de Kalman en considérant comme modèle de bruit du bruit blanc et une marche aléatoire.

[Wu et al., 2015] proposent une méthodologie qui vise à tenir compte d'une légère non-linéarité des processus via l'introduction d'une marche aléatoire correspondant aux paramètres déterministes (vitesse ou signaux saisonniers pas tout à fait constant au cours du temps). Ils proposent un filtre de Kalman avec un lissage RTS (Rauch-Taub-Striebel). La position d'une station est modélisée par :

$$V_k(t_k - t^o) + x_k^o + S_k$$

où k est le numéro de la semaine et (S_k) correspond à un oscillateur périodique. Il propose d'introduire une souplesse en ce qui concerne la linéarité en considérant les positions et les vitesses comme des marches aléatoires. Ils comparent les vitesses obtenues avec celles de l'ITRF2008 et obtiennent un RMS à 0.4, 0.4 et 1.0 mm pour les directions ENU et 0.7, 0.4 et 0.8 pour les axes X, Y et Z. L'aspect non-linéaire de la méthode permet de considérer des trajectoires non

1. Les mesures GPS sont données dans un référentiel donné par 7 stations.

linéaires comme celle de la station AREQ au Pérou qui a subi des déformations post-sismiques ou celle de la station THU3 dont les mouvements dus à la fonte de glace au Groenland sont non linéaires.

3.5 Problématique

Cet état de l'art va nous conduire à élaborer une méthodologie pour répondre à des questions autour de la dépendance spatiale des séries de positions GNSS.

3.5.1 Dépendance spatiale des séries de positions GNSS

L'ITRF2014 considère la dépendance spatiale du bruit blanc à partir des matrices de variance-covariance fournies avec les solutions des quatre techniques de géodésie spatiale, mais pas celle du bruit corrélé. [Amiri-Simkooei et al., 2017] utilisent des corrélogrammes pour évaluer la dépendance spatiale du bruit total. Ils n'utilisent cependant pas cette information de corrélation spatiale pour améliorer la détermination de coordonnées de référence, ce qui est l'objectif de cette thèse. La prise en compte de la dépendance spatiale peut s'effectuer de diverses manières. Nous proposerons donc plusieurs modèles pour la dépendance spatiale et évaluerons leurs performances respectives à l'aide de techniques de sélection de modèles fondées sur des critères statistiques.

3.5.2 Algorithmique : trouver une méthodologie rapide

Nous avons vu à la sous-section 3.2.2 que la littérature cherche des moyens pour réduire la complexité des algorithmes utilisés, qui est principalement due à l'inversion de matrice dans l'algorithme des moindres carrés. Le volume de données de positions GNSS grossit au fil des années : de nouvelles stations sont rajoutées et les anciennes stations produisent toujours des données (certaines stations ont maintenant acquis vingt ans de données). Je reprends le raisonnement de [Collilieux, 2008]. Si on considère 250 stations de 12 ans, nous avons plus de trois millions de positions journalières en comptant chaque composante. Si l'on souhaite tenir compte de la dépendance spatiotemporelle du bruit contenu dans les séries, l'utilisation de la méthode des moindres carrés conduira à l'inversion d'une matrice de variance-covariance de taille de l'ordre de trois millions, ce qui est difficile. L'utilisation d'un filtre de Kalman permet de traiter ce problème efficacement, il permet d'inverser des matrices plus petites mais plus nombreuses, avec, au final, un gain en temps de calcul. Ce filtre requiert néanmoins que le modèle d'évolution temporelle s'écrive sous une forme espace-état. Nous ne pouvons plus utiliser un bruit de scintillation de manière exacte car il ne peut s'implémenter sous la forme d'un modèle espace-état. Nous avons deux possibilités : approximer le bruit de scintillation par un processus stochastique sous la forme espace-état [Dmitrieva et al., 2015] ou considérer un autre processus de bruit implémentable dans un filtre de Kalman, sans référence au bruit de scintillation. Nous avons choisi la deuxième solution, des éléments pour comprendre les avantages de ce choix sont présentés dans la sous-section 4.4.1.

3.5.3 Apport de la dépendance spatiale

La prise en compte de la dépendance spatiale a de multiples intérêts. Elle permet des détections de rupture plus pertinentes via l'utilisation d'un mode commun pour les résidus [Gazeaux et al., 2015]. Nous avons vu que les séries courtes ont des estimateurs de vitesses biaisés [Blewitt and Lavallée, 2002]. Nous allons donc aussi pouvoir évaluer si la prise en compte de la dépendance spatiotemporelle améliore l'estimateur des vitesses. La prise en compte de la dépendance spatiale peut en effet permettre d'estimer les vitesses de séries courtes de manière plus fiable, de l'information entre les stations étant ajoutée. D'après [Collilieux and Bernard, 2012] la prise en compte de la dépendance spatiale peut aussi améliorer l'imputation des valeurs manquantes et la prédiction dans le futur. Ma méthodologie sera parfaitement adaptée pour évaluer cet apport. Le filtre de Kalman a déjà été utilisé pour prendre en compte la dépendance spatiale des séries de positions GNSS [Liu et al., 2018]. La méthodologie que j'ai adoptée ci-dessus est similaire à celle de cet article. [Liu et al., 2018] ne s'intéressait cependant qu'à l'imputation de valeurs manquantes mais pas à la détermination de repères de référence, comme c'est mon cas.

Deuxième partie

Analyse statistique
spatiotemporelle des séries
temporelles de positions
GNSS

Chapitre 4

Choix de la méthodologie statistique

Ce chapitre vise à détailler la méthodologie adoptée pour étudier et caractériser la dépendance spatiotemporelle des séries temporelles de positions GNSS afin de déterminer plus précisément un repère de référence terrestre.

Dans un premier temps, nous décrivons les données de positions GNSS que l'on sépare en une fonctionnelle correspondant à un modèle cinématique et en un bruit. La seconde section décrit le modèle cinématique, qui comporte une tendance linéaire avec des discontinuités de position et des composantes périodiques ; le reste correspond au bruit que nous modéliserons statistiquement. Dans la troisième section, nous montrerons l'existence d'une dépendance spatiale dans le bruit. Nous privilégions donc la modélisation spatiotemporelle du bruit qui a, de plus, un impact sur les incertitudes des termes de la fonctionnelle, et notamment la vitesse, d'où son intérêt dans l'étude des systèmes de référence terrestres. La quatrième section traitera le bruit qui sera modélisé comme la somme d'un bruit autocorrélé et d'un bruit décorrélé. Ce modèle sera utilisable dans un filtre de Kalman utilisant un modèle espace-état, modèle dans lequel on introduira les dépendances spatiales du bruit. Plusieurs modèles de covariances spatiotemporelles seront proposés. L'introduction d'un nouveau système de coordonnées sera nécessaire pour éliminer la singularité aux pôles du système de coordonnées topocentriques locales et pouvoir considérer les corrélations spatiales sur des composantes horizontales proches des pôles. Enfin, dans la dernière partie, nous décrivons la méthodologie pour sélectionner le meilleur modèle fondée principalement sur la validation croisée.

4.1 Choix des données

Nous avons utilisé des données de positions hebdomadaires calculées par l'International GNSS Service (IGS) dans la cadre de sa seconde campagne de retraitement des données GNSS `repro2` [Rebischung, 2015]. Les coordonnées calculées sont issues de la combinaison des produits de huit centres d'analyse. Cette campagne est une mise à jour de la campagne `repro` qui avait été entreprise en vue du calcul de l'ITRF2008. Cette combinaison a été effectuée à l'IGN par l'équipe en charge des systèmes de référence IGS. Les données utilisées

sont composées de 1403 séries temporelles de positions de stations permanentes GNSS. Les plus longues séries de coordonnées s'étendent de 1994 à 2015.

Parmi les améliorations de ce nouveau lot de données, l'échantillonnage journalier a été une avancée notable. Il a permis une meilleure gestion des discontinuités de position présentes en cas de changement d'instruments ou de séismes dans le calcul de l'ITRF2014. Des données GLONASS ont été incluses par rapport à **repro** qui n'en contient pas. La solution est donc dite GNSS, car n'incluant pas uniquement des données GPS. Les modèles physiques utilisés pour l'ajustement des données brutes GNSS sont conformes aux conventions IERS 2010 [Petit and Luzum, 2010]. Ces conventions sont les mêmes que celles utilisées par les centres d'analyse des données des autres techniques de géodésie spatiale permettant à l'ITRF2014 d'être un repère de référence homogène. Les marées terrestres solides sont donc par exemple absentes des coordonnées utilisées ici. Enfin, il est pertinent de mentionner que les positions de stations sont alignées sur le repère de référence IGB08 qui est dérivé de l'ITRF2008 ; les tendances estimées dans les séries dépendent en effet de cette convention. Les termes périodiques dépendent quant à eux de l'inclusion ou non du mouvement du géocentre dans les coordonnées [Collilieux et al., 2012]. Le mouvement du géocentre est défini par le mouvement du centre des masses de la Terre par rapport à la surface moyenne de la Terre représentée par son centre géométrique, le centre de figure. J'utilise des séries temporelles de positions qui n'incluent que la partie linéaire du mouvement du géocentre ; la partie saisonnière des coordonnées est en effet exprimée par rapport au centre de figure à cause de l'alignement des solutions sur l'origine de l'IGB08 [Dong et al., 2003].

Pour des raisons de temps de calcul, j'ai choisi d'utiliser des données hebdomadaires issues d'une moyenne des données journalières réalisée par l'IGN [Rebischung, 2015]. La thèse étant focalisée sur l'étude de la dépendance spatiale, ce choix n'est pas préjudiciable.

La figure 4.1 représente la carte des 1403 stations **repro2** sans présélection. La densité de stations est particulièrement importante en Europe, aux États-Unis, au Japon. Par rapport aux autres techniques de géodésie spatiale, les stations sont néanmoins bien réparties à travers le monde.

4.1.1 Sélection pour l'étude des corrélations spatiales

Afin de disposer d'un lot de séries fiables pour étudier la dépendance spatiale du bruit présent dans les séries temporelles, des stations réparties sur tout le globe terrestre ont été sélectionnées sur les critères suivants :

- les séries temporelles de positions doivent être longues ; une durée minimale a été fixée à 3 ans ce qui correspond à peu près à la limite recommandée par [Blewitt and Lavallée, 2002] pour estimer des vitesses fiables qui vaut 2.5 ans ;
- les résidus des séries temporelles doivent être sans tendance interannuelle, ni événements transitoires (en particulier les stations sélectionnées ne contiennent pas de déformation post-sismique visible) ; éventuellement les parties problématiques de certaines séries ont été éliminées ;
- la durée du segment entre deux discontinuités ne doit pas être trop courte (les segments de quelques points ont été éliminés).

Les 195 stations sélectionnées pour l'étude des dépendances spatiales du bruit sont présentées sur la figure 4.3. Elles sont réparties inégalement à travers

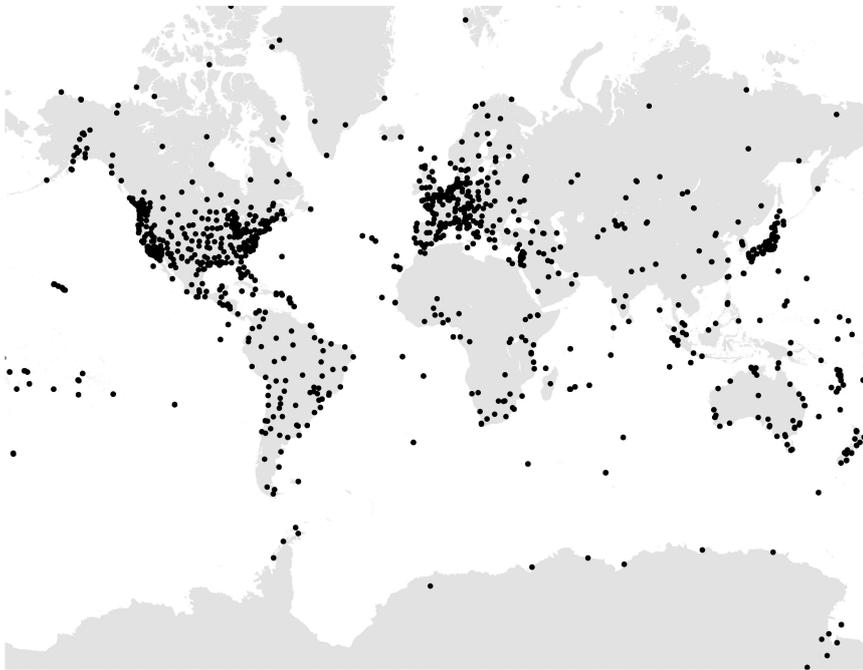


FIGURE 4.1 – Cartes des stations repro2 ayant des séries temporelles de positions hebdomadaires

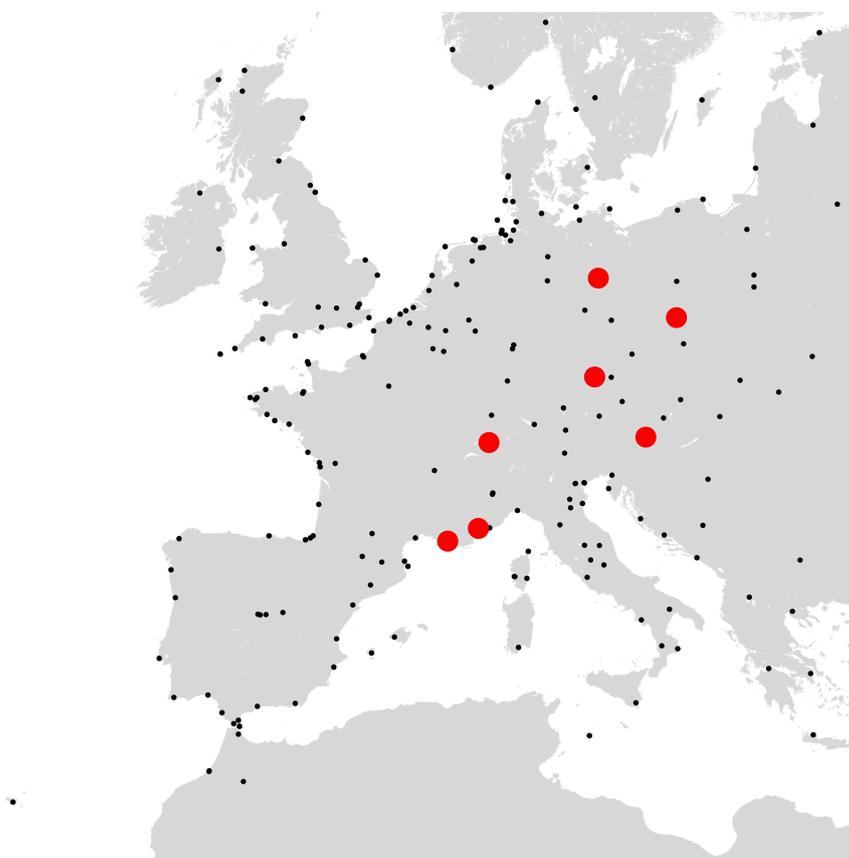


FIGURE 4.2 – Carte des stations `repro2` en Europe en noir. Les stations sélectionnées dont les positions seront intégrées au filtre de Kalman sont représentées en rouge.



FIGURE 4.3 – Carte des stations sélectionnées pour estimer la dépendance spatiale

le monde ; l'Amérique du Nord et l'Europe ont une grande densité de stations, la sélection d'un grand nombre de stations en Europe s'expliquant la grande qualité des stations dans ce continent par rapport, par exemple, aux États-Unis, où les séries temporelles de positions de stations sont souvent beaucoup plus courtes.

4.1.2 Cas d'étude : 10 séries de coordonnées en Europe

Pour mettre en pratique la démarche décrite dans les sections suivantes pour l'évaluation de divers modèles de covariance spatio-temporelle, nous avons par ailleurs choisi un lot de stations situées en Europe du fait de la grande densité de stations à cet endroit. Les séries sont par ailleurs plus longues que celles disponibles aux États-Unis par exemple. J'ai ainsi choisi 10 stations parmi les 195 précédentes. Les critères retenus sont les mêmes que pour les 195 stations en les durcissant ; nous avons choisi des stations dont les séries temporelles s'étendent sur sept années au moins ; de plus, nous avons privilégié la sélection de stations colocalisées. Les stations sélectionnées sont présentées sur la figure 4.2, certaines sont co-localisées. Les plages temporelles des stations sont présentées sur la figure 4.4 ; les séries temporelles ont de grandes plages temporelles de simultanéité ce qui permet d'inférer des dépendances spatiales de manière correcte.

4.2 Modèle cinématique : fonctionnelle

Dans cette section, nous reprenons les paramétrisations classiquement présentées dans la littérature et décrites au chapitre 3. Soit $t_{i,k}$ les temps d'observation ($1 \leq i \leq n_k, 1 \leq k \leq n_{stations}$). Soit t_k^o un temps de référence. Soit $Y_{i,k}$ la série temporelle de positions GNSS pour une composante qui correspond à une station GNSS k . La notation a été définie par un souci de cohérence avec les notations du filtre de Kalman. On utilise un modèle apparenté à un modèle

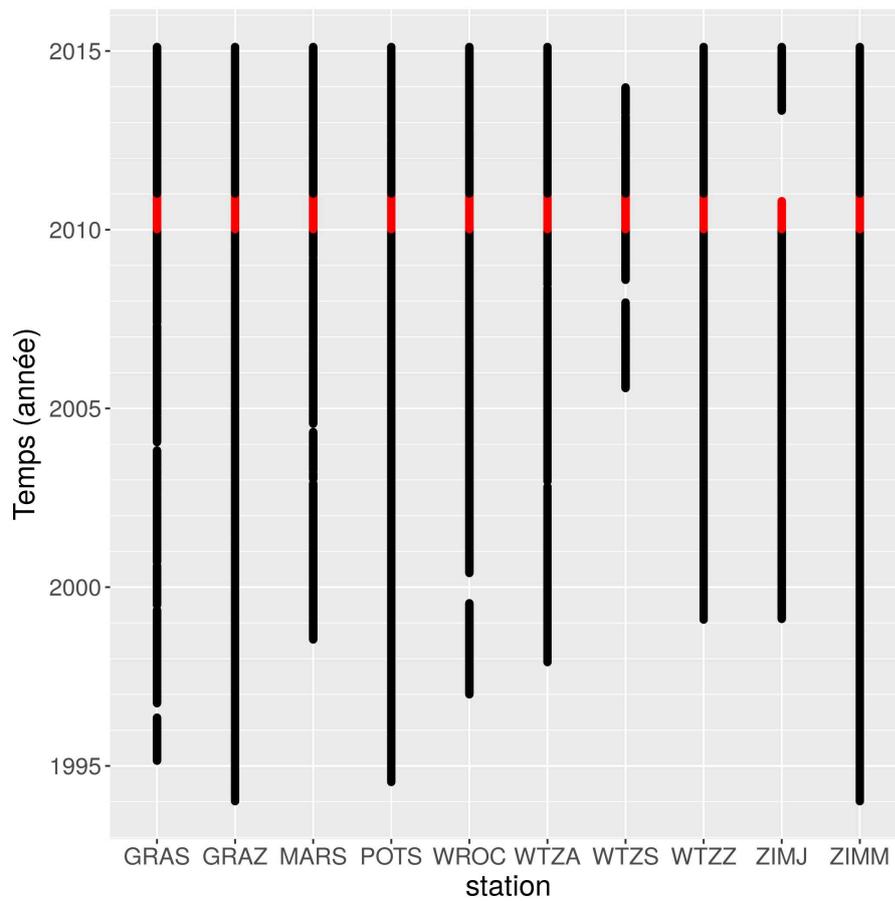


FIGURE 4.4 – Plages temporelles de chacune des 10 stations sélectionnées en Europe. Les zones rouges correspondent aux zones de test de la validation dont la définition sera donnée en section 4.8, elles font partie des données.

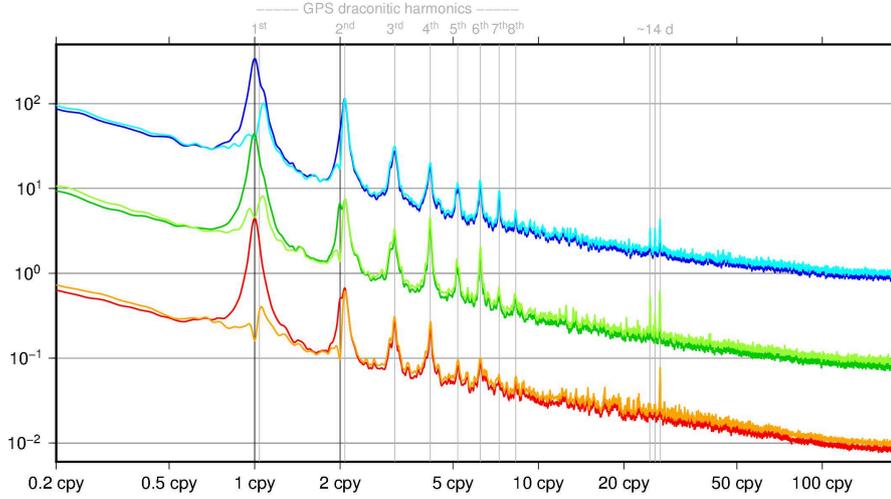


FIGURE 4.5 – Périodogramme cumulé, en échelle logarithmique en abscisses et en ordonnées, des séries temporelles de positions de 1066 stations **repro2** après retrait d'une fonction affine par morceaux, et, pour les couleurs claires, en outre, des périodes annuelles et semi-annuelles. Bleu : Est ; vert : Nord ; rouge : Up. Les périodogrammes sont donnés à un offset près afin d'améliorer la lisibilité. La figure est extraite de [Rebischung et al., 2015].

de régression linéaire :

$$Y_{i,k} = x^{\circ}_{\alpha_{k,pos}(i),k} + \dot{x}_{\alpha_{k,vit}(i),k}(t_{i,k} - t_k^{\circ}) + \sum_{j=1}^m (a_{j,k} \cos(\omega_j(t_{i,k} - t_k^{\circ})) + b_{j,k} \sin(\omega_j(t_{i,k} - t_k^{\circ}))) + \gamma_{i,k},$$

où

- les x° , \dot{x} sont les positions et les vitesses de référence ;
- $a_{j,k}$, $b_{j,k}$ sont les amplitudes des termes périodiques à la pulsation $\omega_j = \frac{2\pi}{T_j}$, T_j étant la période du signal périodique j ,
- t_k° est le temps de référence, généralement fixé un 1er janvier.

Les $x_{j,k}^{\circ}$, les $\dot{x}_{j,k}$, les $a_{j,k}$ et les $b_{j,k}$ constituent les paramètres déterministes ou paramètres de régression.

Pour la composante périodique, nous considérerons :

- les périodes annuelles et semi-annuelles,
- les draconitiques GPS jusqu'à la 7^e harmonique.

Le choix de ces périodes s'appuie sur le périodogramme cumulé présenté sur la figure 4.5 [Rebischung et al., 2015].

Les fonctions croissantes au sens large $\alpha_{k,pos} : \llbracket 1, n \rrbracket \rightarrow \llbracket 1, n_{k,pos} \rrbracket$ et $\alpha_{k,vit} : \llbracket 1, n \rrbracket \rightarrow \llbracket 1, n_{k,vit} \rrbracket$, connues, permettent de tenir compte des discontinuités. L'entier $n_{k,pos}$ (respectivement $n_{k,vit}$) est égal au nombre de portions temporelles délimitées par les discontinuités de position (respectivement de vitesse) pour la station k ; sur chaque segment respectif seront estimées une position de référence et une vitesse. La quantité $\gamma_{i,k}$ représente le bruit total, somme

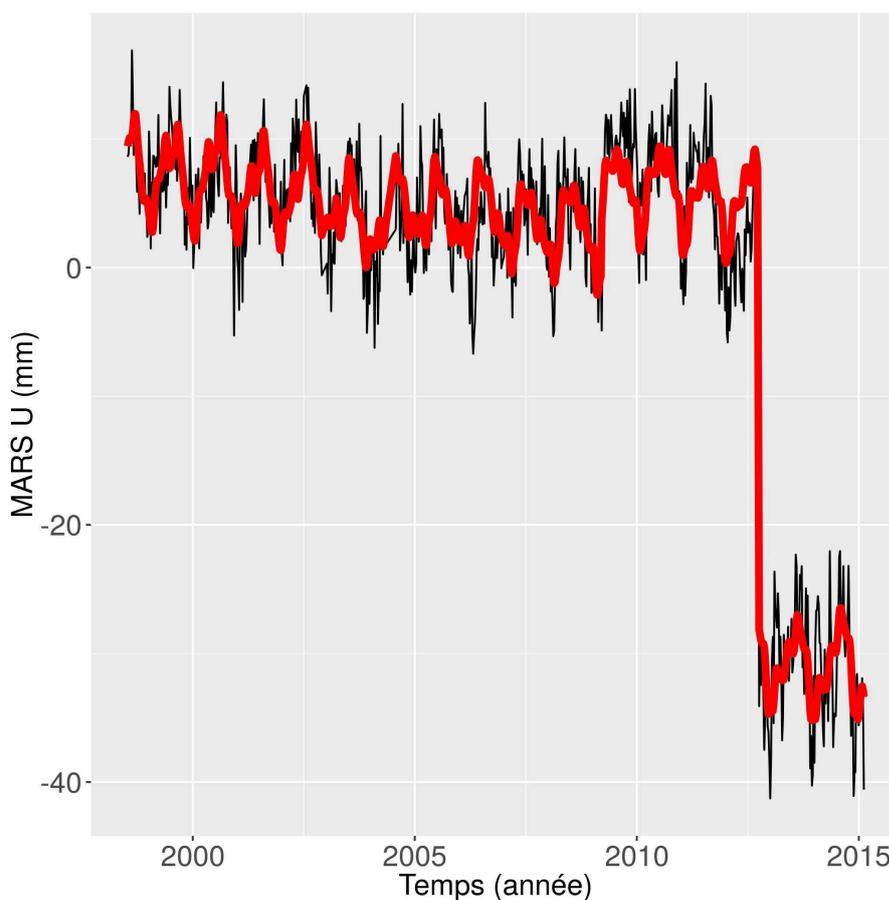


FIGURE 4.6 – Série temporelle de positions GNSS de la station MARS pour la composante verticale. En noir sont représentées les positions fournies par l'IGS et en rouge leur partie déterministe issue de la régression linéaire.

des bruits corrélé et décorréolé temporellement. Les $Y_{i,k}$ peuvent éventuellement correspondre à des valeurs manquantes.

Nous avons choisi des stations non affectées par des déformations post-sismiques. En cas de déformation post-sismique, il est possible de soustraire un modèle paramétrique prédéterminé comme pour l'ITRF2014. Des méthodes plus sophistiquées peuvent être envisagées comme, par exemple, estimer simultanément le déplacement post-sismique à plusieurs stations [Lercier, 2014], mais ne sont pas traitées ici.

À titre d'illustration, la figure 4.6 présente les données de positions verticales de la station MARS (Marseille, France) et son modèle paramétrique estimé.

L'objectif des sections suivantes est de déterminer la covariance spatiotemporelle $\text{Cov}(\gamma_{k,i}, \gamma_{k',i'})$. Dans le cadre d'un modèle stationnaire variographique, une covariance de la forme $\text{Cov}(Y(t, \vec{s}), Y(t', \vec{s}'))$ sera utilisée où \vec{s} et \vec{s}' correspondent aux stations assimilées à un vecteur position.

4.3 Première étude de la dépendance spatiale : dépendance des séries temporelles du bruit total

Nous allons examiner, dans une première approche, la dépendance spatiale des séries de bruits totaux qui sont obtenues en retranchant le modèle paramétrique estimé.

Pour les 10 stations en Europe, j'ai tracé les courbes des positions résiduelles verticales de manière superposée sur la figure 4.7. Nous voyons une corrélation pour les séries temporelles de positions pour les stations colocalisées WTZA/WTZS/WTZZ et ZIMM/ZIMJ. La corrélation pour les stations non colocalisées est plus subtile, mais est, à certains endroits, perceptible, ce qui montre l'existence de corrélations spatiales entre séries temporelles de positions GNSS.

Afin de déterminer à la fois la dépendance spatiale et temporelle des données, nous utiliserons le variogramme spatiotemporel pour les bruits totaux de hauteur pour les 195 stations décrites en début de chapitre ; ce variogramme est présenté sur la figure 4.8 (nous verrons ultérieurement dans la section 4.7 comment considérer les composantes horizontales). La courbe rouge correspond à un variogramme spatial classique ; les autres courbes sont tracées pour des décalages temporels non nuls. Cette courbe rouge n'est pas plate ce qui montre l'existence d'une dépendance spatiale jusqu'à une distance angulaire de 30° ; ce résultat est cohérent avec les corrélogrammes présentés par [Amiri-Simkooei et al., 2017]. Ces informations peuvent être présentées avec le décalage temporel en abscisses comme sur la figure 4.9. Cette figure montre bien l'existence d'une dépendance temporelle dans le bruit total qui se traduit par une dépendance du variogramme en fonction du décalage temporel.

4.4 Choix du modèle temporel

Le modèle temporel choisi est constitué d'un bruit décorréolé (ou blanc) et d'un bruit corrélé, comme présenté au chapitre 3.

4.4.1 Bruit corrélé

Le bruit corrélé généralement utilisé pour décrire les corrélations temporelles des positions GNSS est le bruit de scintillation ; ce choix provient d'une analyse par maximum de vraisemblance pénalisée [Santamaría-Gómez et al., 2011]. Néanmoins, il est impossible dans un filtre de Kalman d'implémenter du bruit de scintillation, nous avons donc choisi d'utiliser un AR(1) qui ne correspond pas au véritable bruit présent dans les séries, mais permet néanmoins d'introduire des corrélations temporelles, de façon aisée, dans un filtre de Kalman. Rappelons la définition de l'AR(1) : un processus stochastique (\tilde{X}_t) est un AR(1) s'il existe $\alpha \in]-1, 1[$ tel que $\tilde{X}_t - \alpha\tilde{X}_{t-1}$ suivent des lois normales centrées indépendantes identiquement distribuées. J'appellerai α le paramètre d'échelle de l'AR(1) et la variance des lois normales centrées sera notée σ_{AR}^2 .

Il est habituel d'écrire α comme $\alpha = e^{-dt/\tau_0}$ avec dt le pas d'échantillonnage des séries et τ_0 un temps de corrélation.

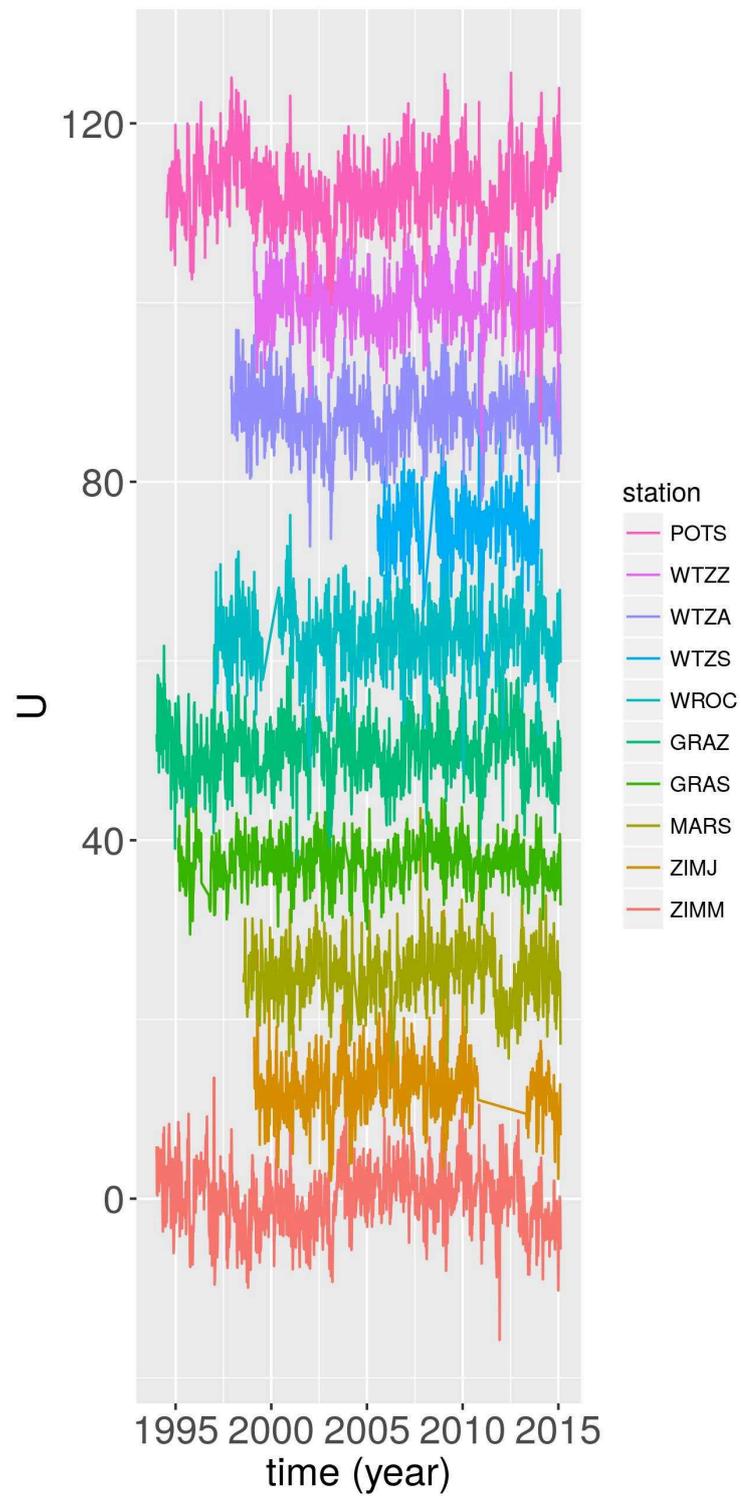


FIGURE 4.7 – Séries de positions résiduelles verticales pour les 10 stations en Europe. Les périodes considérées sont les périodes annuelle et semi-annuelle et les périodes draconitiques GPS jusqu'à la 7^e harmonique. Des offsets ont été ajoutés pour des raisons de lisibilité.

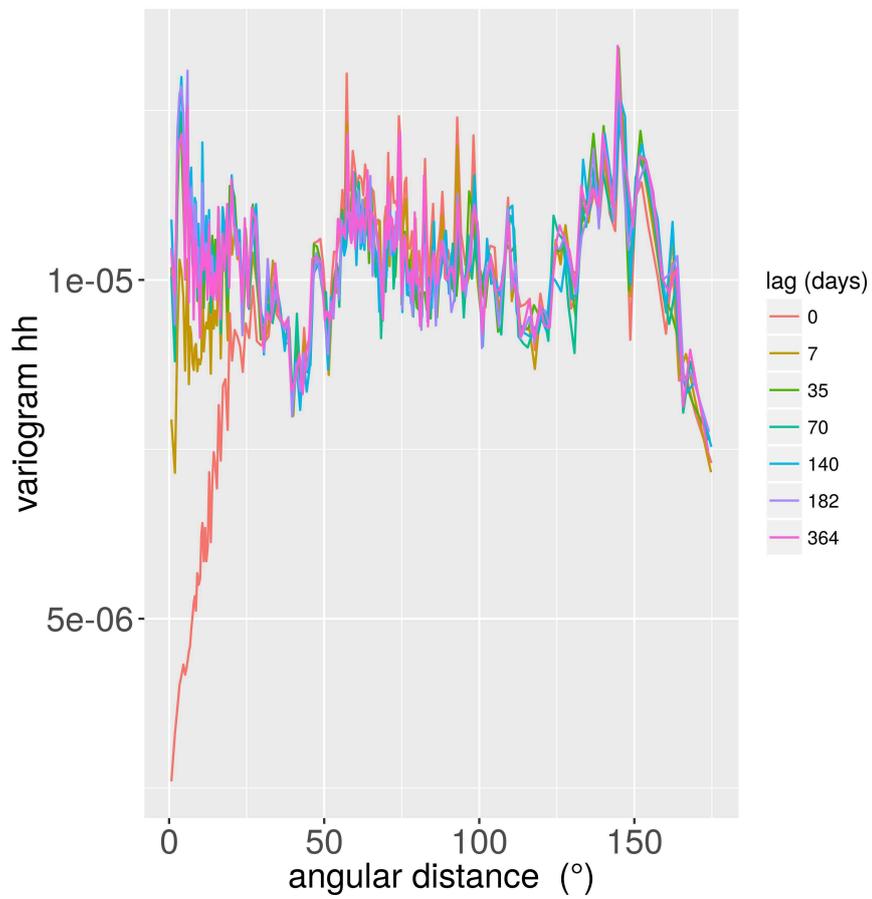


FIGURE 4.8 – Variogramme spatiotemporel (m^2) des séries temporelles de positions résiduelles verticales des 195 stations sélectionnées. Chaque courbe correspond à un décalage temporel.

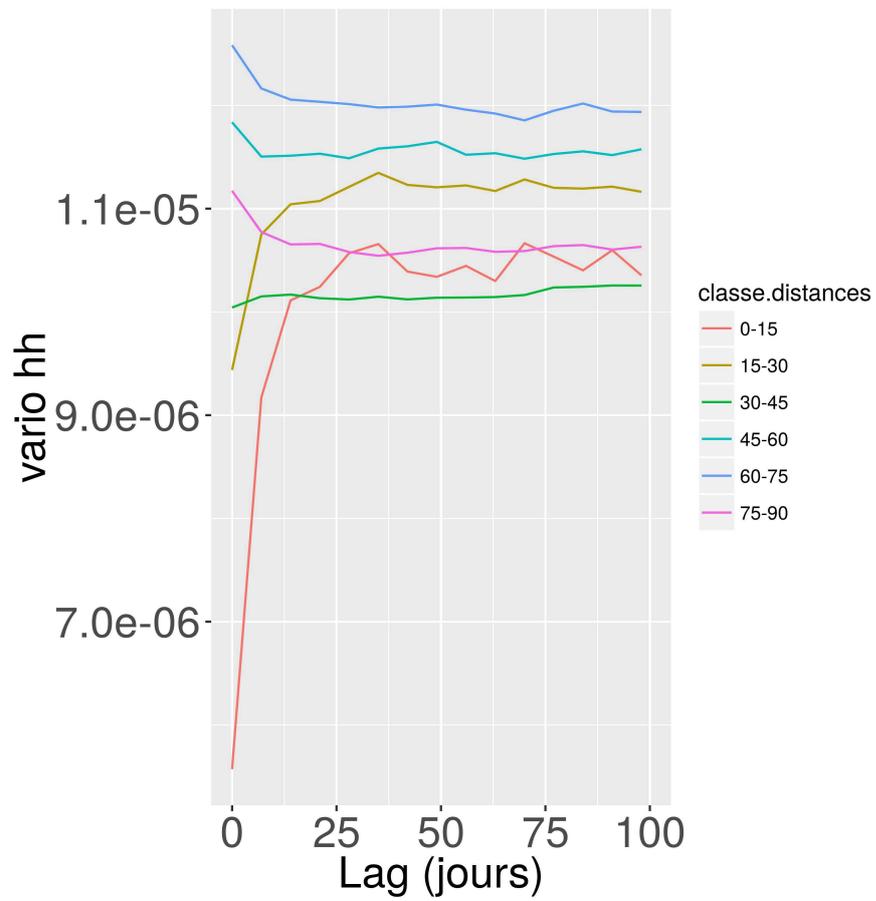


FIGURE 4.9 – Variogramme spatiotemporel (m^2) des séries temporelles de positions résiduelles verticales des 195 stations sélectionnées. Chaque courbe correspond à une classe de distance angulaire (en degré).

La fonction de covariance d'un AR(1) vaut :

$$C(\tau) = \frac{\sigma_{AR}^2}{1 - \alpha^2} \alpha^{|\tau|} = \frac{\sigma_{AR}^2}{1 - \alpha^2} e^{-|\tau|/\tau_0}.$$

La fonction d'autocorrélation de l'AR(1) est $\tau \mapsto e^{-|\tau|/\tau_0}$.

Par rapport au bruit de scintillation, l'AR(1) a de nombreux avantages :

- il est stationnaire, contrairement à la marche aléatoire et au bruit de scintillation, il contient un dispositif de retour vers la moyenne c'est-à-dire que si la série temporelle s'éloigne de 0, elle a tendance à être ramenée vers 0, cette « force de retour » est d'autant plus forte que l'éloignement de 0 est important, ce qui permet d'obtenir des positions moyennes fiables ;
- il s'implémente facilement dans un filtre de Kalman, sans aucune approximation.

Grâce au logiciel Hector [Bos et al., 2013a], j'ai estimé le paramètre d'échelle des AR(1) pour les séries temporelles des 195 stations sélectionnées en considérant chaque série temporelle individuellement et en ajoutant un bruit blanc homogène dans le modèle de bruit. La figure 4.10 présente les histogrammes des paramètres d'échelle avec un histogramme par coordonnées. Les histogrammes sont peu dispersés, en particulier pour les composantes horizontales et concentrés sur des valeurs proches de 1 ce qui est attendu si les temps d'autocorrélation des séries sont significativement plus grands que leur taux d'échantillonnage car, si, en reprenant les notations précédemment utilisées, $\tau_0 \gg 1$ alors $\alpha = e^{-1/\tau_0}$ est proche de 1. Pour avoir une meilleure idée de leur dispersion notamment pour les α proches de 1, nous avons donc représenté sur la figure 4.11 les histogrammes des temps d'autocorrélation estimés. On y note une dispersion relativement faible pour la composante verticale ; pour les composantes horizontales, cette dispersion est un peu supérieure ; l'artefact à 100 jours est lié au logiciel Hector, qui ne peut fournir des α supérieurs à 0.99, ce qui correspond à des temps de corrélation de 100 jours.

Afin de pouvoir obtenir des quantités homogènes sur lesquelles on inférera une dépendance spatiale, j'ai choisi un paramètre d'échelle des AR(1) commun à toutes les stations et à toutes les composantes, la valeur médiane des paramètres d'échelle des AR(1), qui correspond à un temps de corrélation de 19.7 jours.

4.4.2 Bruit blanc variable

À ce bruit corrélé s'ajoute un bruit décorrélé. Le modèle de bruit blanc que nous allons utiliser est le bruit blanc variable, dont l'écart-type est proportionnel aux erreurs formelles des coordonnées fournies avec les solutions IGS avec, comme facteur de proportionnalité, β . Ce choix du bruit blanc variable s'appuie sur les résultats de [Santamaría-Gómez et al., 2011].

Grâce au logiciel CATS [Williams, 2008], j'ai estimé ce facteur β par maximum de vraisemblance en considérant chaque série individuellement et un modèle de bruit composé d'un AR(1) et de bruit blanc variable.

Le choix d'un β unique pour les stations ou d'un β par station et composante se pose. La figure 4.12 présente l'histogramme des β estimés par CATS. La figure 4.12 montre une dispersion des β moyenne qui ne permet pas réellement de conclure ; nous allons donc tester des modèles avec un ou plusieurs β . Notons

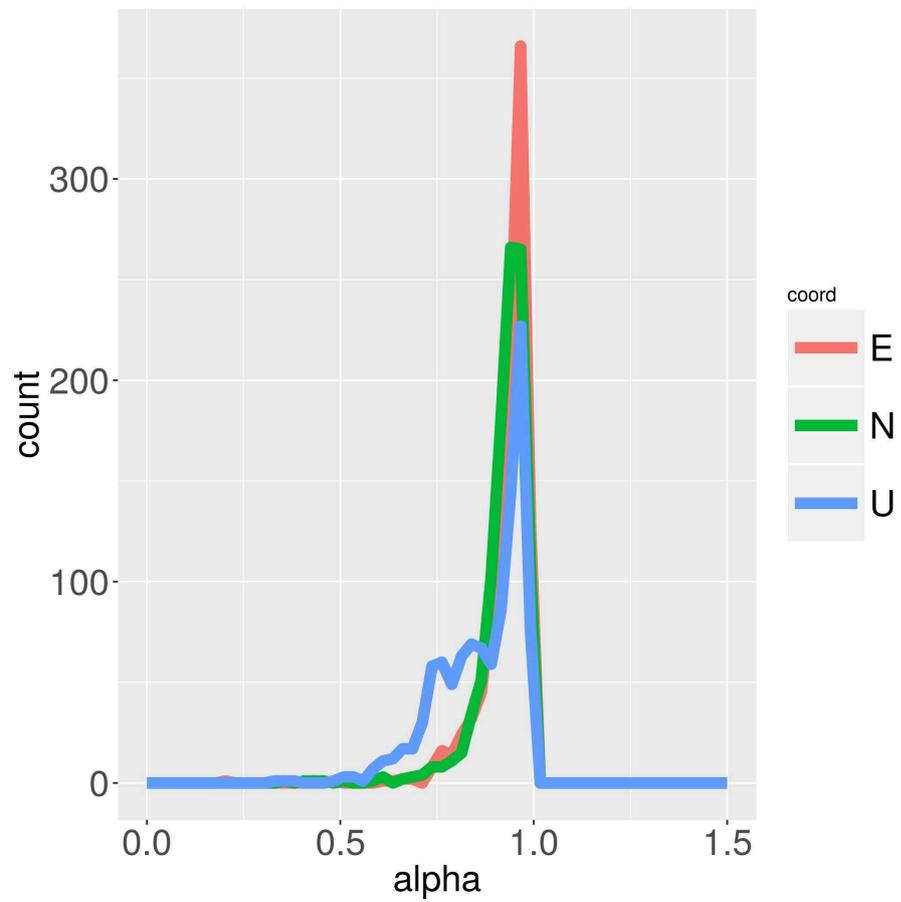


FIGURE 4.10 – Histogrammes des paramètres d'échelle des AR(1) estimés sous un modèle bruit blanc homogène et AR(1) par coordonnées.

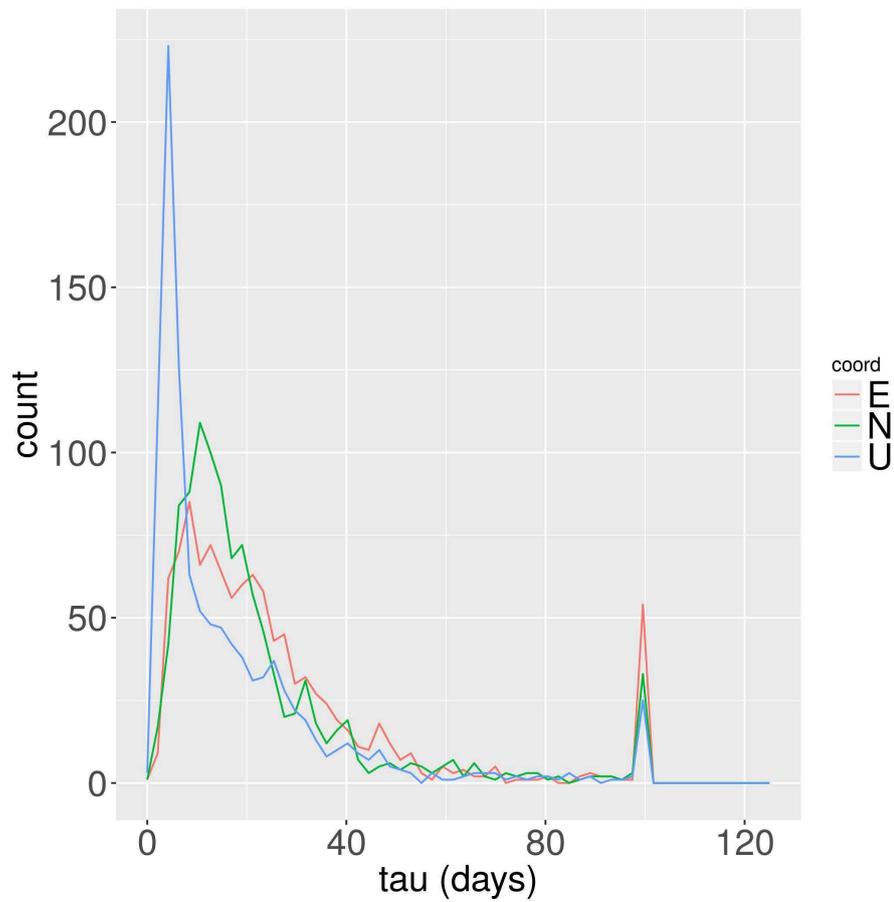
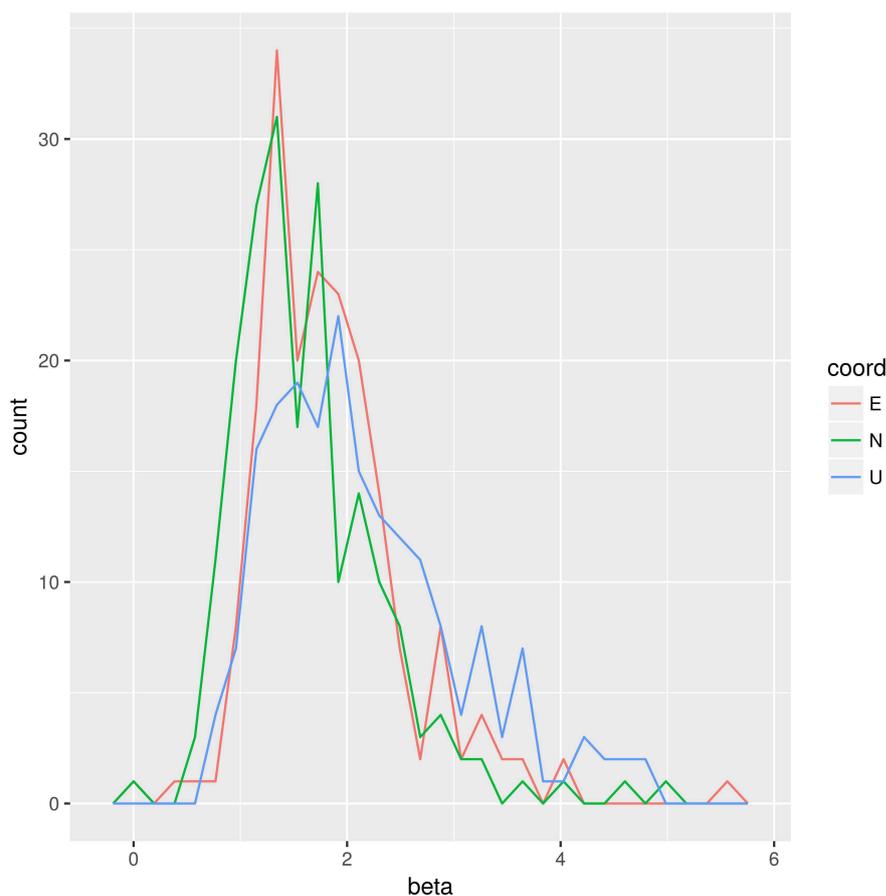


FIGURE 4.11 – Histogrammes des temps de corrélation des AR(1) estimés sous un modèle bruit blanc homogène et AR(1) par coordonnées.

FIGURE 4.12 – Histogramme des $\hat{\beta}$ estimés par maximum de vraisemblance

néanmoins que les β sont légèrement supérieurs à 1 ce qui correspond à une légère sous-estimation des erreurs formelles fournies par l'IGS. De plus, ils sont plus dispersés et légèrement plus élevés pour la composante verticale que pour les composantes horizontales.

4.4.3 Bruit total

Afin de valider le paramètre α choisi, il m'a paru intéressant de comparer, pour le bruit total, la puissance spectrale du modèle de bruit théorique avec celle du bruit expérimental. Pour cela, je m'appuierai sur le périodogramme de Lomb-Scargle défini par l'équation (1.2); je considérerai une moyenne de ces périodogrammes sur les stations ayant plus de 1000 mesures. Pour le périodogramme expérimental, j'ai calculé la moyenne des périodogrammes des bruits totaux normalisés par l'estimateur de leur écart-type; j'appellerai ce périodogramme le périodogramme normalisé cumulé expérimental.

Pour le périodogramme théorique, j'ai choisi un paramètre d'échelle d'AR(1) identique pour toutes les séries. La variance du bruit blanc variable et de l'AR(1) est quant à elle propre à chaque série à travers une estimation effectuée avec

le logiciel CATS. Pour le périodogramme théorique, j'ai considéré la moyenne des périodogrammes théoriques des bruits totaux normalisés par leur écart-type théorique ; j'appellerai ce périodogramme le périodogramme normalisé cumulé théorique. L'utilisation de périodogrammes normalisés permet d'obtenir des quantités comparables avant de faire la moyenne.

Cette comparaison est illustrée par la figure 4.13. Cette dernière montre que l'ajustement de la puissance spectrale du modèle de bruit théorique à celle du bruit expérimental est bon pour les basses fréquences jusqu'à 30 cycles par an. L'ajustement aux hautes fréquences est moins bon qu'aux basses fréquences, en particulier pour la composante verticale (il est impossible d'ajuster le périodogramme à la fois aux fréquences hautes, moyennes et basses ; le maximum de vraisemblance a privilégié les fréquences moyennes et basses au détriment des hautes fréquences). Ce mauvais ajustement n'est pas gênant outre mesure : la prédiction long-terme, en effet, est meilleure lorsque les basses fréquences sont bien modélisées que lorsque les hautes fréquences sont bien modélisées. D'autre part, l'essentiel de la puissance spectrale du bruit se situe aux basses fréquences.

4.5 Modèle spatio-temporel proposé

4.5.1 Forme du modèle espace-état

La dépendance spatiale sera modélisée par les matrices de variance-covariance de bruit d'un modèle espace-état. C'est pourquoi nous allons présenter le modèle espace-état dans cette section. Le modèle espace-état s'appuie sur les équations (1.3) et (1.4) dont nous allons reprendre les notations.

Nous considérons uniquement deux stations \vec{s}_1 et \vec{s}_2 pour avoir les notations les plus légères possible en confondant les stations et leurs positions (\vec{s}_1 et \vec{s}_2 étant les positions moyennes des stations et non leur position à l'échelle millimétrique qui varie au cours du temps) ; ces expressions se généralisent facilement à un nombre quelconque de stations. Nous allons décrire les éléments du filtre de Kalman. Les observations du filtre de Kalman sont les positions GNSS brutes :

$$Y_i = \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix}.$$

Les états contiennent les paramètres déterministes et les AR(1) normalisés :

$$X_i = (x_{1,1}^\circ, \dots, x_{1,n_1,pos}^\circ, x_{1,2}^\circ, \dots, x_{2,n_2,pos}^\circ, \dot{x}_{1,1}, \dots, \dot{x}_{1,n_1,vit}, \\ \dot{x}_{2,1}, \dots, \dot{x}_{2,n_2,vit}, a_{1,1}, a_{1,2}, \dots, a_{m,1}, a_{m,2}, \\ b_{1,1}, b_{1,2}, \dots, b_{m,1}, b_{m,2}, \tilde{\varepsilon}_{i,1}^{AR}, \tilde{\varepsilon}_{i,2}^{AR})^T,$$

où $\tilde{\varepsilon}_{i,k}^{AR}$ est la valeur de l'AR(1) de la k -ième station au temps t_i normalisée par son amplitude.

Le processus AR(1) à la date t_i pour la station k est égal à $\sigma_{AR} \tilde{\varepsilon}_{i,k}^{AR}$ où $\tilde{\varepsilon}_{i+1,k}^{AR} = \alpha \tilde{\varepsilon}_{i,k}^{AR} + \tilde{\eta}_{i,k}^{AR}$ et $\tilde{\eta}_{i,k}^{AR}$ est l'innovation normalisée.

Le bruit d'observation correspond aux bruits blancs :

$$B_t = (\varepsilon_{i,1}^{VW}, \varepsilon_{i,2}^{VW}), \quad (4.1)$$

où $\varepsilon_{i,k}^{VW}$ désigne la valeur du bruit blanc au i -ème instant et à la k -ième station.

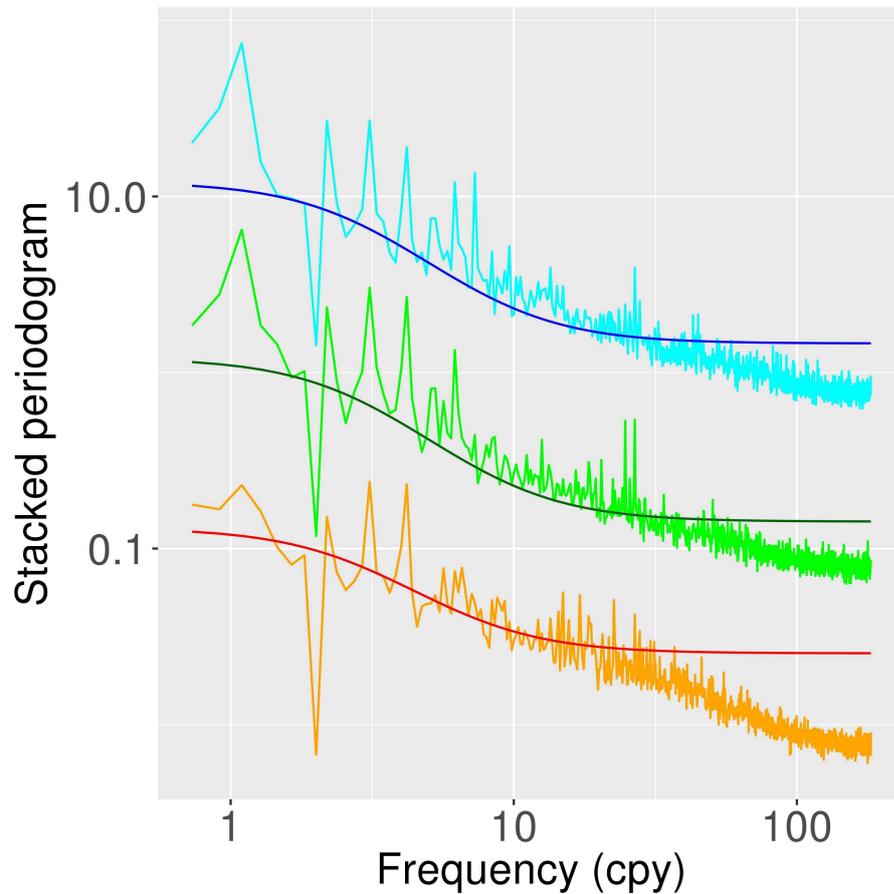


FIGURE 4.13 – Périodogramme normalisé cumulé théorique vs périodogramme normalisé cumulé expérimental des bruits totaux. Les périodogrammes expérimentaux ont des couleurs claires et les périodogrammes théoriques ont des couleurs foncées. Les couleurs associées aux composantes est, nord et verticale sont respectivement le bleu, le vert et le rouge (et l'orange). Des offsets ont été rajoutés pour des raisons de lisibilité.

La matrice d'observation H_i vaut :

$$\begin{pmatrix} \vec{\mathbb{I}}_{\alpha_1, pos(i), n_1, pos}^T & O \\ O & \vec{\mathbb{I}}_{\alpha_2, pos(i), n_2, pos}^T \\ (t_i - t_1^\circ) \vec{\mathbb{I}}_{\alpha_1, vit(i), n_1, vit}^T & O \\ O & (t_i - t_2^\circ) \vec{\mathbb{I}}_{\alpha_2, vit(i), n_2, vit}^T \\ \cos(\omega_1(t_i - t_1^\circ)) & 0 \\ 0 & \cos(\omega_1(t_i - t_2^\circ)) \\ \vdots & \vdots \\ \cos(\omega_m(t_i - t_1^\circ)) & 0 \\ 0 & \cos(\omega_m(t_i - t_2^\circ)) \\ \sin(\omega_1(t_i - t_1^\circ)) & 0 \\ 0 & \sin(\omega_1(t_i - t_2^\circ)) \\ \vdots & \vdots \\ \sin(\omega_m(t_i - t_1^\circ)) & 0 \\ 0 & \sin(\omega_m(t_i - t_2^\circ)) \\ \sigma_{AR,1} & 0 \\ 0 & \sigma_{AR,2} \end{pmatrix}^T,$$

où $\vec{\mathbb{I}}_{l,p}$ est le vecteur de dimension p qui vaut 0 partout sauf au niveau de sa l -ième position où il vaut 1 et $\sigma_{AR,k}^2$ est le paramètre de variabilité de l'AR(1) pour la k -ième station. Pour un même filtre de Kalman, afin de considérer des dépendances spatiales entre quantités homogènes, le paramètre d'échelle de toutes les stations et de toutes les coordonnées sera le même, seul le paramètre de variabilité pourra varier.

L'équation d'observation s'écrit :

$$\begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} = \begin{pmatrix} \bar{\mathbb{I}}_{\alpha_1, pos}^{T(i), n_1, pos} & O \\ O & \bar{\mathbb{I}}_{\alpha_2, pos}^{T(i), n_2, pos} \\ (t_i - t_1^\circ) \bar{\mathbb{I}}_{\alpha_1, vit}^{T(i), n_1, vit} & O \\ O & (t_i - t_2^\circ) \bar{\mathbb{I}}_{\alpha_2, vit}^{T(i), n_2, vit} \\ \cos(\omega_1(t_i - t_1^\circ)) & 0 \\ 0 & \cos(\omega_1(t_i - t_2^\circ)) \\ \vdots & \vdots \\ \cos(\omega_m(t_i - t_1^\circ)) & 0 \\ 0 & \cos(\omega_m(t_i - t_2^\circ)) \\ \sin(\omega_1(t_i - t_1^\circ)) & 0 \\ 0 & \sin(\omega_1(t_i - t_2^\circ)) \\ \vdots & \vdots \\ \sin(\omega_m(t_i - t_1^\circ)) & 0 \\ 0 & \sin(\omega_m(t_i - t_2^\circ)) \\ \sigma_{AR,1} & 0 \\ 0 & \sigma_{AR,2} \end{pmatrix}^T \begin{pmatrix} x_{1,1}^\circ \\ \vdots \\ x_{1, n_1, pos} \\ x_{2,1}^\circ \\ \vdots \\ x_{1, n_2, pos} \\ \dot{x}_{1,1} \\ \vdots \\ \dot{x}_{1, n_2, vit} \\ a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{m,1} \\ a_{m,2} \\ b_{1,1} \\ b_{1,2} \\ \vdots \\ b_{m,1} \\ b_{m,2} \\ \tilde{\varepsilon}_{i,1}^{AR} \\ \tilde{\varepsilon}_{i,2}^{AR} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i,1}^{VW} \\ \varepsilon_{i,2}^{VW} \end{pmatrix},$$

où les $\sigma_{AR,k}$ utilisés sont préalablement estimés à l'aide de CATS.

Le bruit de transition A_i contient des zéros aux positions correspondant aux paramètres déterministes, les dernières positions contiennent l'innovation des AR(1) :

$$A_i = (0, \dots, 0, \tilde{\eta}_{i,1}^{AR}, \tilde{\eta}_{i,2}^{AR})^T. \quad (4.2)$$

La matrice de transition L_i est la matrice diagonale d'éléments diagonaux 1 sauf pour les éléments diagonaux correspondant à la partie aléatoire qui valent le paramètre d'échelle α de l'AR(1) :

$$\begin{pmatrix} 1 & & & \\ & \ddots & & (0) \\ & & 1 & \\ (0) & & & \alpha \\ & & & & \alpha \end{pmatrix}.$$

L'équation de transition sera de la forme :

$$X_{i+1} = \begin{pmatrix} 1 & & & \\ & \ddots & & (0) \\ & & 1 & \\ (0) & & & \alpha \\ & & & & \alpha \end{pmatrix} X_i + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \tilde{\eta}_{i,1}^{AR} \\ \tilde{\eta}_{i,2}^{AR} \end{pmatrix}.$$

Les matrices de covariance du modèle espace-état seront détaillées ci-dessous, car plusieurs modèles seront proposés pour être évalués.

Valeur et variance a priori du modèle espace-état La variance a priori des paramètres déterministes (ou paramètres de régression) est fixée proportionnelle à la variance issue des résultats de CATS. Ce facteur multiplicatif est fixé à 20.

Pour évaluer l'impact de ce facteur multiplicatif, une étude de la série de hauteur de la station MARS (Marseille, France) a été conduite. Le passage de ce facteur multiplicatif de 20 à 1000 engendre une différence de $2\mu\text{m}/\text{an}$ pour la vitesse et de $6\mu\text{m}/\text{an}$ pour son incertitude, les calculs ayant été effectués dans un filtre de Kalman pour MARS uniquement. Cet impact négligeable valide le choix de ce coefficient multiplicatif. La variance a priori d'un AR(1) est fixée de manière naturelle à la variance d'un AR(1) :

$$\frac{\sigma_{AR}^2}{1 - \alpha^2}.$$

Les états sont initialisés aux valeurs estimées par maximum de vraisemblance pour les paramètres déterministes et fixés à zéro pour les AR(1).

Fonction de covariance du bruit total Pour un modèle quelconque, calculons la covariance entre $\varepsilon_{t,k_1,AR}$ et $\varepsilon_{t+\tau,k_2,AR}$ en supposant $\tau \geq 0$. En appliquant la définition par récurrence de l'AR(1), on a :

$$\text{Cov}(\varepsilon_{t,k_1,AR}, \varepsilon_{t+\tau,k_2,AR}) = \text{Cov}(\varepsilon_{t,k_1,AR}, \alpha^\tau \varepsilon_{t,k_2,AR} + f(\eta_{t+1}, \dots, \eta_{t+\tau-1}))$$

où f est une fonction. Donc, on a :

$$\text{Cov}(\varepsilon_{t,k_1,AR}, \varepsilon_{t+\tau,k_2,AR}) = \sigma_{k_1,AR} \sigma_{k_2,AR} \alpha^\tau t_{i,j}^{AR}$$

où $t_{i,j}^{AR}$ est le terme (i, j) de la matrice T^{AR} .

Pour le bruit total, la covariance vaut :

$$\text{Cov}(\varepsilon_{t,k_1}^{tot}, \varepsilon_{t+\tau,k_2}^{tot}) = \sigma_{k_1,AR} \sigma_{k_2,AR} \alpha^\tau t_{k_1,k_2}^{AR} + \sigma_{t,k_1,k_2,WN}^2 \delta_{\tau,0} \quad (4.3)$$

où $\delta_{\cdot,\cdot}$ est le symbole de Kronecker et $\sigma_{t,k_1,k_2,WN}^2$ est la covariance du bruit blanc à l'instant t entre les stations k_1 et k_2 .

4.6 Définition des différents modèles

Nous allons proposer différents modèles de covariance spatiale obtenus en jouant sur la matrice de variance-covariance M_t du bruit blanc variable B_t et sur la matrice de variance-covariance T_t du bruit de transition A_t :

- modèle SCS ("sans corrélation spatiale") : sans corrélation spatiale de l'AR(1) et du bruit blanc VW ; la matrice de variance-covariance du bruit blanc variable est diagonale ; l'amplitude du bruit blanc variable VW est remise à l'échelle par un facteur β_k différent pour chaque station et pour chaque composante ;
- modèle ACS-PROCHE-ITRF ("avec corrélation spatiale proche ITRF") : sans corrélation spatiale de l'AR(1) et avec corrélation spatiale du bruit blanc VW ; contrairement au modèle SCS, la matrice de variance-covariance du bruit blanc variable est pleine afin d'avoir un modèle de bruit blanc qui correspond à la pratique actuelle du calcul de l'ITRF ; l'amplitude du bruit blanc variable VW est remise à l'échelle par un facteur β_k différent pour chaque station ;

- modèle ACS-EMP1 ("avec corrélation spatiale empirique 1") : spatiotemporel avec utilisation de la covariance empirique pour la corrélation spatiale de l'AR(1) et avec corrélation spatiale du bruit blanc VW ; un facteur d'échelle différent β_k pour l'amplitude du bruit blanc variable VW est proposé ;
- modèle ACS-EMP2 ("avec corrélation spatiale empirique 2") : spatiotemporel avec utilisation de la covariance empirique pour la corrélation spatiale de l'AR(1) et avec corrélation spatiale du bruit blanc VW ; un facteur d'échelle unique β pour l'amplitude du bruit blanc variable VW est utilisé ;
- modèle ACS-VARIO ("avec corrélation spatiale variogramme") : spatiotemporel avec utilisation de la covariance basée sur un modèle de variogramme pour l'AR(1) et avec corrélation spatiale du bruit non corrélé VW. Une amplitude β_k du bruit blanc variable différente pour chaque station a été utilisée.

Notons ici que pour les modèles ACS-EMP1 et ACS-EMP2, deux alternatives sont testées pour la modélisation de la dépendance spatiale du bruit temporellement décorrélé. Nous verrons au chapitre suivant que cette option a peu d'effet sur les résultats. De façon arbitraire, la remise à l'échelle par un facteur β_k différent pour chaque station du bruit décorrélé VW a été adoptée pour les modèles ACS-PROCHE-ITRF et ACS-VARIO.

Pour la partie temporelle, pour tous les modèles, T est diagonale par bloc :

$$\text{Cov}(A_t) = T = \begin{pmatrix} O & O \\ O & T_{AR} \end{pmatrix}$$

où l'écriture par bloc sépare les paramètres déterministes des paramètres de l'AR(1).

Pour rester cohérents avec la section précédente, nous expliciterons les équations pour deux stations uniquement, la généralisation à plusieurs stations étant aisée.

Modèle SCS

La matrice de variance-covariance des innovations vaut la matrice identité :

$$T_{AR} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (4.4)$$

et la matrice de variance-covariance du bruit blanc est diagonale :

$$M_t = \text{Cov}(B_t) = \begin{pmatrix} \beta_1^2 v_{1,1,t} & 0 \\ 0 & \beta_2^2 v_{2,2,t} \end{pmatrix}, \quad (4.5)$$

où $v_{i,j,t}$ est la covariance du bruit blanc fournie dans les solutions IGS à l'époque t entre les stations i et j ; la matrice de variance-covariance de la solution IGS de l'instant t sera notée $V_t = (v_{i,j,t})_{i,j}$.

Modèle ACS-PROCHE-ITRF

Le modèle ACS-PROCHE-ITRF a pour but de se rapprocher de la méthodologie actuellement utilisée pour le calcul de l'ITRF. La matrice de variance-covariance

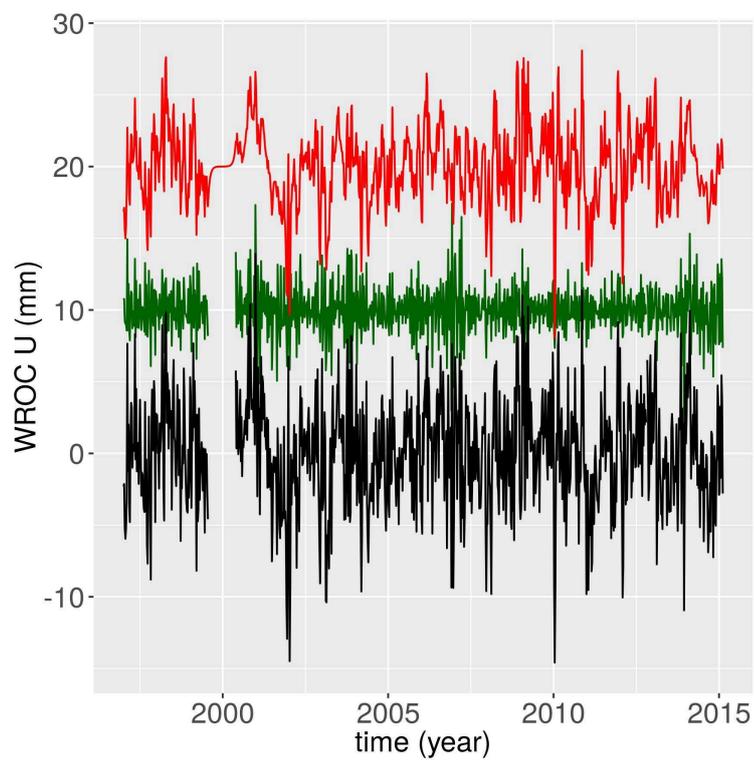


FIGURE 4.14 – Sont présentés le bruit total en noir de la série de positions verticales de la station WROC (Wroclaw, Pologne) ainsi que l'AR(1) estimé par filtre de Kalman en rouge et le bruit blanc estimé par filtre de Kalman en vert. Les offsets ont été ajoutés pour des raisons de lisibilité.

des innovations est, comme pour le modèle SCS, donnée par l'équation (4.4), mais la matrice de variance-covariance des bruits blancs est pleine et non uniquement diagonale :

$$M_t = \text{Cov}(B_t) = \hat{\beta}^T S_t \hat{\beta} = \begin{pmatrix} \hat{\beta}_1^2 v_{1,1,t} & \hat{\beta}_1 \hat{\beta}_2 v_{1,2,t} \\ \hat{\beta}_1 \hat{\beta}_2 v_{2,1,t} & \hat{\beta}_2^2 v_{2,2,t} \end{pmatrix}, \quad (4.6)$$

où $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^T$ est le vecteur des amplitudes du bruit blanc variable estimées par maximum de vraisemblance (CATS). Cette formulation permet de remettre à l'échelle les variances tout en conservant les corrélations.

Modèle ACS-EMP1

Le filtre de Kalman du modèle SCS fournit des séries de bruit blanc et d'AR(1) comme présenté sur la figure 4.14, mais aussi des séries d'innovation d'AR(1). On dispose ainsi d'une estimation des innovations des différents AR(1) pour chaque date. Les covariances du bruit de transition T_{k_1, k_2}^{AR} seront déduites de ces séries préalablement estimées : ce principe aura cours aussi dans les modèles ACS-EMP2 et ACS-VARIO.

Le taux de valeurs manquantes sur les stations considérées étant faibles (5,9 %), l'inférence de la matrice de variance-covariance des erreurs de transition pourra se faire sur l'ensemble des estimations des innovations fournies par le filtre, y compris les valeurs interpolées.

Dans le modèle ACS-EMP1, la matrice $(T_{k_1, k_2}^{AR})_{1 \leq k_1, k_2 \leq n_{stations}}$ est, sauf pour la diagonale, la covariance empirique des innovations des AR(1) normalisées estimées par le modèle SCS des stations k_1 et k_2 ; la diagonale est prise égale à 1.

Si $\hat{\varepsilon}_{i,k,AR}$ est l'AR(1) et $\hat{\varepsilon}_{i,k,AR} / \sigma_{k,AR}$ est l'AR(1) normalisé où k est le numéro de la station, alors $(T_{k_1, k_2}^{AR})_{k_1, k_2}$ ($1 \leq k_1, k_2 \leq n_{stations}$) est la covariance empirique entre $\hat{\varepsilon}_{\cdot, k_1, AR}$ et $\hat{\varepsilon}_{\cdot, k_2, AR}$:

$$\hat{\rho}_{k_1, k_2} = \frac{1}{\text{Card}(\Omega) - 1} \sum_{\Omega} (\hat{\varepsilon}_{i, k_1, AR} - \bar{\varepsilon}_{\cdot, k_1, AR}) (\hat{\varepsilon}_{j, k_2, AR} - \bar{\varepsilon}_{\cdot, k_2, AR}), \quad (4.7)$$

où $\Omega = \{(i, j) | t_{i, k_1} = t_{j, k_2}\}$, et $\bar{\varepsilon}_{\cdot, k, AR}$ correspondent à des moyennes empiriques. Ainsi la matrice de covariance-covariance de l'innovation des AR(1) qui est symétrique s'écrit-elle :

$$T^{AR} = \begin{pmatrix} 1 & \hat{\rho}_{1,2} \\ \hat{\rho}_{2,1} & 1 \end{pmatrix}. \quad (4.8)$$

La matrice M_t est prise en utilisant le produit (4.6).

Lorsqu'il n'y a pas de chevauchement dans les séries temporelles des stations k_1 et k_2 , le terme T_{k_1, k_2} est fixé à 0.

Modèle ACS-EMP2

Ce modèle diffère du modèle ACS-EMP1 uniquement par la matrice de covariance du bruit des observations.

La matrice M_t est prise en utilisant le produit

$$M_t = \bar{\beta}^2 V_t, \quad (4.9)$$

où V_t est la matrice de variance-covariance de la solution IGS au temps t . $\bar{\beta}$ est la médiane des coefficients du bruit blanc variable sur l'ensemble des stations du filtre.

Tout comme le modèle précédent, la matrice de variance-covariance des AR(1) est donnée par l'équation (4.8).

Les modèles ACS-EMP1 et ACS-EMP2 ont l'avantage de pouvoir bien s'adapter à des covariances quelconques, mais possèdent des limites théoriques : il n'est pas impossible d'obtenir des incertitudes négatives sur les paramètres estimés du fait que T_{AR} n'est pas forcément définie positive. L'utilisation des variogrammes dans le modèle ACS-VARIO permet de se protéger d'un tel risque. Ce risque est néanmoins théorique ; dans tous les calculs que j'ai exécutés, ce cas ne s'est jamais présenté.

Modèle ACS-VARIO

Le modèle ACS-VARIO possède la même matrice de covariance du bruit des observations que les modèles ACS-PROCHE-ITRF et ACS-EMP1. Elles sont définies par l'équation (4.6).

La matrice T^{AR} est quant elle basée sur une modélisation des dépendances spatiales basée sur les variogrammes. Pour une composante donnée, la matrice de variance-covariance des innovations normalisées des AR(1) est remplie selon une fonction de la distance h_{k_1, k_2} qui sépare les stations k_1 et k_2 , et qui est ajustée sur des variogrammes empiriques.

Le modèle proposé est le modèle exponentiel, auquel s'ajoute un effet pépite ; nous avons choisi le modèle exponentiel car les processus gaussiens (spatiaux) sous-jacents sont continus, mais pas forcément dérivables ce qui permet de s'adapter à des données réelles. De plus, nous verrons sur la figure 5.2 que l'ajustement de ce variogramme théorique au variogramme empirique est correct. La covariance entre deux séries sur une même composante séparées de h se calculera comme :

$$\Gamma \delta_{h,0} + \sigma_{exp}^2 \exp(-h/h_0),$$

où Γ représente l'effet pépite et h_0 et σ_{exp}^2 représentent respectivement la portée et le palier du variogramme exponentiel.

Les paramètres du variogramme sont estimés par moindres carrés non linéaires à partir du variogramme empirique, lui-même calculé à partir de la nuée variographique (cf. section 5.1).

Ce modèle a l'avantage de rendre les matrices de variance-covariance définies positives, mais fait l'hypothèse d'une covariance isotrope exponentielle du bruit corrélé que les données ne vérifient pas parfaitement.

4.7 Modèle ACS-VARIO : choix d'un nouveau système de coordonnées

Dans la section précédente, le modèle ACS-VARIO est présenté pour une des trois composantes, mais les composantes horizontales du repère topocentrique local posent problème au voisinage des pôles. Par exemple, la composante nord d'une station d'un côté du pôle nord et cette même composante d'une station située de l'autre côté du pôle nord pointeront l'une vers l'autre. Pour des stations situées en Europe par contre, les deux vecteurs nord seront orientés dans la

même direction. Afin de lever cette singularité, il est nécessaire d'utiliser un nouveau système de coordonnées [Collilieux et al., 2012]. Ce repère va permettre de comparer les déplacements horizontaux de deux points quelles que soient leurs positions sur le globe. Les covariances seront associées à une paire de points : il nous suffit donc de considérer les covariances pour deux points A et B sur une sphère assimilée à la surface de la Terre. Les longitudes de A et B (respectivement les latitudes de A et B) seront notées λ_A et λ_B (respectivement φ_A et φ_B). Nous définissons au point A un repère orthonormal direct $(\vec{u}_A, \vec{v}_A, \vec{h}_A)$ où

- \vec{u}_A est le vecteur dirigé de A vers B le long du grand cercle reliant les deux points,
- \vec{h}_A est le vecteur unitaire dirigé selon la verticale en A ,
- \vec{v}_A complète le trièdre orthonormal direct.

Il en va de même pour le point B , on définit un repère orthonormal direct $(\vec{u}_B, \vec{v}_B, \vec{h}_B)$ où :

- \vec{u}_B est le vecteur dirigé de B vers la direction opposée à A de manière tangente au grand cercle reliant les deux points,
- \vec{h}_B est le vecteur unitaire dirigé selon la verticale en B ,
- \vec{v}_B complète le trièdre orthonormal direct.

Cette disposition n'annihile pas les singularités, il faut être d'autant plus prudent si la distance angulaire de deux points est proche de 180° : cette construction n'a mathématiquement pas de sens pour des points diamétralement opposés. Néanmoins, pour des points très proches, cette construction a toujours un sens, contrairement à l'utilisation des coordonnées est et nord qui comportent une singularité aux pôles. L'astuce qui fait que cette construction fonctionne, sous réserve que la portée soit petite, est que, pour des points éloignés, la covariance est proche de 0, ce qui fait que la singularité n'a pas d'importance.

Notons ψ l'angle au centre entre A et B ; il se déduit des longitudes et latitudes de A et B en s'appuyant sur la trigonométrie sphérique et la figure 4.15 :

$$\psi = \arccos(\sin \varphi_A \sin \varphi_B + \cos \varphi_A \cos \varphi_B \cos(\lambda_A - \lambda_B)).$$

Les azimuts en A et B qui sont a priori différents sont donnés par les formules suivantes :

$$\sin(Az_A) = \frac{\cos(\varphi_B) \sin(\lambda_B - \lambda_A)}{\sin \psi},$$

$$\sin(Az_B) = -\frac{\cos(\varphi_A) \sin(\lambda_B - \lambda_A)}{\sin \psi}.$$

Ces deux formules ne sont valables que si A et B sont orientés d'ouest en est et si la quantité $\lambda_B - \lambda_A$ est positive. Il faut donc modifier les angles et éventuellement intervertir A et B :

- si $\lambda_B - \lambda_A \in [0, \pi]$, il ne faut rien modifier ;
- si $\lambda_B - \lambda_A > \pi$, on remplace licitement λ_B par $\lambda_B + 2\pi$ et on permute A et B ;
- si $\lambda_B - \lambda_A \in]-\pi, 0[$, on permute A et B ;
- si $\lambda_B - \lambda_A < -\pi$, on remplace licitement λ_A par $\lambda_A - 2\pi$;
- si $\lambda_B = \lambda_A$ et $\varphi_A > \varphi_B$, on permute A et B et on choisit $Az_A = Az_B = 0$;
- si $\lambda_B = \lambda_A$ et $\varphi_A = \varphi_B$, on choisit, par exemple, $Az_A = Az_B = \frac{\pi}{2}$.

Ainsi les coordonnées (u, v, h) se déduisent-elles linéairement des coordonnées (e, n, h) grâce à la formule :

$$\begin{pmatrix} u_A \\ v_A \\ h_A \end{pmatrix} = R(Az_A) \begin{pmatrix} e_A \\ n_A \\ h_A \end{pmatrix},$$

où

$$R(\theta) = \begin{pmatrix} \sin \theta & \cos \theta & 0 \\ -\cos \theta & \sin \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

La formule pour le point B est similaire en remplaçant $R(Az_A)$ par $R(Az_B)$:

$$\begin{pmatrix} u_B \\ v_B \\ h_B \end{pmatrix} = R(Az_B) \begin{pmatrix} e_B \\ n_B \\ h_B \end{pmatrix}.$$

La matrice de variance-covariance est transformée des coordonnées (e, n) vers les coordonnées (u, v) de la manière suivante :

$$\text{Cov} \begin{pmatrix} u_A \\ v_A \\ h_A \\ u_B \\ v_B \\ h_B \end{pmatrix} = \begin{pmatrix} R(Az_A) & O \\ O & R(Az_B) \end{pmatrix} \text{Cov} \begin{pmatrix} e_A \\ n_A \\ h_A \\ e_B \\ n_B \\ h_B \end{pmatrix} \begin{pmatrix} R(Az_A)^T & O \\ O & R(Az_B)^T \end{pmatrix}.$$

4.8 Choix de modèle : validation croisée

Savoir a priori quel modèle est le meilleur est difficile. L'ajout de paramètres n'est pas la panacée au-delà de la surcharge de calculs qu'il implique : l'ajout d'un trop grand nombre de paramètres diminue l'erreur sur la base d'apprentissage, mais peut détériorer la prédiction (en particulier l'extrapolation, mais aussi l'intrapolation dans une moindre mesure). Ce trop grand nombre de paramètres conduit à l'ajustement du bruit statistique, ce qui est appelé surapprentissage. D'autre part, l'utilisation de peu de paramètres induit des hypothèses souvent implicites ce qui peut être gênant dans le cadre de données réelles. Dans notre cas, par exemple, la covariance empirique a l'avantage d'utiliser peu d'hypothèses (on considère la covariance des innovations gaussienne, en plus d'être constante au cours du temps), ce qui permet de s'adapter à davantage de situations, avec, comme revers de la médaille, le risque de surapprentissage. En revanche, l'utilisation du variogramme a l'avantage d'expliquer les dépendances spatiales de manière simple et lisible avec un risque moindre de surapprentissage, mais on peut douter de sa véracité. Mais, comme dirait George Box, « tous les modèles sont faux, mais certains sont utiles ».

La validation croisée consiste à retirer une partie des données de la base d'apprentissage d'une méthode statistique et à réestimer les données à partir de la base d'apprentissage réduite puis à comparer les données réelles et les données réestimées. Dans la pratique, nous supprimons un an de données, ce

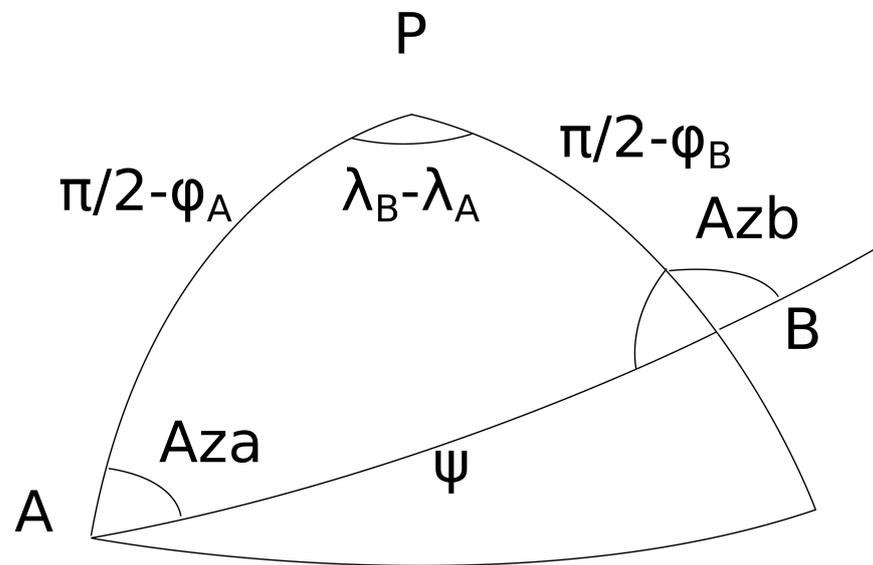


FIGURE 4.15 – Représentation du triangle formé par le pôle Nord et les deux points A et B , triangle dans lequel nous appliquerons les formules de géométrie du triangle sur une sphère

qui correspond à un ensemble d'instantanés temporels I_k de mesures, pour chaque station successivement. Pour chaque modèle proposé, nous réestimons la position (notons sa valeur vraie $x_{t,k}$) par $\hat{x}_{t,k}$ (qui correspond à la somme de l'estimation de la tendance linéaire et périodique et de l'AR(1)) où k désigne le numéro de la station et t l'instant considéré. Nous utilisons un indicateur appelé RMS pour mesurer l'écart entre les valeurs vraies et les valeurs réestimées :

$$\sqrt{\text{mean} \{((x_{t,k} - \hat{x}_{t,k})^2) | k, t \in I_k\}}. \quad (4.10)$$

Sont éliminées du RMS les valeurs correspondant aux instants entre le premier saut appartenant à la zone de test et le dernier saut, car la réestimation de ces valeurs s'appuie sur des ordonnées à l'origine qui ne peuvent être déduites de la base d'apprentissage (elles ne peuvent être déduites que de la base de test). Lorsqu'il n'y a qu'un saut ou qu'il n'y a pas de saut, aucune valeur n'est éliminée du RMS.

Lorsque l'on considère plusieurs composantes, le RMS correspondant vaut la moyenne quadratique des RMS de chacune des composantes.

La validation croisée prend pour critère de sélection de modèle la minimisation de l'erreur de prédiction, qui est un estimateur naturel du bruit blanc. Cette minimisation de l'estimateur du bruit blanc est un choix classique dans les statistiques, mais interrogeable, car il existe peut-être du signal géophysique dans le bruit blanc. Ce choix est bon lorsqu'il s'agit de faire de la prédiction dans le futur ou d'imputer des valeurs manquantes, car il est impossible de prédire un bruit décorrélé à la fois temporellement et spatialement dans le futur en l'absence de données extérieures. Le RMS ne peut être inférieur à un seuil qui correspond à la moyenne quadratique d'un bruit décorrélé à la fois temporellement et spatialement. La méthode sélectionnée par la validation croisée n'est cependant pas adaptée lorsque l'on veut séparer de manière exacte le bruit blanc du bruit corrélé, car minimiser le bruit blanc, ce que réalise cette méthode, n'est pas le bon critère pour cet objectif.

Les figures 4.16 et 4.17 illustrent la validation croisée. Les positions pour la station POTS (Postdam, Allemagne) entre 2010 et 2011 sont éliminées de la base d'apprentissage et ensuite réestimées par filtre de Kalman sur la base de différents modèles de covariance spatiotemporelle.

4.9 Vérification de l'amélioration de l'estimation des vitesses

On peut vérifier si la prise en compte de la dépendance spatiale pour les séries courtes améliore l'estimation des vitesses. En effet, dans le cas de séries courtes, les signaux temporellement corrélés viennent d'autant plus fausser les vitesses estimées. Or ces signaux corrélés pourront être déterminés à l'aide des séries de stations proches.

Pour cela, nous considérerons une série longue d'une station qui aura une vitesse r considérée comme sa valeur vraie. Ensuite, nous tronquerons la série à quelques années de données. Nous pouvons estimer la vitesse de deux manières différentes :

- on peut estimer la vitesse en considérant la station individuellement ; notons cet estimateur \hat{r}_0 ;

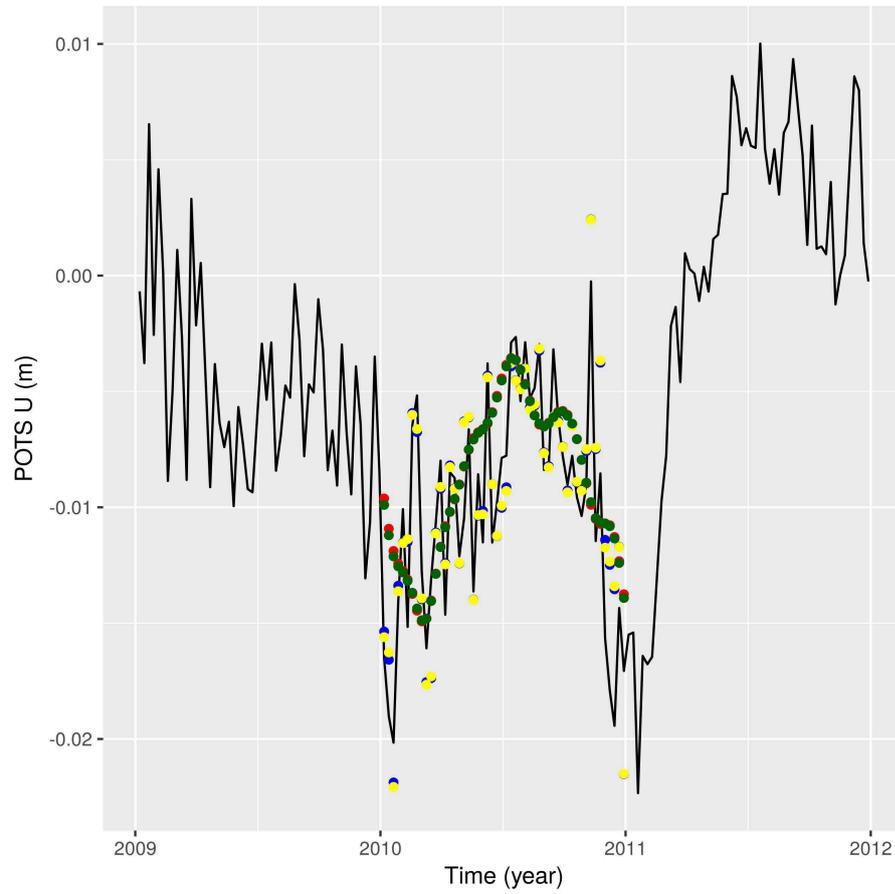


FIGURE 4.16 – Série temporelle de positions verticales (noir) de la station POTS (Potsdam, Allemagne) avec leurs estimations par validation croisée (rouge : modèle SCS ; bleu : mod. ACS-EMP1 ; jaune : modèle ACS-EMP2 ; vert : modèle ACS-VARIO).

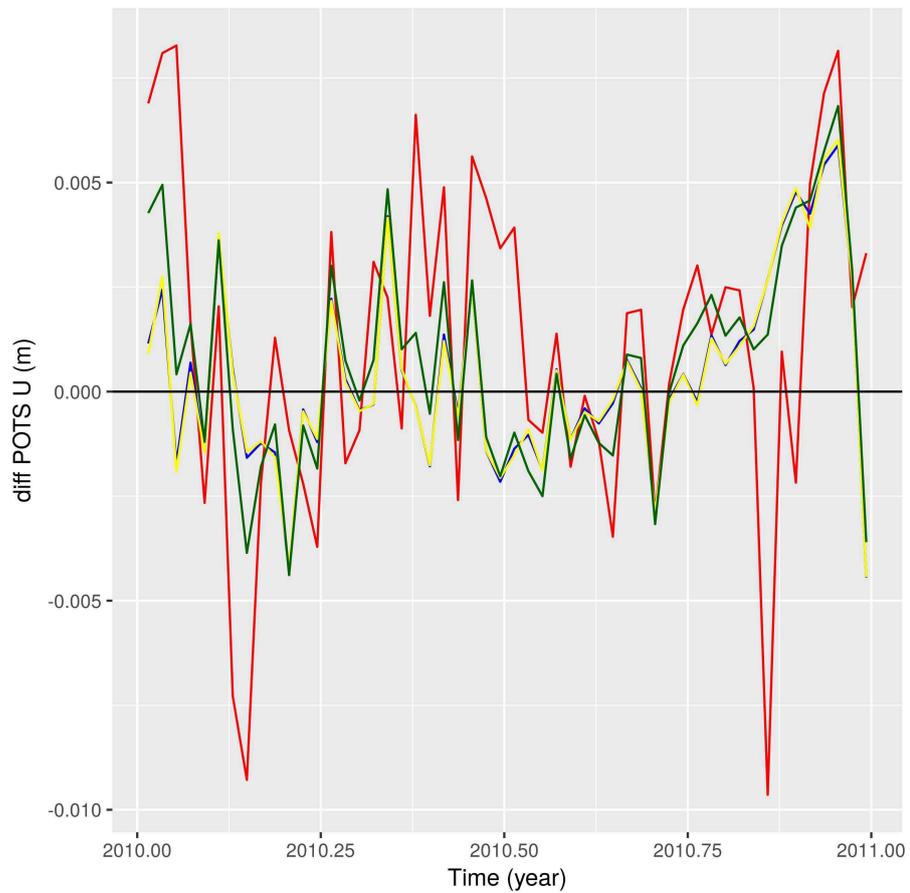


FIGURE 4.17 – Séries temporelles des erreurs sur les positions verticales de la station POTS (Potsdam, Allemagne) issues de la validation croisée (rouge : modèle SCS ; bleu : mod. ACS-EMP1 ; jaune : modèle ACS-EMP2 ; vert : modèle ACS-VARIO).

— on peut estimer la vitesse en utilisant la dépendance spatiotemporelle selon une des méthodes ci-dessus, les autres stations n'ayant pas de données supprimées ; notons cet estimateur \hat{r}_{st} .

L'utilisation de la dépendance spatiale sur la série courte devrait réduire l'erreur sur l'estimateur de la vitesse : $|r - \hat{r}_{st}|$ devrait être inférieure à $|r - \hat{r}_0|$.

Le chapitre a donné l'ensemble de la méthodologie pour montrer l'apport de la prise en compte de la dépendance spatiotemporelle, méthodologie qui sera mise en œuvre dans le chapitre suivant.

Chapitre 5

Résultats et discussion

Ce chapitre présente les résultats qui ont été obtenus à partir de la méthodologie décrite au chapitre 4. Le chapitre fournira les variogrammes des innovations des AR(1) indispensables pour exécuter le filtre de Kalman du modèle ACS-VARIO. Ensuite, nous présenterons les résultats de la validation croisée pour les 10 stations en Europe (représentées en rouge sur la figure 4.2) ; nous vérifierons ces résultats sur deux autres régions. Enfin, nous comparerons les vitesses estimées avec et sans prise en compte la dépendance spatiale.

5.1 Modèle ACS-VARIO : établissement du modèle à partir des variogrammes

La dépendance spatiale entre séries est modélisée par la donnée d'une matrice de variance-covariance des innovations de l'AR(1). Dans une première étape, les innovations ont été estimées pour chaque station individuellement. Les variogrammes empiriques ont été estimés dans une seconde étape à partir des séries obtenues et sont ensuite modélisés paramétriquement. On utilise la moyenne et la médiane pour déduire le variogramme de la nuée variographique. Les graphiques montrent que ce choix n'influe pas beaucoup les résultats. Les intervalles de confiance sont calculés par bootstrap.

Variogrammes empiriques

Ces variogrammes expérimentaux des innovations des AR(1) normalisés sont présentés sur la figure 5.1 avec un variogramme par paire de composantes (l'ordre des composantes est indifférent) ; cette figure présente également les variogrammes pour les composantes croisées ; cela est similaire à [Amiri-Simkooei et al., 2017] à la différence que je considère les innovations des AR(1) qui correspondent à peu près aux différences premières. Parlons des variogrammes non croisés. Pour les composantes v et h , le variogramme est régulier et présente un palier à environ 60° . La composante u est un peu plus atypique, elle présente une légère bosse à une distance angulaire de 60° . L'allure générale des variogrammes ressemble à celles des covariogrammes de déformations de surcharge obtenus à partir des données de gravimétrie GRACE présentés dans [Rebischung et al., 2017]. L'ensemble du bruit corrélé spatialement dans les séries GNSS n'est

	pépite	portée	palier (°)
uu	0.23 (0.03)	0.43 (0.03)	18.9 (2.0)
vv	0.24 (0.03)	0.37 (0.02)	15.6 (1.3)
hh	0.06 (0.02)	0.42 (0.02)	13.8 (1.0)

TABLE 5.1 – Paramètres des variogrammes théoriques et leurs incertitudes. La portée et le palier désignent les paramètres du variogramme exponentiel indépendamment de la pépite.

pas dû à des déformations de surcharge. Mais cette ressemblance indique que le ou les phénomènes causant les corrélations spatiales des séries GNSS ont une structure spatiale similaire à celle des déformations de surcharge.

Variogrammes théoriques

La figure 5.2 présente les différents variogrammes théoriques, ils sont tous estimés à partir des variogrammes expérimentaux par moindres carrés non linéaires en utilisant le modèle exponentiel. Les paramètres des modèles exponentiels ajustés sont donnés dans la table 5.1. Le variogramme théorique hh est bien ajusté au variogramme empirique. La portée du variogramme ajusté (ou théorique) hh vaut 13.8 ° : lorsque deux points sont trop éloignés pour que le nouveau système de coordonnées uvh ait du sens, la covariance associée est quasi-nulle.

Dans le cadre d'un filtre de Kalman tridimensionnel, nous utilisons pour les composantes horizontales un unique variogramme obtenu en prenant la moyenne des paramètres du variogramme théorique (palier, portée, pépite) des variogrammes uu et vv . Cela se justifie, car les variogrammes théoriques uu et vv sont proches, bien que les variogrammes expérimentaux soient moins proches. Le modèle est aussi plus robuste sur le plan théorique, en particulier, les matrices de variance-covariance sont définies positives. La portée du variogramme théorique commun à uu et vv est 17.3 ° (le variogramme uu et le variogramme vv ont des portées valant respectivement 18.9 ° et 15.6 °). On se rend compte que la portée est légèrement supérieure pour les composantes horizontales que pour la composante verticale.

Covariances croisées

Les variogrammes croisés uv , uh et vh sont quasiment plats, nous avons donc choisi un modèle pépitique. Les paramètres de ce modèle pépitique sont donnés dans la table 5.2.

Pour les covariances croisées, nous allons dévoiler qu'il existe une corrélation quasi-nulle inter-composante. Pour cela, nous considérons x et y deux composantes distinctes quelconques parmi u et v et h . Les variables aléatoires à partir desquels nous calculons les variogrammes sont de moyenne nulle. Il se montre facilement :

$$\text{Cov}_{x,y}(0^+) = -\gamma_{x,y}(0^+) + \frac{1}{2}\gamma_{x,x}(+\infty) + \frac{1}{2}\gamma_{y,y}(+\infty), \quad (5.1)$$

où $\gamma_{.,.}$ est le variogramme éventuellement croisé et où $f(\infty) = \lim_{t \rightarrow +\infty} f(t)$ et $f(0^+) = \lim_{t \rightarrow 0, t > 0} f(t)$.

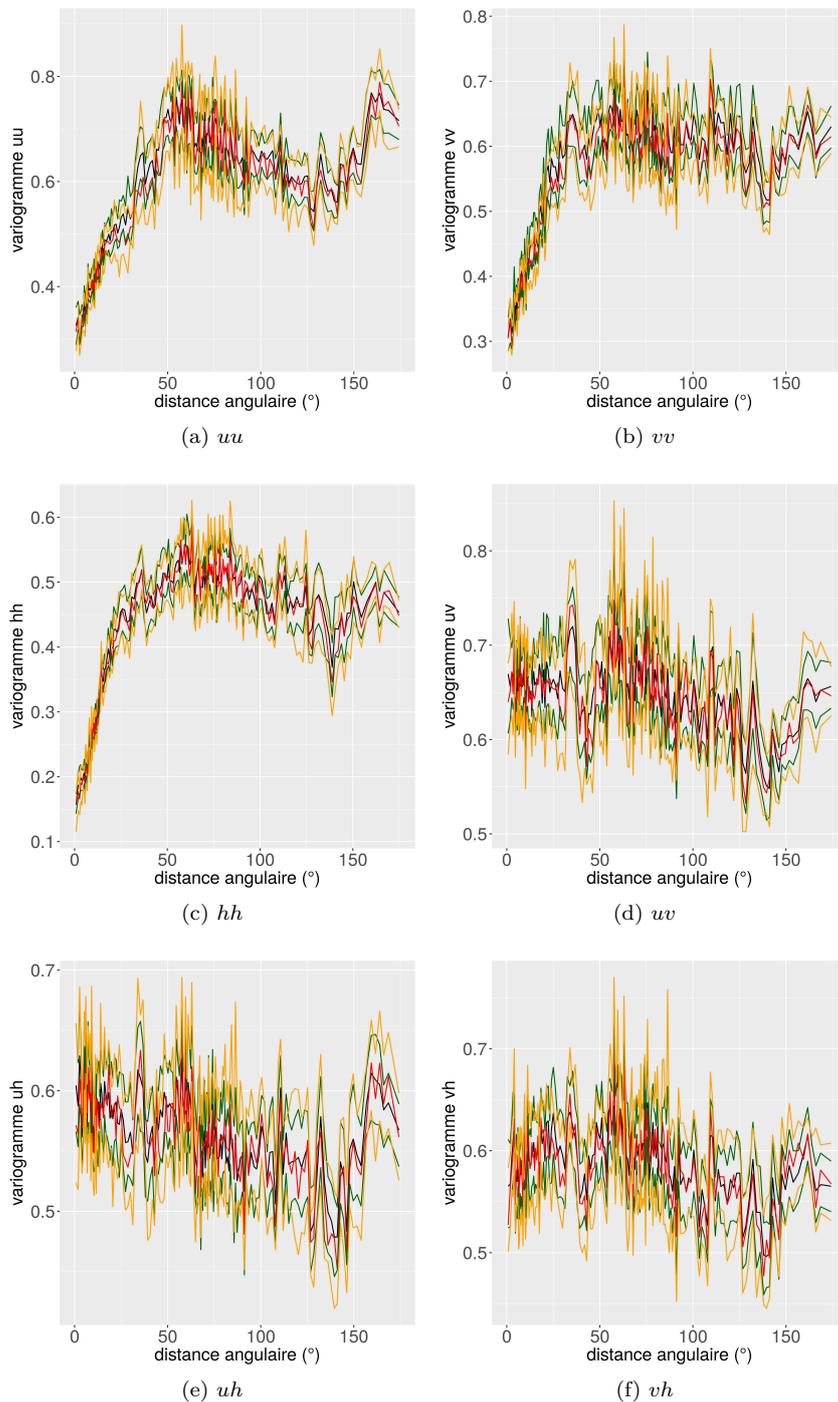


FIGURE 5.1 – Variogramme expérimental des innovations des AR(1) normalisés avec intervalle de confiance ; en rouge, la moyenne est utilisée pour synthétiser la nuée variographique (l'intervalle de confiance à 95 % est en orange) ; en noir, la médiane est utilisée pour synthétiser la nuée variographique (l'intervalle de confiance à 95 % est en vert).

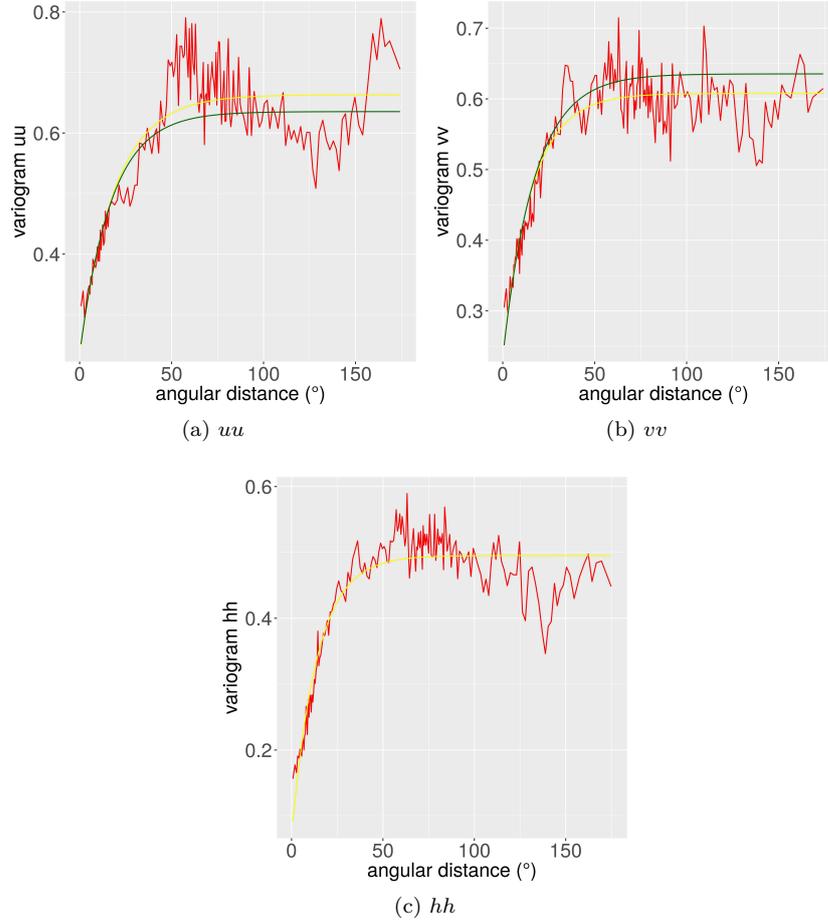


FIGURE 5.2 – Sont représentés en rouge les variogrammes expérimentaux uu , vv et hh des innovations des AR(1) normalisés en fonction de la distance angulaire. En jaune sont représentés les variogrammes théoriques ajustés pour chaque composante individuellement. En vert sont représentés les variogrammes communs aux composantes horizontales uu et vv .

En effet, notons $X_{\vec{s}}$ et $Y_{\vec{s}}$ les processus spatiaux où \vec{s} désigne un point de l'espace. Soit $\Delta\vec{s}$ un vecteur de longueur h .

On s'appuie pour la démonstration sur l'égalité :

$$\mathbb{E} \left[\frac{1}{2} (X_{\vec{s}} - Y_{\vec{s}+\Delta\vec{s}})^2 \right] = \frac{1}{2} \mathbb{E} X_{\vec{s}}^2 + \frac{1}{2} \mathbb{E} Y_{\vec{s}+\Delta\vec{s}}^2 - \mathbb{E} [X_{\vec{s}} Y_{\vec{s}+\Delta\vec{s}}].$$

Il suffit ensuite de faire tendre h vers 0^+ pour obtenir l'égalité (5.1).

Nous appliquons cette formule aux paires de composantes dont les éléments sont distincts. La table 5.3 montre que les covariances intercomposantes sont quasi-nulles, ce qui est confirmé par [Amiri-Simkooei et al., 2017]. Nous les considérerons donc comme nulles dans les paramètres du filtre de Kalman.

Ces modèles de variogramme seront utilisés pour toutes les zones d'étude : Europe, Caraïbes et côte est des États-Unis.

	pépité
uv	0.64 (0.01)
uh	0.56 (0.01)
vh	0.59 (0.01)

TABLE 5.2 – Paramètres et leurs incertitudes du modèle pépitique pour les variogrammes croisés

uv	-0.013
uh	0.015
vh	-0.034

TABLE 5.3 – Covariances croisées obtenues à partir des variogrammes croisés des innovations et de la formule (5.1)

5.2 Validation croisée

Rappelons que la validation croisée oriente le choix du modèle considéré (il permet de sélectionner le meilleur modèle parmi SCS, ACS-PROCHE-ITRF, ACS-EMP1, ACS-EMP2 et ACS-VARIO) ; plus le RMS des erreurs de prédiction est faible, meilleur est le modèle. La validation croisée sera appliquée aux séries de coordonnées de 10 stations en Europe (voir sous-section 4.1.2) de façon détaillée. Pour vérifier si les résultats obtenus sont généralisables à d'autres régions, deux autres lots de séries dans les Caraïbes et la côte est des États-Unis seront utilisés ; les deux autres régions servant plutôt à montrer la généralisabilité de l'algorithme proposé.

5.2.1 Europe

Rappelons que la validation croisée s'appuie sur des bases d'apprentissage et des bases de test ; la figure 4.4 présente celles-ci. En rouge et en noir, nous avons les plages temporelles des stations. Chaque segment rouge correspond à une base de test ; de chacune d'entre elles, les positions sont supprimées de la base d'apprentissage, les paramètres du modèle sont réestimés et les positions sont réestimées. En effet, nous faisons une comparaison entre les positions réestimées et les vraies positions.

La table 5.4 présente les RMS obtenus par validation croisée en Europe (équation (4.10)) en fonction de la modélisation des covariances spatiales adoptée. Sont présentés les RMS par composante. La colonne « NEU » présente les RMS de l'algorithme où les trois composantes sont traitées simultanément, il est notamment pris en compte les covariances inter-composantes pour le bruit blanc. La colonne « NEU mean » correspond à une moyenne quadratique des RMS afférents aux trois composantes, les covariances inter-composantes du bruit blanc ne sont pas prises en compte. D'une manière générale, les RMS sont plus élevés pour la composante verticale que pour les composantes horizontales. Cela peut être dû :

- aux signaux de surcharge qui sont plus importants pour la composante verticale¹,

1. Ces signaux sont en partie absorbés par les signaux périodiques estimés mais imparfai-

	E	N	U	NEU mean	NEU
Mod. SCS	0.97	0.97	3.72	2.29	2.29
Mod. ACS-PROCHE-ITRF	0.97	0.97	3.75	2.31	2.28
Mod. ACS-EMP1	0.70	0.70	2.68	1.66	1.61
Mod. ACS-EMP2	0.70	0.70	2.68	1.65	1.63
Mod. ACS-VARIO	0.70	0.71	2.52	1.56	1.63

TABLE 5.4 – RMS (mm) de la validation croisée relative à l’erreur de prédiction pour les stations en Europe. La colonne « NEU mean » présente les moyennes quadratiques des colonnes « E », « N » et « U ». La colonne « NEU » est issue d’un calcul unique intégrant la covariance spatiale du bruit blanc issue des solutions IGS.

- à des erreurs de modélisation (par exemple, de la troposphère) dont l’impact est plus important en verticale que pour les composantes horizontales.

Les valeurs de RMS sur les 3 composantes (NEU) pour les modèles SCS et ACS-PROCHE-ITRF ne diffèrent que de 0.01 mm. On en conclut que l’utilisation d’une matrice pleine pour le bruit blanc comme cela se fait pour l’ITRF ne modifie pas sensiblement le RMS par rapport à l’utilisation d’une matrice diagonale. Les modèles ACS-EMP1 et ACS-EMP2 qui utilisent un β ou plusieurs β ont des RMS identiques aux erreurs d’arrondis près.

Toutes les prises en compte de la dépendance spatiotemporelle étudiées conduisent à une réduction du RMS, sauf le modèle ACS-PROCHE-ITRF, qui n’inclut que la corrélation spatiale du bruit blanc. Pour la composante verticale, le RMS est plus petit dans le cadre du modèle paramétrique ACS-VARIO par rapport aux modèles ACS-EMP1 et ACS-EMP2 utilisant les covariances empiriques ; pour les composantes horizontales, les RMS des modèles ACS-EMP1, ACS-EMP2 et ACS-VARIO ne diffèrent que 0.01 mm. Considérer les covariances intercomposantes du bruit blanc (colonne « NEU ») aboutit un RMS plus petit que de ne pas le considérer (colonne « NEU mean ») [1.63 mm versus 1.56 mm].

La figure 4.17 associée à la figure 4.16 trace les différentes erreurs de prédiction de la position verticale de la station POTS (Potsdam, Allemagne). Elle confirme l’équivalence des modèles ACS-EMP1 et ACS-EMP2 et la supériorité des modèles tenant compte de la dépendance spatiotemporelle. Néanmoins, la supériorité du modèle ACS-VARIO par rapport aux autres modèles n’est pas claire : la différence de RMS entre ACS-VARIO et ACS-EMP1 pour la composante verticale ne s’élève qu’à 0.16 mm et le RMS porte sur dix stations.

Les tables 5.5, 5.6 et 5.7 présentent les RMS détaillés par station. La réduction du RMS dépend beaucoup de la station ; elle est néanmoins systématique pour toutes les stations pour les modèles spatiotemporels ACS-EMP1, ACS-EMP2 et ACS-VARIO. Pour la composante verticale, le modèle paramétrique ACS-VARIO est meilleur que les modèles ACS-EMP1 et ACS-EMP2 pour les stations colocalisées (WTZA/WTZS/WTZZ d’une part et ZIMJ/ZIMM d’autre part) et les trois modèles sont à peu près équivalents pour les autres stations. Pour les composantes horizontales, les réductions de RMS induites par l’utilisation des modèles ACS-EMP1, ACS-EMP2 et ACS-VARIO par rapport au modèle SCS sont similaires.

tement car il existe un signal de surcharge interannuel.

Station	SCS	ACS-PROCHE-ITRF	ACS-EMP1	ACS-EMP2	ACS-VARIO
GRAS	2.59	2.59 (-0.1)	1.81 (-30.1)	1.80 (-30.7)	1.77 (-31.7)
GRAZ	3.77	3.80 (0.7)	3.26 (-13.3)	3.27 (-13.2)	3.24 (-14.0)
MARS	3.81	3.92 (2.9)	3.49 (-8.3)	3.54 (-7.1)	3.51 (-8.0)
POTS	4.09	4.06 (-0.7)	2.23 (-43.8)	2.29 (-43.9)	2.33 (-43.0)
WROC	5.00	5.00 (0.2)	3.24 (-35.2)	3.25 (-34.8)	3.38 (-32.1)
WTZA	3.20	3.24 (1.3)	1.90 (-40.6)	1.89 (-40.8)	1.57 (-50.9)
WTZS	4.40	4.40 (0.0)	3.81 (-13.4)	3.83 (-13.0)	2.74 (-37.6)
WTZZ	3.42	3.44 (0.4)	1.93 (-43.6)	1.93 (-43.6)	1.61 (-52.9)
ZIMJ	3.17	3.15 (-0.8)	2.40 (-24.2)	2.38 (-24.8)	2.20 (-30.5)
ZIMM	3.42	3.44 (0.5)	2.10 (-38.4)	2.09 (-38.9)	1.89 (-44.8)
toutes stations	3.73	3.75 (0.5)	2.68 (-28.3)	2.68 (-28.0)	2.53 (-32.2)

TABLE 5.5 – RMS (mm) de la validation croisée en fonction de la station et du modèle pour la composante verticale. Entre parenthèses est indiqué le pourcentage de variation par rapport au modèle SCS.

La prise en compte d'une matrice de variance-covariance du bruit blanc ne réduit pas considérablement le RMS global. Pour les composantes horizontales est et nord ; si la réduction des RMS diffère d'une station à l'autre entre les modèles SCS et ACS-EMP1 (c'est aussi valable si on compare les modèles SCS et ACS-EMP2), la réduction moyenne du RMS est similaire pour les composantes est et nord.

Nous avons décidé d'ouvrir l'étude à d'autres régions du monde pour montrer la généralisabilité de la méthode.

5.2.2 Caraïbes

Nous allons prendre une région particulièrement complexe d'un point de vue géophysique : les Caraïbes. En effet, nous sommes dans une zone tectonique active, et on peut supposer d'autre part que la dynamique saisonnière et inter-annuelle soit différente suivant les îles. Comme nous sommes en zone océanique, on peut s'attendre à ce que les déformations de surcharge soient plus faibles et introduisent moins de corrélations spatiales entre séries qu'en zone continentale.

Le réseau de stations ainsi que les bases d'apprentissage et de test sont présentés sur les figures 5.3 et 5.4 ; comme l'indique la carte, les stations sont situées sur des îles différentes ; les stations ont moins de plages de données simultanées et sont un peu moins longues par rapport à l'Europe.

La figure 5.5 présente les résidus des séries temporelles de positions des stations sélectionnées aux Caraïbes. Nous remarquons que les séries temporelles sont plus courtes qu'en Europe et elles ne se chevauchent pas toujours temporellement.

La table 5.8 présente les RMS de la validation croisée en fonction du modèle de covariance et par composante. La réduction des RMS lorsque l'on utilise les méthodes prenant en compte la dépendance spatiotemporelle, réduction qui ne dépasse pas 0.04 mm, est faible bien que toujours existante. Les modèles ACS-EMP1 et ACS-EMP2 aboutissent à des RMS à peu près équivalents (dont l'écart ne dépasse pas 0.01 mm) ce qui montre que considérer une ou plusieurs

Station	SCS	ACS-PROCHE-ITRF	ACS-EMP1	ACS-EMP2	ACS-VARIO
GRAS	0.99	0.98 (-2.2)	0.77 (-22.4)	0.79 (-20.9)	0.77 (-23.3)
GRAZ	0.87	0.87 (0.4)	0.55 (-35.9)	0.55 (-36.1)	0.60 (-30.6)
MARS	0.74	0.73 (-0.4)	0.61 (-16.8)	0.61 (-17.0)	0.61 (-17.1)
POTS	0.90	0.92 (1.5)	0.86 (-5.1)	0.86 (-5.0)	0.81 (-9.5)
WROC	1.01	1.02 (0.6)	0.62 (-39.5)	0.61 (-39.2)	0.73 (-28.7)
WTZA	0.79	0.80 (0.8)	0.32 (-59.4)	0.33 (-59.0)	0.36 (-53.9)
WTZS	1.06	1.10 (3.5)	1.03 (-2.7)	1.04 (-1.6)	0.95 (-10.4)
WTZZ	1.13	1.11 (-0.9)	0.40 (-64.2)	0.41 (-63.7)	0.15 (-56.8)
ZIMJ	1.21	1.21 (-0.0)	0.91 (-18.7)	0.91 (-24.9)	0.90 (-25.6)
ZIMM	0.92	0.93 (0.7)	0.78 (-15.4)	0.77 (-15.2)	0.75 (-18.0)
toutes stations	0.97	0.97 (0.3)	0.70 (-27.5)	0.70 (-27.0)	0.70 (-27.0)

TABLE 5.6 – RMS (mm) de la validation croisée en fonction de la station et du modèle pour la composante Est. Entre parenthèses est indiqué le pourcentage de variation par rapport au modèle SCS.

Station	SCS	ACS-PROCHE-ITRF	ACS-EMP1	ACS-EMP2	ACS-VARIO
GRAS	0.75	0.77 (1.7)	0.59 (-22.7)	0.58 (-22.8)	0.60 (-21.1)
GRAZ	0.79	0.81 (2.0)	0.55 (-31.1)	0.55 (-30.5)	0.59 (-25.5)
MARS	1.01	1.02 (-0.1)	0.79 (-22.0)	0.79 (-22.0)	0.81 (-20.6)
POTS	0.90	0.90 (0.8)	0.65 (-26.4)	0.65 (-26.4)	0.66 (-25.8)
WROC	1.21	1.23 (1.6)	1.21 (-29.4)	0.85 (-29.4)	0.92 (-24.1)
WTZA	0.96	0.95 (-0.7)	1.59 (-38.4)	0.59 (-38.4)	0.66 (-31.1)
WTZS	1.39	1.35 (-2.5)	1.09 (-21.1)	1.09 (-20.9)	0.93 (-32.7)
WTZZ	0.73	0.74 (1.3)	0.34 (-54.1)	0.33 (-54.0)	0.36 (-50.5)
ZIMJ	0.93	0.94 (-0.4)	0.87 (-7.3)	0.87 (-7.4)	0.86 (-7.5)
ZIMM	0.96	0.96 (-0.5)	0.61 (-36.5)	0.61 (-36.4)	0.66 (-31.2)
toutes stations	0.97	0.97 (0.2)	0.70 (-27.5)	0.70 (-27.4)	0.71 (-26.2)

TABLE 5.7 – RMS (mm) de la validation croisée en fonction de la station et du modèle pour la composante Nord. Entre parenthèses est indiqué le pourcentage de variation par rapport au modèle SCS.



FIGURE 5.3 – Carte des stations sélectionnées pour les Caraïbes

	E	N	U	NEU mean
Mod. SCS	0.96	1.15	3.66	2.28
Mod. ACS-EMP1	0.92	1.11	3.61	2.24
Mod. ACS-EMP2	0.92	1.11	3.61	2.24
Mod. ACS-VARIO	0.95	1.07	3.61	2.24

TABLE 5.8 – RMS (mm) de la validation croisée relative à l’erreur de prédiction pour les stations aux Caraïbes

amplitudes de bruit blanc variable conduit à des RMS quasi-égaux. Contrairement à l’Europe, l’utilisation d’une méthode paramétrique (modèle ACS-VARIO) et l’utilisation des méthodes utilisant la covariance empirique aboutit à des RMS quasi-égaux (leur écart ne dépasse pas 0.04 mm quelle que soit la composante), ce qui montre des capacités de prédiction très similaires en moyenne. Le modèle ACS-VARIO n’a pas un avantage net sur les autres modèles tenant en compte la dépendance spatiotemporelle; cela peut s’expliquer par le fait que la majorité des stations ayant été utilisée pour établir le variogramme utilisé dans ACS-VARIO se situe aux États-Unis et en Europe et non aux Caraïbes, ce qui peut expliquer les sous-performances de ce modèle. Dans les Caraïbes, l’utilisation de la dépendance spatiale réduit le RMS mais de manière très légère, du moins relativement à l’Europe.

5.2.3 Côte est des États-Unis

L’Amérique du Nord est une région dense en stations, elle a néanmoins l’inconvénient de posséder des stations dont les séries IGS `repro2` sont plus courtes qu’en Europe; l’autre zone géographique d’étude choisie est la côte est des États-Unis. Le réseau de stations ainsi que les bases d’apprentissage et de

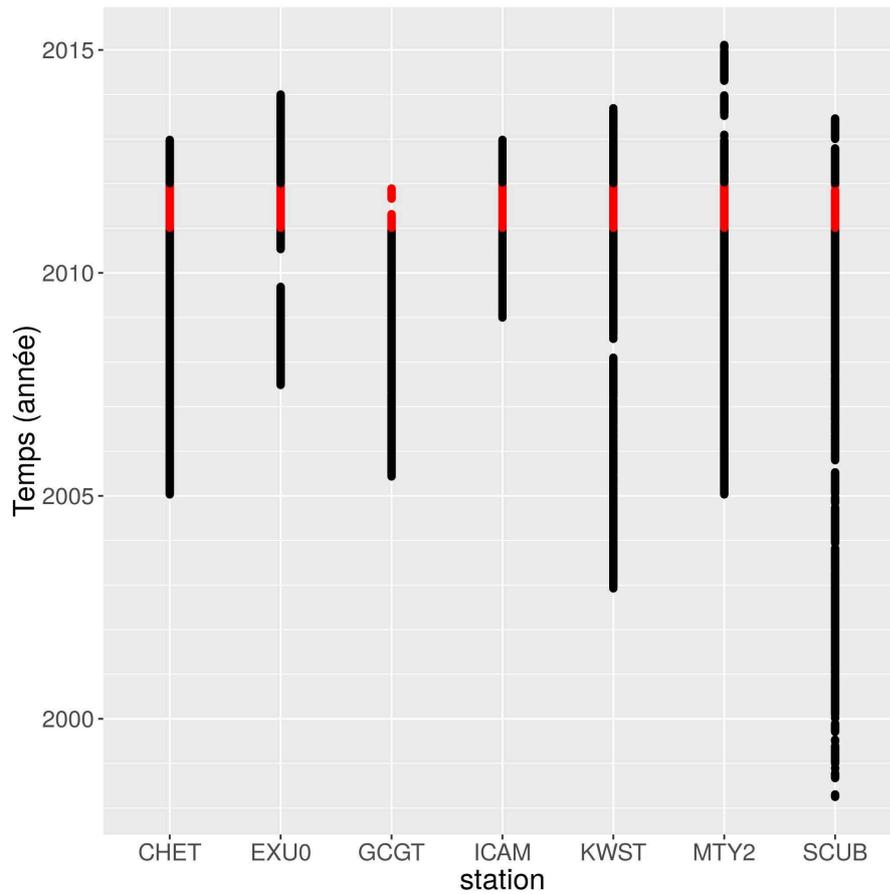


FIGURE 5.4 – Plages temporelles des données de positions pour les différentes stations utilisées aux Caraïbes. Successivement pour chaque station, la zone rouge est supprimée de la base d'apprentissage et nous réutilisons le filtre de Kalman ; la concordance entre les vraies données et les données réestimées est exprimée via un RMS.

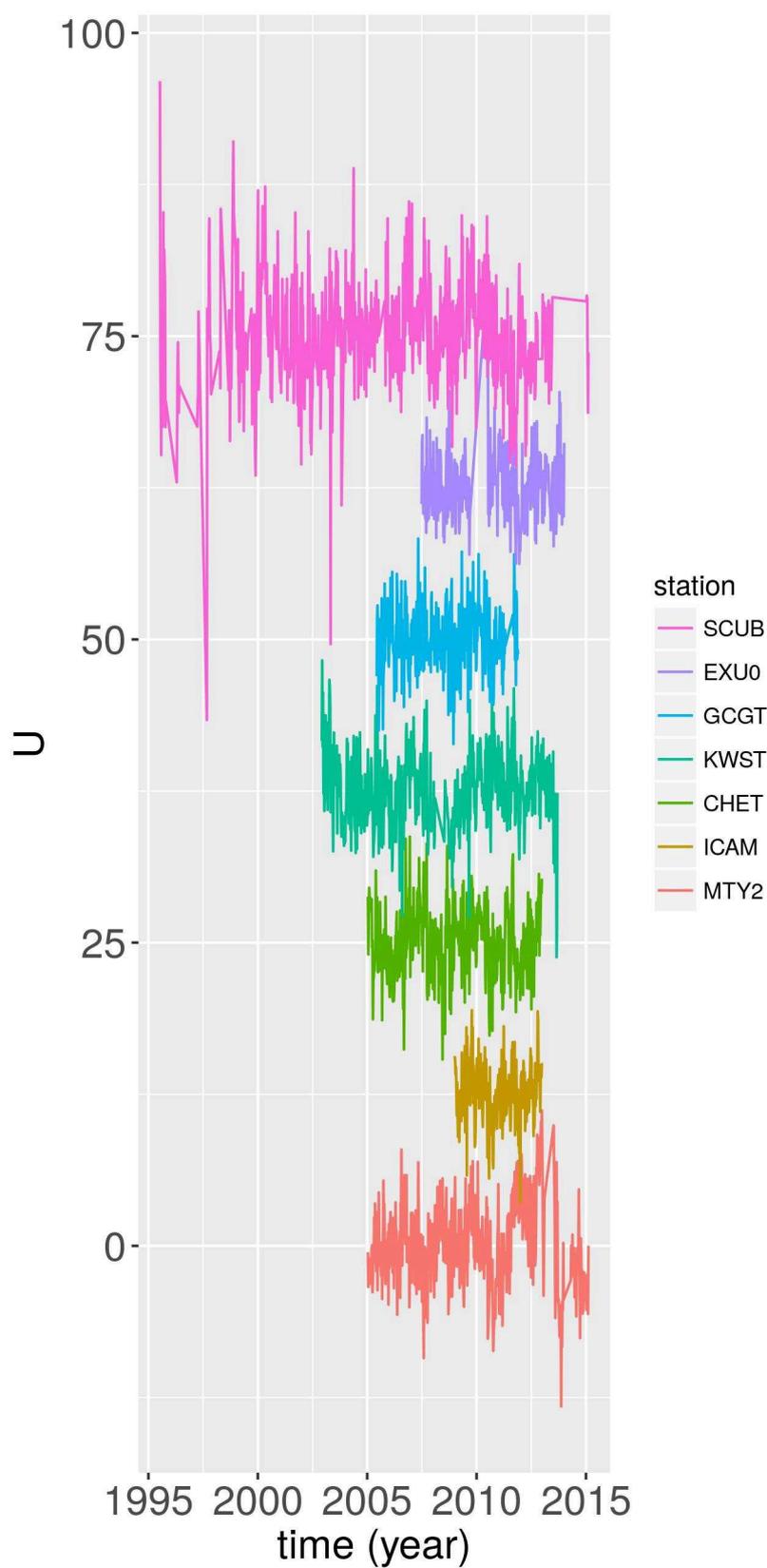


FIGURE 5.5 – Séries de positions résiduelles verticales pour les stations sélectionnées de l'archipel caribéen. Les périodes considérées sont les périodes annuelle et semi-annuelle et les périodes draconitiques GPS jusqu'à la 7^e harmonique. Des offsets ont été ajoutés pour des raisons de lisibilité.

	E	N	U	NEU mean
Mod. SCS	0.81	1.01	3.21	2.00
Mod. ACS-EMP	0.73	0.89	2.71	1.70
Mod. ACS-EMP2	0.73	0.89	2.71	1.70
Mod. ACS-VARIO	0.66	0.82	2.34	1.48

TABLE 5.9 – RMS (mm) de la validation croisée relative à l’erreur de prédiction pour les stations sur la côte est des États-Unis

test sont présentés sur la figure 5.6 et 5.7.

La figure 5.8 présente les séries temporelles des bruits totaux de positions verticales GNSS pour toutes les stations sélectionnées. Des corrélations sont visibles sur cette figure entre séries de bruit de différentes stations.

La prise en compte de la dépendance spatiale réduit le RMS de 0.08, 0.22 et 0.50 mm respectivement pour les composantes est, nord et verticale pour les modèles ACS-EMP1 et ACS-EMP2; elle améliore donc la prédiction (imputation des valeurs manquantes notamment). L’utilisation d’une ou de plusieurs amplitudes de bruits blancs variables (modèles ACS-EMP1 et ACS-EMP2) ne modifie pas le RMS à plus de 0.01 mm et donc vraisemblablement la qualité de prédiction. Dans toutes les composantes, l’utilisation du modèle paramétrique (modèle ACS-VARIO) fournit, comme en Europe, une meilleure performance en termes de réduction du RMS que celle de la covariance empirique (la différence de RMS vaut 0.07, 0.07 et 0.37 mm respectivement en est, nord et verticale), cette bonne performance est rendue possible par le fait que l’étendue spatiale des stations de l’ordre de 13.5° est de l’ordre du palier qui vaut 13.8° , ceci nous permet d’avoir des covariances déduites du variogramme non nulles entre stations. La supériorité de ACS-VARIO peut s’expliquer l’existence d’un surapprentissage dans les modèles ACS-EMP1 et ACS-EMP2.

Nous avons montré que la prise en compte de la dépendance spatiale pour le bruit corrélé dans le filtre de Kalman améliorerait la capacité prédictive du modèle ce qui est montré par une diminution des RMS, contrairement à la prise en compte de la dépendance spatiale pour le bruit blanc qui est sans effet sur le RMS. Pour la composante verticale, le modèle ACS-VARIO semble meilleur que les modèles ACS-EMP1 et ACS-EMP2; pour les composantes horizontales, cette supériorité est moins nette. Ce constat est peut-être trompeur dans la mesure où la base d’apprentissage utilisée pour établir le variogramme contient certes des stations situées partout à travers le globe, mais beaucoup de stations situées en Europe ou aux États-Unis; l’utilisation d’une moyenne ou d’une médiane pondérée (pondération qui se fera en fonction de la densité de stations) pour estimer le variogramme à partir de la nuée variographique pourrait permettre d’obtenir des résultats d’une qualité meilleure sur toute la Terre. Pour les modèles ACS-EMP1 et ACS-EMP2, l’introduction de covariances inter-composantes n’améliore pas (ni ne dégrade) la capacité prédictive de la méthode, mais augmente la charge de calcul. Cette méthode a l’avantage de pouvoir s’adapter à des régions du monde variées, même pour les régions où la géophysique est complexe comme les Caraïbes, mais présente en revanche un risque de surapprentissage.



FIGURE 5.6 – Carte des stations sélectionnées pour la côte est des États-Unis. La position de certaines stations a été légèrement modifiée pour éviter les chevauchements.

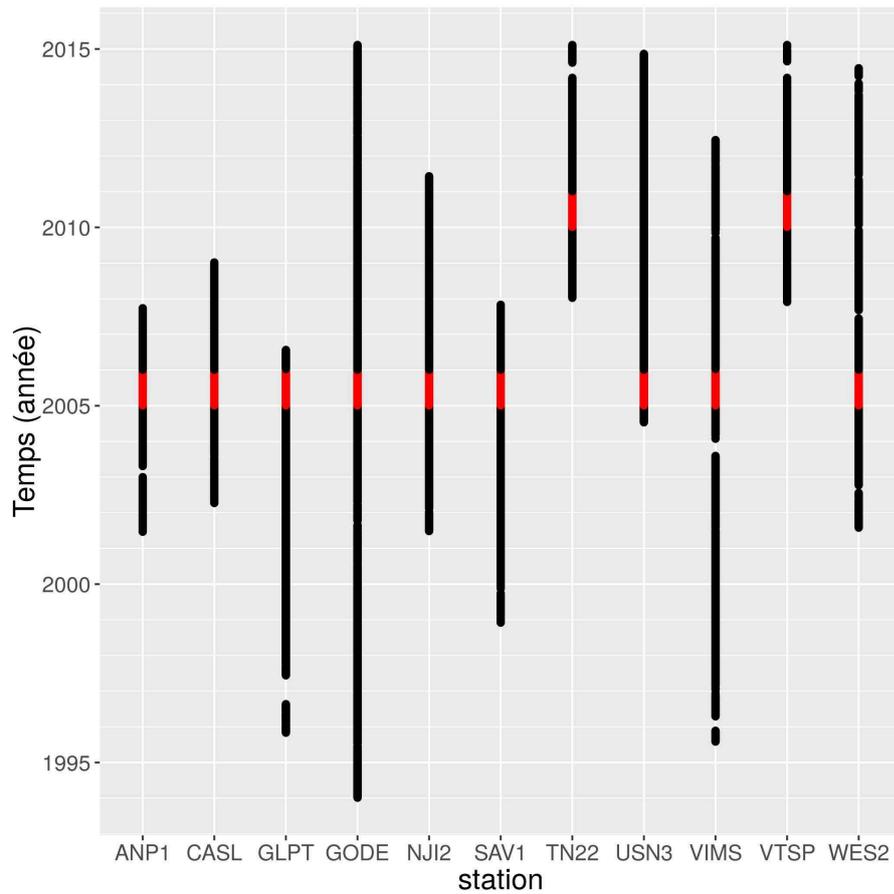


FIGURE 5.7 – Plages temporelles des données de positions pour les différentes stations sélectionnées de la côte est des États-Unis. Successivement pour chaque station, la zone rouge est supprimée de la base d'apprentissage et nous réutilisons le filtre de Kalman ; la concordance entre les vraies données et les données réestimées est exprimée via un RMS.

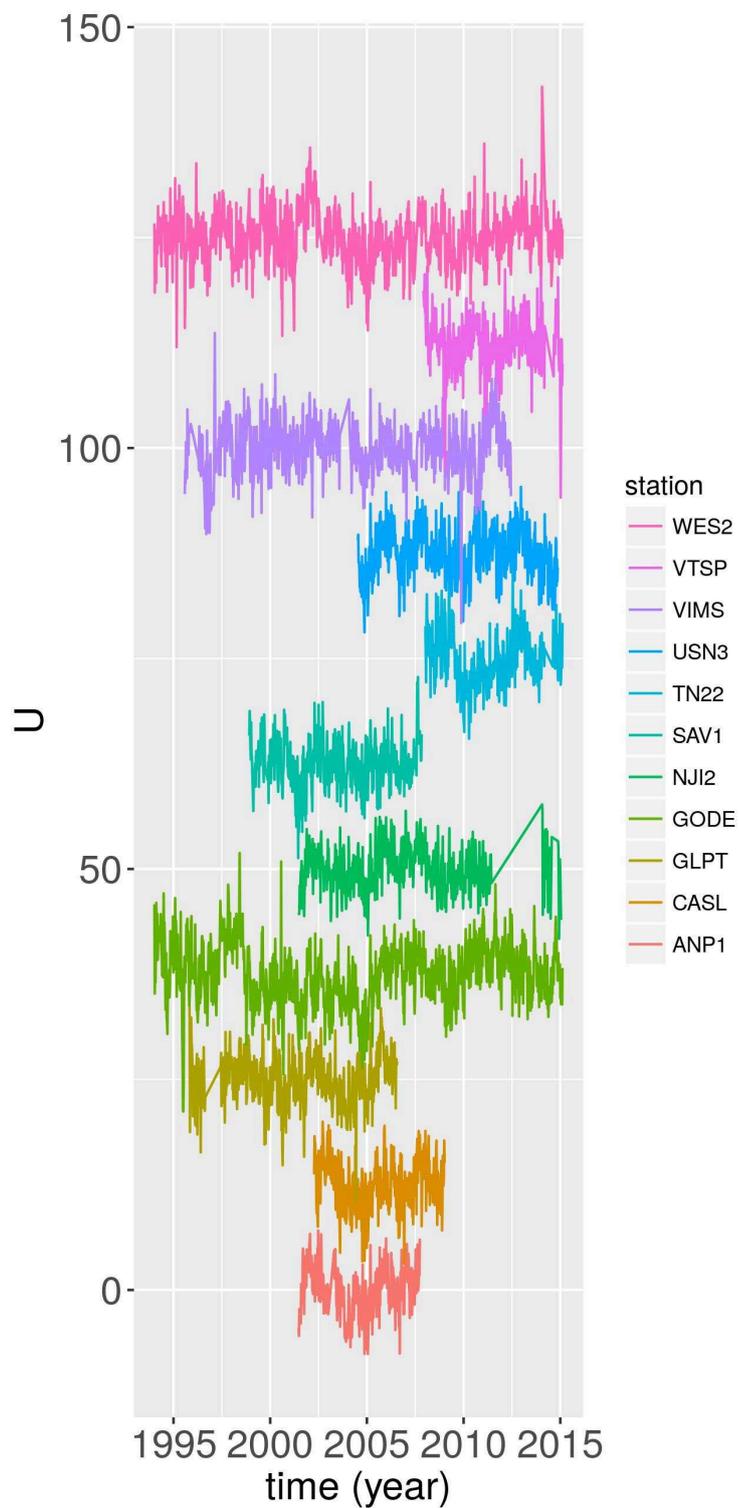


FIGURE 5.8 – Séries de positions résiduelles verticales pour les stations sélectionnées de la côte est des États-Unis. Les périodes considérées sont les périodes annuelle et semi-annuelle et les périodes draconitiques GPS jusqu'à la 7^e harmonique. Des offsets ont été ajoutés pour des raisons de lisibilité.

5.3 Amélioration de l'estimation des vitesses

Il peut arriver que certaines séries temporelles soient de durées trop courtes pour déterminer les vitesses de manière fiable. Nous allons appliquer la méthodologie vue à la section 4.9 à la composante verticale des séries sélectionnées en Europe et sur la côte est des États-Unis. Nous n'avons gardé que les 100 premiers points des séries hebdomadaires pour chaque série. Comme au chapitre précédent, nous avons considéré que la vraie vitesse r est celle donnée par filtre de Kalman sur la série complète. La vitesse \hat{r}_{st}^{COV} (respectivement \hat{r}_{st}^{VARIO}) correspond à la vitesse obtenue grâce au filtre de Kalman de la méthode ACS-EMP1 (respectivement, de la méthode ACS-VARIO) en tronquant, dans la base d'apprentissage, la série de la station dont on calcule la vitesse. Nous avons comparé ces résultats au filtrage spatial, technique qui consiste à retirer aux séries de positions une moyenne sur toutes les stations des résidus obtenus par moindres carrés, avant d'estimer de nouveau par moindres carrés la vitesse ; ce filtrage spatial a été effectué avec les mêmes données tronquées que ci-dessus ; pour chaque station, sa série temporelle est tronquée, les autres restent identiques, la vitesse de cette station est estimée.

La table 5.10 présente les différentes vitesses de cette étude pour les deux zones précitées. La très forte augmentation relative de l'erreur d'estimation sur les vitesses pour la station GLPT est due au fait que l'erreur de départ est probablement par hasard très faible. L'utilisation de la dépendance spatiale permet sur l'ensemble des stations une réduction de l'erreur sur la vitesse de 1 % et 37 % respectivement pour les modèles ACS-EMP1 et ACS-VARIO par rapport à l'erreur lorsque l'on n'utilise pas la dépendance spatiale (modèle SCS). Le modèle ACS-VARIO semble par conséquent meilleur que le modèle ACS-EMP1 pour diminuer l'erreur d'estimation de la vitesse des stations à séries courtes ; cela peut s'expliquer par, dans le cas des modèles ACS-EMP1 et ACS-EMP2, les faibles recouvrements des séries temporelles de positions GNSS entre stations (en particulier lorsque les séries temporelles sont tronquées) qui peuvent fausser l'estimation de la matrice de variance-covariance, ce qui n'est pas le cas pour le modèle ACS-VARIO. Ce constat est valable pour les stations européennes où l'utilisation de la dépendance spatiale permet sur l'ensemble des stations une réduction de l'erreur sur la vitesse de 4 % et 19 % respectivement pour les modèles ACS-EMP1 et ACS-VARIO par rapport à l'erreur lorsque l'on n'utilise pas la dépendance spatiale (modèle SCS). Ce constat est amplifié aux États-Unis où le modèle ACS-EMP1 empire l'erreur de l'estimation de la vitesse de 4% par rapport au modèle SCS alors que le modèle ACS-VARIO l'améliore de 65 % ! Nous allons maintenant comparer les erreurs induites par le filtrage spatial à celles induites par le modèle ACS-VARIO. Le modèle de filtre spatial fournit de meilleures estimations de vitesses que ACS-VARIO en Europe (la réduction de l'erreur moyenne atteint 27 % contre 19 %), mais pas sur la côte est des États-Unis (la réduction de l'erreur moyenne atteint 24 % contre 65 %). Le filtrage spatial semble plus dépendant de la présence d'un grand nombre de données simultanément comme c'est le cas en Europe. Les erreurs moindres du filtrage spatial par rapport à ACS-VARIO peuvent s'expliquer par le fait que ACS-VARIO utilise des séries de positions de stations GNSS à travers le monde pour établir un variogramme unique ; les paramètres des variogrammes théoriques utilisés dans ce modèle ne s'adaptent pas à la région spécifique étudiée.

	r	\hat{r}_0	\hat{r}_{st}^{COV}	\hat{r}_{st}^{VARIO}	\hat{r}_{st}^{SF}	$ \hat{r}_0 - r $	$ \hat{r}_{st}^{COV} - r $	$ \hat{r}_{st}^{VARIO} - r $	$ \hat{r}_{st}^{SF} - r $
GRAS	-0.32	0.38	0.14	0.17	-2.38	0.70	0.46 (34%)	0.49 (30%)	2.06 (-240%)
GRAZ	-0.04	-3.78	-3.74	-2.98	-2.38	3.74	3.70 (1 %)	2.94 (21 %)	2.34 (37%)
MARS	-0.57	-0.85	-0.72	-0.87	-0.46	0.28	0.15 (46 %)	0.30 (-7 %)	0.11 (61%)
POTS	-0.40	5.73	5.85	5.85	2.88	6.13	6.25 (-2 %)	6.25 (-2 %)	3.28 (43%)
WROC	-0.27	1.15	1.09	0.64	0.82	1.42	1.36 (4 %)	0.91 (36 %)	1.09 (23%)
WTZA	-0.36	-0.88	-0.73	0.03	1.56	0.52	0.37 (29 %)	0.39 (25 %)	1.20 (-57%)
WTZS	0.52	2.18	2.11	1.05	0.92	1.66	1.59 (4 %)	0.53 (68 %)	0.40 (76%)
WTZZ	-0.53	1.29	0.91	0.93	1.18	1.82	1.44 (21 %)	1.46 (20 %)	1.71 (6%)
ZIMJ	0.19	-0.64	-0.80	-0.97	-1.20	0.83	0.99 (-19 %)	1.16 (-40 %)	0.56 (-67%)
ZIMM	0.89	-0.52	-0.39	0.91	0.90	1.41	1.28 (9 %)	0.02 (99 %)	0.01 (99 %)
Europe	1.75	1.68 (4%)	1.41 (19%)	1.28 (27%)
ANP1	-1.43	-1.82	-1.72	-1.52	-0.55	0.39	0.29 (26%)	0.09 (77%)	0.88 (55%)
CASL	-0.79	-3.84	-3.90	-2.85	-1.12	3.05	3.11(- 2%)	2.06 (32%)	0.33 (89%)
GLPT	-2.39	-2.42	-2.85	-2.23	-3.96	0.03	0.46 (-1433%)	0.16 (-433%)	1.61 (-5267%)
GODE	-1.25	-2.20	-2.13	-1.22	-1.07	0.95	0.88 (7%)	0.03 (97%)	0.18 (81%)
NJI2	-1.17	-1.40	-1.30	-1.19	-0.29	0.23	0.13 (48%)	0.02 (91%)	0.88 (-120%)
SAV1	-0.30	-0.70	-0.53	-0.33	0.89	0.40	0.23 (42.5%)	0.03 (92.5%)	1.19 (-198%)
TN22	-0.24	-1.58	-1.54	-0.38	0.27	1.34	1.30 (3%)	0.14 (90%)	0.51 (62%)
USN3	-0.56	2.22	1.94	0.56	1.53	2.78	2.50 (10%)	1.12 (60 %)	2.09 (25%)
VIMS	-1.39	-0.15	-0.06	-1.35	-0.76	1.24	1.33 (-7%)	0.04 (97%)	0.63 (49%)
VTSP	0.16	-2.32	-2.79	-0.06	-1.09	2.48	2.95 (-19%)	0.22 (91%)	1.25 (50%)
WES2	-0.03	-1.80	-1.79	-1.35	-0.81	1.77	1.76 (1%)	1.32 (25%)	0.78 (56%)
USA	1.19	1.24 (-4%)	0.41 (65%)	0.94 (24%)
moy.	1.49	1.48 (1%)	0.94 (37%)	1.10 (26%)

TABLE 5.10 – Compilation des résultats obtenus à la suite des calculs en vue de débiaiser les vitesses. Nous avons repris les notations de la section 4.9; \hat{r}_0 correspond à la vitesse obtenue grâce à la série non tronquée; \hat{r}_{st}^{COV} (respectivement \hat{r}_{st}^{VARIO}) correspond à la vitesse obtenue grâce à la méthode ACS-EMP1 (respectivement à la méthode ACS-VARIO) appliquée aux séries tronquées. \hat{r}_{st}^{SF} correspond à la vitesse obtenue grâce au filtrage spatial appliqué aux séries tronquées. Les lignes « Europe » et « USA » présentent les moyennes respectives des erreurs en Europe et aux États-Unis. Les vitesses sont exprimées en mm/y. Les pourcentages entre parenthèses montrent la réduction de l'erreur sur les vitesses par rapport à celle sans prise en compte de la dépendance spatiotemporelle.

Conclusions et perspectives

En raison du faible coût des stations GNSS permanentes et de la grande précision de la technique GNSS, un nombre croissant de stations GNSS permanentes sont déployées dans le monde. Les plus anciennes appartenant au réseau IGS disposent par ailleurs de plus de 24 ans de données. La disponibilité de telles données (du fait du grand nombre de stations et de la longueur des séries de positions) rend aujourd'hui possible une étude poussée des corrélations entre séries temporelles de différentes stations. Or, comme nous l'avons montré dans cette thèse, cette corrélation est clairement détectable jusqu'à une distance d'environ 5000 km. Elle s'explique par l'origine commune d'une partie du déplacement physique du sol, en particulier les phénomènes de surcharges, et probablement par des erreurs systématiques d'origine variée comme des erreurs de modélisation des orbites de satellites GNSS. Jusqu'à maintenant, cette corrélation entre séries de coordonnées de stations proches était éliminée par l'intermédiaire de la méthode dite de filtrage spatial : elle consiste à retirer une moyenne du signal de déplacement résiduel sur un ensemble de sites proches de la station dont la série de coordonnées est à filtrer.

Ma thèse a développé une méthodologie alternative afin de tenir compte de la dépendance spatiale des séries temporelles de positions GNSS en vue d'estimer des vitesses de déplacement plus fiables et de permettre la prédiction de positions de stations (imputation de valeurs manquantes et prédiction dans le futur) ; la méthodologie proposée s'appuie sur un filtre de Kalman, ce qui ouvre des perspectives d'intégration plus rigoureuse de la corrélation spatiale dans le processus d'estimation de l'ITRF. Elle peut se placer plus généralement dans le contexte de l'estimation d'un repère de référence sous la forme de séries temporelles de coordonnées, comme proposé par [Dong et al., 1998], et, plus récemment, par [Wu et al., 2015]. Chaque composante de série temporelle de coordonnées intègre ainsi un modèle de bruit égal à la somme d'un bruit blanc variable et d'un processus autorégressif d'ordre 1, utilisable dans un filtre de Kalman. Le filtre de Kalman permet alors de paramétrer une dépendance spatiale entre les innovations des processus autorégressifs précités, possibilité que nous avons utilisée.

Plusieurs modèles de dépendance spatiale et le modèle n'en tenant pas compte ont été comparés. Parmi les modèles tenant compte de la dépendance spatiale, un modèle est basé sur l'utilisation de la covariance empirique et un second sur l'utilisation de variogrammes. Au-delà des considérations théoriques, la covariance empirique a l'avantage d'être plus facilement maniable. La validation croisée, technique statistique de sélection de modèles, permet de comparer ces trois modèles dans une visée prédictive au sens large (imputation des valeurs manquantes, prédiction dans le futur), tout en évitant une complexification

inutile du modèle (surapprentissage). Nous utilisons trois jeux de données composés d'une dizaine de stations ; chaque jeu de données est associé à une des trois régions du monde : l'Europe, les Caraïbes et la côte est des États-Unis. Pour les stations en Europe et aux États-Unis, les deux modèles tenant compte de la dépendance spatiale sont largement meilleurs en termes de validation croisée que le modèle ne la prenant pas en compte ; dans ces régions, l'utilisation de variogrammes aboutit de plus à de meilleurs résultats pour la composante verticale par rapport à l'utilisation de la covariance empirique (pour les composantes horizontales, cette supériorité est un peu moins nette). Cependant, la prise en compte de la dépendance spatiale aux Caraïbes fournit un résultat plus contrasté, bien que très légèrement meilleur qu'un modèle sans prise en compte de la corrélation spatiale. Il semble ainsi que la prise en compte des dépendances spatiales fonctionne mieux dans les zones continentales que des zones océaniques (notamment lorsque les stations sont sur plusieurs îles différentes) ce qui est conforme à l'intuition physique. Ce travail de validation gagnerait toutefois à être complété sur des zones plus étendues.

Un autre résultat principal de ce travail est donc la production d'un modèle générique de corrélations spatio-temporelles de séries de coordonnées GNSS basé sur un variogramme. Ce modèle pourra être repris dans des travaux futurs. Un second résultat pertinent est l'amélioration de la détermination de la vitesse de séries courtes lorsque la dépendance spatiale est prise en compte à hauteur de 37 %.

Une des difficultés rencontrées dans cette thèse est l'impossibilité d'implémenter du bruit de scintillation dans un filtre de Kalman. Si nous avons choisi un modèle simple sous la forme d'un AR(1), une perspective de travail pertinente serait l'étude d'une meilleure approximation du bruit de scintillation sous la forme d'un modèle espace-état. L'objectif est de fournir le meilleur compromis en termes de nombre de paramètres explicatifs pour modéliser de manière fine le bruit de scintillation sans pour autant tomber dans le piège du surapprentissage.

Troisième partie

Annexes

Annexe A

Posters et présentations

Les posters et présentations que j'ai présentés sont les suivants :

- Benoist, C., X. Collilieux, P. Rebischung, Z. Altamimi, L. Bel, K. Charnard, O. Jamet and L. Métivier (2018). *Spatial correlations in IGS station position time series*, European Geosciences Union General Assembly 2018, apr. 2018.
- Benoist, C., P. Rebischung, L. Métivier, X. Collilieux, Z. Altamimi and O. Jamet (2017). *Analyse statistique des données GNSS en vue d'améliorer l'ITRF*, Journées de la Recherche IGN, mars 2017.
- Benoist, C., P. Rebischung and Z. Altamimi (2016). *Vers la prise en compte de la dépendance spatio-temporelle des séries de position GNSS dans leur analyse*, poster présenté aux Journées de la Recherche de l'IGN, 24-25 mars 2016, Marne la Vallée, France, mars 2016.
- Rebischung, P., J. Ray, C. Benoist, L. Métivier and Z. Altamimi (2015). *Error analysis of the IGS repro2 station position time series*, présenté à l'AGU Fall Meeting 2015, San Francisco, 14-18 December (abstract G23B-1065)

Spatial correlations in IGS station position time series

C. Benoist ⁽¹⁾, X. Collilieux ^(1,2), P. Rebischung ⁽¹⁾, Z. Altamimi ⁽¹⁾, L. Bel ⁽³⁾, K. Chanard ⁽¹⁾, L. Métivier ⁽¹⁾, O. Jamet ⁽¹⁾

(1) IGN France, Université Paris Diderot-LAREG, PARIS CEDEX 13, France, (2) IGN France, ENSG, Champs-sur-Marne, France, (3) UMR MIA-Paris, AgroparisTech, INRA, Université Paris-Saclay, 75005 Paris

Objectives

- Input data are repro2 IGS/GNSS position time series used in ITRF2014 [Altamimi et al., 2016].
- Aims of this poster are :
 - to prove that taking into account dependencies between station coordinates improves GNSS stacking of station position time series,
 - to obtain more realistic uncertainties for GNSS stations velocities
- For that purpose, we used maximum likelihood estimation and Kalman filter.

Temporal dependencies

- It is known that the noise of GNSS position time series is correlated ; best found correlated noise model is, on average, flicker noise, whose power spectrum density is in $1/f$, as shown by its stacked periodograms of coordinate residuals.
- Periodic signals appear in the time series. Annual and semi-annual signals are due to, in part, loading effects. Signals at GPS draconic periods are attributed to satellite orbit errors and multipath.

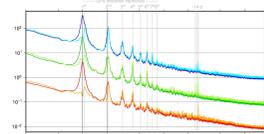


Fig. 1: Stacked periodogram of the residuals from a long-term stacking of the IGS rep2 solutions (East: blue; North: green; Up: red). Light colors do not include annual and semi-annual signals, dark colors do.

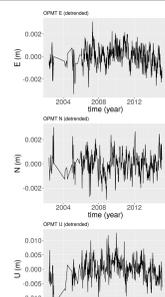


Fig. 2 : Residual time series of the station OPMT (Paris, France)

Temporal model

- Our kinematic temporal model for station position time series includes:
 - linear trends with position discontinuities,
 - annual and semi-annual signals,
 - draconitic signals up to the 7th harmonic,
 - a correlated noise model composed of variable white noise + first-order autoregressive process AR(1): $X_{t+1} = \alpha X_t + \varepsilon_t$ where $\varepsilon_t \sim \mathcal{N}(0, \sigma_{AR}^2)$ where α is a real between -1 and 1 and the ε_t are innovations.
 - We have chosen an autoregressive process instead of flicker noise, because flicker noise cannot be used in a Kalman filter.
 - Variable white noise is decorrelated noise whose standard deviations are proportional to formal errors of station position time series.

Methodology

- Our algorithm is composed of three steps:
- using CATS (Williams, 2008), we estimate noise parameters by maximum likelihood and kinematic parameters by least-squares regression to initialize Kalman filter,
 - we apply Kalman filter to each station position time series independently,
 - we compute the empirical covariance matrix of the innovations of AR(1) from the previous step and introduce it in a second network Kalman filter.
- The choice of covariance estimator has the advantage of requesting few hypotheses on the data to apply the methodology, what is practical when we work with real data.



Fig. 3: Distribution of the ten stations used in this study (some stations are co-located)

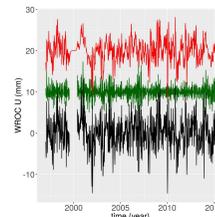


Fig. 4: White noise (green) and AR(1) (red) estimated by Kalman filter and total noise (black) in vertical position time series of station WROC (Wrocław, Poland)

Contribution of modeling dependencies between stations

To understand the contribution of taking into account dependencies between station coordinates, we:

- withdrew, for each station successively, a one-year time range from the input of Kalman filter
- compared the true data (black) with Kalman filter estimator with (blue) and without (red) taking into account dependencies between the station coordinates.

As seen in figures 5 and 6, taking into account dependencies between station coordinates is relevant.

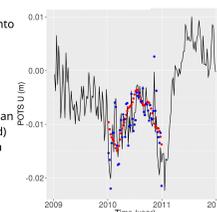


Fig. 5: Vertical position time series of time series of station POTS (Postdam, Germany) Kalman estimator with (blue) and without (red) taking into account dependencies between station coordinates

Results

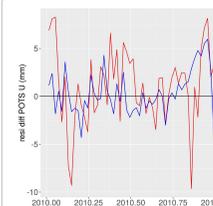


Fig. 6: Estimator of Kalman error of station POTS Up with (blue) and without (red) taking into account dependencies between station coordinates

stations	Station-by-station		Variations (%)
	Kalman filter	Spatiotemporal Kalman filter	
GRAS	0.056	0.050	-12.2
GRAZ	0.138	0.113	-19.9
MARS	0.102	0.118	+8.3
POTS	0.079	0.068	-16.9
WROC	0.173	0.143	-19.4
WTZA	0.041	0.036	-11.2
WTZS	0.210	0.182	-14.1
WTZZ	0.119	0.102	-15.6
ZIMJ	0.058	0.053	-8.3
ZIMM	0.045	0.042	-8.2
mean	0.108	0.094	-14.3
median	0.099	0.085	-15.2

Tab. 1: Uncertainties (mm/y) on the velocities without and with taking into account spatial correlations and corresponding variations

station	Station-by-station		Variation (%)
	Kalman filter	Spatiotemporal Kalman filter	
GRAS	2.59	1.81	-35.5
GRAZ	3.77	3.26	-14.2
MARS	3.81	3.49	-8.7
POTS	4.09	2.23	-56.1
WROC	5.00	3.24	-42.6
WTZA	3.20	1.90	-50.9
WTZS	4.40	3.81	-14.4
WTZZ	3.42	1.93	-55.8
ZIMJ	3.17	2.40	-27.5
ZIMM	3.42	2.10	-47.6
all stations	3.73	2.68	32.9

Tab. 2: RMS (mm) of signal estimation error without and with taking into account spatial correlations and corresponding variations

Conclusions

- Taking into account spatial dependencies:
- improves the predictive capacity of the model, which is shown by cross validation: predictions fit better data,
 - reduces formal errors of velocities, the reduction is more sensible for co-located stations,
 - provides more reliable interpolation of time series.

Reference

- Altamimi, Z., Rebischung, P., Métivier, L., & Collilieux, X. (2016). ITRF2014: A new release of the International Terrestrial Reference Frame modeling nonlinear station motions. *Journal of Geophysical Research: Solid Earth*, 121(8), 6109-6131.
- Williams, S. D. (2008). CATS: GPS coordinate time series analysis software. *GPS solutions*, 12(2), 147-153.



INSTITUT NATIONAL DE L'INFORMATION GÉOGRAPHIQUE ET FORÊSTIÈRE

Dépendances spatio-temporelles des séries de positions GNSS

Clément Benoist
clement.benoist@ign.fr



LAREG
LABORATOIRES DE RECHERCHE EN GÉODÉSIE

Résumé

Prise en compte des dépendances spatio-temporelles dans les séries de positions GNSS pour une meilleure détermination du mouvement des plaques tectoniques.

Les séries temporelles de positions de stations GNSS sont modélisables comme la somme d'une tendance linéaire, de signaux périodiques d'amplitudes à estimer et de fréquences fixes et de bruit. Une modélisation magnoprior du bruit peut conduire à des incertitudes irréalistes sur les vitesses estimées des stations, et donc nuire à leur interprétation géophysique.

Dans le but de tenir compte des corrélations temporelles du bruit lors de l'estimation des vitesses des stations, les paramètres de différents modèles de bruit ont été estimés par maximum de vraisemblance pour une série de positions, ce qui permettra ultérieurement de choisir le modèle de bruit le plus adapté.

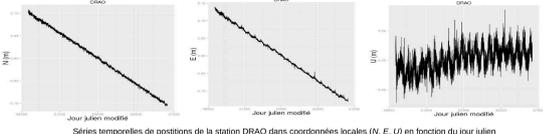
Une partie ultérieure de la thèse consistera à tenir compte, lors de l'estimation des vitesses, des dépendances spatiales qui existent entre séries de stations proches.

Données

Les données sont constituées de positions GNSS dans le repère local (N,E,U) [Nord, Sud, Haut] avec leur incertitudes exprimées par des écarts-types. Sont aussi indiquées les discontinuités estimées via l'indice SOLN.

Il y a une dichotomie entre les composantes horizontales E et N et la composante verticale U. Les composantes horizontales correspondent en général à une tendance linéaire bruitée alors que dans la composante verticale dominent des signaux périodiques dus à des phénomènes de surcharge.

La difficulté est concilier un traitement automatisé des données tout en tenant compte des particularités des stations.



Séries temporelles de positions de la station DRAO dans coordonnées locales (N, E, U) en fonction du jour julien modifié.

Vers le débruitage

Nous considérons que les séries temporelles de positions sont la somme d'une tendance linéaire correspondant au mouvement des plaques tectoniques, de termes périodiques de fréquences fixes et d'un terme de bruit.

Nous utilisons plusieurs modèles de bruit qui peuvent s'additionner :

- bruit de loi puissance : sa densité spectrale est proportionnelle à $1/f^\alpha$ où α est l'indice spectral.
- bruit blanc : le bruit à un instant est indépendant du bruit à un autre instant, le spectre de puissance est constant en fréquence, c'est un bruit de loi puissance d'indice spectral 0.
- bruit de marche aléatoire : somme cumulée de bruits blancs, c'est un bruit de loi puissance d'indice spectral 2 ; dans le cas gaussien, on parle de mouvement brownien.
- bruit de scintillation : c'est un bruit de loi puissance d'indice spectral 1, il est fréquemment rencontré dans de nombreux domaines dont la géodésie.

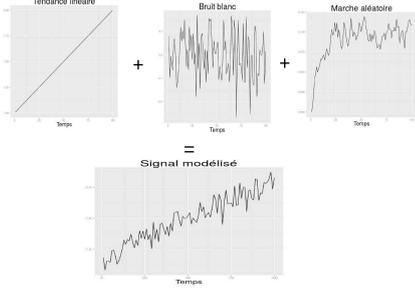
Dans les modèles, nous considérons quasi-exclusivement les versions gaussiennes de ces bruits.

En général, nous considérons dans le modèle une somme de bruit blanc et d'un ou plusieurs autres bruits.

Le débruitage s'effectue en deux étapes :

- le maximum de vraisemblance gaussien permet l'estimation des paramètres de bruit.
- Santamaría-Gómez et al. [2011] a comparé des modèles de bruit grâce, entre autres, au maximum de vraisemblance ; il propose des méthodes de choix de modèles pour le bruit.
- Bos et al. [2008] permet d'automatiser le calcul dans les séries temporelles de positions continues.
- le filtre de Kalman permet de débruité le signal, il a comme avantage d'être simple d'un point de vue théorique, de supporter les valeurs manquantes, cette méthode est connue largement dans le monde de la recherche scientifique, elle peut être effectuée en grande dimension ; une amélioration peut être d'utiliser le lisseur de Kalman. La difficulté réside dans l'introduction d'un bruit de corrélation dans le filtre.

J'ai implémenté le maximum de vraisemblance pour quelques modèles de bruit, il me sera possible de les compléter en fonction des besoins. J'ai traité des séries réelles. Le filtre de Kalman a été testé sur des données synthétiques.



Exemple de modélisation de séries temporelles de positions GNSS

Ouvertures

Nous travaillons pour l'instant station par station. Cela ne nous permet pas d'évaluer la vitesse de la station avec précision, on ne peut inférer valablement que l'incertitude sur celle-ci. Pour l'estimation des vitesses proprement dites, il faudra travailler à l'échelle d'un réseau global de stations.

Il faudra ensuite tenir compte des discontinuités présentes dans les séries de positions dues pour la plupart à des tremblements de terre et des changements d'équipement. Néanmoins, la détection des discontinuités n'entre pas dans le sujet de ma thèse. Nous nous contenterons donc d'une liste de discontinuités prédéfinies.

Une partie ultérieure de la thèse consistera à tenir compte, non seulement des dépendances temporelles, mais aussi des dépendances spatiales qui existent entre séries de stations proches.

Applications

La connaissance précise des vitesses des stations GNSS et de leurs incertitudes est nécessaire à un certain nombre d'applications :

- détermination du mouvement relatif des plaques tectoniques
- modification du rebond post-glaciaire
- observation de la montée globale du niveau des mers par marégraphe ou altimétrie satellitaire

Merci à X. Collilieux, O. Jamet, L. Métyvier pour leurs conseils.

Cadre de la thèse

- Commencée en 2014
- Réalisée au laboratoire LAREG (IGN)
- Direction : Z. Altamimi
- Encadrement : P. Rebischung
- École doctorale : Observatoire de Paris

Références

- Bos, M., Fernandes, R., Williams, S., & Bastos, L. (January 01, 2008). Fast error analysis of continuous GPS observations. *Journal of Geodesy*, 82, 3, 157-166.
- Santamaría-Gómez, A., Bouin, M.-N., Collilieux, X., & Wöppelmann, G. (January 01, 2011). Correlated errors in GPS position time series: Implications for velocity estimates. *Journal of Geophysical Research: Solid Earth*, 116.

Journées de la recherche de l'IGN 2015



FIGURE A.2 – Poster *Vers la prise en compte de la dépendance spatio-temporelle des séries de position GNSS dans leur analyse* présenté aux Journées de la Recherche en 2016

Introduction A

Eight Analysis Centers (AC) of the International GNSS Service (IGS) have completed a second reanalysis campaign (repro2) of the GNSS data collected by the IGS global tracking network back to 1994, using the latest available models and methodology (<http://acc.igs.org/repro2.html>). The AC repro2 contributions include in particular daily terrestrial frame solutions, the first time with sub-weekly resolution for the full IGS history. In order to study the random and systematic errors present in the AC repro2 station position time series, we stacked the daily repro2 solutions of each individual AC into a long-term linear frame, and analyzed the stacking residuals.

Processing strategy:

- pre-process daily AC solutions as described in Rebischung et al. (2015)
- reject stations with less than 500 days of observations or large non-linear motions
- correct for post-seismic deformations using ITRF2014P models
- use ITRF2014P discontinuity list (with marginal AC-specific adjustments)
- annual + semi-annual signals estimated jointly with long-term positions and velocities
- stacking iterations performed until < 0.01% of observations have normalized residuals larger than 5

Computations were performed using the new-generation CATREF software.

Temporal domain B

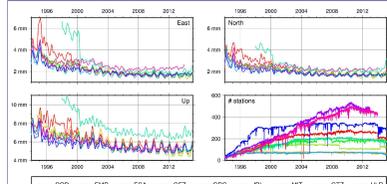


Figure 1 Top and bottom left: Daily WRMS of the stacking residuals. Bottom right: numbers of stations in the pre-processed, cleaned daily AC solutions. With a few exceptions, the non-linear, non-seasonal station motions have similar amplitudes amongst ACs, with daily WRMS oscillating around 1.7 mm in horizontal and 5.5 mm in vertical (after 2004). The main exceptions are:

- COD: higher noise level before 1999, especially in East – ambiguity resolution?
- GRG: higher noise level before 2000 – phase ambiguities not fixed
- GRG: substantially higher WRMS in Up (> 7 mm after 2004) – under investigation
- ULR: higher noise level in horizontal – sub-optimal combination of station clusters

Frequency domain C

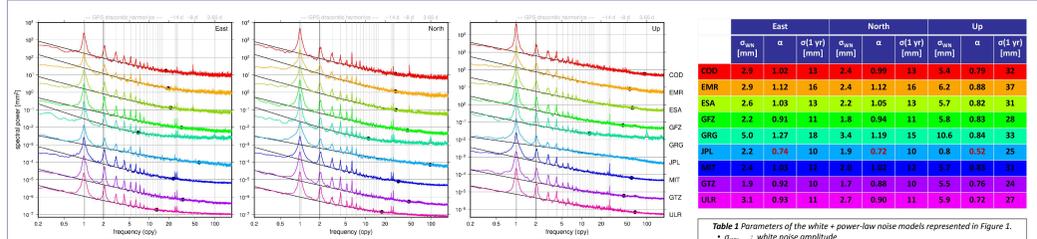


Figure 3 Color lines: stacked periodograms of the residual time series from the individual AC stackings, with annual and semi-annual signals restored. Black lines: results from the fits of white + power-law noise models (i.e., $a + b/f^\alpha$ functions) to the periodograms. Black dots: crossover frequencies between white and power-law noises.

	East			North			Up		
	σ_{white} [mm]	α	$a(1/yr)$ [mm]	σ_{white} [mm]	α	$a(1/yr)$ [mm]	σ_{white} [mm]	α	$a(1/yr)$ [mm]
COD	2.9	1.02	13	2.4	0.99	13	5.4	0.79	12
EMR	2.9	1.12	16	2.4	1.12	16	6.2	0.88	17
ESA	2.6	1.03	13	2.2	1.05	13	5.7	0.82	11
GFZ	2.2	0.91	11	1.8	0.94	11	5.3	0.83	7.8
GRG	5.0	1.27	18	3.4	1.19	15	10.6	0.94	13
JPL	2.2	0.74	10	1.5	0.72	10	0.8	0.57	25
MIT	2.8	1.09	12	2.0	1.03	12	5.2	0.81	21
GTZ	1.9	0.92	10	1.7	0.88	10	5.5	0.76	24
ULR	3.1	0.92	11	2.7	0.90	11	5.9	0.72	27

Table 1 Parameters of the white + power-law noise models represented in Figure 1.
 • σ_{white} : white noise amplitude
 • α : spectral index
 • $a(1/yr)$: total noise amplitude at 1 cpy

Background noise: According to the stacked periodograms shown in Figure 3, the background noises present in the AC repro2 station position time series can be well described by the sum of white and power-law noises. JPL excepted, the estimated spectral indices (Table 1) are close to 1 for the horizontal components, but closer to 0.8 for the vertical component. JPL's spectral indices are noticeably lower, and JPL's periodograms hardly show trace of white noise flattening at high frequencies.

Spectral lines: Spectral peaks at harmonics of the GPS draconic year (1.04 cpy) are visible for all ACs and components, generally up to the 8th. Some features evidenced by Ray et al. (2013) are additionally confirmed:

- GRG: broad peaks around 3.65 and 2.2 days – could be due to date rounding errors; unexplained peaks at 13.15 days in East and North
- COD & ESA: spectral peaks at 7.8 days in East and North – likely due to the use of GLODAS data
- 14.76 days (M, alias): spectral peaks visible in the horizontal periodograms of all ACs – seem more pronounced for GFZ/GTZ and MIT/ULR
- 14.19 days (O, alias): spectral peaks visible in the horizontal periodograms of some ACs – clearer for MIT/ULR
- consistent with a rotational origin (e.g. errors in the IGS subdaily COP tide model)
- 13.63/13.66 days (M, tide): spectral peaks visible for all ACs and components – faint for JPL, stronger for GFZ/GTZ and MIT/ULR

Geographic distribution of periodic signals D

We fitted sinusoids of different periods (annual, semi-annual, first 8 harmonics of the GPS draconic year, 14.77, 14.19 and 13.63 days) to the residual time series from the individual AC stackings and drew amplitude/phase maps of the most significant signals of each period, some of which are shown below. Inter-AC agreement is visually good for annual signals, moderate for semi-annual signals. Most draconic signals show strong spatial correlations, implying a main satellite-related source. Inter-AC agreement is good for the 2 first draconic harmonics, variable for the higher harmonics. The horizontal 14.77 and 14.19-day signals also show strong spatial correlations and similarities amongst ACs.

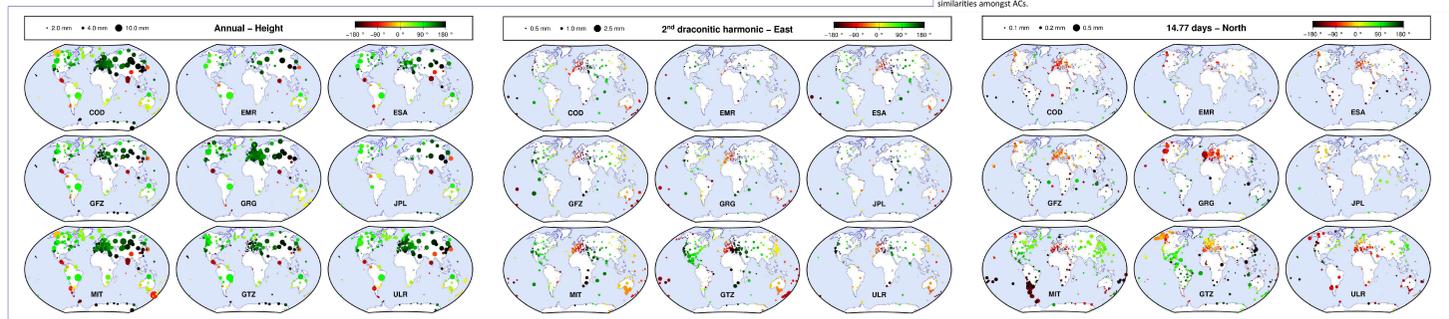


FIGURE A.3 – Poster Error analysis of the IGS repro2 station position time series présenté à l'AGU en 2015

Bibliographie

- [Abraham, 2009] Abraham, C. (2009). Method and apparatus for processing a satellite positioning system signal using a cellular acquisition signal. US Patent App. 10/926,792.
- [Altamimi et al., 2007] Altamimi, Z., Collilieux, X., Legrand, J., Garayt, B., and Boucher, C. (2007). ITRF2005 : A new release of the international terrestrial reference frame based on time series of station positions and earth orientation parameters. *Journal of Geophysical Research : Solid Earth*, 112(B9).
- [Altamimi et al., 2011] Altamimi, Z., Collilieux, X., and Métivier, L. (2011). ITRF2008 : an improved solution of the international terrestrial reference frame. *Journal of Geodesy*, 85(8) :457–473.
- [Altamimi et al., 2016] Altamimi, Z., Rebischung, P., Métivier, L., and Collilieux, X. (2016). ITRF2014 : A new release of the international terrestrial reference frame modeling nonlinear station motions. *Journal of Geophysical Research : Solid Earth*, 121(8) :6109–6131.
- [Altamimi et al., 2002] Altamimi, Z., Sillard, P., and Boucher, C. (2002). ITRF2000 : A new release of the international terrestrial reference frame for earth science applications. *Journal of Geophysical Research : Solid Earth*, 107(B10).
- [Amiri-Simkooei et al., 2017] Amiri-Simkooei, A. R., Mohammadloo, T. H., and Argus, D. F. (2017). Multivariate analysis of GPS position time series of JPL second reprocessing campaign. *Journal of Geodesy*, 91(6) :685–704.
- [Baire et al., 2014] Baire, Q., Bruyninx, C., Legrand, J., Pottiaux, E., Aerts, W., Defraigne, P., Bergeot, N., and Chevalier, J. (2014). Influence of different gps receiver antenna calibration models on geodetic positioning. *GPS solutions*, 18(4) :529–539.
- [Beckley et al., 2007] Beckley, B., Lemoine, F., Luthcke, S., Ray, R., and Zelensky, N. (2007). A reassessment of global and regional mean sea level trends from TOPEX and jason-1 altimetry based on revised reference frame and orbits. *Geophysical Research Letters*, 34(14).
- [Bélisle, 1992] Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on \setminus^d . *Journal of Applied Probability*, 29(4) :885–895.
- [Bellard et al., 2012] Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., and Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology letters*, 15(4) :365–377.
- [Bellard et al., 2013] Bellard, C., Leclerc, C., and Courchamp, F. (2013). Potential impact of sea level rise. *Nature Conservation*.

- [Blewitt and Lavallée, 2002] Blewitt, G. and Lavallée, D. (2002). Effect of annual signals on geodetic velocity. *Journal of Geophysical Research : Solid Earth*, 107(B7).
- [Bogusz and Klos, 2016] Bogusz, J. and Klos, A. (2016). On the significance of periodic signals in noise analysis of gps station coordinates time series. *GPS Solutions*, 20(4) :655–664.
- [Bos et al., 2013a] Bos, M., Fernandes, R., and Bastos, L. (2013a). Hector user manual version 1.1.
- [Bos et al., 2008] Bos, M., Fernandes, R., Williams, S., and Bastos, L. (2008). Fast error analysis of continuous GPS observations. *Journal of Geodesy*, 82(3) :157–166.
- [Bos et al., 2013b] Bos, M. S., Fernandes, R. M. S., Williams, S. D. P., and Bastos, L. (2013b). Fast error analysis of continuous GNSS observations with missing data. *Journal of Geodesy*, 87(4) :351–360.
- [Bossler, 2017] Bossler, P. (2017). GNSS : Systèmes globaux de positionnement par satellites.
- [Boucher and Altamimi, 2001] Boucher, C. and Altamimi, Z. (2001). ITRS, PZ-90 and WGS 84 : current realizations and the related transformation parameters. *Journal of Geodesy*, 75(11) :613–619.
- [Collilieux, 2008] Collilieux, X. (2008). *Analyse des séries temporelles de positions des stations de géodésie spatiale : application au Repère International de Référence Terrestre (ITRF)*. PhD thesis, Observatoire de Paris.
- [Collilieux et al., 2014] Collilieux, X., Altamimi, Z., Argus, D., Boucher, C., Dermanis, A., Haines, B., Herring, T., Kreemer, C., Lemoine, F., Ma, C., et al. (2014). External evaluation of the terrestrial reference frame : report of the task force of the iag sub-commission 1.2. In *Earth on the Edge : Science for a Sustainable Planet*, pages 197–202. Springer.
- [Collilieux et al., 2007] Collilieux, X., Altamimi, Z., Coulot, D., Ray, J., and Sillard, P. (2007). Comparison of very long baseline interferometry, gps, and satellite laser ranging height residuals from itr2005 using spectral and correlation methods. *Journal of Geophysical Research : Solid Earth*, 112(B12).
- [Collilieux et al., 2017] Collilieux, X., Altamimi, Z., Rebischung, P., and Métiévier, L. (2017). Coordinate kinematic models in the international terrestrial reference frame release. In *Honory Volume for A. Dermanis*. International Association of Geodesy.
- [Collilieux and Bernard, 2012] Collilieux, X. and Bernard, E. (2012). Modélisation spatiotemporelle pour la détermination des vitesses de déplacements d’un réseau global de stations GPS.
- [Collilieux et al., 2012] Collilieux, X., Bernard, E., and Jamet, O. (2012). Covariance spatiotemporelle. Communication personnelle.
- [Collilieux and Wöppelmann, 2011] Collilieux, X. and Wöppelmann, G. (2011). Global sea-level rise and its relation to the terrestrial reference frame. *Journal of Geodesy*, 85(1) :9–22.
- [Dee et al., 2011] Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al. (2011). The era-interim reanalysis : Configuration and performance of the

- data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656) :553–597.
- [Dmitrieva et al., 2015] Dmitrieva, K., Segall, P., and DeMets, C. (2015). Network-based estimation of time-dependent noise in gps position time series. *Journal of Geodesy*, 89(6) :591–606.
- [Dong et al., 2002] Dong, D., Fang, P., Bock, Y., Cheng, M., and Miyazaki, S. (2002). Anatomy of apparent seasonal variations from GPS-derived site position time series. *Journal of Geophysical Research (Solid Earth)*, pages 2075–+.
- [Dong et al., 2006] Dong, D., Fang, P., Bock, Y., Webb, F., Prawirodirdjo, L., Kedar, S., and Jamason, P. (2006). Spatiotemporal filtering using principal component analysis and karhunen-loeve expansion approaches for regional GPS network analysis. *Journal of Geophysical Research : Solid Earth*, 111(B3).
- [Dong et al., 1998] Dong, D., Herring, T., and King, R. (1998). Estimating regional deformation from a combination of space and terrestrial geodetic data. *Journal of Geodesy*, 72 :200–214.
- [Dong et al., 2003] Dong, D., Yunck, T., and Heflin, M. (2003). Origin of the international terrestrial reference frame. *Journal of geophysical research : solid earth*, 108(B4).
- [Dow et al., 2009] Dow, J. M., Neilan, R. E., and Rizos, C. (2009). The international GNSS service in a changing landscape of global navigation satellite systems. *Journal of Geodesy*, 83(3) :191–198.
- [Elósegui et al., 1995] Elósegui, P., Davis, J., Jaldehag, R., Johansson, J., Niell, A., and Shapiro, I. (1995). Geodesy using the global positioning system : The effects of signal scattering on estimates of site position. *Journal of Geophysical Research : Solid Earth*, 100(B6) :9921–9934.
- [Fish et al., 2005] Fish, M. R., Cote, I. M., Gill, J. A., Jones, A. P., Renshoff, S., and Watkinson, A. R. (2005). Predicting the impact of sea-level rise on caribbean sea turtle nesting habitat. *Conservation biology*, 19(2) :482–491.
- [Fletcher and Reeves, 1964] Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients. *The computer journal*, 7(2) :149–154.
- [Fondation, 2017] Fondation, W. (2017). Wikimedia commons.
- [Fowler, 2004] Fowler, C. M. R. (2004). *The Solid Earth : An Introduction to Global Geophysics*. Cambridge University Press, 2 edition.
- [Gazeaux et al., 2015] Gazeaux, J., Lebarbier, E., Collilieux, X., and Métivier, L. (2015). Joint segmentation of multiple gps coordinate series. *Journal de la Société Française de Statistique*, 156(4) :163–179.
- [Guyader, 2011] Guyader, A. (2011). Régression linéaire. *Université Rennes*, 2 :60–61.
- [Hauer, 2017] Hauer, M. E. (2017). Migration induced by sea-level rise could reshape the US population landscape. *Nature Climate Change*.
- [Heinkelmann et al., 2008] Heinkelmann, R., Freymueller, J. T., and Schuh, H. (2008). A postseismic relaxation model for the 2002 denali earthquake from GPS deformation analysis applied to VLBI. In *IVS 2008 General Meeting, St. Petersburg, Russia*.

- [Herring et al., 1990] Herring, T. A., Davis, J. L., and Shapiro, I. I. (1990). Geodesy by radio interferometry : The application of kalman filtering to the analysis of very long baseline interferometry data. *Journal of Geophysical Research : Solid Earth*, 95(B8) :12561–12581.
- [Hofmann-Wellenhof et al., 2000] Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J. (2000). *GPS, theory and practice*. Springer.
- [Hosking, 1981] Hosking, J. R. (1981). Fractional differencing. *Biometrika*, 68(1) :165–176.
- [IERS, 2006] IERS (2006). Sinex : Solution INdependent EXchange format. Technical report, International Earth Rotation and Reference Systems Service.
- [Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1) :35–45.
- [Khan et al., 2010] Khan, S. A., Wahr, J., Bevis, M., Velicogna, I., and Kendrick, E. (2010). Spread of ice mass loss into northwest greenland observed by grace and gps. *Geophysical Research Letters*, 37(6).
- [King and Watson, 2010a] King, M. A. and Watson, C. S. (2010a). Long GPS coordinate time series : multipath and geometry effects. *Journal of Geophysical Research : Solid Earth*, 115(B4).
- [King and Watson, 2010b] King, M. A. and Watson, C. S. (2010b). Long GPS coordinate time series : multipath and geometry effects. *Journal of Geophysical Research : Solid Earth*, 115(B4).
- [King et al., 2008] King, M. A., Watson, C. S., Penna, N. T., and Clarke, P. J. (2008). Subdaily signals in GPS observations and their effect at semiannual and annual periods. *Geophysical research letters*, 35(3).
- [Lambeck et al., 2014] Lambeck, K., Rouby, H., Purcell, A., Sun, Y., and Sambridge, M. (2014). Sea level and global ice volumes from the last glacial maximum to the holocene. *Proceedings of the National Academy of Sciences*, 111(43) :15296–15303.
- [Langbein, 2017] Langbein, J. (2017). Improved efficiency of maximum likelihood analysis of time series with temporally correlated errors. *Journal of Geodesy*, 91(8) :985–994.
- [Lercier, 2014] Lercier, D. (2014). *Prise en compte des déformations co- et post-sismiques pour la détermination d'un repère de référence*. Thèse de l'école doctorale d'astronomie et d'astrophysique d'île de france, Observatoire de Paris, 61 Avenue de l'Observatoire - 75014 PARIS. Thèse dirigée par MM Z. Altamimi, X. Collilieux et C. Vigny.
- [Levallois, 1989] Levallois, J.-J. (1989). *Mesurer la Terre : 300 ans de géodésie française*. Presses de l'école nationale des ponts et chaussées.
- [Liu et al., 2018] Liu, N., Dai, W., Santerre, R., and Kuang, C. (2018). A MATLAB-based kriged kalman filter software for interpolating missing data in GNSS coordinate time series. *GPS Solutions*, 22(1) :25.
- [Luceri and Pavlis, 2016] Luceri, V. and Pavlis, E. (2016). The ILRS contribution to ITRF2014. Technical report, International ILRS Service.
- [Ma et al., 2013] Ma, C., MacMillan, D., Bolotin, S., Le Bail, K., Gordon, D., and Gipson, J. (2013). Comparison of realizations of the terrestrial reference

- frame. In *Reference Frames for Applications in Geosciences*, pages 51–56. Springer.
- [Mao et al., 1999] Mao, A., Harrison, C. G., and Dixon, T. H. (1999). Noise in GPS coordinate time series. *Journal of Geophysical Research : Solid Earth*, 104(B2) :2797–2816.
- [Marone et al., 1991] Marone, C. J., Scholtz, C., and Bilham, R. (1991). On the mechanics of earthquake afterslip. *Journal of Geophysical Research : Solid Earth*, 96(B5) :8441–8452.
- [McCarthy, 1996] McCarthy, D. D. (1996). IERS conventions (1996). *IERS Technical Note*, 21 :1–95.
- [Menemenlis et al., 2008] Menemenlis, D., Campin, J.-M., Heimbach, P., Hill, C., Lee, T., Nguyen, A., Schodlok, M., and Zhang, H. (2008). Ecco2 : High resolution global ocean and sea ice data synthesis. *Mercator Ocean Quarterly Newsletter*, 31 :13–21.
- [Morel and Willis, 2005] Morel, L. and Willis, P. (2005). Terrestrial reference frame effects on global sea level rise determination from TOPEX/poseidon altimetric data. *Advances in Space Research*, 36(3) :358–368.
- [Nelder and Mead, 1965] Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The computer journal*, 7(4) :308–313.
- [Pearlman et al., 2002] Pearlman, M. R., Degnan, J. J., and Bosworth, J. (2002). The international laser ranging service. *Advances in Space Research*, 30(2) :135–143.
- [Peltier et al., 2015] Peltier, W., Argus, D., and Drummond, R. (2015). Space geodesy constrains ice age terminal deglaciation : The global ice-6g_c (vm5a) model. *Journal of Geophysical Research : Solid Earth*, 120(1) :450–487.
- [Penna and Stewart, 2003] Penna, N. and Stewart, M. (2003). Aliased tidal signatures in continuous GPS height time series. *Geophysical Research Letters*, 30(23).
- [Petit and Luzum, 2010] Petit, G. and Luzum, B. (2010). IERS conventions (2010). Technical report, DTIC Document.
- [Prenat et al., 2010] Prenat, M., Keribin, C., and Rossignol, R. (2010). Séries chronologiques. *Université Paris-Sud*.
- [Rauch et al., 1965] Rauch, H. E., Striebel, C., and Tung, F. (1965). Maximum likelihood estimates of linear dynamic systems. *AIAA journal*, 3(8) :1445–1450.
- [Ray et al., 2008] Ray, J., Altamimi, Z., Collilieux, X., and van Dam, T. (2008). Anomalous harmonics in the spectra of GPS position estimates. *GPS Solutions*, 12(1) :55–64.
- [Rebischung, 2015] Rebischung, P. (2015). Combined repro2 SINEX solutions available. IGS mail.
- [Rebischung et al., 2016] Rebischung, P., Altamimi, Z., Ray, J., and Garayt, B. (2016). The IGS contribution to ITRF2014. *Journal of Geodesy*, 90(7) :611–630.
- [Rebischung et al., 2014] Rebischung, P., Altamimi, Z., and Springer, T. (2014). *Can GNSS contribute to improving the ITRF definition ?* PhD thesis, Observatoire de Paris.

- [Rebischung et al., 2017] Rebischung, P., Chanard, K., Metivier, L., and Altamimi, Z. (2017). Flicker noise in gnss station position time series : How much is due to crustal loading deformations? *AGU Fall Meeting Abstracts*.
- [Rebischung et al., 2015] Rebischung, P., Garayt, B., Altamimi, Z., and Collilioux, P. (2015). The IGS contribution to ITRF2014. 26th IUGG General Assembly.
- [Reichle et al., 2011] Reichle, R. H., Koster, R. D., De Lannoy, G. J., Forman, B. A., Liu, Q., Mahanama, S. P., and Touré, A. (2011). Assessment and enhancement of merra land surface hydrology estimates. *Journal of climate*, 24(24) :6322–6338.
- [Rodriguez-Solano et al., 2012] Rodriguez-Solano, C., Hugentobler, U., Steigenberger, P., and Lutz, S. (2012). Impact of earth radiation pressure on GPS position estimates. *Journal of geodesy*, 86(5) :309–317.
- [Rothacher et al., 2008] Rothacher, M., Beutler, G., Bosch, W., Donnellan, A., Gross, R. S., Hinderer, J., Ma, C., Pearlman, M., Plag, H., Richter, B., Ries, J., Schuh, H., Seitz, F., Shum, C., Smith, D. A., Thomas, M., Velicogna, I., Wah, J., Willis, P., and Woodworth, P. (2008). *GGOS 2020*, chapter 9, pages 142–159. International Association of Geodesy. The future Global Geodetic Observing System (GGOS).
- [Santamaría-Gómez et al., 2011] Santamaría-Gómez, A., Bouin, M.-N., Collilioux, X., and Woepplmann, G. (2011). Correlated errors in GPS position time series : Implications for velocity estimates. *Journal of Geophysical Research : Solid Earth*, 116(B1).
- [Savage and Prescott, 1978] Savage, J. and Prescott, W. (1978). Asthenosphere readjustment and the earthquake cycle. *Journal of Geophysical Research : Solid Earth*, 83(B7) :3369–3376.
- [Scargle, 1982] Scargle, J. D. (1982). Studies in astronomical time series analysis. ii-statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263 :835–853.
- [Schmid et al., 2007] Schmid, R., Steigenberger, P., Gendt, G., Ge, M., and Rothacher, M. (2007). Generation of a consistent absolute phase-center correction model for gps receiver and satellite antennas. *Journal of Geodesy*, 81(12) :781–798.
- [Schuh and Behrend, 2012] Schuh, H. and Behrend, D. (2012). VLBI : a fascinating technique for geodesy and astrometry. *Journal of Geodynamics*, 61 :68–80.
- [Shepherd et al., 2012] Shepherd, A., Ivins, E. R., Geruo, A., Barletta, V. R., Bentley, M. J., Bettadpur, S., Briggs, K. H., Bromwich, D. H., Forsberg, R., Galin, N., et al. (2012). A reconciled estimate of ice-sheet mass balance. *Science*, 338(6111) :1183–1189.
- [Sillard, 1999] Sillard, P. (1999). *Modélisation des systèmes de référence*. PhD thesis, Observatoire de Paris.
- [Steigenberger et al., 2009] Steigenberger, P., Boehm, J., and Tesmer, V. (2009). Comparison of gmf/gpt with vmf1/ecmwf and implications for atmospheric loading. *Journal of Geodesy*, 83(10) :943.

- [Stewart et al., 2005] Stewart, M., Penna, N., and Lichti, D. (2005). Investigating the propagation mechanism of unmodelled systematic errors on coordinate time series estimated using least squares. *Journal of geodesy*, 79(8) :479–489.
- [Tregoning and Herring, 2006] Tregoning, P. and Herring, T. A. (2006). Impact of a priori zenith hydrostatic delay errors on gps estimates of station heights and zenith total delays. *Geophysical Research Letters*, 33.
- [Tufféry, 2012] Tufféry, S. (2012). *Data Mining et statistique décisionnelle : L'intelligence des données*. Technip.
- [Wdowinski et al., 1997] Wdowinski, S., Bock, Y., Zhang, J., Fang, P., and Genrich, J. (1997). Southern california permanent GPS geodetic array : Spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 landers earthquake. *Journal of Geophysical Research : Solid Earth*, 102(B8) :18057–18070.
- [Williams, 2003] Williams, S. (2003). The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. *Journal of Geodesy*, 76(9) :483–494.
- [Williams, 2008] Williams, S. D. (2008). CATS : GPS coordinate time series analysis software. *GPS solutions*, 12(2) :147–153.
- [Willis et al., 2010] Willis, P., Fagard, H., Ferrage, P., Lemoine, F. G., Noll, C. E., Noomen, R., Otten, M., Ries, J. C., Rothacher, M., Soudarin, L., et al. (2010). The international DORIS service (IDS) : toward maturity. *Advances in space research*, 45(12) :1408–1420.
- [Wu et al., 2015] Wu, X., Abbondanza, C., Altamimi, Z., Chin, T. M., Collilioux, X., Gross, R. S., Heflin, M. B., Jiang, Y., and Parker, J. W. (2015). KALREF—a kalman filter and time series approach to the international terrestrial reference frame realization. *Journal of Geophysical Research : Solid Earth*, 120(5) :3775–3802.
- [Xu et al., 2017] Xu, X., Dong, D., Fang, M., Zhou, Y., Wei, N., and Zhou, F. (2017). Contributions of thermoelastic deformation to seasonal variations in gps station position. *GPS Solutions*, 21(3) :1265–1274.
- [Zhang et al., 1997] Zhang, J., Bock, Y., Johnson, H., Fang, P., Williams, S., Genrich, J., Wdowinski, S., and Behr, J. (1997). Southern california permanent GPS geodetic array : Error analysis of daily position estimates and site velocities. *Journal of Geophysical Research : Solid Earth*, 102(B8) :18035–18055.

RÉSUMÉ

Tout positionnement global et précis a besoin d'un repère de référence. Les repères de référence s'appuient sur la donnée de coordonnées de points matérialisés appelés points de référence. En ces points sont situés des capteurs de positionnement tels que les GNSS (dont le GPS est la déclinaison américaine)~; ils fournissent, après un certain nombre de calculs, les positions de manière périodique qui varient par exemple sous l'effet des mouvements tectoniques.

Nous avons évalué l'apport de la prise en compte des dépendances entre stations. Entre stations, nous avons confirmé que les positions entre points de référence (matérialisés par des stations) sont corrélées. La prise en compte de la dépendance entre stations améliore la capacité à prédire des positions. De plus, nous avons montré que cette prise en compte aboutit à des estimations de vitesses meilleures pour les stations comportant peu de positions.

MOTS CLÉS

ITRF, géodésie, statistiques

ABSTRACT

All global and precise positioning needs reference frame. Reference frames provide coordinates of materialized points called reference points. At these points positioning sensors like GNSS (whose GPS is American network) are located; after calculations, they provide periodically positions which vary for example under the effect of tectonic movements.

We evaluated contribution of taking into account of spatial (geographic) dependencies to geodesic issues. Between stations, we confirmed that positions between reference points (materialized by stations) are correlated. Taking into account dependencies between stations led an improvement of predictive capacity. Moreover, we proved that taking into account these improve speed estimators for short-duration stations.

KEYWORDS

ITRF, geodesy, statistics