THÈSE

présentée en vue de l'obtention du grade de

DOCTEUR DE L'UNIVERSITÉ TOULOUSE III

Discipline : Mathématiques appliquées
Spécialité : Statistique

par

Elie Maza

Prévision de trafic routier par des méthodes statistiques

Espérance structurelle d'une fonction aléatoire

Soutenue le 6 décembre 2004
devant le jury composé de Messieurs les Professeurs :

Jean-Marc Azaïs	Université Paul Sabatier	Président
Jean-Marc Bardet	Université Paris I	Directeur de thèse
Christophe Communay	Société IRINCOM	Examinateur
Juan Antonio Cuesta Albertos	Universidad de Cantabria	Rapporteur
Gilles Ducharme	Université Montpellier II	Examinateur
Fabrice Gamboa	Université Paul Sabatier	Directeur de thèse
Gérard Grégoire	Université Grenoble I	Rapporteur
Jean-Michel Loubes	CNRS et Université Montpellier II	Examinateur

Laboratoire de Statistique et Probabilités
UFR MIG - UMR CNRS 5583
Université Paul Sabatier - Toulouse III

A mes parents.
A Myriam.

Remerciements

Je tiens en premier lieu à remercier Jean-Marc Bardet et Fabrice Gamboa pour avoir dirigé mes recherches durant ces trois années de thèse. Je remercie tout particulièrement Fabrice pour ses nombreux conseils, statistiques et non statistiques, et pour avoir su encourager et valoriser mes initiatives. Travailler avec lui a été très enrichissant et très plaisant. J'espère pouvoir continuer à profiter, après ma thèse, de son expérience et de sa sympathie.

Je tiens aussi à remercier Jean-Michel Loubes que j'ai connu dès le début de ma thèse alors qu'il s'apprêtait à soutenir son doctorat. Dès lors, il a su, par ses nombreux conseils, me faire part de son expérience et de ses connaissances. Collaborer avec lui a été pour moi un plaisir et m'a permis d'arriver à bout de nombreux problèmes mathématiques. Merci pour le temps qu'il n'a pas hésité à me consacrer.

Je remercie aussi messieurs Juan Antonio Cuesta Albertos et Gérard Grégoire pour m'avoir fait l'honneur d'être les rapporteurs de ma thèse. Merci pour les idées et les corrections qu'ils ont apportées au manuscrit d'origine.

Je remercie Gilles Ducharme pour m'avoir fait l'honneur de sa présence dans mon jury et pour l'intérêt qu'il a porté à mon travail. Merci pour sa lecture minutieuse de mon manuscrit et pour les remarques et corrections qu'il m'a suggérées.

Je tiens à remercier le Laboratoire de Statistique et Probabilités qui, il y a de ça trois ans, me permettait de m'initier à ce beau métier qu'est la Recherche. Merci donc à Jean-Marc Azaïs qui, les circonstances aidant, a pensé à moi pour ce projet.

Je remercie aussi la société IRINCOM, et son directeur Christophe Communay, pour avoir mis en place cette thèse CIFRE au sein de son projet de prévision de trafic routier, et pour m'avoir permis de faire partie de l'aventure.

Je suis aussi reconnaissant aux dirigeants de l'ACI MIST-R (Modélisation Informatique et Statistique du Trafic Routier): M. Danech-Pajouh (INRETS), F. Gamboa (LSP), J-M. Loubes (CNRS et Université Montpellier II) et M. Sebag (CNRS et LRI Orsay), ainsi qu'à G. Celeux (INRIA FUTURS Orsay) de me permettre de participer activement à ce vaste projet de recherche inter-universitaire.

Merci enfin aux nombreux doctorants qui ont partagé avec moi un bureau, un épineux problème mathématique, une idée, une soirée, un congrès ou une pause café.

Table des matières

Remerciements **5**

Introduction **9**

1 Prévision de trafic routier **15**
1.1 Les données 15
1.2 L'objectif et la méthodologie 20
1.3 Une distance appropriée à nos données 22
1.4 Quelques résultats 27

2 Prévision de trafic routier (article soumis) **29**
2.1 Introduction 30
2.2 Data and methodology 31
2.2.1 Description 31
2.2.2 Data quality 32
2.2.3 Forecasting method 34
2.3 Mixture model 35
2.3.1 Model description 35
2.3.2 Estimation of the number of components of the mixture . . . 38
2.3.3 Results 39
2.4 Classification method 41
2.4.1 Hierarchical classification 41
2.4.2 Choice of the optimal number of clusters 43
2.4.3 Results 45
2.5 Travel time forecasting 46
2.6 Conclusion 51

3 Modèle de translation **53**
3.1 Introduction 53
3.2 Le contraste 55
3.3 Les résultats asymptotiques 55
3.4 Application 56

4 Modèle de translation (article soumis) **59**
4.1 Introduction 60
4.2 Semi-parametric estimation of shifts 62
4.2.1 Model 62

4.2.2 Identifiability 63
4.2.3 Estimation 64
4.3 Asymptotic normality 65
4.4 Applications and simulations 67
4.4.1 Simulations 67
4.4.2 Pinch force data 68
4.4.3 Application to road traffic forecasting 69
4.5 Technical Lemmas 69
4.6 Appendix 70
4.7 Figures 82

5 Espérance structurelle 87
5.1 Le modèle de déformation 91
5.2 Estimation de l'espérance structurelle 92
5.2.1 Modèle continu 93
5.2.2 Modèle discret 97
5.2.3 Estimation de l'espérance structurelle 105
5.2.4 Réduction des hypothèses sur le processus de déformation . . 107
5.3 Cas où la fonction f n'est pas croissante 110
5.3.1 Définition de notre transformation 111
5.3.2 Propriétés 111
5.3.3 Application à notre modèle 114
5.4 Simulations 119
5.4.1 Simulation des fonctions de déformation 119
5.4.2 Simulation des données 122
5.4.3 Estimation 123
5.5 Cas où les mesures sont bruitées 124

Bibliographie 135

Introduction

Le présent travail de thèse est le fruit d'une Convention Industrielle de Formation par la Recherche (CIFRE). Le partenaire académique de cette thèse CIFRE est le Laboratoire de Statistique et Probabilités de l'Université Paul Sabatier (Toulouse III). Quant au partenaire industriel, il s'agit de la société IRINCOM/TrafficFirst (http://www.trafficfirst.com/) dirigée par Christophe Communay. TrafficFirst est une société spécialisée dans l'élaboration de l'information trafic en temps réel. Elle met ces données à disposition des professionnels pour qu'ils puissent les intégrer dans leurs applications. La première partie de cette thèse fait partie intégrante du projet de prévision de trafic routier mené a bien par la société TrafficFirst.

Le recueil de mesures de trafic routier s'est considérablement amélioré, quantitativement et qualitativement, ces dernières années. Cela traduit un besoin croissant des gestionnaires du trafic routier et des utilisateurs professionnels. En effet, ces professionnels accordent une importance croissante aux déplacements effectués dans une infrastructure routière qui est en pleine expansion. Ce besoin est clairement identifié dans de très nombreux projets de grande envergure tels que [DPA94] et [VGDPMW98].

Deux caractéristiques font du projet qui sous-tend cette thèse un projet original. D'une part, jusqu'à présent, les multiples études effectuées sur le sujet étaient destinées aux gestionnaires de la route (comme dans [VGDPMW98]) et non pas, comme ici, aux utilisateurs professionnels. De plus, les différentes études effectuées jusque là portent sur des sections de routes bien précises (comme dans [DPA94]) et non pas sur la totalité du réseau routier, ce qui est ici l'objectif. Plus précisément, aucune méthode de prévision de trafic sur la totalité d'un réseau routier n'est actuellement en place. En effet, par exemple en France, les seules prévisions de trafic portent sur un horizon d'une journée au minimum (*Bison-Futé*). Par contre, les prévisions à horizon d'une à deux heures sont inexistantes ou non développées, ou non développables, à grande échelle.

Les chapitres 1 et 2 sont dédiés à la mise en place d'une méthode de prévision de trafic routier à court terme (de quelques minutes à quelques heures). Plus précisément, nous y développons une méthode de prévision de temps de parcours pour le réseau autoroutier d'Île-de-France. Le chapitre 2 est un article actuellement soumis. Les chapitres 3, 4 et 5 sont dédiés à une problématique issue de l'étude des données de trafic routier: le recalage de courbes. Le chapitre 4 est aussi un article actuellement soumis.

Nous allons maintenant décrire de manière succincte le contenu de la thèse. Les cinq chapitres décrits brièvement ci-dessus peuvent être regroupés en trois parties comme suit:

i) la prévision de temps de parcours,
ii) le modèle de translation,
iii) l'espérance structurelle.

Le contenu de chacune de ces trois parties est le suivant.

i) La prévision de temps de parcours

Dans le chapitre 1, nous décrivons les données qui font l'objet de notre étude et qui sont utilisées dans la méthode de prévision de temps de parcours. Ces données sont issues d'appareils de mesure (ou stations de comptage) situés le long des principaux axes routiers. Ces stations de comptage sont réparties de façon irrégulière, tous les 300 à 500 mètres environ, de manière plus ou moins dense selon les axes routiers. Les mesures issues d'une station de comptage sont: le débit, le taux d'occupation et la vitesse du flux de véhicules détecté. Chacune de ces trois mesures de trafic peut être interprétée comme une fonction du temps. Nous pouvons ainsi présenter un échantillon de taille m de l'une quelconque de ces mesures (notre historique de données) comme les fonctions du temps

$$f_i : [a,b] \to \mathbb{R}, i = 1, \ldots, m. \tag{1}$$

De plus, ces fonctions sont observées de façon discrète sur les $n+1$ points équirépartis

$$t_j = a + j\frac{b-a}{n} \in [a,b], j = 0, \ldots, n. \tag{2}$$

Les deux hypothèses suivantes sont sous-jacentes à notre méthode de prévision de trafic sur un réseau routier. D'une part, nous supposons qu'il existe un certain nombre fini de comportements du réseau routier et que la connaissance de ces comportements types nous permet de prévoir les comportements futurs. D'autre part, nous supposons que le comportement global du réseau peut être déduit des comportements observés sur chaque station de comptage. Ces hypothèses générales, transposées à nos données, reviennent à supposer que, pour chaque station de comptage, nos courbes observées sont les réalisations d'un modèle de mélange du type

$$Y_{ij} = f_i(t_j) + \varepsilon_{ij} = \sum_{k=1}^{K} \mathbb{1}_k(X_i)\, f_k^*(t_j) + \varepsilon_{ij}, i = 1, \ldots, m, j = 0, \ldots, n, \tag{3}$$

où, les fonctions $(f_k^*)_{k=1,\ldots,K}$ sont les comportements types (ou profils types) de la station de comptage considérée, observés sur la grille (2); les variables aléatoires non observées $(X_i)_{i=1,\ldots,m}$ sont supposées i.i.d. et à valeurs dans l'ensemble discret $\{1, \ldots, K\}$, elles correspondent aux variables exogènes inconnues ayant une influence sur le trafic routier (jour de la semaine, calendrier scolaire, météorologie, etc.); les variables aléatoires $(\varepsilon_{ij})_{i=1,\ldots,m,\, j=0,\ldots,n}$ modélisent le bruit de mesure. Ces hypothèses

nous ont naturellement conduit à utiliser une méthode de classification automatique afin d'estimer les profils types $(f_k^*)_{k=1,\ldots,K}$. De plus, nous avons fait appel à des principes de la théorie de l'apprentissage afin d'estimer le nombre optimal de comportements K. La méthode est résumée dans le chapitre 1 et détaillée dans le chapitre 2. D'autre part, dans le chapitre 1, nous définissons la distance de $\mathbb{R}^{n+1}$ qui est utilisée lors de notre classification et en montrons une propriété importante pour notre étude. En effet, nous constatons que nos données, qui sont des fonctions du temps, ne peuvent pas être discriminées avec une distance classique comme la distance euclidienne. Cette problématique est par exemple étudiée dans [DLB96] qui est consacré à l'analyse de fonctions du temps, ou données évolutives, ou données longitudinales. Certaines de nos courbes présentent donc des similarités de structure (des courbes semblables translatées les unes par rapport aux autres) qui sont prises en compte par notre distance

$$\Delta\,(x,y) = \sqrt{(x-y)'\,D\,(x-y)},$$

où, x et y sont les vecteurs $(f_{i_1}(t_j))_{j=0,\ldots,n} \in \mathbb{R}^{n+1}$ et $(f_{i_2}(t_j))_{j=0,\ldots,n} \in \mathbb{R}^{n+1}$, avec des valeurs (i_1,i_2) dans $\{1,\ldots,m\}^2$; la matrice D, de taille $n+1$, est définie par ces éléments

$$D_{kl} = \frac{(n+1)-|k-l|}{n+1}, k = 1,\ldots,n+1, l = 1,\ldots,n+1.$$

La méthode de classification automatique et de prévision est comparée (dans le chapitre 2) à une méthode d'estimation des paramètres du modèle de mélange (3) qui est basée sur la maximisation de la log-vraisemblance. Cette dernière méthode fait appel à une extension stochastique de l'algorithme EM (Expectation and Maximisation). En effet, les variables aléatoires $(X_i)_{i=1,\ldots,m}$ étant non observées, l'algorithme EM nous permet de maximiser une log-vraisemblance complétée par des simulations de ces variables.

ii) Le modèle de translation

Les chapitres 3 et 4 sont dédiés à l'étude du modèle de translation

$$f_i(\cdot) = f\,(\cdot - \theta_i^*) + \varepsilon_i(\cdot), i = 1,\ldots,m, \tag{4}$$

où, $f : \mathbb{R} \to \mathbb{R}$ est une fonction T-périodique inconnue ($T = b - a$); les valeurs $(\theta_i^*)_{i=1,\ldots,m} \in \mathbb{R}^m$ sont les paramètres de translation inconnus. En effet, la classification des courbes issues de (1) nous a fait constater que, dans certaines des classes, les courbes semblaient se déduire les unes des autres par une simple translation. Dans notre contexte, cela peut s'expliquer aisément en supposant que le phénomène aléatoire de formation des congestions de trafic routier garde la même structure d'un jour à l'autre mais varie dans le temps, c'est-à-dire dans la journée. Dans le chapitre 3, nous résumons les résultats théoriques obtenus, détaillons les motivations et donnons les résultats sur notre prévision de temps de parcours. Le chapitre 4 est l'article soumis, nous y effectuons l'estimation des paramètres $(\theta_i^*)_{i=1,\ldots,m} \in \mathbb{R}^m$ par une méthode de M-estimation. Le modèle étudié est en fait déduit du modèle (4)

par un passage à la base de Fourier de nos fonctions (1). Plus précisément, en notant j le nombre complexe tel que $j^2 = -1$, le modèle étudié s'écrit

$$d_{il} = e^{-jl\alpha_i^*} c_l(f) + w_{il}, i = 1, \ldots, m, l = -(n-1)/2, \ldots, (n-1)/2,$$

où, n est supposé impair; les valeurs $(c_l(f))_{l\in\mathbb{Z}}$ sont les coefficients de Fourier de la fonction f; les paramètres $\left(\alpha_i^* = \frac{2\pi}{T}\theta_i^*\right)_{i=1,\ldots,m}$ sont nos paramètres de translation normalisés; les variables aléatoires $(w_{il})_{i=1,\ldots,m,\, l=-(n-1)/2,\ldots,(n-1)/2}$ sont i.i.d. complexes et modélisent le bruit de mesure. Ainsi, la base de Fourier diagonalise l'opérateur de translation et nous permet d'estimer les valeurs des paramètres de translation en minimisant le critère empirique

$$M_n(\alpha) = \frac{1}{m} \sum_{i=1}^{m} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \left|\tilde{c}_{il}(\alpha) - \hat{c}_l(\alpha)\right|^2,$$

où, $\alpha = (\alpha_i)_{i=1,\ldots,m} \in \mathbb{R}^m$;

$$\tilde{c}_{il}(\alpha) = e^{jl\alpha_i} d_{il}, i = 1, \ldots, m, l = -(n-1)/2, \ldots, (n-1)/2,$$

est le coefficient de Fourier rephasé;

$$\hat{c}_l(\alpha) = \frac{1}{m} \sum_{i=1}^{m} \tilde{c}_{il}(\alpha), l = -(n-1)/2, \ldots, (n-1)/2,$$

est la moyenne des coefficients rephasés; la suite $(\delta_l)_{l\in\mathbb{Z}}$ est convenablement choisie. Nous montrons la consistance de l'estimateur obtenu dans ce modèle semi-paramétrique, ainsi que sa normalité asymptotique.

iii) L'espérance structurelle

Dans le chapitre 5, nous élargissons le modèle (4). En effet, l'utilisation de ce modèle s'est avérée être inefficace dans notre cadre de prévision de trafic. Cela peut s'expliquer par le fait que la similarité que nous avons constatée entre certaines de nos courbes ne se déduit par seulement d'une simple translation mais d'une déformation plus générale de l'axe des abscisses. Nous pouvons traduire cela en écrivant

$$f_i(\cdot) = f \circ h_i^{-1}(\cdot), i = 1, \ldots, m,$$

où les fonctions, dites de déformation, $(h_i)_{i=1,\ldots,m}$ sont inconnues. Ce problème est explicité dans [RS97] et [RS02]. Lorsque nos observations sont des fonctions du temps, deux sortes d'aléas peuvent apparaître: un aléa en amplitude et un aléa en temps. D'un point de vue statistique, ce problème apparaît dans [Rao58] et fait référence à la comparaison de courbes de croissance. De nombreuses méthodes de résolution ont été proposées ces dernières années. Notamment dans [WG97] sous la dénomination anglaise *structural averaging*, et sous l'appellation *curve registration* dans [RD91], [RWF95], [RL98], [KG92], [Kne94] et [KLMR00]. Or, les estimations

des fonctions $(h_i)_{i=1,\ldots,m}$ proposées dans les travaux cités ci-dessus ne tiennent pas compte de l'information apportée par un nombre croissant d'observations. En effet, ces méthodes sont principalement basées sur le recalage de deux fonctions. Une fois ce recalage obtenu, il est possible de recaler l'ensemble des m fonctions sur une fonction donnée a priori ou sur une fonction moyenne dépendant elle-même des fonctions $(h_i)_{i=1,\ldots,m}$. Comme nous le détaillerons dans le chapitre 5, le modèle ainsi posé n'est pas identifiable. C'est aussi le cas de notre modèle (4). Notre nouveau modèle s'écrit

$$Y_{ij} = f \circ H_i^{-1}(w,t_j), i = 1, \ldots, m, j = 0, \ldots, n,$$

où, les processus aléatoires $(H_i)_{i=1,\ldots,m}$ sont i.i.d. et la valeur w appartient à un certain espace probabilisé. Ce dernier modèle est lui aussi non identifiable. Par contre, l'avantage d'un tel modèle est qu'il nous permet, sous certaines hypothèses sur le processus de déformation H sous-jacent, de définir de manière unique et intrinsèque, une notion d'*espérance structurelle* qui nous permet de palier à la non identifiabilité. Ainsi, dans le chapitre 5, nous proposons un estimateur de cette espérance structurelle et en montrons la consistance et la normalité asymptotique.

Chapitre 1

Prévision de temps de parcours sur le réseaux autoroutier d'Île-de-France

La section 1.1 est dédiée à la description des données de trafic routier. Dans les sections suivantes, nous résumons la méthode de prévision et donnons quelques résultats. La méthode de prévision est entièrement décrite dans l'article soumis [LML02] qui constitue le corps du chapitre 2. Ainsi, dans la section 1.2, nous rappelons l'objectif de la prévision et la méthodologie utilisée. Dans la section 1.3, nous donnons des résultats obtenus après la soumission de l'article. Enfin, dans la section 1.4, nous donnons quelques résultats.

1.1 Les données

Une grande partie du réseau routier de l'agglomération parisienne est dotée d'une infrastructure de mesure de l'état du trafic. Les éléments premiers de cette infrastructure sont des capteurs, implantés tous les 300 à 800 mètres le long des axes routiers. Le capteur est couplé à un appareil de mesure qui quantifie le phénomène détecté. Le couple capteur/appareil de mesure est appelé *station de comptage*. Le nombre total de stations de comptage sur le réseau autoroutier parisien est actuellement approximativement de 2000. Les mesures ou variables que nous étudions ici ne sont pas directement issues des capteurs. En effet, nos variables d'intérêt sont dites macroscopiques par opposition aux mesures dites microscopiques qui sont, par exemple, les mesures d'écart de temps inter-véhiculaires ou l'espacement entre véhicules. Ces mesures macroscopiques sont donc des moyennes pondérées des mesures microscopiques sur des périodes de temps de 20 secondes, 1 minute ou 6 minutes. Notre période de référence est de 6 minutes. Les principales variables macroscopiques issues des stations de comptage sont:

- le débit, qui correspond à la répartition des véhicules dans le temps,
- la concentration, qui correspond à la répartition des véhicules dans l'espace,
- le taux d'occupation, qui est une grandeur sans dimension, définie à partir de la concentration et caractérisant la proportion de temps durant laquelle le capteur est occupé,
- la vitesse du flot de véhicules.

Le lecteur intéressé par les modes de fonctionnement des différents capteurs utilisés dans le domaine du trafic routier, ainsi que par une description détaillée des variables mesurées, pourra se référer à [Coh90]. Au sein du Service Inter-départemental de l'Exploitation Routière (SIER) de la Direction Régionale de l'Equipement (DRE) d'Île-de-France, les données mesurées ont été sauvegardées depuis plusieurs années. Cela nous à permis de nous intéresser à la période de deux ans allant du 1er janvier 2001 au 1er janvier 2003. Plus précisément, notre échantillon comprend 709 journées de mesures. Nous allons maintenant décrire la structure de ces données sur un tronçon de route particulier de l'A4W. Ce tronçon de route comprend 38 stations de comptage et mesure 21.82 kilomètres. Nous noterons par la suite:

- ζ_s la station de comptage numéro s, avec $s = 1, \ldots, S = 38$,
- ξ_j le jour de la mesure numéro j, avec $j = 1, \ldots, J = 709$.

Pour chaque station ζ_s et chaque journée ξ_j, nous possédons les mesures de vitesse, de débit et de taux d'occupation sur des périodes de 6 minutes, de 5h00 à 23h00, soit 180 mesures par station et par jour. Les mesures situées entre 23h00 et 5h00 sont très chaotiques et n'ont pas été introduites dans l'étude. En effet, l'intérêt est en premier lieu porté sur les données qui vont nous permettre d'obtenir la prévision demandée, à savoir les heures de la journée durant lesquelles les activités économiques liées au trafic routier sont très importantes (voir [LML02]). Nous noterons les mesures de vitesse $V_j^s(p)$ avec $p = 1, \ldots, P = 180$, $s = 1, \ldots, S = 38$ et $j = 1, \ldots, J = 709$. La période p correspond donc à l'intervalle horaire $\left[\frac{p+49}{10}, \frac{p+50}{10}\right[$. Par exemple, pour $p = 1$, la période correspondante est [5.0,5.1[, soit de 5h00 à 5h06. De même, pour les mesures de débit et de taux d'occupation, nous noterons $\left(D_j^s(p)\right)_{jsp}$ et $\left(T_j^s(p)\right)_{jsp}$. Nous avons donc au total 26 942 courbes de vitesse et 4 849 560 mesures de vitesse, et le même nombre de données pour les mesures de débit et de taux d'occupation. La figure 1.1 nous présente, pour une journée j fixée (le 17 janvier 2001) et une station s fixée, les données $\left(V_j^s(p)\right)_{p=1,\ldots,P}$, $\left(D_j^s(p)\right)_{p=1,\ldots,P}$ et $\left(T_j^s(p)\right)_{p=1,\ldots,P}$. Cette figure illustre le lien existant entre ces trois mesures du trafic routier. La chute de la vitesse entre 7h00 et 8h30 est bien évidemment directement liée à une demande accrue des utilisateurs du réseau routier, soit à une augmentation du débit et bien sûr du taux d'occupation.

Une représentation des variables de trafic, de plus en plus utilisée par les gestionnaires du trafic routier, est la *cartographie automatique* (voir [Coh90]). Il s'agit de représenter la variable d'intérêt dans un compact de $\mathbb{R}^2$. Plus précisément, il s'agit d'une représentation temps/espace de la variable étudiée. Par exemple, l'axe des abscisses représente le temps, dans notre cas nous aurons une représentation dans l'intervalle [5h00,23h00], et l'axe des ordonnées représente l'espace, dans notre cas nous aurons la disposition des 38 stations de comptage considérées, donc un intervalle de 21.82 kilomètres. Cette représentation peut, par exemple, prendre la forme de courbes de niveaux et fait dans ce cas appel à des méthodes d'interpolation et d'estimation de contours. Ces cartes fournissent aux exploitants du réseaux des outils permettant, par exemple, de visualiser, de quantifier et de suivre l'évolution des phénomènes de congestion du trafic sur un tronçon de route particulier. A titre d'illustration, la figure 1.2 nous présente une grille irrégulière et rectangulaire de la

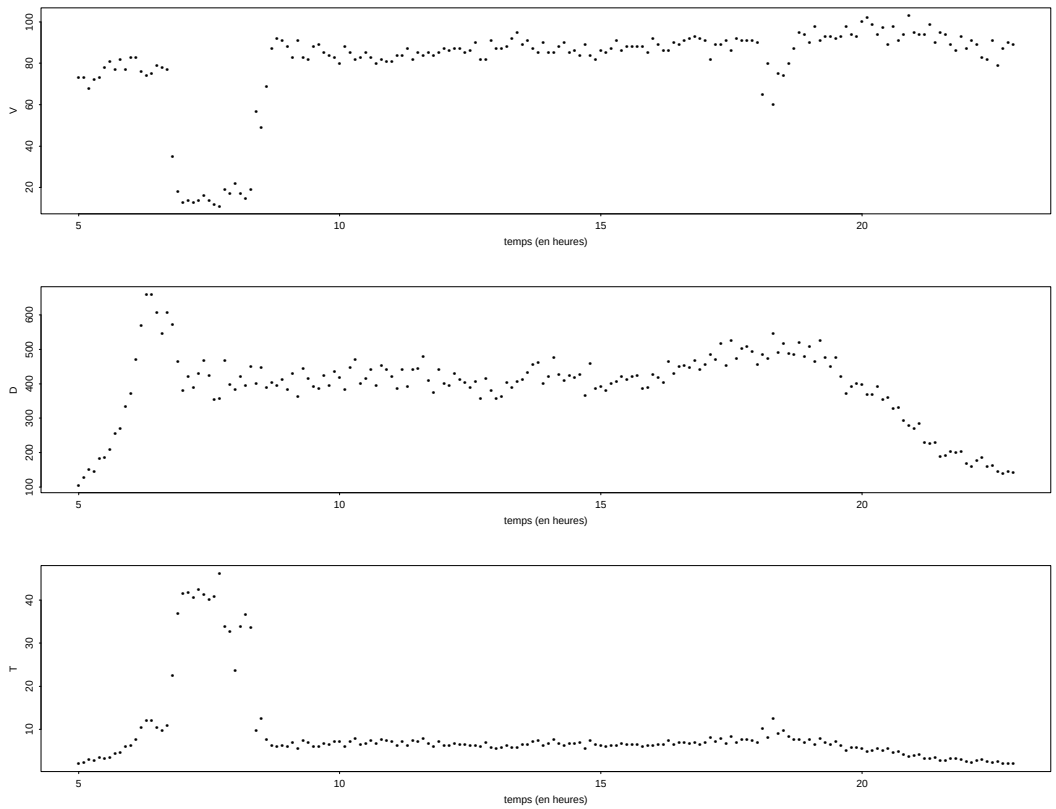

FIG. 1.1 – *De haut en bas, les données de vitesse, de débit et de taux d'occupation, pour la journée du 17 janvier 2001 et pour une station de comptage fixée.*

variable vitesse sur notre tronçon de route A4W, pour la journée du 17 janvier 2001. La grille est dite irrégulière lorsque sur l'un au moins des axes les points d'observation ne sont pas équirépartis, ce qui est bien évidemment le cas ici puisque nos stations de comptage ne sont pas toutes à la même distance. Le terme rectangulaire fait référence à l'orthogonalité des axes du plan de la représentation. Cette carte nous montre donc que les congestions sont apparues ce jour-là entre 7h00 et 10h00 sur les 10 premiers kilomètres du tronçon, et dans l'après-midi, entre 17h00 et 20h00 mais de façon plus sporadique tout le long du tronçon de route.

La figure 1.2 nous permet ici de souligner un point très important concernant les mesures du trafic routier. Les zones claires du graphe correspondent à des vitesses faibles. Or, dans notre base de données, les données manquantes sont notées "-1" et sont donc ici représentées en blanc. Ces données manquantes sont dans notre graphique notées "-" pour ne pas les confondre avec des vitesses très faibles. Le pourcentage de données manquantes est d'environ 14% pour ce 17 janvier 2001. Pour la totalité des données de notre tronçon A4W, le pourcentage de données manquantes est de 25%. Les causes de ces données manquantes peuvent être multiples. La fragilité des dispositifs de mesure et la difficulté d'entretien d'une infrastructure d'une telle ampleur font que le pourcentage de données manquantes est très important.

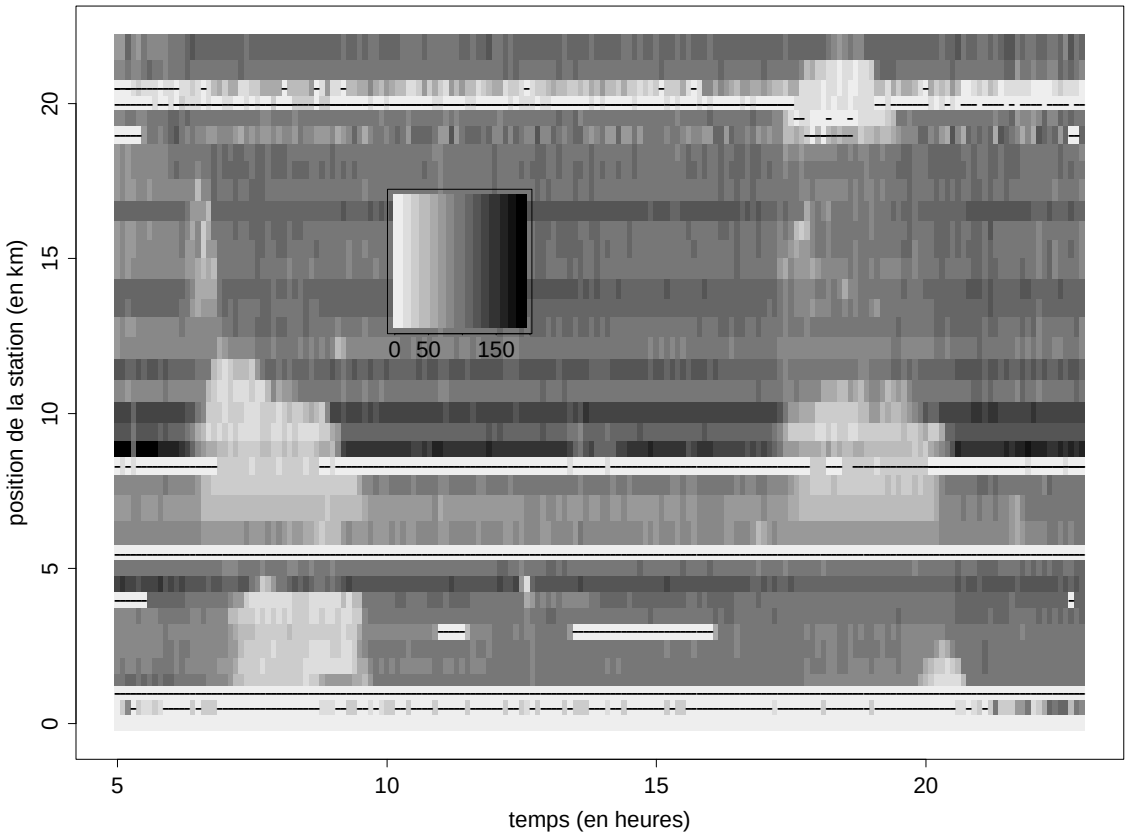

FIG. 1.2 – *Représentation des données de vitesse pour la journée du 17 janvier 2001. Les données manquantes sont marquées avec le signe "-".*

Ce pourcentage est le même pour chacune des trois variables étudiées. En effet, la donnée manquante est, a priori, due à un mauvais fonctionnement de l'appareil de mesure et les trois mesures (de vitesse, de débit et de taux d'occupation) sont ainsi égales à -1 pour les mêmes périodes. La figure 1.3 représente le pourcentage de données manquantes par jour et par station de comptage. Cela nous permet de constater que certaines des stations de comptage ont été hors service durant la quasi totalité de la période de deux ans que nous considérons. Cela nous permet de constater aussi que le pourcentage de données manquantes est rarement strictement compris entre 0% et 100%. Cela signifie tout simplement que lorsque la station de comptage est hors service, elle le reste jusqu'à l'intervention des gestionnaires du réseau. Ainsi, même si l'apparition des données manquantes peut être considérée comme aléatoire, sa propagation dans le temps ne l'est pas. Ces données manquantes ne sont donc pas *missing at random* (MAR). Cette remarque est visuellement validée par la figure 1.4 qui représente le nombre de données manquantes pour chaque journée de notre échantillon et pour chacune des stations de comptage. En effet, les pourcentages de données manquantes des station de comptage semblent bien indé-

pendants. Cette indépendance est liée à l'indépendance des appareils de mesure. Par contre, l'évolution du pourcentage de données manquantes durant nos 709 journées ne semble pas être dans ce cas.

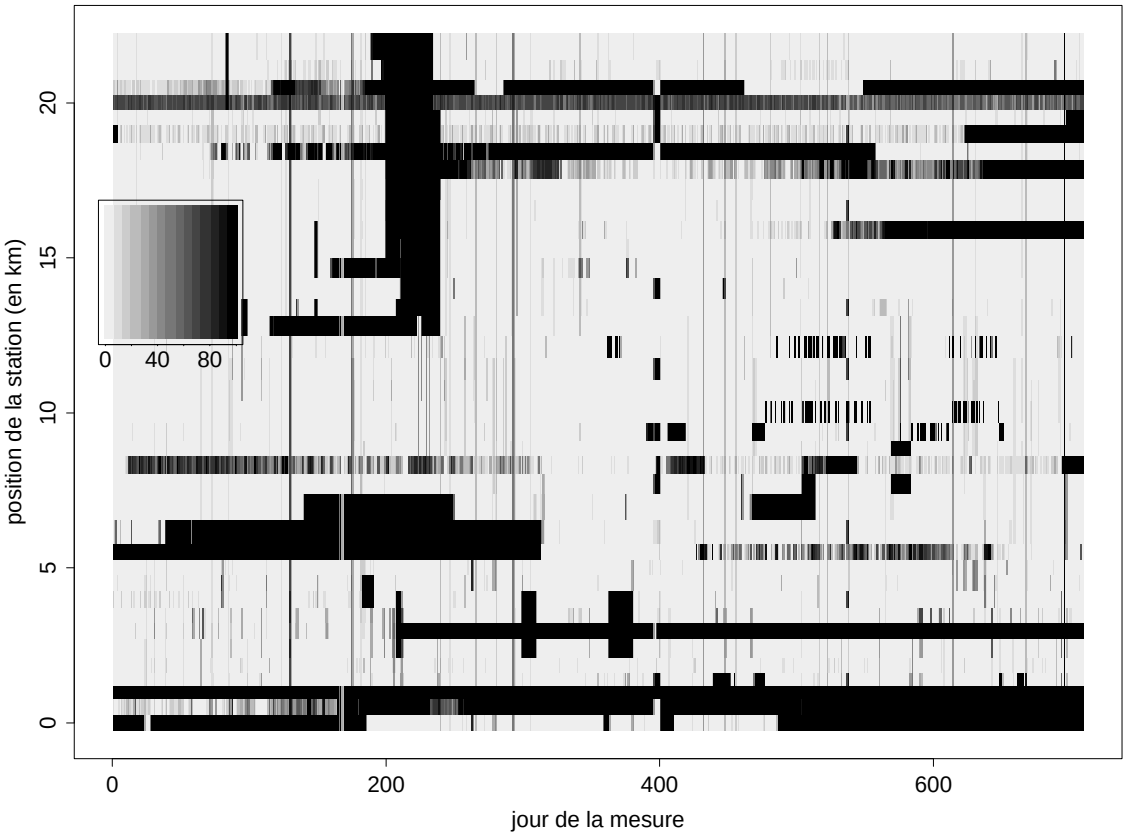

FIG. 1.3 – *Pourcentage de données de vitesse manquantes par station et par jour.*

A ces très nombreuses données manquantes viennent s'ajouter des données dites aberrantes. Ces données sont moins faciles à détecter mais sont aussi très nombreuses. C'est après cette constatation que nous avons mis en place un *protocole qualité* afin de détecter dans un premier temps les données aberrantes et de compléter ensuite au mieux les données. La complétion des données est basée sur la méthode déjà utilisée en cartographie automatique des variables de trafic (voir [Coh90]). Il s'agit d'estimer les données manquantes par une moyenne pondérée des plus proches voisins. Ce protocole est explicité dans le chapitre 2 ([LML02]). Nous en donnons aussi les principaux résultats. A titre d'exemple, la figure 1.5 nous présente les mesures de la journée du 17 janvier 2001 après ce protocole qualité. Le pourcentage de données manquantes n'est plus que de 2%. Pour la totalité de nos données de vitesse, après le protocole qualité, le pourcentage de données manquantes est de 6%.

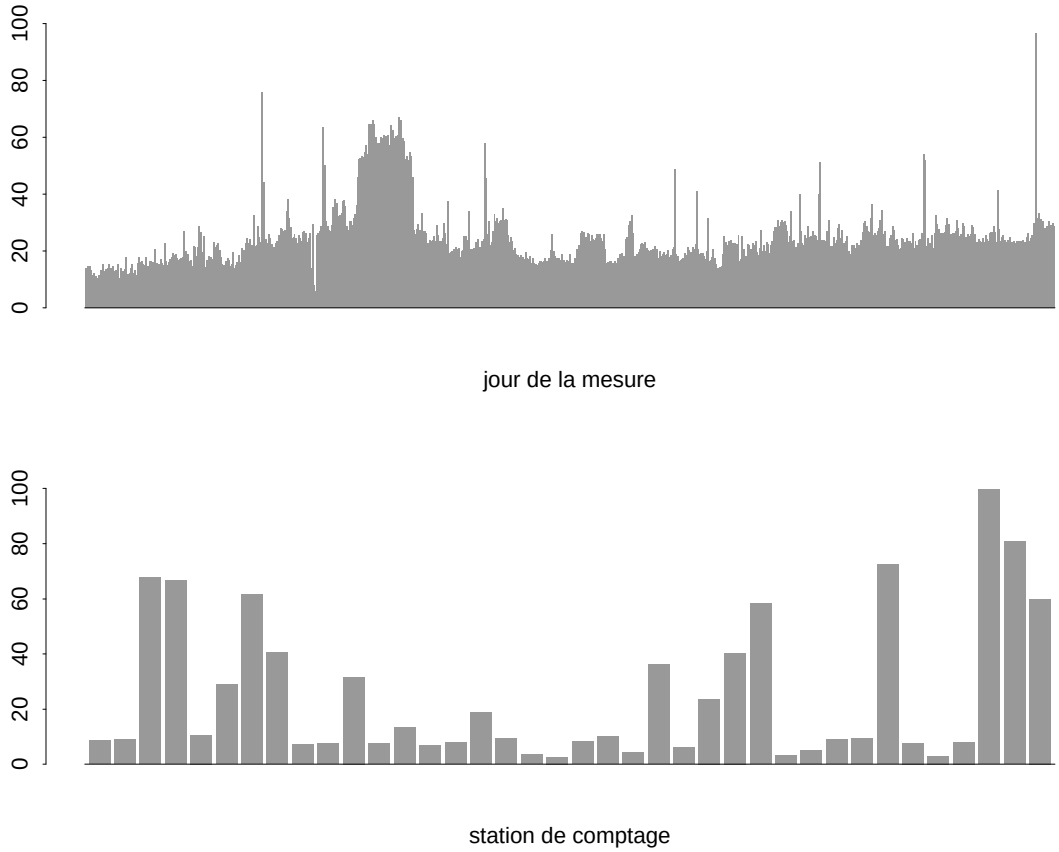

FIG. 1.4 – *Pourcentage de données manquantes pour chaque journée de notre échantillon (figure du haut) et pour chaque station de comptage (figure du bas).*

1.2 L'objectif et la méthodologie

L'objectif de l'étude est la prévision, à l'heure H, du temps de parcours, sur un itinéraire donné, d'un utilisateur qui prendra la route à l'heure $H + h$, avec $h \geq 0$. Nous supposerons que H appartient à la période numéro p_0, c'est-à-dire que $H \in \left[\frac{p_0+49}{10}, \frac{p_0+50}{10}\right[$. Par définition, le temps de parcours prévu (TPP) est le temps que va mettre un véhicule entrant à l'heure $H + h$ dans l'itinéraire donné pour parcourir celui-ci. Pour nous, un itinéraire est un tronçon de route contenant un certain nombre de stations de comptage et délimité par deux d'entre elles. Le temps de parcours prévu est donc un indicateur de l'état du trafic à l'heure $H + h$ sur ce tronçon de route. Le TPP peut être estimé de différentes façon. Traditionnellement, il est estimé à partir des données de débit et de taux d'occupation. Dans notre approche, nous nous sommes contentés des données de vitesse pour l'estimation de cet indicateur de l'état du trafic. Notre algorithme de calcul du temps de parcours (TP) sur un itinéraire donné est basé sur la simulation du suivi d'un véhicule de la première à la dernière station de comptage. En effet, lors de son parcours, le véhicule ne garde pas une vitesse constante. Sa vitesse évolue en fonction de la vitesse

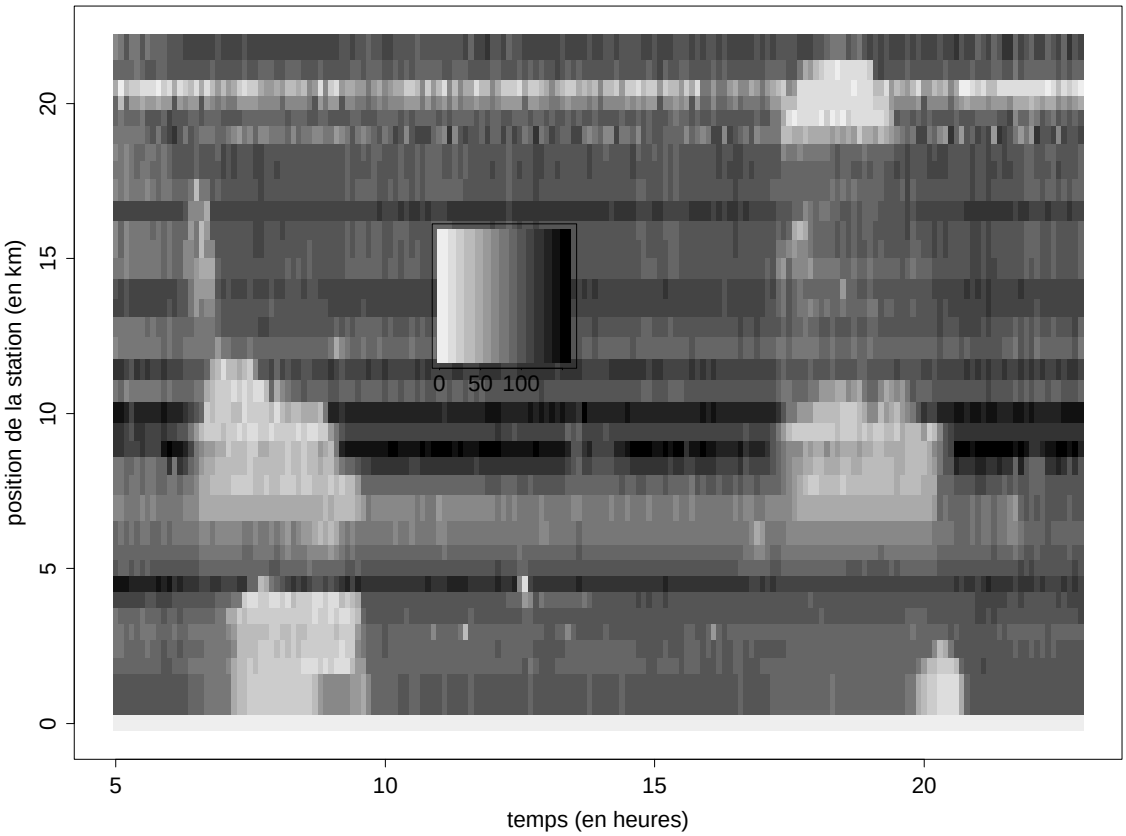

FIG. 1.5 – *Représentation des données de vitesse pour la journée du 17 janvier 2001 après le protocole qualité. Les données manquantes sont marquées avec le signe "-".*

mesurée par la station de comptage la plus proche et en fonction du temps que le véhicule va mettre pour aller d'une station à l'autre. Donc, d'après cette estimation du TP, si nous connaissons les vitesses le jour de la demande de prévision, noté ξ_{j_0}, c'est-à-dire si nous connaissons les vitesses $\left(V_{j_0}^s(p)\right)_{j_0 sp}$ pour toutes les stations de comptage de l'itinéraire $(\zeta_s, s = 1, \ldots, S)$ et pour toutes les périodes de la journée ξ_{j_0} $(p = 1, \ldots, P)$, alors nous pouvons calculer le TPP demandé. Or, bien évidemment, à la période p_0, nous ne connaissons les données de vitesse que pour les périodes strictement inférieures à p_0:

$$V_{j_0}^s(p), p \in \{1, \ldots, p_0 - 1\}, s \in \{1, \ldots, S\}.$$

L'idée générale pour l'estimation du TPP est, d'une part, que l'itinéraire sur lequel doit se dérouler la prévision peut être décomposé en stations de comptage. D'autre part, que la connaissance à un instant H des vitesses

$$\left(V_{j_0}^s(p)\right)_{p \geq p_0}$$

pour toutes les stations de comptage de l'itinéraire, nous permet de faire le calcul du TPP. Ainsi, notre problème de prévision de temps de parcours est un problème

de prévision des courbes de vitesse pour chaque station de comptage. Pour cette prévision, nous faisons l'hypothèse que sur chaque station de comptage, il existe un nombre fini de courbes de vitesse qui représentent, à quelques fluctuations près, les comportements types sur cette station de mesure. Donc, nous supposons que toute nouvelle courbe de vitesse peut être approximée par l'un de ces profils types. Pour l'estimation de ces profils, nous faisons appel à une méthode classique de classification hiérarchique et à un principe d'apprentissage. Dans la section suivante, nous décrivons le principe de la classification. La méthode de prévision à proprement parler est détaillée dans le chapitre 2.

1.3 Une distance appropriée à nos données

Dans notre méthode de prévision, nous faisons appel à une méthode de classification hiérarchique. La classification hiérarchique consiste à classer un groupe d'individus en un nombre réduit de classes. Cela nous permet de résumer l'information contenu dans notre échantillon de départ. La mise en place de la classification hiérarchique nécessite au préalable la définition

- des individus étudiés,
- d'une distance (ou similarité) entre individus,
- d'une distance (ou similarité) entre classes d'individus.

Les individus sont ici des fonctions mesurées sur un nombre fini de points d'un intervalle $[a,b] \subset \mathbb{R}$, plus précisément sur les points $t_i = a + \frac{i-1}{n}$, $i = 1,\ldots,n$. Un individu est donc un vecteur réel $x = (x_1,\ldots,x_n)'$ avec $n \in \mathbb{N}^*$ (le symbole "$'$" étant le symbole de transposition d'une matrice). Dans cette section, nous justifions le choix de la distance entre individus utilisée dans le chapitre 2 ([LML02]). Cette distance, que nous noterons Δ, est définie, pour deux individus $x \in \mathbb{R}^n$ et $y \in \mathbb{R}^n$, par

$$\Delta(x,y) = \sqrt{(x-y)'D(x-y)}, \tag{1.1}$$

où D est une matrice que nous détaillerons plus tard. De plus, afin de montrer que Δ est bien une distance, nous montrerons que la matrice D est bien définie-positive. Pour cela, nous aurons besoin du lemme 1.1. Avant d'énoncer ce lemme, définissons la matrice carrée de Toëplitz A ([GS58]), de taille $m \in \mathbb{N}^*$, par ses éléments

$$\forall i = 1,\ldots,m, \forall j = 1,\ldots,m, a_{ij} = c - |i-j|,$$

où $c \in \mathbb{R}$. Soit, sous forme matricielle,

$$A = \begin{pmatrix} c & c-1 & \ldots & c-m+1 \\ c-1 & c & \ddots & \vdots \\ \vdots & \ddots & \ddots & c-1 \\ c-m+1 & \ldots & c-1 & c \end{pmatrix}. \tag{1.2}$$

Lemme 1.1 *Le déterminant de la matrice A, noté $|A|$, définie par* (1.2) *vaut*

$$|A| = 2^{m-2}(2c-m+1).$$

Preuve 1.2 *En soustrayant, sur la matrice A, la colonne $(l-1)$ à la colonne l, pour tout $l = 2, \ldots, m$, nous obtenons*

$$|A| = \begin{vmatrix} c & -1 & \ldots & -1 \\ c-1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & -1 \\ c-m+1 & 1 & \ldots & 1 \end{vmatrix}.$$

Ainsi, en développant ce déterminant par rapport à la première colonne, nous obtenons

$$|A| = \sum_{l=1}^{m} (-1)^{l+1}(c-l+1)|A_{l1}|, \tag{1.3}$$

où, $|A_{l1}|$ est le déterminant mineur de A associé à l'élément a_{l1}, c'est-à-dire, le déterminant de la matrice carrée A_{l1} obtenue en ôtant la ligne l et la colonne 1 de la matrice A. L'expression (1.3) se simplifie en remarquant que les colonnes $(l-1)$ et l de la matrice A_{l1} sont égales pour tout $l = 2, \ldots, m-1$. Ainsi,

$$\forall l = 2, \ldots, m-1, |A_{l1}| = 0. \tag{1.4}$$

De plus, nous avons $A_{m1} = -A'_{11}$, d'où

$$|A_{m1}| = (-1)^{m-1}|A_{11}|. \tag{1.5}$$

Par conséquent, avec (1.4) et (1.5), nous obtenons l'expression suivante pour (1.3):

$$|A| = (2c-m+1)|A_{11}|.$$

Avec des calculs similaires, faisant appel à (1.4) et (1.5), nous obtenons

$$|A_{11}| = \begin{vmatrix} 1 & -1 & \ldots & -1 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ 1 & \ldots & 1 & 1 \end{vmatrix} = 2|B_{11}|,$$

avec, $|B_{11}|$ le déterminant mineur de A_{11} associé à l'élément situé sur la première ligne et la première colonne de A_{11}. Par récurrence, nous obtenons

$$|A_{11}| = 2^{m-2}.$$

D'où le résultat.

■

Nous définissons maintenant la matrice D, utilisée dans la définition (1.1) de la distance Δ, par ces éléments

$$D_{ij} = \frac{n-|i-j|}{n}, \forall i = 1, \ldots, n, \forall j = 1, \ldots, n.$$

Soit, sous forme matricielle,

$$nD = \begin{pmatrix} n & n-1 & \dots & 1 \\ n-1 & n & \ddots & \vdots \\ \vdots & \ddots & \ddots & n-1 \\ 1 & \dots & n-1 & n \end{pmatrix}. \tag{1.6}$$

Proposition 1.3 *L'application Δ de la définition* (1.1)*, avec la matrice D définie par* (1.6)*, est une distance sur $\mathbb{R}^n$.*

Preuve 1.4 *D'après le lemme 1.1, les déterminants principaux de la matrice nD sont tous strictement positifs, en effet,*

$$\forall m = 1, \dots, n, 2^{m-2}(2n - m + 1) > 0.$$

Ainsi, la matrice nD est définie positive, et l'application Δ, définie avec le produit scalaire $x'Dy$, est bien une distance.

■

La définition de la distance Δ peut être élargie à des fonctions. Nous noterons Δ les deux distances, en se référant, d'après les arguments, à la définition vectorielle ou fonctionnelle. Soient f et g deux fonctions de $\mathcal{C}^0_+([a,b])$, où $\mathcal{C}^0_+([a,b])$ est l'ensemble des fonctions continues intégrables, positives et bornées, définies sur l'intervalle $[a,b]$. La distance Δ entre les deux fonctions f et g est définie par

$$\Delta(f,g) = \sqrt{\int_a^b\int_a^b (f(u) - g(u))(f(v) - g(v))\frac{(b-a) - |u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v}. \tag{1.7}$$

Proposition 1.5 *L'application Δ de la définition* (1.7) *est une distance sur $\mathcal{C}^0_+([a,b])$.*

Preuve 1.6 *Il suffit de montrer que l'application*

$$\langle f,g\rangle = \int_a^b\int_a^b f(u)g(v)\frac{(b-a) - |u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v$$

est un produit scalaire sur $\mathcal{C}^0_+([a,b])$. Ceci est bien le cas, puisque pour toute paire de fonctions f et g de $\mathcal{C}^0_+([a,b])$, nous avons,

- *pour tout $\lambda \in \mathbb{R}$, $\langle \lambda f,g\rangle = \lambda\langle f,g\rangle$,*
- *$\langle f,g\rangle = \langle g,f\rangle$,*
- *$\langle f,f\rangle \geq 0$, en effet, $f \geq 0$ et pour tout $(u,v) \in [a,b]^2$, $0 \leq \frac{(b-a)-|u-v|}{(b-a)} \leq 1$,*
- *$\langle f,f\rangle = 0 \Rightarrow f = 0$, en effet, puisque $f \in \mathcal{C}^0_+([a,b])$, nous avons*

$$\begin{aligned} \langle f,f\rangle = 0 \quad &\Rightarrow \quad \forall (u,v) \in [a,b]^2, f(u)f(v)\frac{(b-a) - |u-v|}{(b-a)} = 0 \\ &\Rightarrow \quad f = 0. \end{aligned}$$

D'où le résultat.

■

Dans notre contexte de classification de courbes de vitesse, le principal intérêt de cette distance Δ réside dans la proposition suivante.

Proposition 1.7 *Soient* $f : [a,b] \to \{0,1\}$ *et* $g_\epsilon : [a,b] \to \{0,1\}$ *définies par*

$$f(t) = \mathbb{1}_{[a,b]}(t) - \mathbb{1}_{]\alpha,\beta[}(t),$$

et

$$g_\epsilon(t) = \mathbb{1}_{[a,b]}(t) - \mathbb{1}_{]\alpha+\epsilon,\beta+\epsilon[}(t),$$

avec, $a < \alpha < \beta < b$ *et* $\epsilon \in I = [a-\alpha, b-\beta]$. *Alors,*

$$\Delta^2(f,g_\epsilon) = \begin{cases} \frac{-2}{3(b-a)}|\epsilon|^3 + \frac{2(\beta-\alpha)}{(b-a)}|\epsilon|^2 & \forall \epsilon \in [\alpha-\beta, \beta-\alpha] \cap I, \\ \frac{2(\alpha-\beta)^3}{3(b-a)} + \frac{2(\alpha-\beta)^2}{(b-a)}|\epsilon| & \forall \epsilon \in (]-\infty, \alpha-\beta[\cup]\beta-\alpha, +\infty[) \cap I. \end{cases}$$

Preuve 1.8 *Nous pouvons ici, sans perte de généralité, nous restreindre aux valeurs positives de* ϵ. *Aussi, nous supposons que* $\beta-\alpha < b-\beta$, *et nous calculons séparément* $\Delta^2(f,g_\epsilon)$ *avec* $\epsilon \in [0, \beta-\alpha]$ *et avec* $\epsilon \in]\beta-\alpha, b-\beta]$. *Ainsi, dans le cas* $0 \leq \epsilon \leq \beta-\alpha$, *nous avons*

$$f(t) - g_\epsilon(t) = -\mathbb{1}_{]\alpha,\alpha+\epsilon]}(t) + \mathbb{1}_{[\beta,\beta+\epsilon[}(t),$$

donc

$$\begin{aligned}
\Delta^2(f,g_\epsilon) &= \int_a^b\int_a^b (f(u)-g_\epsilon(u))(f(v)-g_\epsilon(v))\frac{(b-a)-|u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v \\
&= \int_a^b\int_a^b (-\mathbb{1}_{]\alpha,\alpha+\epsilon]}(u) + \mathbb{1}_{[\beta,\beta+\epsilon[}(u))(-\mathbb{1}_{]\alpha,\alpha+\epsilon]}(v) + \mathbb{1}_{[\beta,\beta+\epsilon[}(v))\frac{(b-a)-|u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v \\
&= \int_a^b\int_a^b \mathbb{1}_{]\alpha,\alpha+\epsilon]}(u)\mathbb{1}_{]\alpha,\alpha+\epsilon]}(v)\frac{(b-a)-|u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v \\
&\quad - \int_a^b\int_a^b \mathbb{1}_{]\alpha,\alpha+\epsilon]}(u)\mathbb{1}_{[\beta,\beta+\epsilon[}(v)\frac{(b-a)-|u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v \\
&\quad - \int_a^b\int_a^b \mathbb{1}_{[\beta,\beta+\epsilon[}(u)\mathbb{1}_{]\alpha,\alpha+\epsilon]}(v)\frac{(b-a)-|u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v \\
&\quad + \int_a^b\int_a^b \mathbb{1}_{[\beta,\beta+\epsilon[}(u)\mathbb{1}_{[\beta,\beta+\epsilon[}(v)\frac{(b-a)-|u-v|}{(b-a)}\,\mathrm{d}u\,\mathrm{d}v \\
&= \frac{2}{(b-a)}\left(\int_\alpha^{\alpha+\epsilon}\int_\alpha^{\alpha+\epsilon}(b-a)-|u-v|\,\mathrm{d}u\,\mathrm{d}v - \int_\beta^{\beta+\epsilon}\int_\alpha^{\alpha+\epsilon}(b-a)-|u-v|\,\mathrm{d}u\,\mathrm{d}v\right) \\
&= \frac{-2}{3(b-a)}\epsilon^3 + \frac{2(\beta-\alpha)}{(b-a)}\epsilon^2.
\end{aligned}$$

Dans le cas où $\beta-\alpha < \epsilon \leq b-\beta$, *les calculs sont similaires, avec*

$$f(t) - g_\epsilon(t) = -\mathbb{1}_{]\alpha,\beta[}(t) + \mathbb{1}_{]\alpha+\epsilon,\beta+\epsilon[}(t).$$

D'où le résultat.

■

L'utilisation de la distance Δ pour la classification automatique de nos courbes de vitesse est bien justifiée par la proposition 1.7. En effet, la fonction f donnée dans cette proposition simule, de manière très simplifiée, la chute de la vitesse mesurée lors d'une congestion du trafic routier, sur un point donné du réseau, un jour donné. De plus, la fonction g_ϵ simule, le même type de phénomène mais translaté dans la journée. Ainsi, la proposition 1.7 nous dit que plus le phénomène sera éloigné dans le temps, c'est-à-dire, plus ϵ sera grand, plus la distance entre nos deux courbes f et g_ϵ sera grande. La figure 1.6 nous montre un exemple des distances $\Delta(f,g_\epsilon)$ obtenues avec les valeurs de f et de g_ϵ définies par $a = 0, b = 1, \alpha = 0.2$ et $\beta = 0.5$. Nous comparons aussi ces valeurs aux valeurs obtenues avec la distance euclidienne classique. Dès que $\epsilon \geq \beta - \alpha$, c'est-à-dire dès que nos deux chutes de vitesse n'ont plus d'intersection, la distance euclidienne reste constante. Il apparaît donc qu'une méthode de classification basée sur une distance classique ne peut être efficace dans la construction de classes bien séparées.

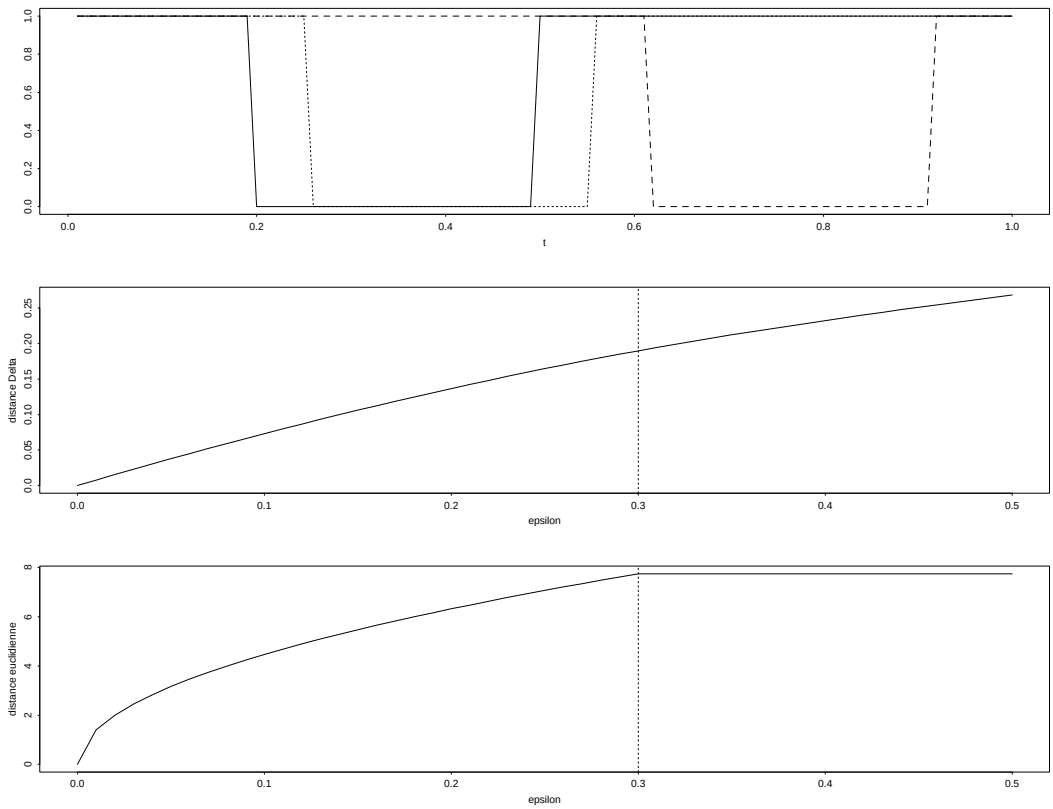

FIG. 1.6 – *En haut, les fonctions f et g_ϵ définies dans la proposition 1.7, avec les valeurs: $a = 0, b = 1, \alpha = 0.2$ et $\beta = 0.5$. La fonction f est en trait plein, la fonction $g_{\epsilon'}$ est en pointillés ($\epsilon' \in [0,\beta - \alpha]$) et la fonction $g_{\epsilon''}$ est en tirés ($\epsilon'' \in]\beta - \alpha, b - \beta]$). Au milieu, les valeurs de $\Delta(f,g_\epsilon)$ pour $\epsilon \in [0,b-\beta]$. En bas, les valeurs de la distance euclidienne entre f et g_ϵ pour $\epsilon \in [0,b-\beta]$. La droite verticale dans les deux graphes du bas a pour abscisse $\beta - \alpha$.*

1.4 Quelques résultats

Les résultats des prévisions de TP sont détaillés dans la section 2.5 du chapitre 2. Nous soulignerons simplement ici que ces prévisions de temps de parcours sont assez satisfaisantes, en ce sens où les congestions de trafic sont assez bien estimées. D'autant plus que la méthode utilisée ici ne considère pas la totalité des données disponibles. En effet, nous pouvons supposer que l'introduction des données de débit et des données de taux d'occupation amélioreront les résultats. Par contre, notre méthode n'est pas directement applicable à l'ensemble des données recueillies. En effet, après différents essais incluant ces données dans notre méthode, nous avons constaté que la prévision n'était pas meilleure. L'utilisation des données de débit et de taux d'occupation s'avère être plus délicate.

Chapitre 2

Road trafficking description and short term travel time forecasting, with a classification method (submitted article)

By authors: Marc Lavielle*, Jean-Michel Loubes* and Elie Maza**.

* CNRS - Laboratoire de Mathématiques
Equipe de Probabilités, Statistique et Modélisation
UMR 8628 - bâtiment 425, Université Paris Sud
91405 Orsay cedex
** Laboratoire de Statistique et Probabilités
UMR C5583 - Université Paul Sabatier
118, route de Narbonne
F-31062 Toulouse cedex 4

Marc.Lavielle@math.u-psud.fr
Jean-Michel.Loubes@math.u-psud.fr
Elie.Maza@math.ups-tlse.fr

Abstract: The purpose of this work is, on the one hand, to study how to forecast road trafficking on highway networks and, on the other hand, to describe future traffic events. Here, road trafficking is measured by the vehicle velocities. We propose two methodologies, the first one is based on an empirical classification method, and the second one, on a probability mixture model. We use an SAEM type algorithm (a Stochastic Approximation of the EM algorithm) to select the densities of the mixture model. Then, we test the validity of our methodologies by forecasting short term travel times.

2.1 Introduction

Long term forecasting of road traffic was developed a long time ago, for example in the *Bison Futé* program (http://www.bison-fute.equipement.gouv.fr). The novel idea of short term forecasting has only appeared recently. Indeed, the emergence of new technologies enables us to obtain the necessary data in real time. Now, road trafficking is not only a qualitative variable describing the state of a car stream (moving, blocking or stopped, as it is done in [CDPB96]), but also a quantitative variable (with measures of speed, flow and occupancy rate).

Our main purpose is to forecast travel time on the Parisian highway network. More precisely, we want to forecast, at time H, the travel time that a vehicle will need, at time $H + h$, $h \geq 0$, to go from one point to another. Then, contrary to previous works, see for example [VGDPMW98] or [DPA94], where forecasts were made only at a specific point of the network, we aim at estimating road trafficking at all the points of the observation grid. This observation grid is composed of all the measurement stations located on the network. Moreover, our methodology builds archetypes of road trafficking whose interpretation is of great interest. Thus, we did not consider an approach using time series. Indeed, such a work was done in [BJS03] but the structure of our data prevents a wide use of these technics, as it will be discussed later.

Two main assumptions are the core of our study. The first one is that, in a short term, road trafficking mostly depends on what just happened. The second one is that there is a fixed number of traffic patterns, and every new observation day can be compared to them. The data used for this study have confirmed these statements. Moreover, such assumptions are considered true by all the professionals of road traffic, such as the INRETS (*Institut National de Recherche sur les Transports et leur Sécurité*). As a consequence, the issue of travel time forecasting has two different aspects. Firstly, we have to determine the representative behaviours or patterns of road trafficking. Secondly, we have to compare the incoming observations to these archetypes, and choose which behaviour this observation belongs to. Here, classification methods enable us to allocate data into representative sets (see [Gor99], [Cel88], [BFOS84] or [Jam78] for more general references). Also, mixture problems consist in studying and identifying mixtures of probability distributions (see [Che95], [LL95] or [CL01] for example). Hence, they are well suited for our purpose.

The article is divided into four main parts. In the first part, Section 2.2, we introduce the data used for this work. We also present the preliminary treatments to detect and eliminate outliers. Then, we present the forecasting methodology. Section 2.3 provides a model for the vehicle speed change by considering a mixture setting. An SAEM type algorithm is used to estimate the different components of the mixture. Section 2.4 is devoted to the study of an empirical classification to construct significant clusters, and to give archetypes of each traffic behaviour. Finally, in Section 2.5, we compare the two different approaches by forecasting travel times with the patterns obtained by the two methodologies.

2.2 Data and methodology

In Section 2.2.1, we first describe the data which our study is based on. After that, in Section 2.2.2, considering the need of establishing a quality protocol for these observations, we clarify the processing of aberrant data detection and of missing data restoration. Lastly, in Section 2.2.3, we explain the forecasting method.

2.2.1 Description

Most of Paris and its suburb road network is equipped with a traffic measurement infrastructure. The main elements of this infrastructure are sensors, set approximately at every 500 meters along main road axes. The sensor is coupled with a measuring device which quantifies the detected phenomenon. The pair sensor/device is called *counting station*. The main variables resulting from these counting stations (see [Coh90]) are, for a fixed period of time, the following ones:

- the flow, defined by the number of vehicles detected by the sensor,
- the occupancy rate, defined by the rate of time in which the counting station is occupied,
- the speed, defined by the mean of vehicle speeds.

The data presented in the paper are provided by the SIER (*Service Interdépartemental d'Exploitation Routière*). This organization is attached to the DRE (*Direction Régionale de l'Equipement*), and manage the road traffic measurement infrastructure on the Paris area. We will note hereafter:

- ζ_s the counting station number s, $s = 1, \ldots, S$, with S the total number of stations on the network (actually $S \approx 2000$),
- ξ_n the day number n, $n = 1, \ldots, N$, with N the total number of days considered in the study.

Data measured by counting stations being safeguarded for several years, it is then possible to choose the optimal size of the database for this study. So, we consider a two years history, from January 1/2001 to January 1/2003 ($N = 709$). This choice was made considering two factors: on the one hand, a too long history will tend to underestimate the latest changes in the road traffic context, like modifications of the road infrastructure or new user behaviours; on the other hand, a too short history would not be informative enough. Hence, according to these considerations, and with road managers (SIER) agreements, we have empirically set the size of the historic database to approximately two years.

For each station ζ_s and each day ξ_n, we know the speed measurements over periods of 6 minutes, from 5 AM to 11 PM. That is to say 180 daily speed measurements per station. We will note $Y_n^s(t)$ the speed measurement, for the station ζ_s and the day ξ_n, for $t = 1, \ldots, T$, with $T = 180$. The period t corresponds to the interval of time in hours $\left[\frac{t+49}{10}, \frac{t+50}{10}\right[$. See Figure 2.2 for examples of such speed curves.

Our study is carried out on a representative axis of the Paris highway (named A4W) where it is difficult to forecast travel times. This road section is 21.82 kilometers long and has 38 counting stations. The representativeness of this road section

was approved by the managers of Paris road network, via the TrafficFisrt company (http://www.trafficfirst.com).

2.2.2 Data quality

The database used for our purpose gathers speed measurements over two years. The rough data, resulting from this collection, cannot be treated directly. Indeed, the aberrant data and the missing data are too numerous (see Figure 2.1). Such erroneous measurements are due, on the one hand, to the errors of measurement of counting stations, and on the other hand, to the difficulty of maintaining these numerous stations. To bypass the problem of aberrant and missing measurements, we elaborated a filtering and completion protocol. This protocol is composed of two parts. The first part is based on the detection of the aberrant data, and was elaborated in collaboration with the SIER managers. The second part deals with the completion of the missing data. The structure of the quality protocol is as follows (with missing data noted -1):

(A) the aberrant data detection is based on the three following points:

  (a) detection of excessive speed measures: if a speed measurement is higher than 160 km/h, then this measure is considered as aberrant, and is put to -1,

  (b) detection of speed measures that are too low: if speed measurements are lower than 5 km/h during more than 36 periods, i.e. 3.6 hours, then these measurements are considered aberrant, and are put to -1,

  (c) detection of constant speed measurements: if speed measurements are constant for more than 5 periods, i.e. 0.5 hour, then these measurements are considered aberrant, and are set to -1,

(B) the missing measurement completion is carried out by calculating a space and time average with non missing data: if $Y_n^s(t) = -1$, i.e. if the measurement is missing, then we replace it by

$$Y_n^s(t) = \frac{Y_n^{s-1}(t) + Y_n^{s+1}(t) + Y_n^s(t-1) + Y_n^s(t+1)}{4},$$

or by the average of the non missing values if any one is also equal to -1. Obviously, if all are equal to -1, then we let $Y_n^s(t) = -1$. This completion is repeated twice.

The three points presented in (A) correspond to the sources of error known by the road traffic managers. For example, the constant speed measurements are due to stations that have not been re-initialized after a measure and that automatically repeat this same measure over several consecutive periods.

After this quality protocol was performed, the number of days used for the study is reduced for some counting stations. Indeed, we exclusively use in our study the curves Y_n^s which are complete after the protocol, i.e. without missing data. Figure 2.1 presents the percentage of complete days before and after the protocol for each counting station. We can notice that some counting stations have no complete days before the protocol. So, the EM algorithm, used in Section 2.3, allows to work with MAR data (i.e. Missing At Random data). However, in many counting stations, missing data are not MAR, and too numerous. Thus, it was not possible to use such a solution to complete data. Indeed, the protocol must be used automatically, and many counting stations do not possess the required assumptions.

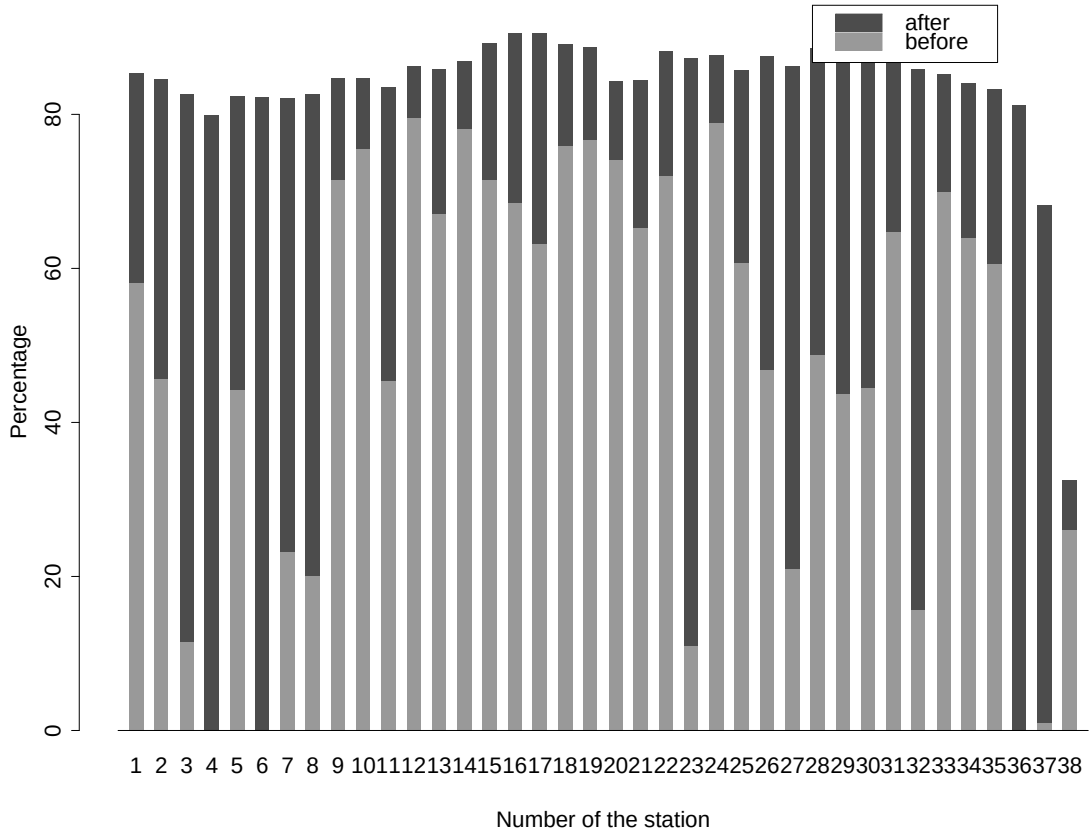

FIG. 2.1 – *Percentage of days without missing data, for each station, before and after the quality protocol. We have, after the protocol, approximately* 80% *of the database which can be used in the forecasting method.*

Figure 2.2 presents 3 speed curves of the counting station number 12, before the protocol (dotted line) and after the protocol (solid line). In this example, only the third curve has missing data and is equal to -1 for some periods. So, we can see that not all the missing values are completed by the protocol.

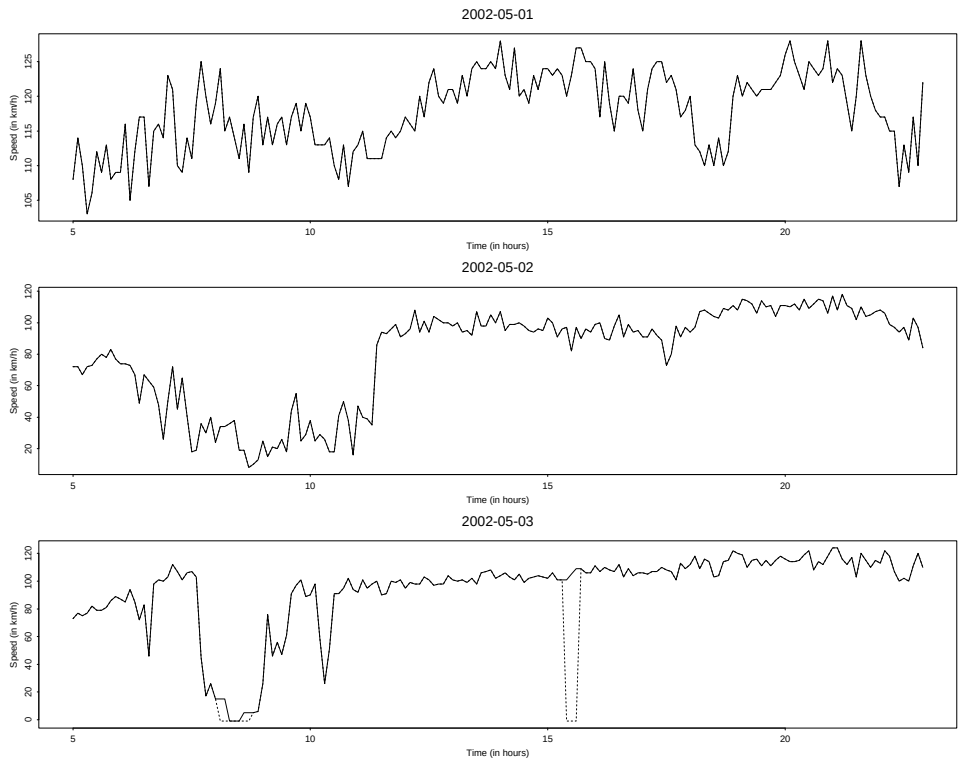

FIG. 2.2 – *Plots of 3 speed curves of 3 consecutive days for the counting station number 12, before the protocol (dotted line) and after the protocol (solid line). Only the third curve has missing data. We can see that the protocol does not complete all missing values (noted* -1*).*

2.2.3 Forecasting method

The aim of the study is the following. At a new day ξ_{n_0}, at time $H \in \left[\frac{t_0+49}{10}, \frac{t_0+50}{10}\right[$, corresponding to period t_0, we observe the speed measurements

$$Y_{n_0}^s(t), \forall t < t_0, \forall s = 1, \ldots, S,$$

thus, we need to forecast

$$Y_{n_0}^s(t), \forall t \geq t_0, \forall s = 1, \ldots, S,$$

in order to predict the travel time on a given itinerary. Indeed, once we get all the values $Y_{n_0}^s(t)$, for all $t \geq t_0$ and for all counting stations ζ_s, $s = 1, \ldots, S$, then we can simply deduce the travel time from one point to another at time $H + h$, for any $h \geq 0$. Hence, the interest is to find the tendency of the new day ξ_{n_0}, and not variations around this tendency. That is not the case in time series applications, and that's why we didn't use them.

We assume that for each counting station ζ_s, there is a number m_s of representative behaviours of road trafficking, noted $f_1, \ldots, f_{m_s}$. Hence, the forecasting method

can be divided in two main parts:

- estimate the standard profiles, $f_1, \ldots, f_{m_s}$, for each counting station ζ_s, $s = 1, \ldots, S$,
- use these standard profiles to forecast travel time.

For sake of simplicity, even if these standard profiles depend on the counting station ζ_s, we will note m the number of behaviours and $f_1, \ldots, f_m$, the associated standard profiles. Indeed, since the study is carried out for each and every counting station, it is not necessary to specify the station index s.

To get estimators of standard profiles $f_1, \ldots, f_m$, we use two different methods: a mixture model in Section 2.3, and an empirical classification method in Section 2.4. Also, in order to make any comparisons between travel time forecasts using these two methods, we get a *test sample* of $N_T = 19$ days. This test sample will be used in Section 2.5 to forecast simulated travel times. Obviously, we chose the days of this test sample randomly in all complete days in all consecutive counting stations. Indeed, we need complete days everywhere in the road section to be able to forecast.

Our classification method (Section 2.4) can be seen as a learning theory methodology. The learning theory is based on learning from the past to predict the future. In order to proceed in a similar fashion, each station database is divided into two samples:

- a *model sample*, used to estimate standard profiles, with N_M complete days (80% of the data),
- a *learning sample*, used to estimate the optimal number of standard profiles, with N_L complete days (20% of the data).

In Section 2.3, in order to estimate the unknown parameters of the mixture model, we use the model sample.

2.3 Mixture model

In this part, we aim at representing road traffic with a mixture model. As a matter of fact, we assume that the daily evolution of the vehicle speed is drawn from a mixture of probability. So, it is necessary to estimate the components of the mixture, as well as the optimal number of components. Such method has already been used but for a qualitative study of road traffic in [Doc95] or [CDPB96].

2.3.1 Model description

Set ζ_s, $s = 1, \ldots, S$, the counting station we are studying. For this station, each day $\xi_n, n = 1, \ldots, N$, we observe the vehicle speed at discrete times $t = 1, \ldots, T$, with $T = 180$. Set, for each $n = 1, \ldots, N$, $y_n = {}^t(y_n(1), \ldots, y_n(T)) \in \mathbb{R}^T$ the vector of daily observed speeds, and $Y_n = {}^t(Y_n(1), \ldots, Y_n(T))$ the corresponding random vector. We assume, as quoted in Section 2.2.3, that there are m different archetypes, $f_1, \ldots, f_m$, where for all $j = 1, \ldots, m$, $f_j = {}^t(f_j(1), \ldots, f_j(T)) \in \mathbb{R}^T$. The assumption behind this modelling is that highway traffic phenomena are independent from the previous

days, and that there are exogenous variables that determine to which pattern the observed data belong. So, the measure of the vehicle velocity at one point comes from one of the previous patterns, with an observation noise, which we assume to be Gaussian. Hence, the model can be written:

$$Y_n = \sum_{j=1}^{m} \mathbb{1}_j(X_n) f_j + \epsilon_n, n = 1, \ldots, N, \tag{2.1}$$

where

- X_n, $n = 1, \ldots, N$, are i.i.d. non observable variables, taking values in the discrete set $\{1, \ldots, m\}$,
- $\epsilon_n \in \mathbb{R}^T, n = 1, \ldots, N$, is a Gaussian vector, independent from the observations, with variance $\sigma^2 I_T$, with I_T the $T \times T$ identity matrix. The observations come from counting stations which are all the same, satisfying to the quality controls. Hence the variance is taken constant equal to σ^2.

The unknown parameters are: the number of components m; the archetypes $f_j, j = 1, \ldots, m$; the noise variance σ^2; as well as the parameters of the law of X_n. The discrete law of X_n is entirely characterized by the probabilities $\pi_j = \mathbf{P}(X_n = j), j = 1, \ldots, m$.

In a first approach, we assume that m is known. Selecting the right number of models is the topic of Section 2.3.2. Hence, the parameters to be estimated are:

$$\Psi = {}^t(\pi_1, \ldots, \pi_m, f_1, \ldots, f_m, \sigma). \tag{2.2}$$

Set $y = (y_1, \ldots, y_N)$ the observed values of the random sample $Y = (Y_1, \ldots, Y_N)$. Set also $x = (x_1, \ldots, x_N)$ the non observed values of the random sample $X = (X_1, \ldots, X_N)$.

For estimating Ψ, we consider the maximum likelihood estimator. So, the log-likelihood can be written in the following form:

$$L(y; \Psi) = \sum_{n=1}^{N} \log \left(\sum_{j=1}^{m} \pi_j \phi(y_n; f_j, \sigma) \right), \tag{2.3}$$

where $\phi(\cdot; \mu, \sigma)$ is the density of a Gaussian vector with mean $\mu \in \mathbb{R}^T$ and variance $\sigma^2 I_T$. The log-likelihood estimator of Ψ is a root of the equation

$$\nabla_\Psi L(y; \Psi) = 0,$$

where $\nabla_\Psi L(y; \Psi)$ is the gradient of L with respect to the unknown parameters of Ψ.

In a mixture model, analogous to the one studied by McLeish and Small in [MS86], the solution of the previous equation can be computed efficiently with an EM algorithm, as it is stated in the work of Basford and McLachlan in [BM85] or [McL82]. The EM algorithm was created by Dempster, Laird and Rubin (see

[DLR77]) to maximize the log-likelihood with missing data. It enables, with a recursive method, to change the problem of maximizing the log-likelihood into the problem of maximizing the completed log-likelihood of the model:

$$L_C(y,x;\Psi) = \sum_{n=1}^{N}\sum_{j=1}^{m} \mathbb{1}_j(x_n)\log\left(\pi_j\phi(y_n;f_j,\sigma)\right). \tag{2.4}$$

Set $Z_n = (Z_{nj})_{j=1,\ldots,m} = (\mathbb{1}_j(X_n))_{j=1,\ldots,m}$. This variable completes the model since it points out which class the random vector Y_n belongs to. This variable follows a multinomial distribution with unknown parameter $\pi = {}^t(\pi_1,\ldots,\pi_m)$.

Let describe the $p+1$ step of the EM algorithm. Set

$$Q\left(\Psi,\Psi^{(p)}\right) = \mathbb{E}\left[L_C(Y,X;\Psi)|Y=y;\Psi^{(p)}\right], \tag{2.5}$$

the expectancy of the log-likelihood of the complete data, conditionally to the observed data, and with respect to the value of the parameter computed at the step p, written $\Psi^{(p)}$. Then we obtain

$$Q\left(\Psi,\Psi^{(p)}\right) = \sum_{n=1}^{N}\sum_{j=1}^{m}\mathbb{E}\left[Z_{nj}|Y_n=y_n;\Psi^{(p)}\right]\log\left(\pi_j\phi\left(y_n;f_j,\sigma\right)\right).$$

Hence, the step $p+1$ of the EM algorithm is divided into two stages: the expectation stage (E) and the maximization stage (M):

(E) In this stage, the random variable Z_{nj} is replaced by its expectancy, conditionally to the observed data, and with respect to the current value of the parameter:

$$\begin{aligned}\tau_k^{(p)}(y_n) &= \mathbb{E}\left[Z_{nk}|Y_n=y_n;\Psi^{(p)}\right]\\ &= \mathbf{P}\left(Z_{nk}=1|Y_n=y_n;\Psi^{(p)}\right)\\ &= \frac{\pi_k^{(p)}\phi\left(y_n;f_k^{(p)},\sigma^{(p)}\right)}{\sum_{j=1}^{m}\pi_j^{(p)}\phi\left(y_n;f_j^{(p)},\sigma^{(p)}\right)}.\end{aligned}$$

(M) In this stage, the maximization is conducted by choosing the value of the parameter Ψ that maximizes $Q(\Psi,\Psi^{(p)})$. It will be written $\Psi^{(p+1)}$. The estimators are the followings:

$$\begin{aligned}\pi_j^{(p+1)} &= \frac{1}{N}\sum_{n=1}^{N}\tau_j^{(p)}(y_n),\\ f_j^{(p+1)} &= \frac{\sum_{n=1}^{N}\tau_j^{(p)}(y_n)y_n}{\sum_{n=1}^{N}\tau_j^{(p)}(y_n)},\\ \sigma^{(p+1)} &= \left(\frac{1}{NT}\sum_{n=1}^{N}\sum_{j=1}^{m}\sum_{t=1}^{T}\tau_j^{(p)}(y_n)\left(y_n(t)-f_j^{(p)}(t)\right)^2\right)^{1/2}.\end{aligned}$$

The model we use, undergoes the assumptions over the EM algorithm, which ensures its convergence.

In order to avoid local minima, we have used a stochastic approximation of the EM algorithm, the SAEM algorithm. Such algorithm has been developed, and its convergence has been proved, by Delyon, Moulines and Lavielle in [DLM99]. The main advantage for using SAEM algorithm rather than EM algorithm is that the former is less sensitive to the choice of the starting point in the algorithm. For a good choice of the initialization parameters, the outcome of the two algorithms are quite the same, while, for a bad choice, the estimates given by successive applications of EM algorithm can be far the one from the others. On the contrary, SAEM provides the same results. For more about the comparison between stochastic versions of the EM algorithm, we refer to Celeux and *al.* in [BCD83], [CD92] or [CCD95].

The step $p+1$ of the SAEM algorithm comes from the step $p+1$ of the EM algorithm in the following way:

(E) The E stage is replaced by a simulation stage. In this stage, we draw $K(p+1)$ realizations of the multinomial variable Z_{nj}, written z_{nj}^k, $k=1,\ldots,K(p+1)$, according to the distribution given by the values of the parameters at the step p, $\Psi^{(p)}$. The log-likelihood is then modified in the following way:

$$\hat{Q}_p(\Psi) = \hat{Q}_{p-1}(\Psi) + \gamma_{p+1}\left(\frac{1}{K(p+1)}\sum_{k=1}^{K(p+1)}\sum_{n=1}^{N}\sum_{j=1}^{m} z_{nj}^k \log\left(\pi_j^{(p)}\phi\left(y_n; f_j^{(p)}, \sigma^{(p)}\right)\right) - \hat{Q}_{p-1}(\Psi)\right),$$

where $(\gamma_p)_{p\geq 1}$ is a sequence of positive reals.

(M) The M stage of the algorithm takes place as previously.

We have used this modified algorithm in our work, with a numerical good choice of the sequences $(\gamma_p)_{p\geq 1}$ and $(K(p))_{p\geq 1}$, which leads to the results presented in Section 2.3.3.

2.3.2 Estimation of the number of components of the mixture

The aim of this study is to find the optimal number m of components of the mixture (2.1). This estimation justifies the results of the previous section 2.3.1. For this, we use a methodology close to model selection approach. For a theoretical approach of these technic, we refer to the work of Baraud, Birgé and Massart in [BM98], [BBM99] or [Bar00] for instance.

For each value $m \geq 1$, we consider the set $\mathcal{F}_m = \{g_1,\ldots,g_m, g_i \in \mathbb{R}^T, \pi_1,\ldots,\pi_m, \sigma\}$, and we write $\mathcal{F} = \cup_{m\geq 1}\mathcal{F}_m$ the collection of all the different models. For a fixed m, we have seen in Section 2.3.1 that it was possible to estimate the unknown parameters

of the model, $\hat{\Psi}^{(m)}$. Hence, it is now possible to compute the estimated log-likelihood of the chosen model $L\left(\hat{\Psi}^{(m)}; y,m\right)$. The idea is hence given by the following remark.

Remark 2.1 *The best choice for m, noted m^*, is the one, such that the function $m \mapsto L\left(\hat{\Psi}^{(m)}; y,m\right)$ does not increase in a significant way for values greater than m^*.*

So, set $J(\Psi,y) = -L(\Psi; y)$. We use the following notations:

$$\hat{\Psi}^{(m)} = \arg\min_{\psi \in \mathcal{F}_m} J(\Psi,y), \tag{2.6}$$

$$J_m = J\left(\hat{\Psi}^{(m)},y\right).$$

Finally, for all $\beta > 0$ and for all $1 \leq m \leq M$, where M is an upper bound for the maximum number of components, set

$$\hat{m}(\beta) = \arg\min_{1 \leq m \leq M} \left(J_m + \beta m\right).$$

The following proposition is due to Lavielle in [Lav02].

Proposition 2.2 *There is a sequence $m_1 = 1 < m_2 < \ldots$, and a sequence $\beta_0 = +\infty > \beta_1 > \ldots$, with*

$$\forall i \geq 1, \beta_i = \frac{J_{m_i} - J_{m_{i+1}}}{m_{i+1} - m_i},$$

such that

$$\forall \beta \in]\beta_i,\beta_{i-1}[, \hat{m}(\beta) = m_i.$$

As a consequence, the estimation procedure of the optimal number of components of the mixture is given by:

1. For $m = 1, \ldots, M$, compute $\hat{\Psi}^{(m)}$ and J_m.

2. Then compute the sequence $(\beta_i)_{i=1,\ldots,M}$, as well as l_i the length of the intervals $]\beta_i,\beta_{i-1}[$, for all $i = 1, \ldots, M$.

3. Keep the largest value of the m_i such that $l_i >> l_j$, for all $j > i$.

Actually, the previous procedure is a model selection approach with a stability criterion that replaces the trade off between bias and variance, as it is quoted by Birgé and Massart in [BM01]. This proposition provides an automatic criterium that mimics the main idea developed in Remark 2.1.

2.3.3 Results

With the criterion presented in Section 2.3.2, we conclude that, for counting station number 12, the values of $J\left(\hat{\Psi}^{(m)},y\right)$ do not decrease in a significant way for

values greater than $m^* = 13$. So, with this criterion, for all $\zeta_s, s = 1, \ldots, S$, we chose the optimal number of representatives.

For the counting station number 12, we present, in Figure 2.3, the values of the log-likelihood at each step of the algorithm. We also graph the values of the parameters $(\pi_j)_{j=1,\ldots,m^*}$ and σ^2. We see that all values have converged before approximately 15 iterations.

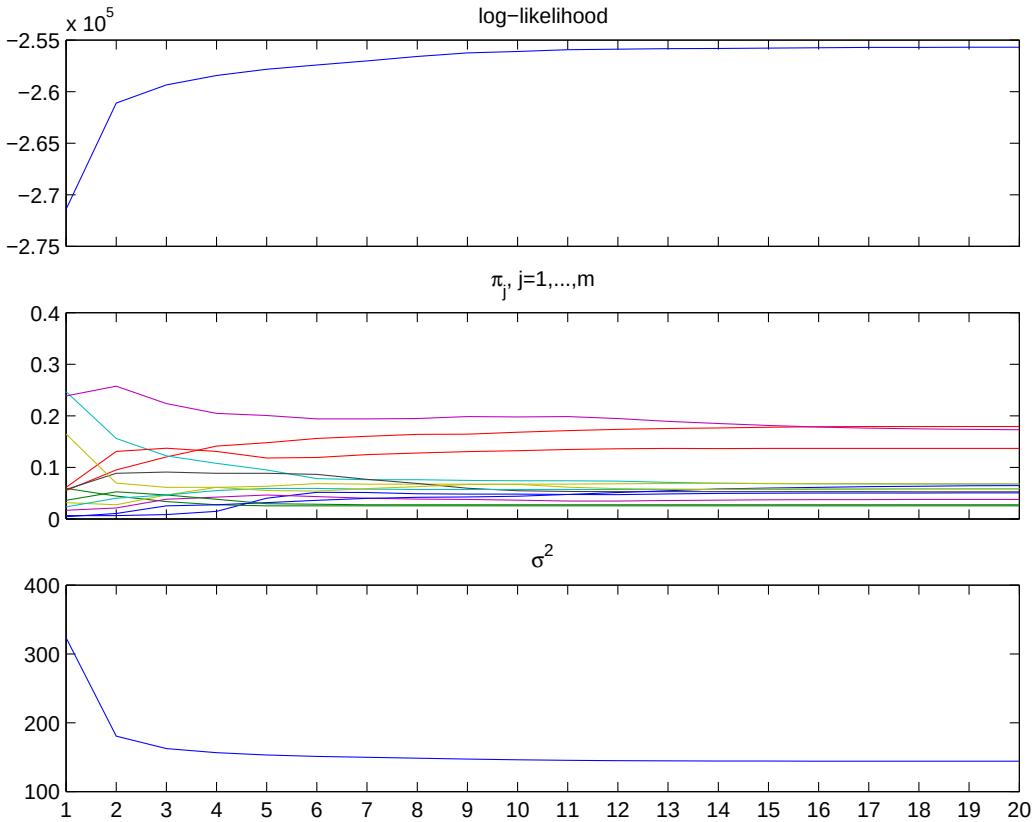

FIG. 2.3 – *Values of log-likelihood and of parameters, at each step of the SAEM algorithm, for the counting station number 12, with* $m = 13$.

The figure 2.4 represents the 13 archetypes of the station number 12. The different curves obtained show distinct typical behaviours. Some well known behaviours appear, like traffic jams at peak hours and traffic keeping moving otherwise.

This method is based on mixture model (2.1). An alternative approach is given by the method of automatic classification, in Section 2.4, which makes it possible to be closer to the data observed.

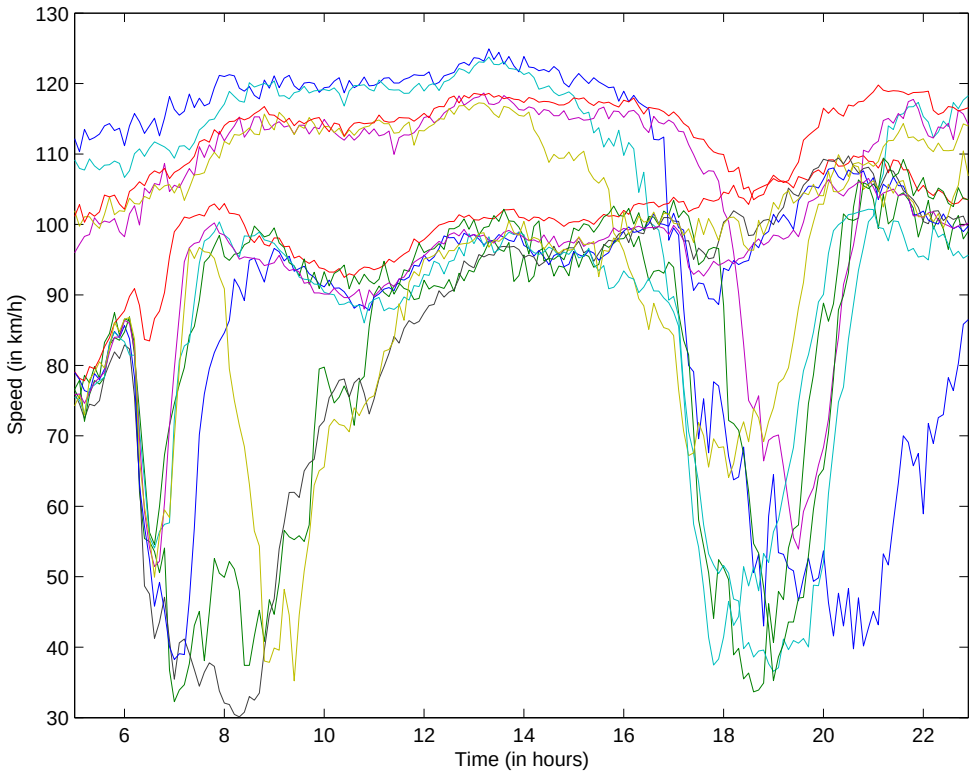

FIG. 2.4 – *The $m^* = 13$ standard profiles of the counting station number 12.*

2.4 Classification method

A classification method aims at gathering individuals in a restricted number of representative classes. Representative classes are such that two individuals taken inside the same class are *similar* (homogeneous class), and such that two individuals taken in two different classes are *distinct* (heterogeneous classes). The introduction of an appropriate distance index for speed curves (distance, dissimilarity index, variation, ultra-metric variation, etc.) will enable us to quantify the qualitative terms *similar* and *distinct*. The distance indices used in the classification domain are numerous and many researches are carried out on the indices associated with longitudinal data, see for example in [DLB96].

Thus, with a suitable distance for our speed curves Y_n^s, we have carried out a hierarchical classification of our data, in Section 2.4.1. In Section 2.4.2, with a methodology close to learning theory, we have estimate the m^* standard profiles, $f_1, \ldots, f_{m^*}$, for each counting station ζ_s. Results are presented in Section 2.4.3.

2.4.1 Hierarchical classification

The result of the hierarchical classification strongly depends on the choice of between-individuals and between-clusters distance indices. Classical distance indices, see for example in [Gor99] and [DLB96], were not appropriate for this kind of tempo-

ral data. Indeed, the study of the road traffic implies taking into account the temporal aspect of our speed curves. For example, the figure 2.5 presents three simplified speed curves, X, Y and Z, one resulting from both others by a simple translation. These three curves are characterized by a constant speed, 90 km/h, from 5 AM to 11 PM, except over a 2 hours period during which speed is 30 km/h, respectively at 8 AM, 11 AM and 2 PM. The Euclidean distance, noted d, calculated on these three curves gives obviously the same result: $d(X,Y) = d(Y,Z) = d(X,Z) = 389$. But, a suitable distance index must make the difference between a deceleration which occurs at 8 AM, at 11 AM or at 2 PM. Thus, we built a distance, which we will note Δ, that takes into account this shift effect.

Definition 2.3 *Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. Set $\Delta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ as*

$$\Delta(x,y) = \sqrt{{}^t(x-y)W(x-y)},$$

with W a $n \times n$ matrix defined by $W_{ij} = \frac{n-|i-j|}{n}, \forall i = 1,\ldots,n, \forall j = 1,\ldots,n$.

We point out that Δ is a distance on $\mathbb{R}^n$. For the preceding example, Δ gives the following results: $\Delta(X,Y) = \Delta(Y,Z) = 637$ and $\Delta(X,Z) = 967$. Thus, Δ enables to differentiate the translation speed curves. This property, on such curves, can be proved by straightforward calculations.

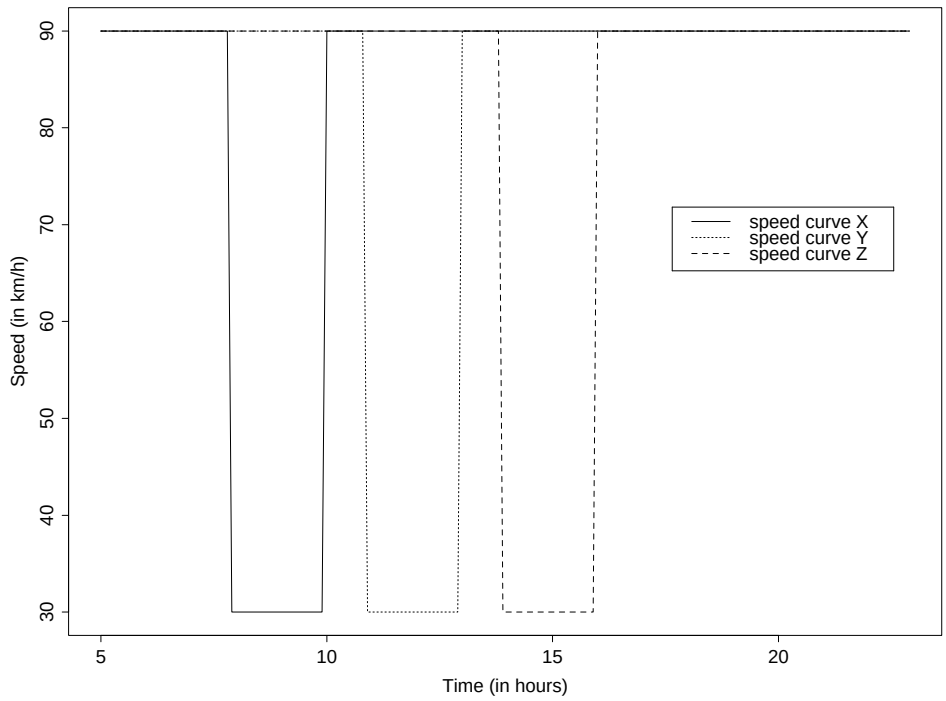

FIG. 2.5 – *Simplified speed curves.*

So, because this shift phenomenon is frequently observed in speed curve behaviours the between-individuals distance is here the Δ distance. The between-clusters

distance index is the distance index of the maximum variation, noted D, and defined, for two clusters A and B, by

$$D(A,B) = \max_{x \in A,\, y \in B} \Delta(x,y).$$

The criterion of the maximum variation guarantees to obtain homogeneous classes, loosing between classes heterogeneity (for examples, see [Gor99]).

The figure 2.6 gives the result of hierarchical classification on the counting stations number 12. For this, we use the Johnson agglomerative algorithm. This simple algorithm gathers together, at each step, the closest clusters.

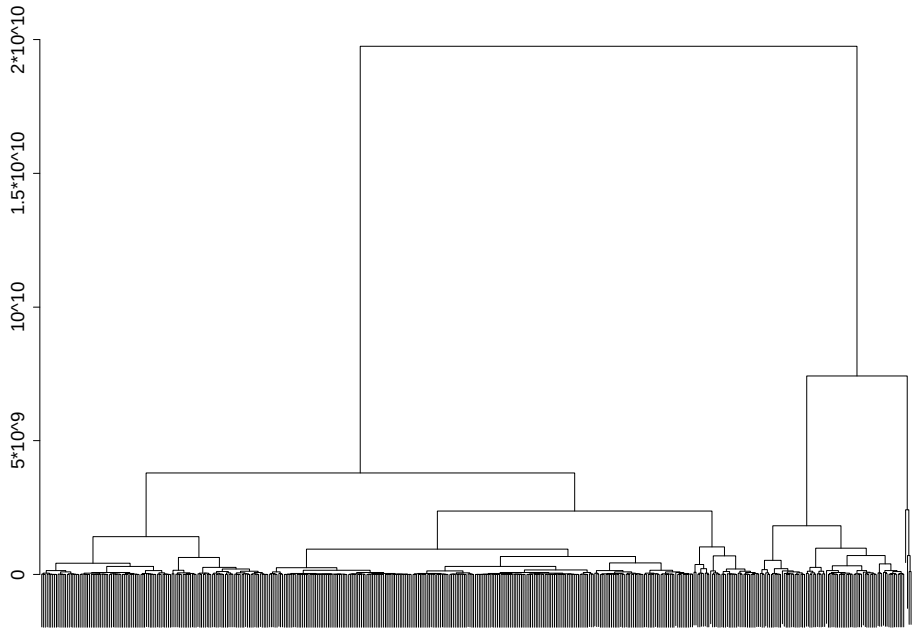

FIG. 2.6 – *Hierarchical classification of the counting station number 12 (Johnson algorithm).*

2.4.2 Choice of the optimal number of clusters

Once the hierarchical classification is carried out, we want to keep only a number m^* of significant classes, for each counting station ζ_s. Obtaining a classification amounts to cut the classification tree at a given height. So, a hierarchical classification heads to a unique classification with m^* clusters. For this purpose, it is necessary to determine a criterion of optimality to choose the number of clusters. In our learning context, this optimality depends on the result we want to obtain. Thus, that amounts conducting forecasts on the learning sample with $m = 2, \ldots, M$, with

M a fixed big enough integer, and then, choosing the number of clusters minimizing a forecasting error criterion.

Definition 2.4 *m^* is the optimal number of clusters, of the counting station number s, ζ_s, if it minimizes the absolute forecasting error over two hours.*

Hence, we write

$$m^* = \arg \min_{m=2,\dots,M} \sum_{n=1}^{N_L} \sum_{t=11}^{161} \sum_{p=t}^{t+19} \left| Y_n^s(p) - f_{m,j(n,t)}^s(p) \right| ,$$

with

$$f_{m,j(n,t)}^s = \arg \min_{f_{m,\tilde{m}}^s,\, \tilde{m}=1,\dots,m} \Delta' \left((Y_n^s)_1^{t-1}, (f_{m,\tilde{m}}^s)_1^{t-1} \right) ,$$

where $f_{m,1}^s, \dots, f_{m,m}^s$, are the standard profiles obtained with m clusters. Hence, $f_{m,j(n,t)}^s$ is the closest profile to Y_n^s in the sense of the distance Δ' defined in Section 2.5, at period t, when we chose m standard profiles.

This process enables us to find the optimal number of classes, i.e., a sufficiently big number to be able to represent the various possible traffic states, and a sufficiently small one not to reveal too many classes and/or too similar classes, which harms the forecast.

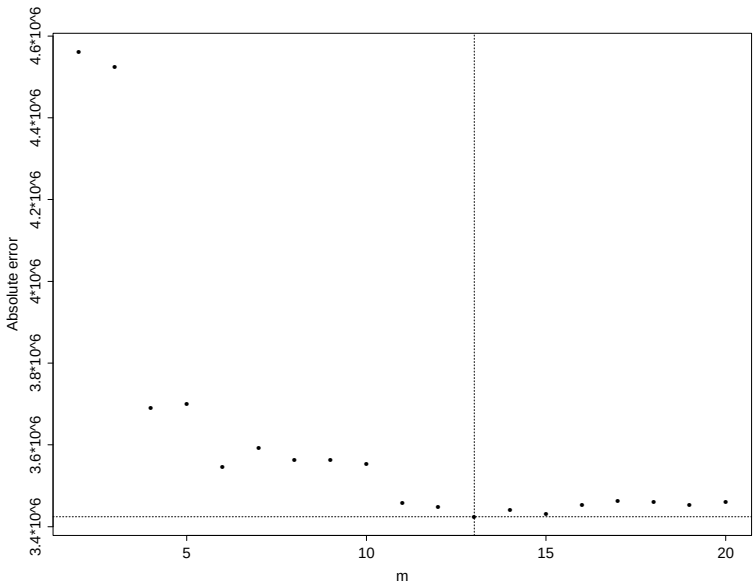

FIG. 2.7 – *Absolute forecasting error for the station number 12, with m standard profiles, $m = 2, \dots, 20$. The optimal value is reached for $m^* = 13$.*

For example, the figure 2.7 shows the absolute errors calculated for the station number 12, for $m = 2, \dots, 20$. We can see that a small number of profiles isn't enough

representative, but also we can see that there is an over fitting phenomenon when the number of profiles increases. So, all counting stations have the same behaviour.

2.4.3 Results

The number of standard profiles is not the same, according to the counting station we consider. The figure 2.8 gives the number of standard profiles obtained on each counting station. The upper bound value of m, M, was chosen equal to 15. Indeed, after some examples with higher values, we saw that $M = 15$ was big enough.

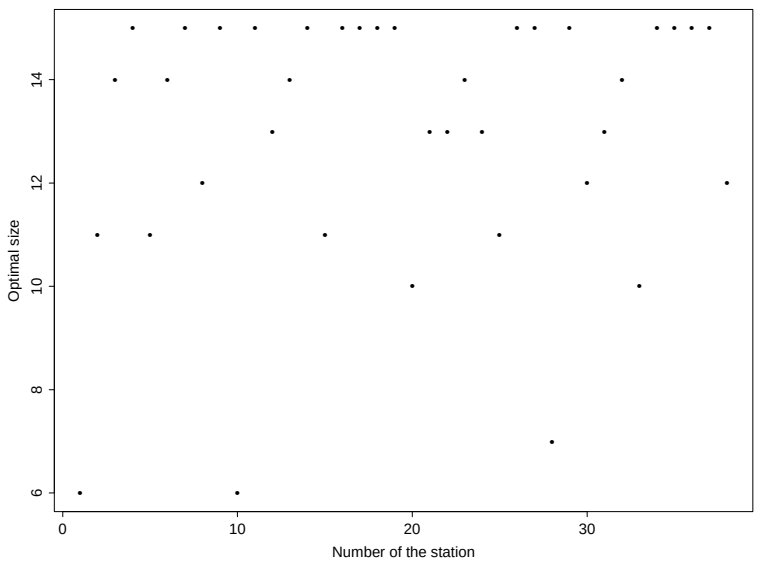

FIG. 2.8 – *Number of standard profiles obtained on each counting station.*

Until now, we classified the data in m^* representative classes, for each station ζ_s. We have now to determine the standard profiles, $\hat{f}_1, \ldots, \hat{f}_{m^*}$, for each cluster. In order to obtain a representative curve of the speed behaviour in each class, we chose to use a robust estimator: the median of the speed curves of each class. The figure 2.9 presents the standard profiles obtained for the station number 12.
The results are qualitatively the same as those we have for the mixture model. Thus, some known behaviours appear, like traffic jams at peak hours. But, there is an important difference. Indeed, the hierarchical classification let appear some curves that seem to be outliers or rare events. Hence, in figure 2.9, there are two curves with larger and deeper traffic jams in the morning, that don't appear in figure 2.4. This is a classical over-smoothing effect of the SAEM algorithm.

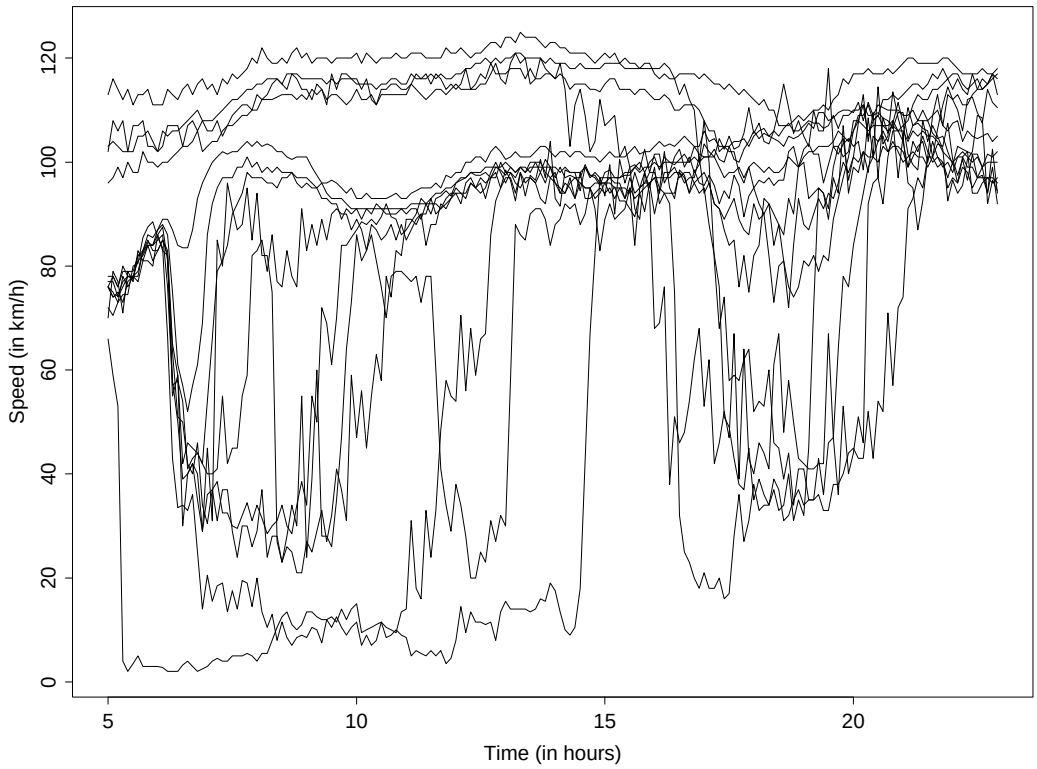

FIG. 2.9 – *Standard profiles of the station number 12.*

2.5 Travel time forecasting

In the two previous parts, we have constructed, for each observation station ζ_s, $s = 1, \ldots, S$, and using two different methods, the standard profiles

$$f_j^{(i)} \in \mathcal{F}^i, i \in \{1,2\}, j = 1, \ldots, m_i.$$

These two sets, $\mathcal{F}^1$ and $\mathcal{F}^2$, represent the archetypes of the daily vehicle speed resulting, respectively, from the mixture model ($i = 1$) and from the empirical classification ($i = 2$). Our aim is now to use such profiles in order to forecast a customer travel time, for a given itinerary, one hour before he leaves (at the time $H + 1$). Indeed, for sake of simplicity, we only present here forecasting results with $h = 1$ hour.

Set ξ_{n_0} the day of the trip, and t_0 such that $H \in \left[\frac{t_0+49}{10}, \frac{t_0+50}{10}\right[$. In order to perform the forecast, we estimate, using the measurements of the day, on all the stations of the itinerary, the speeds $f^s(t), \forall s \in \mathcal{S}, \forall t \geq t_0$, where $\mathcal{S}$ is the set of all stations of the chosen itinerary. Once speed evolutions are known, the estimation of the travel time is easy. As a consequence, the main issue is, for each station, the estimation of the traffic velocity. It is done by comparing the incoming data of the day ξ_{n_0} before the time H, i.e. $Y_{n_0}^s(t), \forall s \in \mathcal{S}, \forall t < t_0$, with all the curves of $\mathcal{F}^1$ or $\mathcal{F}^2$, and by choosing the closest.

More precisely, for a fixed counting station $\zeta_s, s \in \mathcal{S}$, consider d a distance. Set also, $\forall i \in \{1,2\}, \forall g \in \mathcal{F}^i$, $g_1^{t_0-1} = {}^t(g(1),\ldots,g(t_0-1))$ and $Y_1^{t_0-1} = {}^t(Y_{n_0}^s(1),\ldots,Y_{n_0}^s(t_0-1))$. After having chosen one of the two models, $\mathcal{F} = \mathcal{F}^1$ or $\mathcal{F}^2$, the estimator will be $\forall t \geq t_0$, $\hat{f}(t)$, with

$$\hat{f} = \arg\min_{g \in \mathcal{F}} d\left(g_1^{t_0-1}, Y_1^{t_0-1}\right). \tag{2.7}$$

According to the remarks pointed out in Section 2.4, we have chosen $d = \Delta'$, with Δ' a modified restriction of Δ to the subset $\mathbb{R}^{t_0-1} \times \mathbb{R}^{t_0-1}$, defined more precisely hereafter.

Definition 2.5 *Set $Y_1^{t_0-1}$ the collected data, the day ξ_{n_0}, on the station ζ_s, before the time H, i.e. for $t < t_0$ (for sake of simplicity we have omitted station and day indexes). We define the distance between $Y_1^{t_0-1}$ and the different archetypes of the station ζ_s, restricted to the values of $t < t_0$, and written $(f_j)_1^{t_0-1}$, with $f_j \in \mathcal{F}, \forall j = 1,\ldots,m$, by*

$$\Delta'\left(Y_1^{t_0-1}, (f_j)_1^{t_0-1}\right) = \frac{\Delta\left(PY_1^{t_0-1}, P(f_j)_1^{t_0-1}\right)}{\sqrt{\pi_j}}, \tag{2.8}$$

where π_j, $j = 1,\ldots,m$, is the size of the cluster number j, and P is a $t_0 - 1 \times t_0 - 1$ matrix, defined by $P_{ij} = \begin{cases} \frac{1}{t_0-i} & \text{if } i = j \text{ and } i \geq t_0 - 10, \\ 0 & \text{in any other case.} \end{cases}$

Hence, we have used the test sample to forecast travel time on the A4W road section, using the standard profiles obtained with the two methodologies described in Section 2.3 and Section 2.4.

We compare the results with the estimations given by the stationary model, defined as the simplest model, which estimates the speed by the last observed speed, i.e. $Y_{n_0}^s(t_0-1)$. This model is important since, on the one hand it is the only reference we have, and on the other hand, the forecasting results with such a model gives us an indicator of the traffic behaviour on the considered road section. Indeed, good forecasts point out that the traffic is moving freely. On the contrary, bad forecasts with the stationary model show us that the itinerary is often congested.

Table 2.1 present some characteristics of travel time errors obtained with simulated itineraries on the test sample, for the three models: the stationary model, the classification model and the mixture model. The error (in minutes) is defined by

$$\text{error} = \text{estimated travel time} - \text{real travel time}.$$

characteristics	**mean**	**std dev.**	**min.**	**max.**	$P(\textbf{error} \geq 45\textbf{mn})$
stationary model	0.52	8.18	-41	57	0.003
classification model	-1.53	5.40	-41	28	0
mixture model	-2.00	5.65	-41	18	0

TAB. 2.1 – *Results.*

Figure 2.10 presents the histograms of the travel time errors. The three other figures (Figure 2.11, Figure 2.12 and Figure 2.13) are examples of the evolution of the forecast errors throughout the day. More precisely, for each of these days, in our fixed itinerary, we consider a traveler with that itinerary beginning at each of the periods, and then we calculate actual and model predicted travel times.

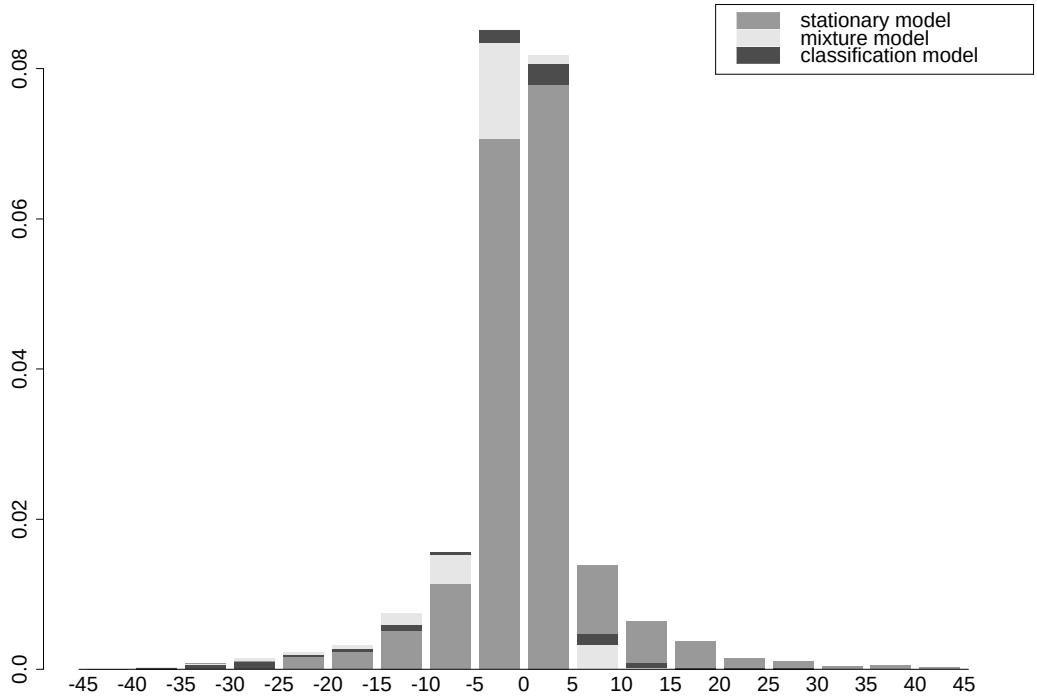

FIG. 2.10 – *Histograms of travel time errors.*

These results enable us to compare the three procedures:

P_0 the rough stationary model,

P_1 the classification model, with an optimal number of clusters chosen with a learning sample,

P_2 the model of the loglikelihood estimator.

Studying these outcomes, we can draw the following conclusions.

Comparison between P_0, P_1 and P_2

First of all, both predictors improve the estimate provided by the stationary model, with smaller prediction variances. This increase depends on the road which

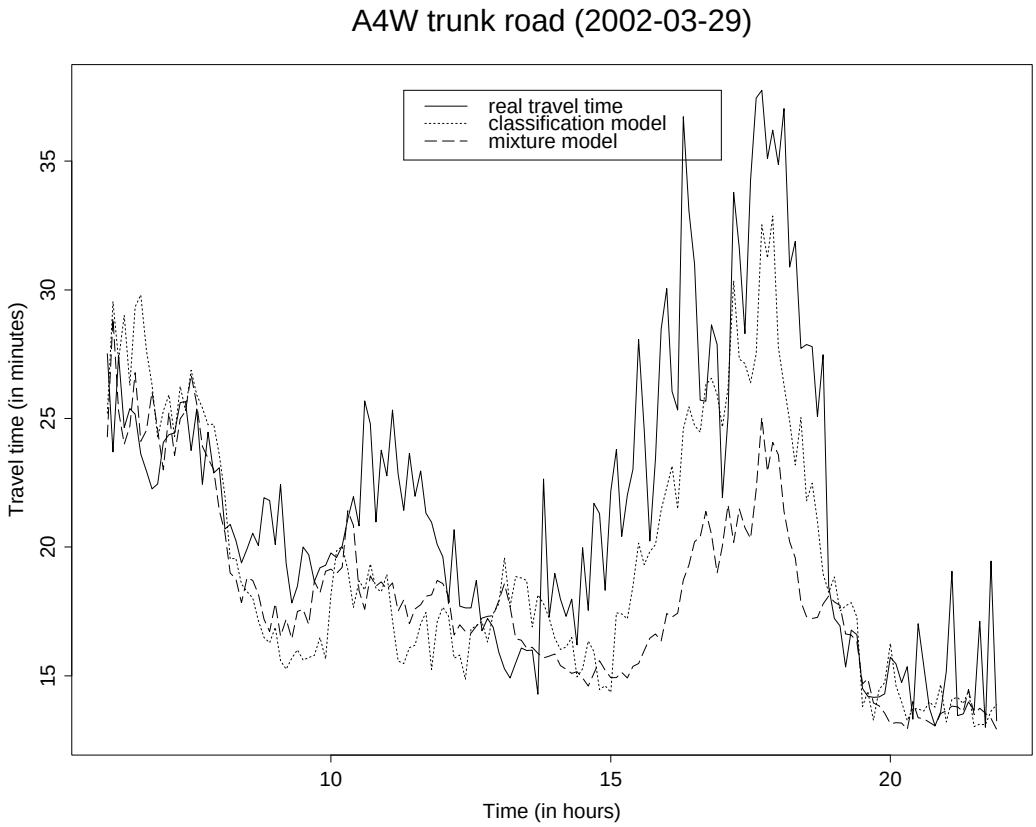

FIG. 2.11 – *Predicted travel times on a test sample day.*

is studied. The more different behaviours of road trafficking can be found on that road, the better is the gain. This is easy to explain since the stationary model provides a mean pattern which is far from the real feature when there are many. Yet, this number of representative patterns is a measure of the complexity of the road, standing for its variability, with respect to changes day after day.

Comparison between P_1 and P_2

We also point out that both models underestimate the real travel time, so for practical applications, this bias can be taken into account. When comparing the performance of our estimators, we can see that the log-likelihood estimator is slightly outperformed by the classification type estimator. Indeed, the mean of the errors with the classification model is close to 0, and the variance is smaller. The reasons for this difference are the following:

- First, the distance chosen to evaluate the performance of the estimator, is the same that is used to classify the data in the methodology described in Section 2.4, since this distance best matches the prediction goals. But the optimal choice of models is achieved, via a learning process, by minimizing the prediction distance over a learning sample. Hence, this choice induces a bias in favor of the classification method.

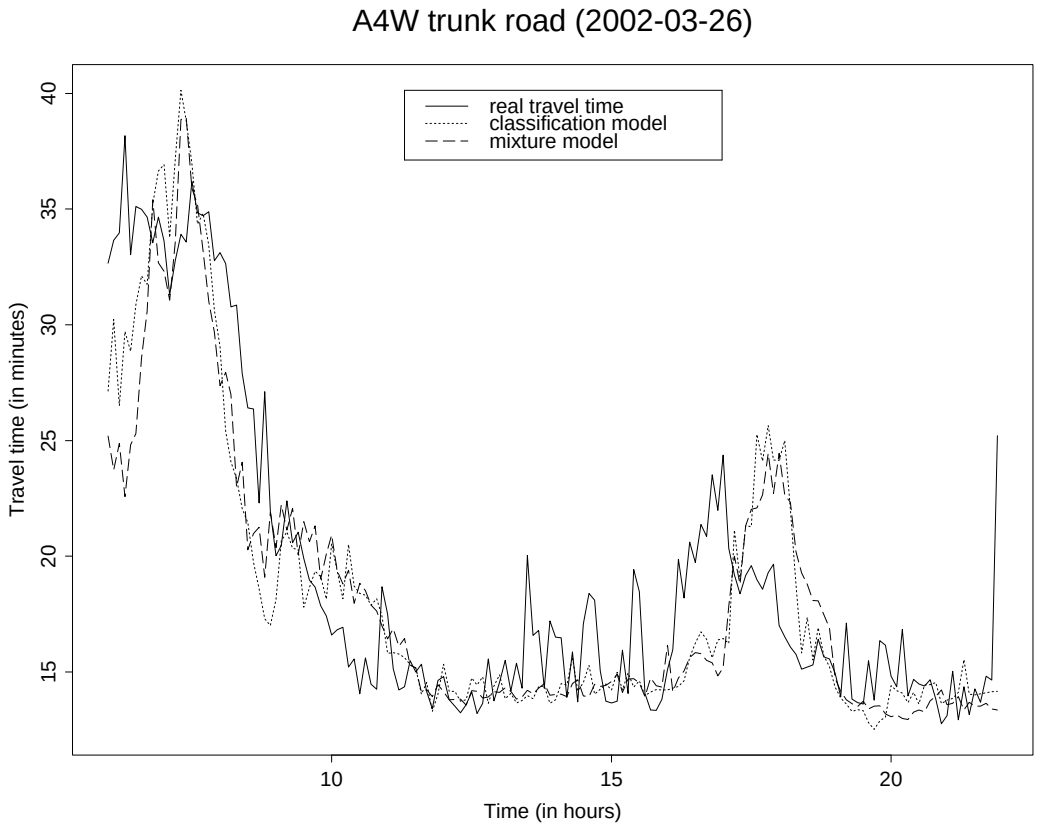

FIG. 2.12 – *Predicted travel times on a test sample day.*

- Outliers and rare events play also an important role in this study. On the one hand, the mixture model is very sensitive to outliers. Indeed, the loglikelihood estimator uses all the data with the same weights to build an average representative, while a classification method tends to isolate such outliers in special classes. Hence, the blurring effect of outliers is stronger for the mixture model since they add a deviation term to the estimates. On the other hand, rare events are more easily catched by the classification method. Indeed, we have point out in Section 2.4.3, that standards profiles given by the classification method contain more rare event profiles. Hence, for example in Figure 2.13, which is one of the worst case, the congested phenomena are slightly better detected: the travel times with traffic jams of the mornings are better estimated. Unlike theoretical study in model selection, we found the optimal number of models by minimizing the prediction error but without a penalty term. As a matter of fact, we did not want to discard the rare events that can be alone in a cluster, but represent a real behaviour. Hence, it enables the predictor based on the classification model to keep in mind some unlikely events, and to give an adequate response when the observations do not follow a typical pattern. The price to pay is that the collection of cluster for the classification model is slightly larger than the number of features kept by the mixture procedure.

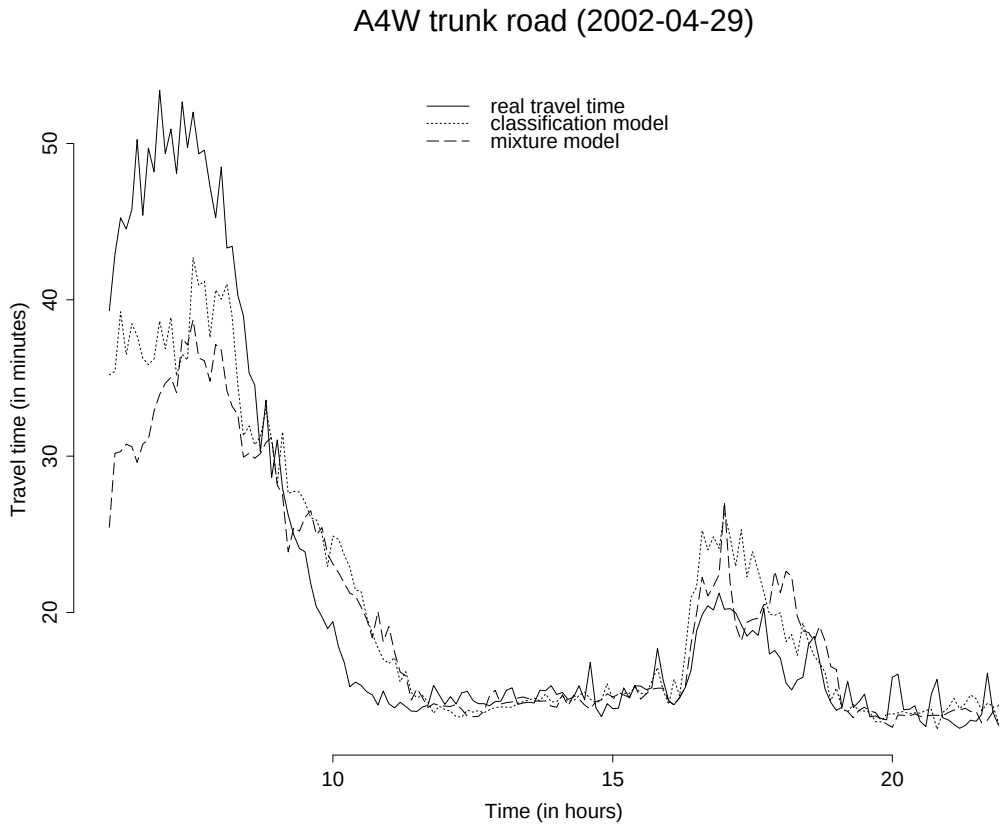

FIG. 2.13 – *Predicted travel times on a test sample day.*

- Nevertheless, there are two advantages for using the loglikelihood model. First, it is a very efficient method on a computational point of view, which is much faster than the computations necessary to perform the learning process of the model selection method. Moreover, for small values of the number of models, loglikelihood estimators provide a better description of the data. But, increasing the number of models for the loglikelihood does not improve the estimation error. Indeed the additional selected patterns are redundant, since, as we have already said, the loglikelihood estimator does not put the stress on rare events, while the classification type predictor isolates such features in single classes.

2.6 Conclusion

The results we have obtained are encouraging and are far better than the results given by the usual global forecasting methods (*Sytadin*, http://www.sytadin.tm.fr, or *Bison Futé* for instance), which only rely on the rough model. Both models are interesting: the mixture model by its simplicity and the good performance, and the classification model which is the more accurate but, at the same time, the more complicated on a computational point of view.

However, it is possible to improve the performances of the method, for example in the choice of the archetypes that stand for a whole cluster. Indeed, in a class, the functions are similar but they may be translation shift between them. As a consequence, taking the median of all the functions for the representative of the class often leads to an over smoothing effect. Methods able of keeping the structure of the functions group, as it is done in [KG92], [Kne94] or [RD91], are developed by Gamboa, Loubes and Maza in [GLM03] in the setting of high dimensional data.

Moreover, other modelling attempts can be conducted. It seems rather natural to take into account the dependency of all the stations which are considered in this work as independent. A method using the spatial links between the observation cells is taken into account by the authors in a forthcoming work.

Finally aggregating the estimators should also improve the performance of the procedure. Indeed, in this work, we have considered separately the prediction given by each methodology. An alternative should be to use a linear combination of such predictors to improve our results. Such a work is still in progress.

Chapitre 3

Estimation dans un modèle de translation

Les progrès effectués en matière de recueil de données font que, de plus en plus souvent, le résultat d'une expérience n'est plus une simple valeur numérique ou qualitative, mais une fonction. C'est le cas dans les chapitres 1 et 2 où nos données sont principalement constituées de courbes de vitesse. Les méthodes statistiques s'adaptent à ces nouvelles données depuis quelques années. Par exemple, l'ouvrage [Sil95] intitulé *Functionnal Data Analysis* propose différentes extensions des méthodes statistiques classiques comme l'Analyse en Composantes Principales ou la régression linéaire. Dans ce chapitre, nous nous plaçons donc dans ce contexte de données fonctionnelles. Dans la section 3.1, nous décrivons le modèle étudié et explicitons les motivations qui nous ont amené à poser un tel modèle. La section 3.2 est dédiée à la description de la méthode de résolution et aux propriétés qui en découlent. Dans la section 3.3, nous donnons les résultats de consistance et de normalité asymptotique obtenus. Ces résultats font l'objet d'un article soumis. Le chapitre 4 suivant est constitué de cet article. Enfin, dans la section 3.4 nous donnons rapidement les résultats obtenus sur nos données de trafic après la soumission de l'article. Ainsi, le présent chapitre résume le chapitre suivant et explicite les résultats obtenus dans le cadre de la prévision de trafic routier.

3.1 Introduction

Pour chacune des stations de comptage, sur les courbes de vitesse, le phénomène de congestion du trafic routier se traduit souvent par une chute très rapide de la vitesse mesurée, puis par des mesures de vitesse approximativement constantes, puis enfin par un retour à la normale des mesures. La classification des courbes, effectuée dans les chapitres précédents, nous a amené à constater que ce phénomène de *bouchon* semblait se répéter à l'identique dans certaines des classes obtenues. Plus précisément, certaines des classes étaient constituée de courbes de vitesse qui semblaient se déduire les unes des autres par une simple translation. Ce phénomène se conçoit aisément si l'on considère que, sur un tronçon de route donné, le flux de trafic routier est constant au fil des jours. Ainsi, le phénomène du bouchon garde la même amplitude et ne se différencie d'un jour sur l'autre que par le moment de

son apparition. Nous avons aussi constaté que les profils types extraits de certaines des classes de notre classification n'avaient plus la même structure que les courbes des classes. En effet, la moyenne des courbes d'une classe donnée tend à perdre en amplitude lorsque le phénomène sous-jacent est un phénomène aléatoire de déformation de l'axe des temps. Ce problème est décrit plus en détails dans [GLM03] (chapitre 4) et dans le chapitre 5. Ces constatations nous ont donc amené à étudier le modèle suivant.

Pour chaque courbe $\mathcal{C}_j, j = 1, \ldots, J$, des mesures $y_{ij}, i = 1, \ldots, n$, sont observées sur des points équidistants $t_i = \frac{i-1}{n}T \in [0, T[$, avec $T \in \mathbb{R}_+^*$. Nous modélisons ces données par

$$Y_{ij} = f\left(t_i - \theta_j^*\right) + \varepsilon_{ij}, j = 1, \ldots, J, i = 1, \ldots, n, \tag{3.1}$$

où

- $f : \mathbb{R} \to \mathbb{R}$ est une fonction T-périodique inconnue,
- $\theta^* = \left(\theta_j^*\right)_{j=1,\ldots,J} \in \mathbb{R}^J$ est un ensemble de paramètres de translation inconnu, avec θ_j^* correspondant à la translation de la courbe numéro j,
- pour tout $j = 1, \ldots, J$, $(\varepsilon_{ij})_{i=1,\ldots,n}$ est un bruit blanc gaussien de variance 1.

Nous sommes donc ici dans un cadre semi-paramétrique. En effet, notre but est l'estimation des paramètres de translation $\theta_j^*, j = 1, \ldots, J$, avec la fonction f inconnue considérée comme un paramètre de nuisance.

Notre estimation est basée sur une décomposition de nos fonctions sur une base de Fourier. En effet, la base de Fourier diagonalise l'opérateur de translation, puisque, pour tout $j = 1, \ldots, J$, les coefficients de Fourier associés aux fonctions translatées $f(\cdot - \theta_j^*)$ sont tels que, pour tout $l \in \mathbb{Z}$,

$$\begin{aligned} c_l\left(f\left(\cdot - \theta_j^*\right)\right) &= \frac{1}{T}\int_0^T f\left(t - \theta_j^*\right) e^{-i2\pi\frac{tl}{T}} \mathrm{d}t \\ &= e^{-i2\pi\frac{\theta_j^* l}{T}} c_l(f), \end{aligned}$$

où i est le nombre complexe tel que $i^2 = -1$. Ainsi, par l'intermédiaire de la transformée de Fourier discrète, en supposant que n est impair, nous transformons le modèle (3.1) en

$$d_{jl} = e^{-il\alpha_j^*} c_l(f) + w_{jl}, j = 1, \ldots, J, l = -(n-1)/2, \ldots, (n-1)/2,$$

où

- $(c_l(f))_{l\in\mathbb{Z}}$ sont les coefficients de Fourier de f,
- les valeurs $\alpha_j^* = \frac{2\pi}{T}\theta_j^* \in \mathbb{R}$, pour tout $j = 1, \ldots, J$, sont nos paramètres de translation,
- pour tout $j = 1, \ldots, J$, $(w_{jl})_{l=-(n-1)/2,\ldots,(n-1)/2}$ est un bruit blanc complexe, de variance $\frac{1}{n}$, avec les parties réelle et imaginaire indépendantes.

Nous sommes donc bien dans un cadre semi-paramétrique. Nos paramètres d'intérêt sont les valeurs $\alpha^* = \left(\alpha_j^*\right)_{j=1,\ldots,J}$ qui dépendent d'un paramètre de nuisance

fonctionnel $(c_l(f))_{l\in\mathbb{Z}}$. Nous remarquons tout d'abord que notre modèle n'est pas identifiable. Ce problème est détaillé dans le chapitre 4. Nous noterons juste ici que pour retrouver l'identifiabilité, nous devrons nous placer dans un espace de paramètres particulier que nous noterons A.

3.2 Le contraste

Face à ce problème d'estimation semi-paramétrique, si nous voulons des hypothèses assez faibles sur la fonction inconnue f, le cadre des M-estimateurs est bien approprié. De plus, nous remarquons que pour tout $\alpha = (\alpha_j)_{j=1,\ldots,J} \in A$, les coefficients rephasés

$$\tilde{c}_{jl}(\alpha) = e^{il\alpha_j} d_{jl}, j = 1, \ldots, J, l = -(n-1)/2, \ldots, (n-1)/2,$$

et la moyenne de ces coefficients

$$\hat{c}_l(\alpha) = \frac{1}{J}\sum_{j=1}^{J} \tilde{c}_{jl}(\alpha), l = -(n-1)/2, \ldots, (n-1)/2,$$

sont tels que

$$\tilde{c}_{jl}(\alpha^*) = c_l(f) + e^{il\alpha_j^*} w_{jl},$$

et

$$\hat{c}_l(\alpha^*) = c_l(f) + \frac{1}{J}\sum_{j=1}^{J} e^{il\alpha_j^*} w_{jl}.$$

Ainsi, la valeur $|\tilde{c}_{jl}(\alpha) - \hat{c}_l(\alpha)|^2$ devrait être petite dès que α est proche de α^*. Après cette remarque, pour une suite quelconque $(\delta_l)_{l\in\mathbb{Z}}$ telle que

$$\sum_{l\in\mathbb{Z}} \delta_l^2 < +\infty,$$

nous définissons le critère empirique

$$M_n(\alpha) = \frac{1}{J}\sum_{j=1}^{J} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \left|\tilde{c}_{jl}(\alpha) - \hat{c}_l(\alpha)\right|^2.$$

Ainsi, le M-estimateur de α, pour un entier n fixé, est donné par

$$\hat{\alpha}_n = \arg\min_{\alpha\in A} M_n(\alpha).$$

3.3 Les résultats asymptotiques

Les deux théorèmes suivants nous assurent la consistance et la normalité asymptotique de notre M-estimateur.

Théorème 3.1 *Sous les hypothèses suivantes sur la fonction f et sur la suite des poids $(\delta_l)_{l\in\mathbb{Z}}$:*

$$\sum_{l\in\mathbb{Z}} |c_l(f)|^2 < +\infty,$$
$$\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 < +\infty,$$

nous avons

$$\hat{\alpha}_n \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \alpha^*.$$

Pour des raisons de simplicité, la normalité asymptotique est démontrée dans l'espace des paramètres

$$A_1 = \left\{\alpha \in [-\pi,\pi[^J;\ \alpha_1 = 0\right\}.$$

En conséquence, nous montrons la normalité asymptotique dans l'espace des paramètres

$$\tilde{A}_1 = [-\pi,\pi[^{J-1},$$

en posant, pour tout $n \in \mathbb{N}$,

$$\hat{\alpha}_n = \arg\min_{\alpha\in\tilde{A}_1} M_n \begin{pmatrix} 0 \\ \alpha \end{pmatrix},$$

avec $\alpha = (\alpha_j)_{j=2,\dots,J}$.

Théorème 3.2 *Sous les hypothèses du théorème 3.1 et sous l'hypothèse supplémentaire suivante:*

$$\sum_{l\in\mathbb{Z}} \delta_l^4 l^4 < +\infty,$$

nous avons

$$\sqrt{n}(\hat{\alpha}_n - \alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}(0,\Gamma),$$

avec

$$\Gamma = \frac{\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2}{\left(\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2\right)^2} \left(I_{J-1} + U_{J-1}\right),$$

où, I_{J-1} est la matrice identité de dimension $(J-1)$ et U_{J-1} est la matrice carrée de dimension $(J-1)$ dont chaque élément est égal à 1.

3.4 Application

Outre les simulations effectuées avec cette modélisation (voir chapitre 4), nous avons mis en place cette estimation dans le cadre de notre prévision de temps de parcours. Dans le chapitre 2 (section 2.4.2 et section 2.4.3) nous calculions les profils types associés à chacune des classes de notre classification par la médiane des courbes de la classe. Nous avons donc introduit, dans notre algorithme de calcul du nombre optimal de classes, l'estimation des translations en utilisant la transformée

de Fourier discrète. Cette nouvelle approche ne s'est pas avérée concluante. En effet, nous avons constaté que l'estimation du nombre optimal de classes par station de comptage était approximativement le même. Par contre, les erreurs calculées lors de l'apprentissage étaient supérieures à celles déjà obtenues. La raison est simplement que les courbes de vitesse qui ont la propriété de se déduire les unes des autres par une simple translation sont peu nombreuses. Cette remarque reste valable pour le modèle étudié dans [Vim04] qui ajoute aux paramètres de translation de notre modèle, des paramètres de dilatation de chacune des courbes. Notre constatation initiale reste tout de même valable, à savoir que la structure des courbes est approximativement la même dans certaines classes. Par contre, le modèle de translation (et de dilatation) s'avère être irréaliste. Cela nous a amené, tout naturellement, à nous intéresser à des transformations plus générales. Ces transformations sont définies et étudiées dans le chapitre 5.

Chapitre 4

Semi-parametric estimation of shifts (submitted article)

By authors: Fabrice Gamboa**, Jean-Michel Loubes* and Elie Maza**.

* CNRS - Laboratoire de Mathématiques
Equipe de Probabilités, Statistique et Modélisation
UMR 8628 - bâtiment 425, Université Paris Sud
91405 Orsay cedex
** Laboratoire de Statistique et de Probabilités
UMR C5583 - Université Paul Sabatier
118, route de Narbonne
F-31062 Toulouse cedex 4

Fabrice.Gamboa@math.ups-tlse.fr
Jean-Michel.Loubes@math.u-psud.fr
Elie.Maza@math.ups-tlse.fr

Abstract: We observe a large number of functions differing from each other only by a translation parameter. While the main pattern is unknown, we propose to estimate the shift parameters using M-estimators. Fourier series enables to transform this statistical problem into a semi-parametric framework. We study the convergence of the estimator and provide its asymptotic behavior. Moreover, we use the method in the applied case of velocity curve forecasting.

4.1 Introduction

More and more often, the outcome of an experiment is not a random variable, but a noisy function. This leads to a sample of curves (see for examples [RS91] or [KG92]). Studying this data might involve separating the functions into different homogeneous groups. Indeed, when observing functional data, we may usually see a few typical patterns which are representative of a mass behavior. So, the observations differ slightly from these archetypes. The individuals usually experience similar events, which are explained by a pattern, but the starting time of the event occurs sooner or later. Classification methods, like repeated measures ANOVA or principal component analysis of curves (see for instance [Rao58] or [Sil95]) ignore this type of variability. Hence, computing a representative curve for each group, severely distorts the analysis of the data. Indeed, the average curve (usually the mean or the median) oversmooths the studied phenomenon, and is not a good description of the reality. Some work has been done to find a representative of a large sample of close enough functions (see for examples [RS91] or [KG92]). An homogeneous group of functions may be modeled as follows. We observe, for each curve $\mathcal{C}_j$, $j = 1, \ldots, J$, at consecutive times t_{ij}, $i = 1, \ldots, n_j$, a noisy measure y_{ij}. We assume that there exists functions $(f_j)_{j=1,\ldots,J}$ defined on a closed interval, say $[0,T]$, such that measures $(y_{ij})_{j=1,\ldots,J,\, i=1,\ldots,n_j}$ are observations from model

$$Y_{ij} = f_j(t_{ij}) + \varepsilon_{ij}, j = 1, \ldots, J, i = 1, \ldots, n_j, \tag{4.1}$$

where $(\varepsilon_{ij})_{j=1,\ldots,J,\, i=1,\ldots,n_j}$ are i.i.d. random variables, representing the observation noise. So, functions $(f_j)_{j=1,\ldots,J}$ are close from each other in the sense that there exists an unknown archetype f and unknown warping functions $(h_j)_{j=1,\ldots,J}$ such that

$$\forall j \in \{1, \ldots, J\}, \forall t \in [0,T], f_j(t) = f \circ h_j(t).$$

Examples of such data might be growth curves, longitudinal data in medicine, speech signals, traffic data or expenditure curves for some goods in the econometric domain. Our main motivation is the analysis of the vehicle speed evolution on a motorway. The data are curves, describing the evolution, on observation cells, of the daily vehicle speed. After performing classification procedures (see for instance [LML02] for a complete study), we obtain clusters of functions where in each subgroup appears a typical common behavior. Indeed, all the curves can be deduced one from another by a shift parameter.

This kind of issue led several statisticians to apply transformation to functions in order to get rid of the shifts and to align the curves. If a parametric model would be available a priori, the analysis would be made easier. But, if the data are numerous, there is not generally enough knowledge to build such a model. Thus, they turn into a non parametric framework. When the pattern is known, the problem turns to align a noisy observation with a fixed feature. Piccioni, Scarlatti and Trouvé in [PST98], Kneip, Li, MacGibbon and Ramsay in [KLMR00], or Ramsay and Li in [RL98] proposed curve registration methods. Their main idea is to align each curve on a target curve f_0, which means finding, for all $j \in \{1, \ldots, J\}$, the warping function

h_j minimizing

$$F_\lambda(f_0,f_j;h_j) = \int \|f_j \circ h_j(t) - f_0(t)\|^2 \,\mathrm{d}t + \lambda \int w_j^2(t)\mathrm{d}t,$$

where h_j belongs to a particular smooth monotone family defined by the solution of the differential equation $D^2h_j = w_j Dh_j$. Hence, w_j is simply D^2h_j/Dh_j, the relative curvature of h_j. Thus, penalizing w_j yields both smoothness and monotonicity of h_j (see [RL98] for more details). The main drawback of such methods is that they assume that the archetype f_0 is known, which is a reasonable assumption in pattern recognition, but which is unrealistic when the observed phenomenon is not well known as in our study. Alternatively, in a non parametric point of view, the pattern is replaced by its estimate. In this case, the issue is a matter of synchronizing sample curves. Wang and Gasser in [WG97] and [WG99] use kernel estimators, as Boularan, Ferré and Vieu in [BFV94] or Núñez-Antón, Rodríguez-Póo and Vieu in [NARPV99]. In another work, Gasser and Kneip, in [GK95], align the curves by aligning the local extrema of the functions, which are estimated as zeroes of the non parametric estimate of the derivative. In all cases, the issue of estimating the shifts is blurred by the estimation of the curves, which leads to non parametric rates of convergence. Hence, it seems natural to study the problem in a semi-parametric framework: the shifts are the parameters of interest to be estimated, while the pattern stands for an unknown nuisance functional parameter.

In this paper, we will assume that the observations can then be written as a regression model, where we observe at discrete times t_{ij}, $i = 1,\ldots,n_j$, an unknown function f translated by a parameter $\theta_j^* \in \mathbb{R}$, and this for all $j = 1,\ldots,J$. Further, an observation noise ε_{ij} is added:

$$Y_{ij} = f\left(t_{ij} - \theta_j^*\right) + \varepsilon_{ij}. \tag{4.2}$$

The difficulty of the work is that the estimation must not rely on the pattern, even if the quantities are deeply linked. That is the reason why we will use an M-estimator built on the Fourier series of the data. In this paper, we will consider the estimation problem (4.2) in the frequency domain (Fourier series). Under identifiability assumptions, we provide a consistent method to estimate $\left(\theta_j^*\right)_{j=1,\ldots,J}$ when f is unknown. Further, we show that fluctuations of our estimates are asymptotically Gaussian. This study can be linked first with the study of Golubev in [Gol88], dealing with the semi-parametric efficiency in the estimation of shifts in a continuous observation scheme, and also with the study of Gassiat and Lévy-Leduc in [GLL], dealing with the estimation of the periodicity of a signal.

The present paper falls into seven parts. Section 4.2 is devoted to the definition of the model and to the description of our estimation method. In Section 4.3, we provide asymptotic properties of the estimators. As a matter of fact, we show that the estimators are convergent and asymptotically normal. The estimating method is effectively performed in Section 4.4, on some simulated data, and then used to analyze road traffic data. The technical lemmas and the proofs are gathered in Section 4.5 and Section 4.6. All figures are postponed to Section 4.7.

4.2 Semi-parametric estimation of shifts

4.2.1 Model

For every curve $\mathcal{C}_j, j = 1, \ldots, J$, we get n observations $y_{ij}, i = 1, \ldots, n$, measured at equispaced times $t_i = \frac{i-1}{n}T \in [0, T[$, with $T \in \mathbb{R}_+^*$. We model these observations in the following way:

$$Y_{ij} = f(t_i - \theta_j^*) + \varepsilon_{ij}, j = 1, \ldots, J, i = 1, \ldots, n, \tag{4.3}$$

where

- $f : \mathbb{R} \to \mathbb{R}$ is an unknown T-periodic function,
- $\theta^* = \left(\theta_j^*\right)_{j=1,\ldots,J} \in \mathbb{R}^J$ is an unknown shift parameter, θ_j^* is the shift of the j-th curve,
- for all $j \in \{1, \ldots, J\}$, $(\varepsilon_{ij})_{i=1,\ldots,n}$ is a Gaussian white noise, with variance 1. For sake of simplicity, we get an unitary variance, but all our results are still valid for a general variance.

Our aim is to estimate the translation factors $\theta_j^*, j = 1, \ldots, J$, without the knowledge of the pattern f. Due to the special structure of the model, Fourier analysis is well suited to conduct such a study. Indeed, the Fourier basis diagonalizes any translation. Then, using a Discrete Fourier Transform (see [BtMvdBvdV03] for more details), and with the following condition:

$$\forall j = 1, \ldots, J, \theta_j^* = \frac{T}{n} k_j^*, k_j^* \in \mathbb{Z}, \tag{4.4}$$

we may transform the model (4.3) into the following one (supposing n is odd):

$$d_{jl} = e^{-il\alpha_j^*} c_l^n(f) + w_{jl}, j = 1, \ldots, J, l = -(n-1)/2, \ldots, (n-1)/2, \tag{4.5}$$

where

- $(c_l^n(f))_{l=-(n-1)/2,\ldots,(n-1)/2}$ are the discrete Fourier coefficients of f:

$$c_l^n(f) = \frac{1}{n} \sum_{m=1}^{n} f(t_m) e^{-i2\pi \frac{ml}{n}}, \forall l = -(n-1)/2, \ldots, (n-1)/2,$$

- $\alpha_j^* = \frac{2\pi}{n} k_j^* \in \frac{2\pi}{n}\mathbb{Z}, j = 1, \ldots, J$, are the phase factors,
- for all $j \in \{1, \ldots, J\}$, $(w_{jl})_{l=-(n-1)/2,\ldots,(n-1)/2}$ is a complex Gaussian white noise, with variance $\frac{1}{n}$, and with independent real and imaginary parts.

Like previously, our goal is to estimate the phase factors $\alpha_j^*, j = 1, \ldots, J$, without the knowledge of the discrete Fourier coefficients of function f. In this paper, we will consider the simplified version of the model (4.5):

$$d_{jl} = e^{-il\alpha_j^*} c_l(f) + w_{jl}, j = 1, \ldots, J, l = -(n-1)/2, \ldots, (n-1)/2, \tag{4.6}$$

where

- $(c_l(f))_{l \in \mathbb{Z}}$ are the Fourier coefficients of f:

$$c_l(f) = \frac{1}{T} \int_0^T f(t) e^{-i2\pi \frac{tl}{T}} \mathrm{d}t, \forall l \in \mathbb{Z},$$

- $\alpha_j^* = \frac{2\pi}{T}\theta_j^* \in \mathbb{R}$, $j = 1,\dots,J$, are the phase factors,
- for all $j \in \{1,\dots,J\}$, $(w_{jl})_{l=-(n-1)/2,\dots,(n-1)/2}$ is a complex Gaussian white noise, with complex variance $\frac{1}{n}$, and with independent real and imaginary parts.

Hence, using the Fourier coefficients of f, $(c_l(f))_{l\in\mathbb{Z}}$, and parameter $\theta^* = \left(\theta_j^*\right)_{j=1,\dots,J} \in \mathbb{R}^J$, instead of the discrete Fourier coefficients of f, $(c_l^n(f))_{l=-(n-1)/2,\dots,(n-1)/2}$, and parameter $\theta^* = \left(\theta_j^*\right)_{j=1,\dots,J} \in \frac{T}{n}\mathbb{Z}^J$, leads to a biased estimation. This drawback is similar to the bias induced by any discretization in regression, which vanishes under some regularity assumptions. This issue is tackled in [DJ94].

We point out that we are facing a semi-parametric model. As a matter of fact, we aim at estimating the parameter $\alpha^* = \left(\alpha_j^*\right)_{j=1,\dots,J}$, which depends on an unknown nuisance functional parameter $(c_l(f))_{l\in\mathbb{Z}}$, the Fourier coefficients of the unknown function f.

4.2.2 Identifiability

We notice that the model (4.6) is not identifiable. Obviously, replacing α^* by

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_J^* \end{pmatrix} + c \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + 2\pi \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix}, c \in \mathbb{R}, \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix} \in \mathbb{Z}^J, \tag{4.7}$$

and replacing $f(\cdot)$ by $f(\cdot - c)$, let invariant the equation (4.6). Thus, we must restrict our parameter space. Hence, from now, we assume that the parameter space A satisfies:

$$\begin{aligned} &\text{i)} \quad A \text{ is compact,} \\ &\text{ii)} \quad \alpha^* \in A, \\ &\text{iii)} \quad \text{if } \alpha \in A \text{ and (4.7) holds for } \alpha, \text{ then } \alpha = \alpha^*. \end{aligned} \tag{4.8}$$

For example, the two following parameter spaces are suitable:

- $A_1 = \left\{\alpha \in [-\pi,\pi[^J \colon \alpha_1 = 0\right\}$. Here, we have that $c = 0$, hence $\alpha = \alpha^*$.
- $A_2 = \left\{\alpha \in [-\pi,\pi[^J \colon \sum_{j=1}^J \alpha_j = 0 \text{ and } \alpha_1 \in \left[0,\frac{2\pi}{J}\right[\right\}$. Here, with the condition $\sum_{j=1}^J \alpha_j = 0$, we have that $c = -\frac{2\pi}{J}\sum_{j=1}^J k_j$. Hence, with equation (4.7), we can write that

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_J^* \end{pmatrix} + \frac{2\pi}{J} \begin{pmatrix} (J-1) & -1 & \cdots & -1 \\ -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -1 \\ -1 & \cdots & -1 & (J-1) \end{pmatrix} \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix}.$$

Hence, we get J different solutions in $[-\pi,\pi[^J \subset \mathbb{R}^J$, and a unique solution with the additional condition $\alpha_1 \in \left[0,\frac{2\pi}{J}\right[$.

4.2.3 Estimation

Since we want to assume only weak regularity conditions on the function f, we will consider a method relying on an M-estimation procedure that does not need any knowledge of the function f. For this, define, for any $\alpha = (\alpha_j)_{j=1,\ldots,J} \in A$, the rephased coefficients

$$\tilde{c}_{jl}(\alpha) = e^{il\alpha_j} d_{jl}, j = 1, \ldots, J, l = -(n-1)/2, \ldots, (n-1)/2,$$

and the mean of these rephased Fourier coefficients

$$\hat{c}_l(\alpha) = \frac{1}{J} \sum_{j=1}^{J} \tilde{c}_{jl}(\alpha), l = -(n-1)/2, \ldots, (n-1)/2.$$

Obviously, we have that

$$\tilde{c}_{jl}(\alpha^*) = c_l(f) + e^{il\alpha_j^*} w_{jl},$$

for all $j \in \{1, \ldots, J\}$, and

$$\hat{c}_l(\alpha^*) = c_l(f) + \frac{1}{J} \sum_{j=1}^{J} e^{il\alpha_j^*} w_{jl}.$$

Hence, $|\tilde{c}_{jl}(\alpha) - \hat{c}_l(\alpha)|^2$ should be small when α is close to α^*. As a consequence, considering a sequence $(\delta_l)_{l\in\mathbb{Z}}$ such that

$$\sum_{l\in\mathbb{Z}} \delta_l^2 < +\infty, \tag{4.9}$$

we define the following criterion function:

$$M_n(\alpha) = \frac{1}{J} \sum_{j=1}^{J} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \left|\tilde{c}_{jl}(\alpha) - \hat{c}_l(\alpha)\right|^2. \tag{4.10}$$

This random function is obviously positive. Furthermore, its minimum value should be reached close to the true parameter α^*. As a matter of fact, the following theorem gives the consistency of our M-estimator, defined by

$$\hat{\alpha}_n = \arg\min_{\alpha\in A} M_n(\alpha).$$

Theorem 4.1 *Under the following assumptions on f and on the weight sequence $(\delta_l)_{l\in\mathbb{Z}}$:*

$$\sum_{l\in\mathbb{Z}} |c_l(f)|^2 < +\infty, \tag{4.11}$$

$$\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 < +\infty, \tag{4.12}$$

we have that

$$\hat{\alpha}_n \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \alpha^*.$$

Proof 4.2 (Proof of Theorem 4.1) *The proof of this theorem follows the classical guidelines of the convergence of M-estimators (see for example [GCP93]). Indeed, it suffices to show that the following conditions hold for the criterion function to ensure consistency of $\hat{\alpha}_n$:*

i) Convergence to a contrast function:

$$M_n(\alpha) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} K(\alpha), \tag{4.13}$$

where $K(\cdot)$ has a unique minimum at α^*.

ii) Set the modulus of uniformly continuity

$$W(n,\eta) = \sup_{\|\alpha-\beta\|\leq\eta} |M_n(\alpha) - M_n(\beta)| .$$

There exists two sequences $(\eta_k)_{k\in\mathbb{N}}$ and $(\epsilon_k)_{k\in\mathbb{N}}$, decreasing to zero, such that for a large enough k, we have

$$\lim_{n\to+\infty} \mathbf{P}_{\alpha^*} \left(W(n,\eta_k) > \epsilon_k\right) = 0. \tag{4.14}$$

These two conditions are fulfilled, as it is proved in Section 4.6.

■

Remark 4.3 *The extra term $(\delta_l)_{l\in\mathbb{Z}}$ used in the definition* (4.10) *enables to smooth the criterion function $M_n(\alpha)$. Indeed, the criterion function must not only converge punctually towards the contrast function (Condition i)), but also uniformly (Condition ii)). Without this smoothing effect, $M_n(\alpha)$ behaves in a too irregular way, and the estimator does not converge. As a matter of fact, without regularization, i.e. $(\delta_l)_{l\in\mathbb{Z}} \equiv 1$, the criterion function converges to a non deterministic process.*

Remark 4.4 *Some links could be established between the estimator we consider and a Bayesian maximum likelihood estimator, where the weights $(\delta_l)_{l\in\mathbb{Z}}$ stand for a particular choice of a prior over the unknown function f. This Bayesian point of view is tackled in [DGT].*

4.3 Asymptotic normality

In this section, we prove that the estimator built in the previous section is asymptotically normal, and we give its asymptotic covariance matrix. In general, the asymptotic covariance matrix hardly depends on the geometric structure of the parameter space A. So, for sake of simplicity, we study the asymptotic normality for the parameter space A_1 defined by

$$A_1 = \left\{\alpha \in [-\pi,\pi[^J;\ \alpha_1 = 0\right\} .$$

Hence, our parameter space has dimension $(J-1)$, and we rewrite this as

$$\tilde{A}_1 = [-\pi,\pi[^{J-1},$$

and, any element in $\tilde{A}_1$ as $\alpha = (\alpha_j)_{j=2,\dots,J}$. Also, for sake of simplicity, in this section and in the proofs of Theorem 4.5, we will write $M_n(\alpha)$ instead of $M_n \begin{pmatrix} 0 \\ \alpha \end{pmatrix}$. So, we consider any estimator defined by

$$\hat{\alpha}_n = \arg\min_{\alpha \in \tilde{A}_1} M_n(\alpha).$$

Theorem 4.5 *Under the assumptions of Theorem 4.1, and the following additional assumption on the weight sequence $(\delta_l)_{l\in\mathbb{Z}}$:*

$$\sum_{l\in\mathbb{Z}} \delta_l^4 l^4 < +\infty, \tag{4.15}$$

we get that

$$\sqrt{n}(\hat{\alpha}_n - \alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}(0,\Gamma), \tag{4.16}$$

with

$$\Gamma = \frac{\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2}{\left(\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2\right)^2} \left(I_{J-1} + U_{J-1}\right),$$

where, I_{J-1} is the identity matrix of dimension $(J-1)$ and U_{J-1} is the square matrix of dimension $(J-1)$ whose all entries are equal to one.

Remark 4.6 *Obviously, if the white noise in the model (4.3) has a variance equal to σ^2, then the limit distribution in the previous theorem has a covariance matrix equal to $\sigma^2\Gamma$.*

The $(J-1)\times(J-1)$ matrix $I_{J-1} + U_{J-1}$ is a positive definite matrix. Indeed, its eigenvalues are positive: J of multiplicity 1 and 1 of multiplicity $J-2$. We notice that $\frac{1}{J-1}U_{J-1}$ is the projection matrix on subspace

$$\left\{(x_i)_{i=1,\dots,J-1} \in \mathbb{R}^{J-1};\ x_1 = \dots = x_{J-1}\right\}.$$

Proof 4.7 (Proof of Theorem 4.5) *Recall that the M-estimator is defined as the minimum of the criterion function $M_n(\alpha)$. Hence, we get*

$$\nabla M_n(\hat{\alpha}_n) = 0,$$

where ∇ is the gradient operator. A second order decomposition leads to: there exists $\bar{\alpha}_n$ in a neighborhood of α^* such that

$$\sqrt{n}\left(\hat{\alpha}_n - \alpha^*\right) = -\left[\nabla^2 M_n\left(\bar{\alpha}_n\right)\right]^{-1} \sqrt{n}\nabla M_n\left(\alpha^*\right), \tag{4.17}$$

where ∇^2 is the Hessian operator. Now, using the two asymptotic results from Proposition 4.9 and from Proposition 4.10, we get

$$\sqrt{n}\nabla M_n(\alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}(0,\Gamma_0),$$

$$\left[\nabla^2 M_n\left(\bar{\alpha}_n\right)\right]^{-1} \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} V,$$

where V is a non negative symmetric matrix of dimension $(J-1)$. Hence, if we set $\Gamma = V'\Gamma_0 V$, the result of Theorem 4.5 follows easily.

■

The problem of choosing the weights $(\delta_l)_{l\in\mathbb{Z}}$ in the definition of the criterion function is crucial (see Remark 4.3). Since we work with L^2 functions, the assumption (4.15) is satisfied as soon as

$$|\delta_l| = O\left(|l|^{-5/4-\nu}\right), \forall \nu > 0.$$

When looking at the asymptotic variance, we see that there is a trade-off which leads to a lower bound for the smoothing sequence. Hence, the optimal choice for $(\delta_l)_{l\in\mathbb{Z}}$ should be given by semi-parametric efficiency. Using Cauchy-Schwarz's inequality, we get that

$$\frac{\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2}{\left(\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2\right)^2} \geq \left(\sum_{l\in\mathbb{Z}} l^2 |c_l(f)|^2\right)^{-1}.$$

This case, corresponding to the least favorable case in the semi-parametric efficiency framework, is obtained for the optimal choice of coefficients $(\delta_l)_{l\in\mathbb{Z}} \equiv 1$. In this case, we would obtain the following asymptotic behavior:

$$\sqrt{n}\sqrt{\sum_{l\in\mathbb{Z}} l^2 |c_l(f)|^2}\,(\hat{\alpha}_n - \alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}\left(0, I_{J-1} + U_{J-1}\right).$$

Nevertheless, the choice of the weight sequence $(\delta_l)_{l\in\mathbb{Z}} \equiv 1$ yields to a non convergent estimator, see Remark 4.3. Non optimality as regards asymptotic efficiency is the price to pay both to deal with a discretized version of the regression model and to handle simultaneous estimation for all the observed curves. Maybe, a different way of estimation could get rid of this drawback but the work is still in progress.

Remark 4.8 *Throughout all the work, we assume that the observation noise in the model* (4.3) *is Gaussian. Nevertheless, we could get rid of this assumption with moment conditions for the errors.*

4.4 Applications and simulations

In this section, we present some numerical applications of the method. The first one, in Section 4.4.1, gives results on simulated data. The second one, in Section 4.4.2, is based on an experiment on human fingers force. The last one, in Section 4.4.3, is carried out with traffic data. All figures are postponed to Section 4.7.

The optimization algorithm used in any resolution is based on a Krylov method (the conjugate gradient method). Indeed, minimizing an L^2 criterion function with a conjugate gradient algorithm yields a reduced step number, and hence, a small complexity.

4.4.1 Simulations

Simulated data are carried out as follows:

$$y_{ij} = f(t_i - \theta_j^*) + \varepsilon_{ij}, j = 1, \ldots, J, i = 1, \ldots, n,$$

with the following choice of parameters:

- $J = 10$,
- $n = 100$,
- for all $j \in \{1, \ldots, J\}$, values $\left(t_i = -\pi + \frac{i-1}{n} 2\pi\right)_{i=1,\ldots,n}$ are equally spaced points on $[-\pi, \pi[$,
- $f(t) = 15 \sin(4t)/(4t)$,
- $\left(\theta_j^*\right)_{j=2,\ldots,J}$ are simulated with a uniform law on $[-\pi/4, \pi/4]$ and $\theta_1^* = 0$,
- for all $j \in \{1, \ldots, J\}$, for all $i \in \{1, \ldots, n\}$, values ε_{ij} are simulated from a Gaussian law with mean 0 and standard deviation 1.

Results are given on Figure 4.1. The target function f is considered as a 2π-periodic function ($T = 2\pi$), hence $\alpha_j^* = \theta_j^*$. The function f is plotted by a solid line in figure 4.1(d). The figure 4.1(a) shows simulated data $(y_{ij})_{j=1,\ldots,J,\, i=1,\ldots,n}$. The mean curve of these data is given on Figure 4.1(d) by the dotted line. We can see that the mean function is representativeness of data. Indeed, the amplitude of higher optimum is reduced and smallest ones disappeared. Figure 4.1(c) shows shifted curves. The mean function of these shifted curves is given on figure 4.1(d) by dashed line. Figure 4.1(b) plots $\left(\alpha_j^*\right)_{j=1,\ldots,J}$ on abscissa axis against $(\hat{\alpha}_j)_{j=1,\ldots,J}$ on ordinate axis. Estimations are very close to truth parameters. Comparison between mean curves, before and after the shift estimation, is straightforward.

A second simulated data set is carried out with $J = 3$, $\theta_1^* = 0$, $\theta_2^* = 1.47$ and $\theta_3^* = 0.24$. The figure 4.2 presents the criterion function $M_n(\cdot)$, with an unique minimum. The figure 4.3 shows shifted curves.

4.4.2 Pinch force data

Data presented here are extracted from an experiment described in [RWF95], and studied in [RL98] with a Curve Registration methodology. Data represent the force exerted by the thumb and forefinger on a force meter during 20 brief pinches. These 20 force measurements having arbitrary beginning, Ramsay and Li in [RL98] begin their study by a landmark alignment of curve maxima (with single shifts). These aligned data are shown in figure 4.4(a).

Our propose is to study these data with the shift estimation methodology. Shift estimations and shifted curves are respectively shown in figure 4.4(b) and figure 4.4(c). We note that shift parameters are almost all close to zero. That means that, in this case, landmark alignment shift quite well the data. Hence, in figure 4.4(d), the mean curves of shifted curve (solid line) and of primary curves (dotted line) are almost the same ones.

But, initially, data were shifted arbitrarily. We simulate such an initial data by shifting arbitrarily the pinch force curves. More precisely, we set $\alpha_1^* = 0$ and we simulate $\left(\alpha_j^*\right)_{j=2,\ldots,20}$ with a uniform law on $[-\pi/5, \pi/5]$, and we shift the j-th curve by α_j^*. Pseudo-simulated data are shown in figure 4.5(a). Hence, with shift estimation methodology, we obtain the shift parameter estimates and the shifted

curves shown respectively in figure 4.5(b) and figure 4.5(c). Figure 4.5(d) shows the obtained mean curve with primary curves (dotted line) and with shifted curves (solid line). The shape of each individual curve is correctly estimated with shifted curves mean.

4.4.3 Application to road traffic forecasting

Most of the Parisian road traffic network is equipped with a traffic road measurement infrastructure. The main elements of this infrastructure are counting stations. These sensors are situated approximately every 500 meters on main trunk roads (motorways and speedways principally). Every counting station measures, daily, the average speed of vehicle flow on 6 minutes periods. We consider measurements from 5 AM to 11 PM, hence, the length of the daily measurement is 180. We note y_{ij} the speed measurement of day $j \in \{1, \ldots, J\}$ and of period $i \in \{1, \ldots, n\}$, with $n = 180$.

Our purpose is to improve, with shift estimation, an existing forecasting methodology. This forecasting methodology is described in [LML02]. This method is based on a classification method. We dispose of a sample of J speed curves and we want to summarize it by a small number N ($N \ll J$) of standard profiles, representatives of each cluster.

Our study is stimulated by the observation of several clusters of these J speed curves. Indeed, we note frequently that many subgroups are composed by curves describing the same behavior. For example, we observe a speed curve subgroup with a same traffic jam or speed reduction, but with a different start time of the phenomenon for each curve. Thus, figure 4.6(a) represent a particular cluster on a particular counting station. Figure 4.6(b) shows estimated shifts. Shifted curves are plotted on figure 4.6(c). So, in this homogeneous cluster, where only a shift phenomenon appear, the mean curves in figure 4.6(d) of shifted curves (solid line) and of primary curves (dotted line) aren't the same. The shift estimated mean is clearly more representative of individual pattern.

4.5 Technical Lemmas

The two following propositions, Proposition 4.9 and Proposition 4.10, are used in the proof of asymptotic normality (Theorem 4.5). Their proofs are postponed to the appendix.

Proposition 4.9 *Assume that the assumptions of Theorem 4.1 are fulfilled. Then*

$$\sqrt{n}\nabla M_n(\alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}(0,\Gamma), \tag{4.18}$$

where the variance matrix is given by

$$\Gamma = \frac{4}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 \left(I_{J-1} - \frac{1}{J}U_{J-1}\right).$$

Proposition 4.10 *Assume that the assumptions of Theorem 4.1 are fulfilled. Further, assume that* (4.15) *holds. Then, for any sequence* $(\bar{\alpha}_n)_{n\in\mathbb{N}}$ *with* $\|\bar{\alpha}_n - \alpha^*\| \leq \|\hat{\alpha}_n - \alpha^*\|$, *we have*

$$\nabla^2 M_n(\bar{\alpha}_n) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{2}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2 \left(JI_{J-1} - U_{J-1}\right). \tag{4.19}$$

4.6 Appendix

Let z be a complex number and $\bar{z}$ its conjugate. We write $\mathfrak{Re}(z) = \frac{1}{2}(z+\bar{z})$ (the real part of z) and $\mathfrak{Im}(z) = \frac{1}{2i}(z-\bar{z})$ (the imaginary part of z).

Proof 4.11 (Proof of Condition (4.13)) *The criterion function* $M_n(\alpha)$ *can be written as follows:*

$$\begin{aligned}
M_n(\alpha) &= \frac{1}{J}\sum_{j=1}^{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |\tilde{c}_{jl}(\alpha) - \hat{c}_l(\alpha)|^2 \\
&= \frac{1}{J}\sum_{j=1}^{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \left(|\tilde{c}_{jl}(\alpha)|^2 + |\hat{c}_l(\alpha)|^2 - 2\mathfrak{Re}\left(\overline{\tilde{c}_{jl}(\alpha)}\hat{c}_l(\alpha)\right)\right) \\
&= \frac{1}{J}\sum_{j=1}^{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \left(|\tilde{c}_{jl}(\alpha)|^2 - |\hat{c}_l(\alpha)|^2\right) \\
&= (I) + (II).
\end{aligned}$$

The first term gives:

$$\begin{aligned}
(I) &= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)|^2 + \frac{1}{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} |w_{jl}|^2 + \frac{2}{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \mathfrak{Re}\left(\overline{c_l(f)} e^{il\alpha_j^*} w_{jl}\right) \\
&= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)|^2 + \frac{1}{nJ}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \left(\xi_{jl}^{x\,2} + \xi_{jl}^{y\,2}\right) \\
&+ \frac{2}{\sqrt{n}J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \left[\cos(\theta_l - l\alpha_j^*)\xi_{jl}^x + \sin(\theta_l - l\alpha_j^*)\xi_{jl}^y\right],
\end{aligned}$$

where we have used the following notation:

$$\forall j = 1,\ldots,J, \forall l = -(n-1)/2,\ldots,(n-1)/2, w_{jl} = \frac{1}{\sqrt{n}}\xi_{jl} = \frac{1}{\sqrt{n}}\left(\xi_{jl}^x + i\xi_{jl}^y\right).$$

Here, $\left(\xi_{jl}^x\right)_{l=-(n-1)/2,\ldots,(n-1)/2}$ *and* $\left(\xi_{jl}^y\right)_{l=-(n-1)/2,\ldots,(n-1)/2}$ *are independent Gaussian sequences with law* $\mathcal{N}_n(0,I_n)$. *Also, we have seted*

$$\forall l = -(n-1)/2,\ldots,(n-1)/2, c_l(f) = |c_l(f)|e^{i\theta_l}, \text{ with } \theta_l \in [0,2\pi[.$$

The second term has the following form:

$$(II) = -\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\alpha_j - \alpha_j^*)} \delta_l c_l(f) \right|^2 - \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{\delta_l}{J} \sum_{j=1}^{J} e^{il\alpha_j} w_{jl} \right|^2$$
$$- \frac{2}{J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{j=1}^{J} \sum_{k=1}^{J} \mathfrak{Re} \left(\delta_l^2 e^{il(\alpha_j - \alpha_j^*)} e^{-il\alpha_k} c_l(f) \overline{w_{kl}} \right).$$

As a consequence, after some calculations, we obtain that:

$$M_n(\alpha) = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)|^2 - \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\alpha_j - \alpha_j^*)} \delta_l c_l(f) \right|^2$$
$$+ \frac{1}{nJ} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \left({\xi_{jl}^x}^2 + {\xi_{jl}^y}^2 \right) - \frac{1}{n} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \frac{\delta_l^2}{J^2} \left| \sum_{j=1}^{J} e^{il\alpha_j} \xi_{jl} \right|^2$$
$$+ \frac{2}{\sqrt{n}J} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \left[\cos(\theta_l - l\alpha_j^*) \xi_{jl}^x + \sin(\theta_l - l\alpha_j^*) \xi_{jl}^y \right]$$
$$- \frac{2}{\sqrt{n}J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \sum_{k=1}^{J} \mathfrak{Re} \left(e^{il(\alpha_j - \alpha_j^*)} e^{-il\alpha_k} e^{i\theta_l} \overline{\xi_{kl}} \right).$$

After expanding the terms, we finally get the following expression:

$$M_n(\alpha) = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)|^2 - \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\alpha_j - \alpha_j^*)} \delta_l c_l(f) \right|^2 \tag{4.20}$$
$$+ \frac{J-1}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \left({\xi_{jl}^x}^2 + {\xi_{jl}^y}^2 \right) \tag{4.21}$$
$$- \frac{2}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \sum_{k>j} \left[\cos(l(\alpha_j - \alpha_k)) (\xi_{jl}^x \xi_{kl}^x + \xi_{jl}^y \xi_{kl}^y) + \right.$$
$$\left. \sin(l(\alpha_j - \alpha_k)) (\xi_{jl}^x \xi_{kl}^y - \xi_{jl}^y \xi_{kl}^x) \right] \tag{4.22}$$
$$+ \frac{2(J-1)}{\sqrt{n}J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \left[\cos(\theta_l - l\alpha_j^*) \xi_{jl}^x + \sin(\theta_l - l\alpha_j^*) \xi_{jl}^y \right] \tag{4.23}$$
$$- \frac{2}{\sqrt{n}J^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| \sum_{j=1}^{J} \sum_{k \neq j} \left[\cos(l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l) \xi_{kl}^x + \right.$$
$$\left. \sin(l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l) \xi_{kl}^y \right]. \tag{4.24}$$

So we have split the criterion function into five terms: (4.20), (4.21), (4.22), (4.23) *and* (4.24). *Their asymptotic behaviors are as follows:*

- *The term* (4.20) *is a deterministic one. Let* ψ *the* T*-periodic function defined by the Fourier serie* $\psi(t) = \sum_{l\in\mathbb{Z}} \delta_l e^{2\pi i l t/T}$. *The function* ψ *is well defined in* $L^2([0,T])$ *since* (4.9) *holds. Hence, using Parseval theorem, we have that*

$$(4.20) \xrightarrow[n\to+\infty]{} \int_0^T |(\psi * f)(t)|^2 \frac{dt}{T} - \int_0^T \left| \frac{1}{J} \sum_{j=1}^{J} (\psi * f)(t + \theta_j - \theta_j^*) \right|^2 \frac{dt}{T}.$$

- *The term* (4.21) *is a pure noise term:*

$$\begin{aligned}(4.21) &= \frac{J-1}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \left({\xi_{jl}^x}^2 + {\xi_{kl}^y}^2 \right) \\ &= \frac{J-1}{nJ^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \chi_{2J,l}^2,\end{aligned}$$

where $\left(\chi_{2J,l}^2\right)_{l=-(n-1)/2,\dots,(n-1)/2}$ *is a sequence of i.i.d. chisquared random variables with* $2J$ *degrees of freedom. So that,* (4.21) *will vanish asymptotically in Probability. Indeed,*

$$\mathbb{E}((4.21)) = \frac{2(J-1)}{nJ} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \xrightarrow[n\to+\infty]{} 0,$$

and

$$\mathrm{Var}((4.21)) = \frac{4(J-1)^2}{n^2 J^3} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 \xrightarrow[n\to+\infty]{} 0.$$

- *The term* (4.22) *is also a pure noise term composed of terms of the type:*

$$U_n = \frac{1}{n} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 \sum_{j=1}^{J} \sum_{k>j} \cos\left(l(\alpha_j - \alpha_k)\right) \xi_{jl}^x \xi_{kl}^y,$$

with $\mathbb{E}(U_n) = 0$, *and*

$$\mathrm{Var}(U_n) = \frac{1}{n^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 \sum_{j=1}^{J} \sum_{k>j} \cos^2\left(l(\alpha_j - \alpha_k)\right) \leq \frac{J^2 - J}{2n^2} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 \xrightarrow[n\to+\infty]{} 0.$$

Hence, the Bienaymé-Tchebychev inequality yields that

$$\forall \epsilon > 0, \mathbf{P}_{\alpha^*}(|U_n| > \epsilon) \leq \frac{Var(U_n)}{\epsilon^2} \xrightarrow[n\to+\infty]{} 0.$$

So that

$$(4.22) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0.$$

- *The terms* (4.23) *and* (4.24) *have the same asymptotic behavior as*

$$V_n = \frac{1}{\sqrt{n}} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |c_l(f)| W_l,$$

with

$$W_l = \sum_{j=1}^{J} \sum_{k \neq j} \cos(l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l) \xi_{kl}^x,$$

where, as before, the random sequences $\left(\xi_{jl}^x\right)_{l=-(n-1)/2,\ldots,(n-1)/2}$ *are i.i.d. and following a Gaussian law* $\mathcal{N}_n(0,I_n)$*. Since* $(c_l(f))_{l\in\mathbb{Z}} \in l^2(\mathbb{Z})$*, we have that*

$$\sum_{l\in\mathbb{Z}} \delta_l^4 |c_l(f)|^2 \|W_l\|_2^2 < +\infty.$$

As a result, the random variables $\sqrt{n}V_n$ *converges in* $\mathbf{P}_{\alpha^*}$*-probability. Hence,*

$$V_n \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0.$$

So, we can conclude the convergences of (4.23) *and* (4.24)*:*

$$(4.23) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0 \text{ and } (4.24) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0.$$

In conclusion, we have the criterion function convergence:

$$M_n(\alpha) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} K(\alpha),$$

with

$$K(\alpha) = \int_0^T |(\psi * f)(t)|^2 \frac{\mathrm{d}t}{T} - \int_0^T \left| \frac{1}{J} \sum_{j=1}^{J} (\psi * f)\left(t + \theta_j - \theta_j^*\right) \right|^2 \frac{\mathrm{d}t}{T}.$$

Moreover, Cauchy-Schwartz inequality yields that

$$\begin{aligned} &\int_0^T \left| \frac{1}{J} \sum_{j=1}^{J} (\psi * f)(t + \theta_j - \theta_j^*) \right|^2 \frac{\mathrm{d}t}{T} \\ &\leq \int_0^T \frac{1}{J} \sum_{j=1}^{J} \left|(\psi * f)(t + \theta_j - \theta_j^*)\right|^2 \frac{\mathrm{d}t}{T} = \int_0^T |(\psi * f)(t)|^2 \frac{\mathrm{d}t}{T}, \end{aligned}$$

hence, $K(\cdot) \geq 0$*, and the minimum value is reached for*

$$\int_0^T |(\psi * f)(t)|^2 \frac{\mathrm{d}t}{T} = \int_0^T \left| \frac{1}{J} \sum_{j=1}^{J} (\psi * f)(t + \theta_j - \theta_j^*) \right|^2 \frac{\mathrm{d}t}{T},$$

which is equivalent, using the Parseval theorem to

$$\sum_{l\in\mathbb{Z}} |\delta_l c_l(f)|^2 = \sum_{l\in\mathbb{Z}} \left| \frac{1}{J} \sum_{j=1}^{J} \delta_l c_l(f) e^{il(\alpha_j - \alpha_j^*)} \right|^2 . \tag{4.25}$$

So, we have that

$$\begin{aligned}
(4.25) &\iff \forall l \in \mathbb{Z}, \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\alpha_j - \alpha_j^*)} \right|^2 = 1 \\
&\iff \forall l \in \mathbb{Z}, \forall j = 1, \ldots, J, \forall k = 1, \ldots, J, \cos\left[l \left((\alpha_j - \alpha_j^*) - (\alpha_k - \alpha_k^*) \right) \right] = 1 \\
&\iff \forall j = 1, \ldots, J, \alpha_j = \alpha_j^* + c\,[2\pi], c \in \mathbb{R}.
\end{aligned}$$

In a matrix way, we get the equation (4.7)*, i.e.*

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_J^* \end{pmatrix} + c \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + 2\pi \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix}, c \in \mathbb{R}, \begin{pmatrix} k_1 \\ \vdots \\ k_J \end{pmatrix} \in \mathbb{Z}^J.$$

Hence, since $\alpha \in A$ and A is defined by (4.8)*, we have shown that*

$$\alpha_j = \alpha_j^*, \forall j = 1, \ldots, J.$$

Since $\alpha \mapsto K(\alpha)$ achieves its unique minimum for $\alpha = \alpha^*$, the condition (4.13) *is fulfilled.*

■

Proof 4.12 (Proof of Condition (4.14)) *Set $\alpha = (\alpha_j)_{j=1,\ldots,J}$ and $\beta = (\beta_j)_{j=1,\ldots,J}$ two translation parameters in A. Our aim is to prove that the convergence of the criterion function is uniform, in the sense that for k large enough and two decreasing to zero sequences $(\eta_k)_{k\in\mathbb{N}}$ and $(\epsilon_k)_{k\in\mathbb{N}}$, we have*

$$\lim_{n\to+\infty} \mathbf{P}\left(\sup_{\|\alpha-\beta\|\le\eta_k} |M_n(\alpha) - M_n(\beta)| > \epsilon_k \right) = 0.$$

Using the expression of the criterion function that we have calculate in the proof of Condition (4.13)*, i.e.*

$$M_n(\alpha) = (4.20) + (4.21) + (4.22) + (4.23) + (4.24),$$

we get

$$M_n(\alpha) - M_n(\beta) = A + B + C,$$

where

$$A = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\beta_j - \alpha_j^*)} \delta_l c_l(f) \right|^2 - \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \left| \frac{1}{J} \sum_{j=1}^{J} e^{il(\alpha_j - \alpha_j^*)} \delta_l c_l(f) \right|^2$$

is the deterministic term, and $(B+C)$ *is the stochastic term, with*

$$B = -\frac{2}{nJ^2}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}\sum_{j=1}^{J}\sum_{k>j}\delta_l^2\left[\cos(l(\alpha_j-\alpha_k))-\cos(l(\beta_j-\beta_k))\right](\xi_{jl}^x\xi_{kl}^x+\xi_{jl}^y\xi_{kl}^y)$$
$$-\frac{2}{nJ^2}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}\sum_{j=1}^{J}\sum_{k>j}\delta_l^2\left[\sin(l(\alpha_j-\alpha_k))-\sin(l(\beta_j-\beta_k))\right](\xi_{jl}^x\xi_{kl}^y-\xi_{jl}^y\xi_{kl}^x),$$

and

$$C = -\frac{2}{\sqrt{n}J^2}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}\delta_l^2|c_l(f)|\sum_{j=1}^{J}\sum_{k\neq j}\big[\cos(\theta_l+l(\alpha_j-\alpha_j^*-\alpha_k))-$$
$$\cos(\theta_l+l(\beta_j-\alpha_j^*-\beta_k))\big]\,\xi_{kl}^x$$
$$-\frac{2}{\sqrt{n}J^2}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}\delta_l^2|c_l(f)|\sum_{j=1}^{J}\sum_{k\neq j}\big[\sin(\theta_l+l(\alpha_j-\alpha_j^*-\alpha_k))-$$
$$\sin(\theta_l+l(\beta_j-\alpha_j^*-\beta_k))\big]\,\xi_{kl}^y.$$

As a result, the stochastic term $(B+C)$ *can be split into two categories of terms, composed, for all* $l\in\{-(n-1)/2,\ldots,(n-1)/2\}$*, of the centered independent random variables*

$$U_l(\alpha,\beta) = \delta_l^2\sum_{j=1}^{J}\sum_{k>j}\left[\cos(l(\alpha_j-\alpha_k))-\cos(l(\beta_j-\beta_k))\right]\xi_{jl}^x\xi_{kl}^x,$$
$$V_l(\alpha,\beta) = \delta_l^2|c_l(f)|\sum_{j=1}^{J}\sum_{k\neq j}\big[\cos(\theta_l+l(\alpha_j-\alpha_j^*-\alpha_k))-\cos(\theta_l+l(\beta_j-\alpha_j^*-\beta_k))\big]\,\xi_{kl}^x.$$

All the remaining terms in B *and* C *differ from* U_l *or* V_l *by the trigonometric functions, but their asymptotic behavior can be deduced in a similar way. Hence, our aim is to bound the probabilities*

$$\mathbf{P}\left(\sup_{\|\alpha-\beta\|\leq\eta}\left|\frac{1}{n}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}U_l(\alpha,\beta)\right|>x\right), \tag{4.26}$$
$$\mathbf{P}\left(\sup_{\|\alpha-\beta\|\leq\eta}\left|\frac{1}{\sqrt{n}}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}}V_l(\alpha,\beta)\right|>x\right). \tag{4.27}$$

Using the basic inequality $|\cos(p)-\cos(q)|\leq|p-q|$ *and after some calculations, we*

have that

$$\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} U_l(\alpha,\beta)\right) \leq J^2\eta^2 \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 l^2,$$
$$\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} V_l(\alpha,\beta)\right) \leq J^2\eta^2 \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^4 l^2 |c_l(f)|^2.$$

Moreover, exponential bounded moments of random variables U_l and V_l gives, we get from Bernstein's inequality for independent variables, that there exists a positive constant $M < +\infty$, such that

$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} U_l(\alpha,\beta)\right| > x\right) \leq 2\exp\left(-\frac{n^2x^2}{2\left(\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} U_l(\alpha,\beta)\right) + Mnx\right)}\right),$$
$$\mathbf{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} V_l(\alpha,\beta)\right| > x\right) \leq 2\exp\left(-\frac{nx^2}{2\left(\mathrm{Var}\left(\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} V_l(\alpha,\beta)\right) + M\sqrt{n}x\right)}\right).$$

Hence, using (4.11) *and* (4.12)*, we get that both previous probabilities go to zero when n goes to infinity. Now, since α and β lie in a compact set of $\mathbb{R}^J$ with $J < +\infty$ fixed, a standard chaining argument (see for instance [vdG00]) proves that there exist $(\eta_k)_{k\in\mathbb{N}}$ and $(\epsilon_k)_{k\in\mathbb{N}}$ such that for k large enough we get*

$$\lim_{n\to+\infty} \mathbf{P}\left(\sup_{\|\alpha-\beta\|\leq\eta_k}\left|\frac{1}{n}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} U_l(\alpha,\beta)\right| > \epsilon_k\right) = 0,$$
$$\lim_{n\to+\infty} \mathbf{P}\left(\sup_{\|\alpha-\beta\|\leq\eta_k}\left|\frac{1}{\sqrt{n}}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} V_l(\alpha,\beta)\right| > \epsilon_k\right) = 0.$$

The deterministic part A remains to be studied. We can write, after some calculations, that

$$|A| \leq \eta \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 |l| |c_l(f)|^2. \tag{4.28}$$

But, since (4.11) *and* (4.12) *hold, using Cauchy-Schwartz inequality, we have that*

$$\begin{aligned}\|f\|_\delta^2 &= \sum_{l\in\mathbb{Z}} \delta_l^2 |l| |c_l(f)|^2 \\ &\leq \left(\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2\right)^{\frac{1}{2}} \left(\sum_{l\in\mathbb{Z}} |c_l(f)|^2\right)^{\frac{1}{2}} \\ &< +\infty.\end{aligned}$$

Hence, as soon as we have chosen two decreasing sequences $(\eta_k)_{k\in\mathbb{N}}$ *and* $(\epsilon_k)_{k\in\mathbb{N}}$ *such that* $\epsilon_k/\eta_k > \|f\|_\delta^2$, *we have that*

$$\lim_{n\to+\infty} \mathbf{P}\left[W(n,\eta_k) > \epsilon_k\right] = 0,$$

which concludes the proof.

■

Proof 4.13 (Proof of Proposition 4.9) *After some calculations, we get the following expressions for the first and the second derivatives of the empirical contrast, for all* $k \in \{2,\dots,J\}$, *for all* $m \in \{2,\dots,J\}$:

$$\frac{\partial M_n}{\partial \alpha_k}(\alpha) = \frac{2}{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l \mathfrak{Im}\left(\tilde{c}_{kl}(\alpha)\overline{\hat{c}_l(\alpha)}\right), \tag{4.29}$$

$$\frac{\partial^2 M_n}{\partial \alpha_k^2}(\alpha) = \frac{2}{J^2}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}(\alpha)\sum_{j\neq k}\overline{\tilde{c}_{jl}(\alpha)}\right), \tag{4.30}$$

$$\forall m \neq k,\ \frac{\partial^2 M_n}{\partial \alpha_k \partial \alpha_m}(\alpha) = -\frac{2}{J^2}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}(\alpha)\overline{\tilde{c}_{ml}(\alpha)}\right). \tag{4.31}$$

By straightforward calculations, we get that

$$\sqrt{n}\frac{\partial M_n}{\partial \alpha_k}(\alpha^*) = \frac{2}{J}\sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l\left(|c_l(f)|\left(V_l^k - V_l\right) + W_l^k\right),$$

where, for all $l \in \mathbb{Z}$,

$$W_l^k = \frac{1}{J\sqrt{n}}\sum_{j=1}^{J}\left[\sin(l(\alpha_k^* - \alpha_j^*))(\xi_{kl}^x\xi_{jl}^x + \xi_{kl}^y\xi_{jl}^y) + \cos(l(\alpha_k^* - \alpha_j^*))(\xi_{kl}^y\xi_{jl}^x - \xi_{kl}^x\xi_{jl}^y)\right],$$

$$V_l^k = (\cos(l\alpha_k^* + \theta_l)\xi_{kl}^x - \sin(l\alpha_k^* + \theta_l)\xi_{kl}^y),$$

and

$$V_l = \frac{1}{J}\sum_{j=1}^{J} V_l^j.$$

Let, for $l \in \mathbb{Z}$,

$$Y_l = \left(\xi_{1l}^x\ \xi_{2l}^x \cdots \xi_{Jl}^x\ \xi_{1l}^y\ \xi_{2l}^y \cdots \xi_{Jl}^y\right)',$$

and, let f_l^k *be the vector of length* $2J$, *defined by*

$$\begin{aligned}\left(f_l^k\right)_k &= \cos\left(l\alpha_k^* + \theta_l\right),\\ \left(f_l^k\right)_{J+k} &= -\sin\left(l\alpha_k^* + \theta_l\right),\\ \left(f_l^k\right)_i &= 0 \text{ if } i \notin \{k, J+k\}.\end{aligned}$$

As a consequence, we get the following expression for V_l^k:

$$V_l^k = \langle f_l^k, Y_l \rangle = f_l^{k\prime} Y_l.$$

In a same way, for $l \in \mathbb{Z}$, let $\bar{B}_l^k$ be the $(2J) \times (2J)$ matrix defined by rows by

$$\begin{aligned}
\left(\bar{B}_l^k\right)_k &= (\sin[l(\alpha_k^* - \alpha_1^*)] \cdots \sin[l(\alpha_k^* - \alpha_J^*)] \ -\cos[l(\alpha_k^* - \alpha_1^*)] \cdots -\cos[l(\alpha_k^* - \alpha_J^*)])\,, \\
\left(\bar{B}_l^k\right)_{J+k} &= (\cos[l(\alpha_k^* - \alpha_1^*)] \cdots \cos[l(\alpha_k^* - \alpha_J^*)] \ \sin[l(\alpha_k^* - \alpha_1^*)] \cdots \sin[l(\alpha_k^* - \alpha_J^*)])\,, \\
\left(\bar{B}_l^k\right)_i &= (0 \cdots 0) \ \ \textit{if } i \notin \{k, J+k\}.
\end{aligned}$$

Further, let the symmetric matrix B_l^k be defined by

$$B_l^k = \frac{\bar{B}_l^k + \left(\bar{B}_l^k\right)'}{2}.$$

Hence, we may write

$$W_l^k = \frac{1}{J\sqrt{n}} Y_l' B_l^k Y_l.$$

Now, we define, for $k = 2, \ldots, J$:

$$\begin{aligned}
\tilde{B}_l^k &= \frac{2}{J} B_l^k, \\
\tilde{f}_l^k &= \frac{2}{J} f_l^k.
\end{aligned}$$

Our aim is to study the asymptotic distribution of the gradient $\sqrt{n}\nabla M_n(\alpha^*)$. For this purpose, we consider $u = (u_2, \ldots, u_J)' \in \mathbb{R}^{J-1}$ and $t = (t_2, \ldots, t_J)' \in \mathbb{R}^{J-1}$, and we define the couple of random variables:

$$(R_n, S_n) = \left(\frac{2}{J} \sum_{k=2}^{J} u_k \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l |c_l(f)| (V_l^k - V_l), \frac{2}{J} \sum_{k=2}^{J} t_k \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l W_l^k \right).$$

Using previous notations, we get

$$\begin{aligned}
R_n &= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l |c_l(f)| \langle g_l(u), Y_l \rangle\,, \ \textit{with } g_l(u) = \sum_{k=2}^{J} u_k \left(\tilde{f}_l^k - \frac{1}{J} \sum_{j=1}^{J} \tilde{f}_l^j \right), \\
S_n &= \frac{1}{J\sqrt{n}} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l Y_l' A_l(t) Y_l, \ \textit{with } A_l(t) = \sum_{k=2}^{J} t_k \tilde{B}_l^k.
\end{aligned}$$

The quadratic part is vanishing in probability when n increases. Indeed, there exists a positive constant $C(t)$, depending on t, such that

$$\|A_l(t)\| \leq C(t),$$

and, this leads to an upper bound for the quadratic term:

$$\mathbb{E}\left(|S_n|\right) \leq \frac{C(t)}{J\sqrt{n}}\mathbb{E}\left(\|Y_1\|\right) \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2|l|.$$

Moreover, since (4.15) *holds and using Cauchy-Schwartz inequality, we have that*

$$\sum_{l\in\mathbb{Z}^*} \delta_l^2|l| \leq \left(\sum_{l\in\mathbb{Z}^*} \delta_l^4 l^4\right)^{\frac{1}{2}} \left(\sum_{l\in\mathbb{Z}^*} \frac{1}{l^2}\right)^{\frac{1}{2}} < +\infty.$$

Hence,

$$\mathbb{E}\left(|S_n|\right) \xrightarrow[n\to+\infty]{} 0.$$

So that

$$S_n \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0.$$

For the last term, we have that

$$\langle g_l(u), Y_l\rangle \sim \mathcal{N}\left(0, \|g_l(u)\|_2^2\right),$$

where

$$\|g_l(u)\|_2^2 = \frac{4}{J^2}u'\left(I_{J-1} - \frac{1}{J}U_{J-1}\right)u,$$

with, I_{J-1} the $(J-1)$ identity matrix, and U_{J-1} the $(J-1)\times(J-1)$ matrix which all entries are equal to 1. So, independence of variables $(Y_l)_{l\in\mathbb{Z}}$ yields that

$$R_n \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{4}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 u'\left(I_{J-1} - \frac{1}{J}U_{J-1}\right)u\right).$$

Hence, we have proved that

$$\sqrt{n}\nabla M_n(\alpha^*) \xrightarrow[n\to+\infty]{\mathcal{D}} \mathcal{N}_{J-1}\left(0, \frac{4}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^4 l^2 |c_l(f)|^2 \left(I_{J-1} - \frac{1}{J}U_{J-1}\right)\right).$$

■

Proof 4.14 (Proof of Proposition 4.10) *First, we pay attention to the non diagonal terms of the matrix of the second derivatives. For $m \neq k$, we get after some*

calculations:

$$
\begin{aligned}
-\frac{J^2}{2}\frac{\partial^2 M_n}{\partial\alpha_k\partial\alpha_m}(\alpha) &= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}(\alpha)\overline{\tilde{c}_{ml}(\alpha)}\right) \\
&= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(e^{il(\alpha_k-\alpha_m)} d_{kl}\overline{d_{ml}}\right) \\
&= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)|^2 \cos\left(l[\alpha_k-\alpha_k^*+\alpha_m^*-\alpha_m]\right) && (4.32) \\
&+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_k-\alpha_k^*-\alpha_m)+\theta_l]w_{ml}^x + \right. \\
&\qquad \left. \sin[l(\alpha_k-\alpha_k^*-\alpha_m)+\theta_l]w_{ml}^y\right) && (4.33) \\
&+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_m-\alpha_m^*-\alpha_k)+\theta_l]w_{kl}^x + \right. \\
&\qquad \left. \sin[l(\alpha_m-\alpha_m^*-\alpha_k)+\theta_l]w_{kl}^y\right) && (4.34) \\
&+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \left[\cos(l[\alpha_k-\alpha_m])(w_{kl}^x w_{ml}^x + w_{kl}^y w_{ml}^y) - \right. \\
&\qquad \left. \sin(l[\alpha_k-\alpha_m])(w_{kl}^y w_{ml}^x - w_{kl}^x w_{ml}^y)\right]. && (4.35)
\end{aligned}
$$

We now study the asymptotic behaviour of each term separately. Indeed, the second derivatives are taken at a point $\bar{\alpha}_n$ which converges to α^*: $\bar{\alpha}_n$ is in the neighborhood of α^* with radius $\|\alpha^*-\hat{\alpha}_n\|$. Hence, we need conditions to claim uniform convergence of $\nabla^2 M_n(\cdot)$.

In a similar way as in the proof of Condition (4.14)*, as we assume that* (4.15) *holds and using Cauchy-Schwartz inequality, we have that*

$$
\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2 \leq \left(\sum_{l\in\mathbb{Z}} \delta_l^4 l^4 |c_l(f)|^2\right)^{\frac{1}{2}} \left(\sum_{l\in\mathbb{Z}} |c_l(f)|^2\right)^{\frac{1}{2}} < +\infty,
$$

hence, the deterministic term (4.32) *converges, uniformly in the variable α.*

Since for all $k \in \{1,\ldots,J\}$, the random variables w_{kl}^x and w_{kl}^y follow a Gaussian law $\mathcal{N}(0,1/n)$, we consider the independent variables ξ_{kl}^x and ξ_{kl}^y such that

$$
w_{kl}^x = \frac{1}{\sqrt{n}}\xi_{kl}^x \text{ and } w_{kl}^y = \frac{1}{\sqrt{n}}\xi_{kl}^y.
$$

For the two second terms (4.33) *and* (4.34)*, we write*

$$(4.33) = \frac{1}{\sqrt{n}} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_k - \alpha_k^* - \alpha_m) + \theta_l]\xi_{ml}^x + \right.$$
$$\left. \sin[l(\alpha_k - \alpha_k^* - \alpha_m) + \theta_l]\xi_{ml}^y\right),$$
$$(4.34) = \frac{1}{\sqrt{n}} \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \left(\cos[l(\alpha_m - \alpha_m^* - \alpha_k) + \theta_l]\xi_{kl}^x + \right.$$
$$\left. \sin[l(\alpha_m - \alpha_m^* - \alpha_k) + \theta_l]\xi_{kl}^y\right).$$

Using the assumptions (4.11) *and* (4.15)*, there exists a random variable* Y *such that*

$$Y_n = \sqrt{n}(4.34) \xrightarrow[n\to+\infty]{\|\cdot\|_2} Y.$$

So that $\frac{1}{\sqrt{n}}Y_n$ *converges in probability to* 0*. Hence, in a similar way as in the proof of Condition* (4.14)*, a standard chaining argument proves that, for all* $\lambda > 0$*, we have*

$$\mathbf{P}\left(\sup_{\alpha\in\tilde{A}_1} (4.34) > \lambda\right) \xrightarrow[n\to+\infty]{} 0,$$

which leads to:

$$\sup_{\alpha\in\tilde{A}_1} (4.34) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} 0.$$

In the same way, we may also conclude that (4.33) *and further* (4.35) *both converge uniformly to* 0*.*

The diagonal terms can be written as follows:

$$\frac{J^2}{2}\frac{\partial^2 M_n}{\partial \alpha_k^2}(\alpha) = \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \mathfrak{Re}\left(\tilde{c}_{kl}\sum_{j\neq k}\overline{\tilde{c}_{jl}}\right)$$

$$= \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)|^2 \sum_{j\neq k}\cos\left(l[\alpha_k - \alpha_k^* + \alpha_j^* - \alpha_j]\right) \tag{4.36}$$

$$+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \sum_{j\neq k}\Big(\cos[l(\alpha_k - \alpha_k^* - \alpha_j) + \theta_l]w_{jl}^x + \sin\left[l(\alpha_k - \alpha_k^* - \alpha_j) + \theta_l\right] w_{jl}^y\Big) \tag{4.37}$$

$$+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 |c_l(f)| \sum_{j\neq k}\Big(\cos[l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l]w_{kl}^x + \sin\left[l(\alpha_j - \alpha_j^* - \alpha_k) + \theta_l\right] w_{kl}^y\Big) \tag{4.38}$$

$$+ \sum_{l=-\frac{n-1}{2}}^{\frac{n-1}{2}} \delta_l^2 l^2 \sum_{j\neq k}\left[\cos(l[\alpha_k - \alpha_j])(w_{kl}^x w_{jl}^x + w_{kl}^y w_{jl}^y) - \sin(l[\alpha_k - \alpha_j])(w_{kl}^y w_{jl}^x - w_{kl}^x w_{jl}^y)\right]. \tag{4.39}$$

Using similar arguments as for the previous terms, we can see that, under the same assumptions we get that all the terms (4.36), (4.37), (4.38) *and* (4.39) *converges uniformly, and we get*

$$\frac{\partial^2 M_n}{\partial \alpha_k^2}(\bar{\alpha}_n) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{2(J-1)}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2.$$

As a result, gathering the two previous results leads to the following asymptotic behavior:

$$\nabla^2 M_n(\bar{\alpha}_n) \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{2}{J^2}\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2 \left(JI_{J-1} - U_{J-1}\right),$$

which proves the result.

Moreover, this matrix is invertible. As a result, we have that

$$\left[\nabla^2 M_n\left(\bar{\alpha}_n\right)\right]^{-1} \xrightarrow[n\to+\infty]{\mathbf{P}_{\alpha^*}} \frac{J}{2\sum_{l\in\mathbb{Z}} \delta_l^2 l^2 |c_l(f)|^2}\left(I_{J-1} + U_{J-1}\right).$$

4.7 Figures

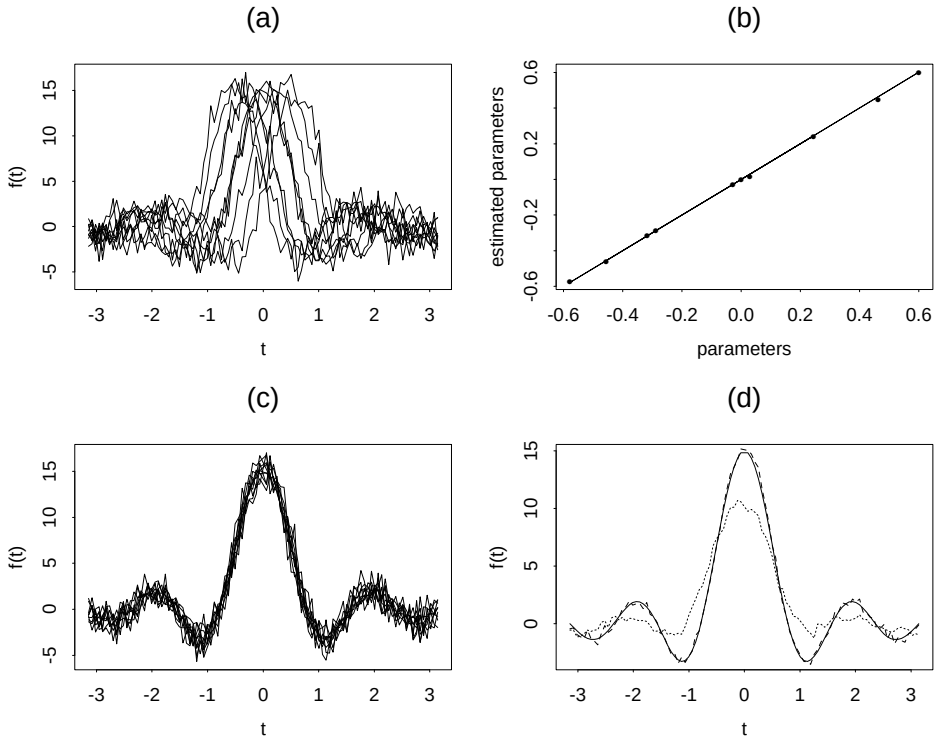

FIG. 4.1 – *Simulation results.*

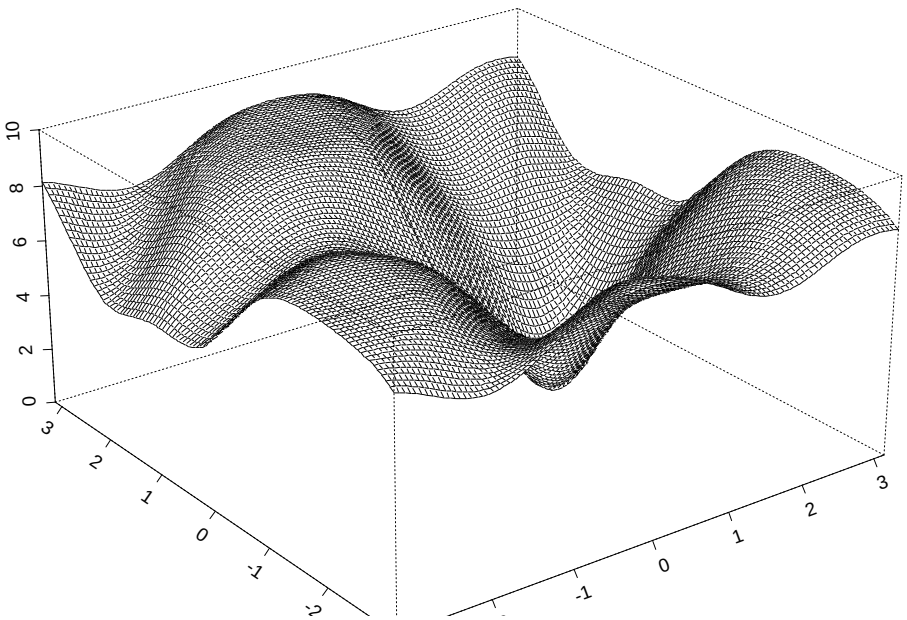

FIG. 4.2 – *Criterion function.*

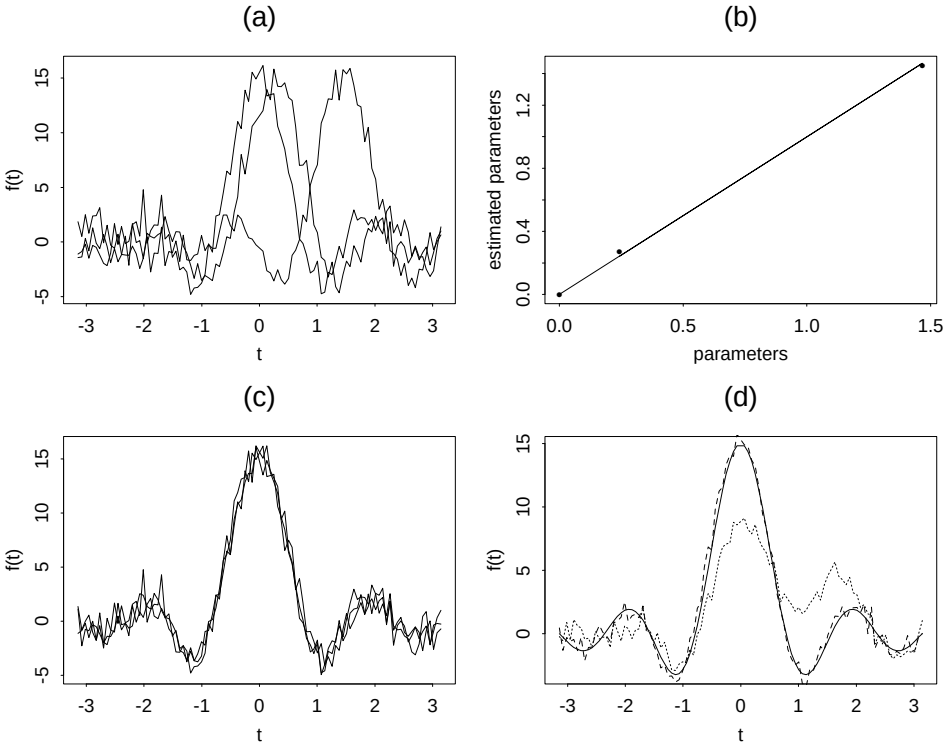

FIG. 4.3 – *Simulation results.*

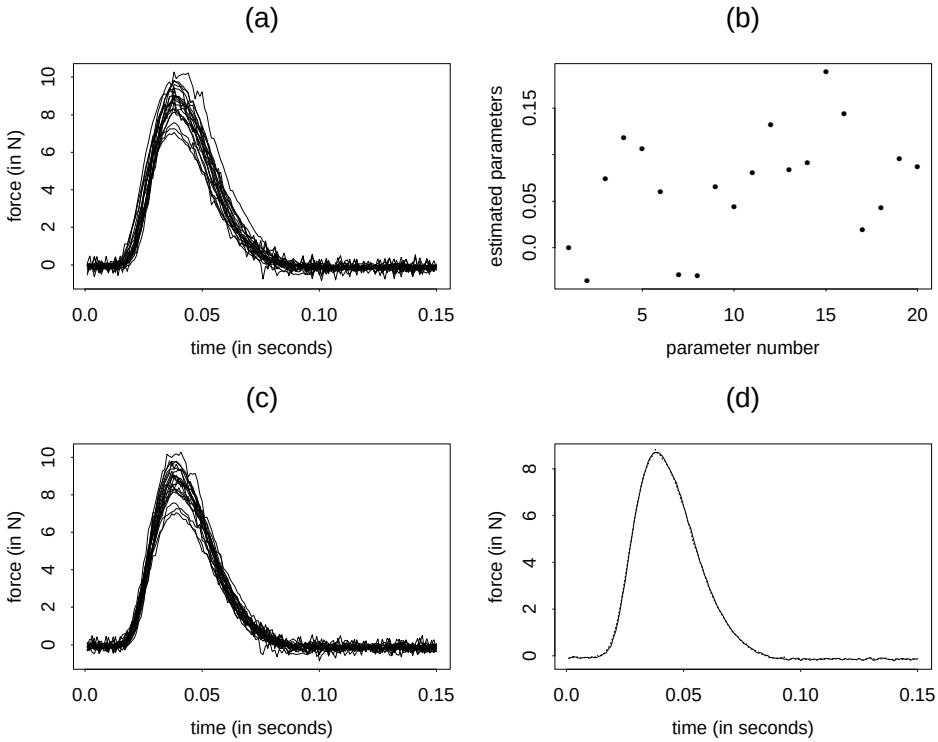

FIG. 4.4 – *Pinch force results.*

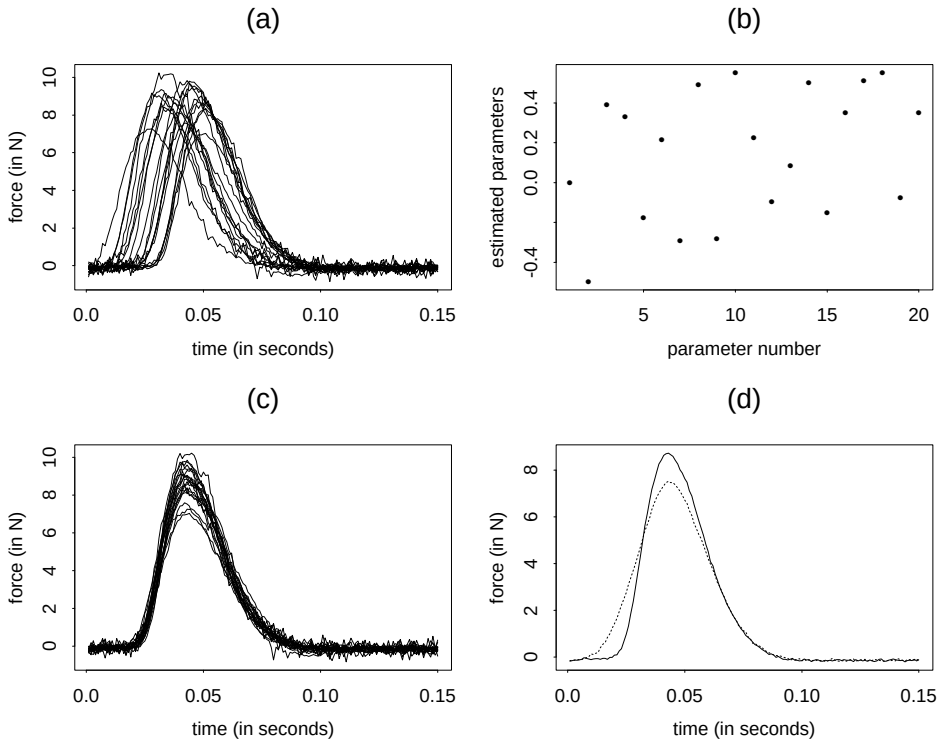

FIG. 4.5 – *Arbitrary shifted pinch force results.*

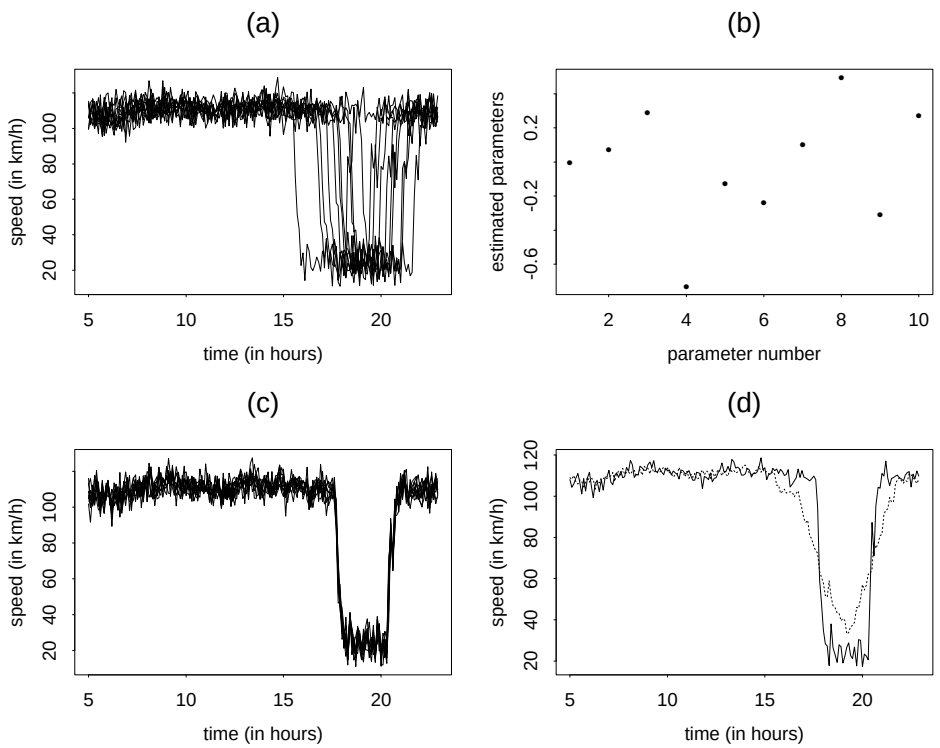

FIG. 4.6 – *Shift estimation results on a particular cluster.*

Chapitre 5

Espérance structurelle d'une fonction aléatoire et problèmes statistiques associés

De façon classique, l'inférence statistique a pour objectif de retrouver certaines des caractéristiques d'une population $\mathcal{P}$ à l'aide des mesures effectuées sur un nombre fini d'individus de cette population. Notons $X \in \mathbb{R}^n$ le vecteur aléatoire modélisant la mesure effectuée sur un individu de $\mathcal{P}$. Notons $(x_i)_{i=1,\ldots,m}$ les valeurs de X observées sur un ensemble de m individus et $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n),P_X)^m$ le modèle statistique associé. Dans le cas d'un échantillon de vecteurs aléatoires indépendants et identiquement distribués, la moyenne empirique $\bar{x}_m = \frac{1}{m}\sum_{i=1}^m x_i$ est une réalisation de l'estimateur $\bar{X}_m = \frac{1}{m}\sum_{i=1}^m X_i$. De plus, si X est intégrable, alors $\bar{X}_m \to \mathbb{E}(X)$ en probabilité. L'espérance $\mathbb{E}(X)$ est une caractéristique centrale qui résume en un seul vecteur, le plus souvent de façon satisfaisante, les mesures effectuées. Autrement dit, l'espérance est souvent un bon représentant de la population $\mathcal{P}$. Malheureusement, lorsque les mesures effectuées ne sont plus des vecteurs, mais des fonctions (par exemple des fonctions du temps) alors l'espérance s'avère être, le plus souvent, un très mauvais représentant de $\mathcal{P}$. En effet, comme cela est expliqué dans [RS97] et [RS02], lorsque la mesure effectuée sur différents individus est une fonction du temps, deux sortes d'aléas apparaissent: un aléa en amplitude et un aléa en temps. Prenons l'exemple des courbes de croissance sur une population de personnes. Le premier effet aléatoire découle du fait que des individus différents auront, par exemple au moment de la puberté, des tailles différentes, donc des observations qui vont varier en amplitude. Le deuxième effet aléatoire est dû au fait que ces mêmes individus n'atteignent pas la puberté au même moment, d'où des observations non synchronisées dans le temps. Dans ce cas, nous observons des courbes dont la moyenne ne nous renseignera pas sur la structure de la courbe de croissance de la population sous-jacente: les deux effets aléatoires (en amplitude et en temps) ne sont pas dissociés. Un exemple simple est illustré par la figure 5.1. Les deux courbes en trait plein représentent les valeurs prises sur l'intervalle [0,1] par deux fonctions de densité gaussienne, d'espérances différentes ($\mu_1 \neq \mu_2$) et de même variance. Nous avons représenté la moyenne de ces deux courbes par des pointillés. Il apparaît très clairement que la moyenne est un très mauvais représentant du phénomène sous-jacent.

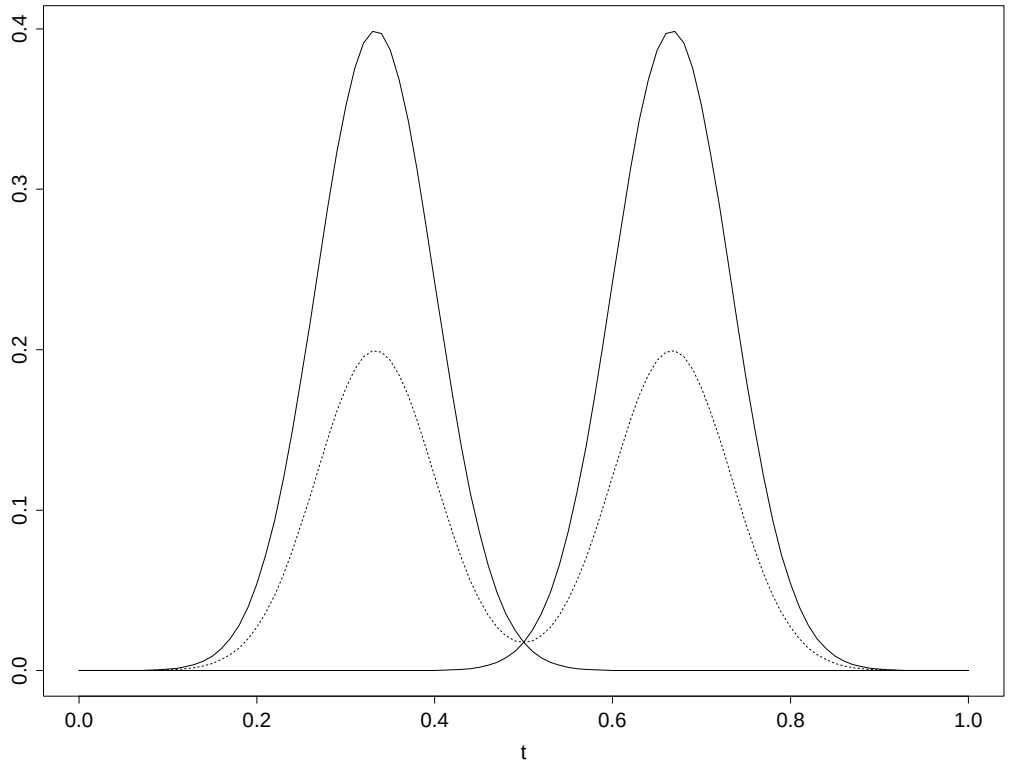

FIG. 5.1 – *Les deux courbes en trait plein représentent les valeurs prises sur l'intervalle* [0,1] *par deux fonctions de densité gaussienne, d'espérances différentes* ($\mu_1 \neq \mu_2$) *et de même variance. La courbe en pointillés est la moyenne de ces deux courbes.*

Historiquement, ce problème apparait dans la littérature d'ingénierie. Plus précisément dans [SC78] sous la dénomination anglaise *time warping*. Le contexte était celui de la reconnaissance de la parole. D'un point de vue statistique, ce problème est présenté dans [Rao58] et fait référence à la comparaison de courbes de croissance. De nombreuses méthodes de résolution ont été proposées ces dernières années. Notamment dans [WG97] et [WG99] sous la dénomination anglaise *structural averaging*. Et aussi, sous l'appellation *curve registration* dans [RD91], [RWF95], [RL98], [KG92], [Kne94] et [KLMR00]. Ce problème est traité dans des domaines aussi divers que la médecine (dans [Boo97]) et la prévision de traffic routier (voir [GLM03] dans les chapitres 3 et 4 de cette thèse).

Le modèle proposé dans la plus grande partie des travaux citées ci-dessus est fondamentalement le même. En effet, de manière simplifiée, le problème est posé de la façon suivante. Soit un ensemble de fonctions réelles

$$f_i : [a,b] \to \mathbb{R}, i = 1, \ldots, m.$$

Il s'agit d'estimer les fonctions suivantes, dites fonctions de déformation (ou *warping functions*, dans la littérature anglaise)

$$h_i : [a,b] \to [a,b],\, i = 1, \ldots, m,$$

continues et strictement croissantes (avec $h_i(a) = a$ et $h_i(b) = b$) minimisant un critère d'ajustement du type

$$\sum_{i=1}^{m} \Psi \left(f_i \circ h_i - \frac{1}{m} \sum_{i=1}^{m} f_i \circ h_i \right), \tag{5.1}$$

où Ψ est une fonction de coût. De plus, ce critère est souvent accompagné d'un terme de pénalité. De manière plus réaliste, il s'agit souvent d'estimer au mieux les fonctions de déformation $(h_i)_{i=1,\ldots,m}$ sur la base des observations bruitées des fonctions $(f_i)_{i=1,\ldots,m}$ sur des points de l'intervalle $[a,b]$, c'est-à-dire, à partir des observations

$$y_{ij} = f_i(t_j) + \varepsilon_{ij},\, i = 1, \ldots, m,\, j = 1, \ldots, n,$$

où, pour tout $i = 1, \ldots, m$, pour tout $j = 1, \ldots, n$, le point $t_j \in [a,b]$ et la variable aléatoire ε_{ij} modélise un bruit de mesure.

Ainsi, le modèle posé n'est pas identifiable. En effet, pour toute fonction

$$\tilde{h} : [a,b] \to [a,b]$$

continue et strictement croissante (avec $\tilde{h}(a) = a$ et $\tilde{h}(b) = b$) l'ensemble des fonctions de déformation

$$\left(h_i \circ \tilde{h} \right)_{i=1,\ldots,m}$$

est aussi solution de (5.1).

Dans ce chapitre, le problème est posé de manière différente. Le modèle que nous étudions est un modèle sans bruit, qui traite seulement de l'aléa dû à une transformation monotone de l'axe des abscisses. Il s'agit d'un modèle de déformation d'une fonction

$$f : [a,b] \to \mathbb{R}$$

inconnue, par un processus aléatoire, dit *processus aléatoire de déformation*,

$$H : [a,b] \to [a,b]$$

inconnu. Ce modèle s'écrit comme suit:

$$Y_{ij} = f \circ H_i^{-1}(w,t_j),\, i = 1, \ldots, m,\, j = 0, \ldots, n,$$

où w appartient à un certain espace probabilisé. Malheureusement, pour les mêmes raisons que précédemment, ce modèle n'est pas identifiable. Par contre, ce modèle nous permet de définir de manière unique et intrinsèque l'*espérance structurelle* associée au modèle de déformation (voir définition en (5.4)). Si nous reprenons l'exemple

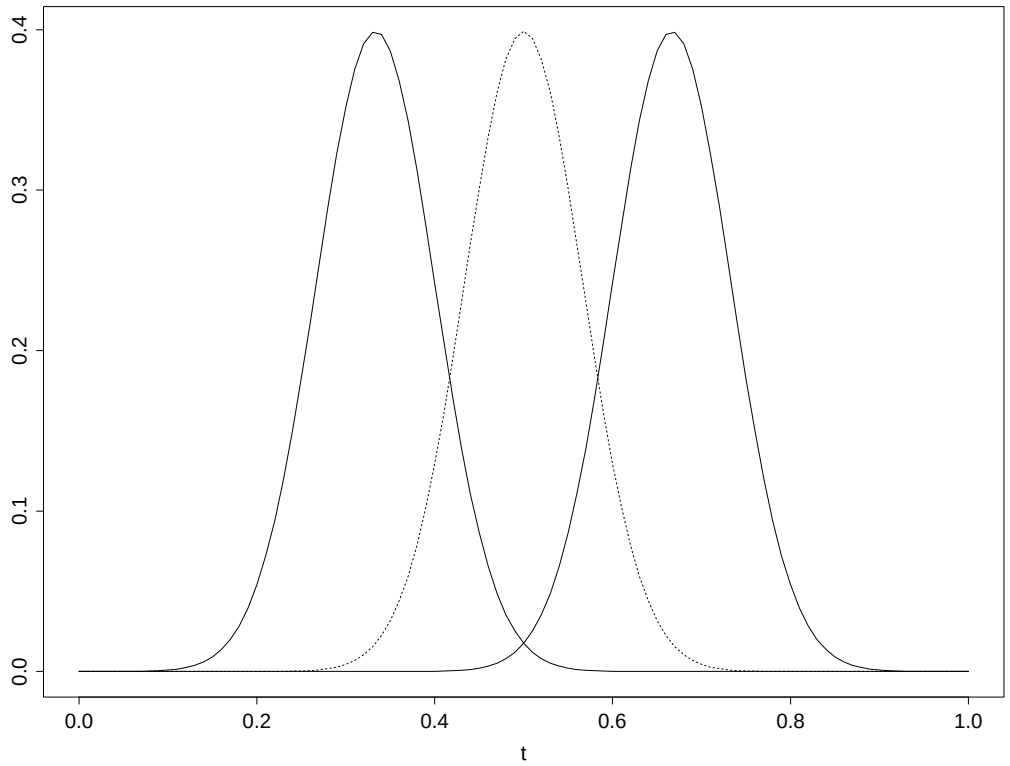

FIG. 5.2 – *Les deux courbes en trait plein représentent deux fonctions de densité gaussienne, d'espérances différentes ($\mu_1 \neq \mu_2$) et de même variance. La moyenne structurelle empirique de ces deux courbes est en pointillés, il s'agit de la fonction de densité gaussienne d'espérance $\frac{\mu_1+\mu_2}{2}$.*

précédent, la figure 5.2 nous montre, en pointillés, la moyenne structurelle empirique des deux densités gaussiennes. Cette moyenne structurelle est la fonction de densité gaussienne d'espérance $\frac{\mu_1+\mu_2}{2}$.

Le modèle que nous étudions est décrit dans la section 5.1. Dans la section 5.2, nous proposons, dans le cas où la fonction f est strictement croissante, des estimateurs empiriques de l'espérance structurelle et du processus de déformation de chacune des fonctions de notre échantillon. De plus, nous montrons, sous certaines hypothèses, la consistance et la normalité asymptotique de ces estimateurs. Dans la section 5.3, nous mettons en place une méthode qui nous permet d'estimer l'espérance structurelle dans le cas où la fonction f n'est pas strictement croissante. La section 5.4 est dédiée à la simulation du processus de déformation H et à un exemple d'estimation pour une fonction f particulière. Pour finir, dans la section 5.5, nous présentons une méthode d'estimation pour des observations bruitées.

5.1 Le modèle de déformation

Le modèle de déformation est défini par

$$Y_{ij} = f \circ H_i^{-1}(w,t_j),\, i = 1,\ldots,m,\, j = 0,\ldots,n, \tag{5.2}$$

où

- $f : [a,b] \to \mathcal{I} \subset \mathbb{R}$ est une fonction réelle continue,
- les processus aléatoires $(H_i)_{i=1,\ldots,m}$ sont i.i.d. et de même loi que le processus aléatoire H défini par

$$\begin{array}{rccl} H: & \Omega & \to & (\mathcal{C}([a,b]), \mathcal{B}(\mathcal{C}([a,b]))) \\ & w & \mapsto & H(w,\cdot), \end{array} \tag{5.3}$$

  où
  - $(\Omega,\mathcal{A},\mathbf{P})$ est l'espace de probabilité sur lequel sont définis nos processus de déformation,
  - $\mathcal{C}([a,b])$ est l'ensemble des fonctions réelles continues, définies sur l'intervalle $[a,b]$ muni de la norme uniforme et de la tribu des boréliens $\mathcal{B}(\mathcal{C}([a,b]))$;

  de plus, nous faisons les hypothèses suivantes sur notre processus de déformation H:
  - i) H est p.s. continu sur $[a,b]$ et p.s. strictement croissant,
  - ii) $H(\cdot,a) \overset{ps}{=} a$ et $H(\cdot,b) \overset{ps}{=} b$,
- pour tout $j \in \{0,\ldots,n\}$, $t_j = a + j\frac{b-a}{n}$.

Afin d'alléger les notations, nous écrirons, pour tout $t \in [a,b]$, $H_i(t)$ au lieu de $H_i(\cdot,t)$. Nous noterons aussi, pour tout $i = 1,\ldots,m$, $F_i = f \circ H_i^{-1}$, et

$$\begin{array}{rccl} f_i: & [a,b] & \to & \mathcal{I} \subset \mathbb{R} \\ & t & \mapsto & f_i(t) = f \circ h_i^{-1}(t), \end{array}$$

où h_i est une réalisation du processus aléatoire H_i. Nous noterons enfin A, l'ensemble des trajectoires $F(w,\cdot) = f \circ H^{-1}(w,\cdot)$ $(w \in \Omega)$ continues, strictement croissantes, et telles que $F(w,a) = f(a)$ et $F(w,b) = f(b)$.

La proposition suivante énonce l'existence, sous les hypothèses (5.3), des moments d'ordre 1 et d'ordre 2 du processus aléatoire H.

Proposition 5.1 *Sous les hypothèses* (5.3)*, les fonctions ϕ (espérance) et γ (moment d'ordre 2), définies par*

$$\begin{array}{rccl} \phi: & [a,b] & \to & [a,b] \\ & t & \mapsto & \phi(t) = \mathbb{E}(H(t)) \end{array}$$

et

$$\begin{array}{rccl} \gamma: & [a,b] & \to & \mathbb{R}_+ \\ & t & \mapsto & \gamma(t) = \mathbb{E}(H^2(t)), \end{array}$$

existent. De plus,

a) ϕ et γ sont continues et strictement croissantes,

b) $\phi(a) = a$, $\phi(b) = b$, $\gamma(a) = a^2$ *et* $\gamma(b) = b^2$. *En particulier, nous avons* $\mathrm{Var}(H(a)) = \mathrm{Var}(H(b)) = 0$.

Nous définissons également la covariance du processus de déformation H par

$$\forall (t,s) \in [a,b]^2, r(t,s) = \mathrm{Cov}\left(H(t),H(s)\right).$$

Preuve 5.2 *Pour tout* $t \in [a,b]$, *la variable aléatoire* $H(t)$ *est bornée. Donc, les fonctions* ϕ *et* γ *existent. De plus, d'après le théorème de Lebesgue et par continuité de* H, *elles sont aussi continues. D'autre part, pour tout* $(t',t'') \in [a,b]^2$ *tels que* $t' < t''$, *nous avons* $H(t') < H(t'')$, *donc* $\mathbb{E}(H(t')) < \mathbb{E}(H(t''))$, *et la fonction* ϕ *est strictement croissante. De même, la fonction* γ *est strictement croissante.*

■

Comme nous l'avons souligné dans l'introduction de ce chapitre, le modèle (5.2) n'est pas identifiable. Par contre, il est naturel de définir l'espérance structurelle (relativement à H) de la fonction aléatoire $f \circ H^{-1}$, notée $\mathbb{ES}_H$, par

$$\mathbb{ES}_H\left(f \circ H^{-1}\right) = f \circ \phi^{-1}, \tag{5.4}$$

où le processus aléatoire H est défini par (5.3). Ainsi, l'espérance structurelle d'une fonction aléatoire n'est pas la fonction f, mais la composée de f par ϕ^{-1}, c'est-à-dire, la composée de f par l'inverse de l'espérance du processus de déformation H. L'espérance structurelle s'interprète donc comme la déformation moyenne de la fonction f par le processus H. Dans la section suivante, nous construisons, dans le cas où la fonction f est strictement croissante, dans le cadre du modèle d'observation (5.2), un estimateur consistant de l'inverse de l'espérance structurelle $f \circ \phi^{-1}$, et en montrons la normalité asymptotique. Nous construisons ensuite, à partir de cet estimateur, un estimateur consistant de l'espérance structurelle elle-même. De plus, nous construisons un estimateur consistant associé à la fonction de déformation $h_i^{-1}, i = 1, \ldots, m$, et en montrons également la normalité asymptotique. En effet, en pratique, il est souvent intéressant de connaître la déformation subie par un individu particulier de notre échantillon.

5.2 Estimation de l'espérance structurelle

Nous nous plaçons ici dans le cas où la fonction f est strictement croissante:

$$\forall (t',t'') \in [a,b]^2, t' < t'' \Rightarrow f(t') < f(t''). \tag{5.5}$$

Cette section se subdivise en quatre parties. Dans la partie 5.2.1, nous traitons le modèle (5.2) avec l'hypothèse supplémentaire que les fonctions $(f_i)_{i=1,\ldots,m}$ sont connues sur tout l'intervalle $[a,b]$. Nous généralisons ensuite les résultats obtenus avec ce modèle *continu*, dans la partie 5.2.2, au modèle d'observation (5.2). Dans la partie 5.2.3, nous construisons l'estimateur de l'espérance structurelle et en montrons la consistance. Dans la partie 5.2.4, nous affaiblissons les hypothèses (5.3.ii) en supposant que $H^{-1}(a)$ et $H^{-1}(b)$ sont des variables aléatoires et ont des lois à support compact.

5.2.1 Modèle continu

En pratique, l'évaluation d'une fonction quelconque est effectuée sur un intervalle discrétisé de $\mathbb{R}$. Ainsi, comme nous l'avons décrit dans la section 5.1, le modèle de déformation est observé sur une grille de points $(t_j)_{j=0,\ldots,n}$ de $[a,b]$. Dans une première approche, nous considérons le modèle théorique

$$Y_i(t) = f \circ H_i^{-1}(t), i = 1, \ldots, m, t \in [a,b]. \tag{5.6}$$

Nous considérons donc que nos fonctions sont mesurées sur tout l'intervalle $[a,b]$. Les résultats obtenus pour ce modèle théorique seront ensuite utilisés pour l'estimation dans le modèle (5.2).

Lorsque la fonction $f : [a,b] \to \mathcal{I} \subset \mathbb{R}$ est strictement croissante, la fonction réciproque $f^{-1} : \mathcal{I} \to [a,b]$ existe et est strictement croissante. De plus, une déformation en phase de la fonction f (c'est-à-dire une déformation sur l'axe des abscisses) correspond à une déformation en amplitude de la fonction f^{-1} (c'est-à-dire à une déformation de l'axe des ordonnées). Il est donc naturel de considérer la moyenne de l'ensemble des fonctions $\left(f_i^{-1}\right)_{i=1,\ldots,m}$ pour estimer $\phi \circ f^{-1}$, soit l'inverse de l'espérance structurelle. Nous proposons donc l'estimateur empirique de l'inverse de l'espérance structurelle

$$\overline{\phi \circ f^{-1}}_m = \frac{1}{m} \sum_{i=1}^{m} F_i^{-1}. \tag{5.7}$$

Dans le théorème suivant, nous montrons la consistance de cet estimateur empirique dans $(\mathcal{C}(\mathcal{I}), \|\cdot\|_\infty)$.

Théorème 5.3 *Sous les hypothèses* (5.3) *et* (5.5)*, l'estimateur* $\overline{\phi \circ f^{-1}}_m$ *converge presque sûrement vers* $\phi \circ f^{-1}$*:*

$$\left\| \overline{\phi \circ f^{-1}}_m - \phi \circ f^{-1} \right\|_\infty \xrightarrow[m \to +\infty]{ps} 0.$$

Démonstration 5.4 *Cette démonstration est une application directe du corollaire 7.10 page 189 de [LT91]. Il s'agit d'une extension de la loi forte des grands nombres à des espaces de Banach.*

Pour tout $i = 1, \ldots, m$*, la fonction* $F_i = f \circ H_i^{-1}$ *est la composée de deux fonctions strictement croissantes, donc* F_i^{-1} *est bien définie et est strictement croissante. De même, la fonction*

$$\overline{\phi \circ f^{-1}}_m = \frac{1}{m} \sum_{i=1}^{m} F_i^{-1}$$

est bien définie et est strictement croissante. De plus, nous avons

$$\begin{aligned}\overline{\phi \circ f^{-1}}_m =& \frac{1}{m}\sum_{i=1}^{m} F_i^{-1}\\ =& \frac{1}{m}\sum_{i=1}^{m} \left(f \circ H_i^{-1}\right)^{-1}\\ =& \frac{1}{m}\sum_{i=1}^{m} H_i \circ f^{-1}.\end{aligned}$$

Posons

$$X_i = H_i \circ f^{-1} - \phi \circ f^{-1}, i \in \mathbb{N}^*$$

et

$$S_n = \sum_{i=1}^{n} X_i, n \in \mathbb{N}^*.$$

Les variables aléatoires $(X_i)_{i=1,\ldots,m}$ *sont par définition boréliennes à valeurs dans* $B = \mathcal{C}(\mathcal{I})$ *où* $\mathcal{I}$ *est un compact de* $\mathbb{R}$*, et* B *est un espace de Banach séparable. De plus, le dual de* B *est l'ensemble des mesures bornées sur* $\mathcal{I}$ *([Rud87]). Nous sommes donc bien dans le cadre du chapitre 7 de [LT91] et nous pouvons appliquer le corollaire 7.10. En effet,*

$$\mathbb{E}\left(\|X_1\|_\infty\right) < +\infty \text{ et } \mathbb{E}\left(X_1\right) = 0.$$

Donc

$$\frac{S_n}{n} \xrightarrow[m\to+\infty]{ps} 0.$$

D'où le résultat.

■

Il est certainement possible de montrer un théorème central limite dans l'ensemble $(\mathcal{C}(\mathcal{I}),\|\cdot\|_\infty)$ en utilisant les résultats généraux de [LT91]. Nous nous contenterons dans le théorème suivant de montrer une version vectorielle de la normalité asymptotique de l'estimateur empirique $\overline{\phi \circ f^{-1}}_m$.

Théorème 5.5 *Sous les hypothèses* (5.3) *et* (5.5)*, pour tout vecteur* $(y_1,\ldots,y_k)'$ *de* $\mathcal{I}^k$, $k \in \mathbb{N}^*$*, nous avons*

$$\sqrt{m}\left(\begin{pmatrix}\overline{\phi \circ f^{-1}}_m(y_1)\\ \vdots \\ \overline{\phi \circ f^{-1}}_m(y_k)\end{pmatrix} - \begin{pmatrix}\phi \circ f^{-1}(y_1)\\ \vdots \\ \phi \circ f^{-1}(y_k)\end{pmatrix}\right) \xrightarrow[m\to+\infty]{\mathcal{L}} \mathcal{N}_k\left(0,\Gamma(y)\right),$$

où $\Gamma(y)$ *est la matrice de variance-covariance de taille* k*, définie pour tout* $(i,j) \in \{1,\ldots,k\}^2$*, par*

$$\Gamma(y)_{ij} = r\left(f^{-1}(y_i),f^{-1}(y_j)\right).$$

En particulier, nous avons pour tout $i \in \{1,\ldots,k\}$, la variance

$$\begin{aligned}\Gamma(y)_{ii} &= r\left(f^{-1}(y_i),f^{-1}(y_i)\right)\\ &= \mathbb{E}\left(\left(H\circ f^{-1}(y_i)\right)^2\right) - \left(\mathbb{E}\left(H\circ f^{-1}(y_i)\right)\right)^2\\ &= \gamma\circ f^{-1}(y_i) - \left(\phi\circ f^{-1}(y_i)\right)^2.\end{aligned}$$

Démonstration 5.6 *Pour tout $i = 1,\ldots,m$, comme nous l'avons vu dans la démonstration du théorème 5.3, la fonction*

$$\overline{\phi\circ f^{-1}}_m = \frac{1}{m}\sum_{i=1}^{m} F_i^{-1}$$

est bien définie et est strictement croissante. En notant, de façon abusive, pour tout vecteur $y = (y_1,\ldots,y_k)'$ de $\mathcal{I}^k$, $k \in \mathbb{N}^*$,

$$\overline{\phi\circ f^{-1}}_m(y) = \begin{pmatrix}\overline{\phi\circ f^{-1}}_m(y_1)\\ \vdots\\ \overline{\phi\circ f^{-1}}_m(y_k)\end{pmatrix}, \phi\circ f^{-1}(y) = \begin{pmatrix}\phi\circ f^{-1}(y_1)\\ \vdots\\ \phi\circ f^{-1}(y_k)\end{pmatrix}$$

et, pour tout $i \in \mathbb{N}^*$,

$$H_i\circ f^{-1}(y) = \begin{pmatrix}H_i\circ f^{-1}(y_1)\\ \vdots\\ H_i\circ f^{-1}(y_k)\end{pmatrix},$$

nous avons

$$\overline{\phi\circ f^{-1}}_m(y) = \frac{1}{m}\sum_{i=1}^{m} H_i\circ f^{-1}(y).$$

Or, les vecteurs aléatoires $\left(H_i\circ f^{-1}(y)\right)_{i=1,\ldots,m}$ sont i.i.d. et ont, comme espérance le vecteur $\phi\circ f^{-1}(y)$ et comme matrice de variance-covariance $\Gamma(y)$. D'où le résultat, par application du théorème central limite.

■

En pratique, dans le cadre d'un modèle de déformation, il est souvent intéressant de connaître, en particulier, la déformation subie par la fonction f_i ($i \in \{1,\ldots,m\}$) c'est-à-dire la fonction de déformation h_i^{-1} qui a été composée avec f. Comme nous l'avons déjà souligné, le modèle (5.2) n'est pas identifiable. Ainsi, comme pour l'espérance structurelle, nous ne pourrons pas estimer la fonction de déformation h_i^{-1}, $i = 1,\ldots,m$. Par contre nous pouvons estimer, en utilisant la même idée que pour la construction de l'estimateur (5.7), la fonction de déformation $\phi\circ h_i^{-1}$, $i = 1,\ldots,m$. Pour cela, nous définissons pour tout $i_0 \in \{1,\ldots,m\}$ fixé, conditionnellement à $F_{i_0} = f_{i_0} \in A$, l'estimateur empirique de la fonction de déformation $\phi\circ h_{i_0}^{-1}$

$$\overline{\phi\circ h_{i_0}^{-1}}_m = \frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m} F_i^{-1}\circ F_{i_0}(t). \tag{5.8}$$

Dans le corollaire suivant, nous montrons la consistance de cet estimateur dans $(\mathcal{C}([a,b]), \|\cdot\|_\infty)$.

Corollaire 5.7 *Nous nous plaçons sous les hypothèses* (5.3) *et* (5.5). *Soit* $f_{i_0} \in A$ *avec* $i_0 \in \{1,\ldots,m\}$ *fixé. Alors, conditionnellement à* $F_{i_0} = f_{i_0}$, *l'estimateur empirique* $\overline{\phi \circ h_{i_0}^{-1}}_m$ *converge presque sûrement vers* $\phi \circ h_{i_0}^{-1}$:

$$\left\| \overline{\phi \circ h_{i_0}^{-1}}_m - \phi \circ h_{i_0}^{-1} \right\|_\infty \xrightarrow[m\to+\infty]{ps} 0.$$

Preuve 5.8 *Soit* $i_0 \in \{1,\ldots,m\}$ *fixé, alors, conditionnellement à* $F_{i_0} = f_{i_0} \in A$, *nous avons*

$$\overline{\phi \circ h_{i_0}^{-1}}_m = \frac{1}{m-1} \sum_{\substack{i=1 \\ i\neq i_0}}^{m} F_i^{-1} \circ F_{i_0} = \frac{1}{m-1} \sum_{\substack{i=1 \\ i\neq i_0}}^{m} F_i^{-1} \circ f_{i_0},$$

ainsi, d'après le théorème 5.3,

$$\overline{\phi \circ h_{i_0}^{-1}}_m \xrightarrow[m\to+\infty]{ps} \phi \circ f^{-1} \circ f_{i_0} = \phi \circ f^{-1} \circ f \circ h_{i_0}^{-1} = \phi \circ h_{i_0}^{-1}.$$

■

Comme pour le théorème 5.5, nous nous contenterons dans le corollaire suivant, de montrer une version vectorielle de la normalité asymptotique de l'estimateur $\overline{\phi \circ h_{i_0}^{-1}}_m$.

Corollaire 5.9 *Nous nous plaçons sous les hypothèses* (5.3) *et* (5.5). *Soit* $f_{i_0} \in A$ *avec* $i_0 \in \{1,\ldots,m\}$ *fixé. Alors, conditionnellement à* $F_{i_0} = f_{i_0}$, *pour tout vecteur* $t = (t_1,\ldots,t_k)'$ *de* $[a,b]^k$, *nous avons*

$$\sqrt{m}\left(\begin{pmatrix} \overline{\phi \circ h_{i_0}^{-1}}_m(t_1) \\ \vdots \\ \overline{\phi \circ h_{i_0}^{-1}}_m(t_k) \end{pmatrix} - \begin{pmatrix} \phi \circ h_{i_0}^{-1}(t_1) \\ \vdots \\ \phi \circ h_{i_0}^{-1}(t_k) \end{pmatrix} \right) \xrightarrow[m\to+\infty]{\mathcal{L}} \mathcal{N}_k\left(0, \Gamma_0(t)\right),$$

où $\Gamma_0(t)$ *est la matrice de variance-covariance de taille* k, *définie pour tout* $(i,j) \in \{1,\ldots,k\}^2$, *par*

$$\Gamma_0(t)_{ij} = r\left(h_{i_0}^{-1}(t_i), h_{i_0}^{-1}(t_j)\right).$$

En particulier, nous avons pour tout $i \in \{1,\ldots,k\}$, la variance

$$\begin{aligned} \Gamma_0(t)_{ii} &= r\left(h_{i_0}^{-1}(t_i), h_{i_0}^{-1}(t_i)\right) \\ &= \mathbb{E}\left(\left(H \circ h_{i_0}^{-1}(t_i)\right)^2\right) - \left(\mathbb{E}\left(H \circ h_{i_0}^{-1}(t_i)\right)\right)^2 \\ &= \gamma \circ h_{i_0}^{-1}(t_i) - \left(\phi \circ h_{i_0}^{-1}(t_i)\right)^2. \end{aligned}$$

Preuve 5.10 *La preuve de ce corollaire découle directement du théorème 5.5. En effet, il suffit de remplacer dans sa démonstration, le vecteur* $(y_1,\ldots,y_k)'$ *par le vecteur* $(f_{i_0}(t_1),\ldots,f_{i_0}(t_k))'$.

■

5.2.2 Modèle discret

Dans ce qui précède, nous nous sommes placés dans le cadre peu réaliste du modèle (5.6), dans lequel les réalisations $(f_i)_{i=1,\dots,m}$ sont observées sur tout l'intervalle $[a,b]$. Dans ce qui suit, nous étendons les théorèmes 5.3 et 5.5 et les corollaires 5.7 et 5.9 au modèle échantillonné (5.2), soit

$$Y_{ij} = f \circ H_i^{-1}(w,t_j),\, i = 1,\dots,m,\, j = 0,\dots,n.$$

Nous construisons donc des estimateurs équivalents à (5.7) et (5.8), basés sur les seules mesures des fonctions $(f_i)_{i=1,\dots,m}$ effectuées sur les points $(t_j)_{j=0,\dots,n}$ de $[a,b]$.

Avant de donner la définition des estimateurs, nous définissons une notion de convergence d'une suite indicée par le couple $(m,n) \in \mathbb{N}^2$. Cette notion nous sera utile par la suite.

Définition 5.11 *Une suite $(U_{mn})_{(m,n)\in\mathbb{N}^2}$ converge vers la constante $c \in \mathbb{R}$, noté*

$$U_{mn} \xrightarrow[m,n\to+\infty]{} c,$$

si, et seulement si,

$$\forall \varepsilon > 0, \exists (m_0,n_0) \in \mathbb{N}^2 \,/\, \forall (m,n) \in \mathbb{N}^2, m > m_0, n > n_0 \Rightarrow |U_{mn} - c| < \varepsilon.$$

Dans le modèle (5.6), l'estimateur empirique (5.7) de $\phi \circ f^{-1}$ a été défini, pour tout $y \in \mathcal{I}$, par

$$\overline{\phi \circ f^{-1}}_m(y) = \frac{1}{m}\sum_{i=1}^{m} F_i^{-1}(y).$$

Il est alors naturel, dans le modèle (5.2), où les fonctions sont seulement mesurées sur la grille $(t_j)_{j=0,\dots,n}$ de $[a,b]$, de remplacer, pour tout $i = 1,\dots,m$, la variable aléatoire $F_i^{-1}(y)$ par la valeur la plus proche dans l'ensemble $\{t_0,\dots,t_n\}$. Soit $y \in \mathcal{I}$ fixé, nous définissons l'estimateur empirique de l'inverse de l'espérance structurelle en y par

$$\widehat{\phi \circ f^{-1}}_{mn}(y) = \frac{1}{m}\sum_{i=1}^{m} T_{j_i}, \tag{5.9}$$

où

$$\forall i = 1,\dots,m, T_{j_i} = \arg\min_{t_j \in \{t_0,\dots,t_n\}} |F_i(t_j) - y|.$$

Dans le théorème suivant, nous montrons la consistance de cet estimateur dans $(\mathcal{C}(\mathcal{I}), \|\cdot\|_\infty)$.

Théorème 5.12 *Sous les hypothèses* (5.3) *et* (5.5)*, l'estimateur empirique $\widehat{\phi \circ f^{-1}}_{mn}$ converge presque sûrement vers $\phi \circ f^{-1}$:*

$$\left\| \widehat{\phi \circ f^{-1}}_{mn} - \phi \circ f^{-1} \right\|_\infty \xrightarrow[m,n\to+\infty]{ps} 0.$$

Démonstration 5.13 *Pour tout $y \in \mathcal{I}$, pour tout $i = 1, \ldots, m$, nous avons*

$$\begin{aligned} T_{j_i} &= \arg \min_{t_j \in \{t_0, \ldots, t_n\}} |F_i(t_j) - y| \\ &= \arg \min_{t_j \in \{t_0, \ldots, t_n\}} \left| F_i(t_j) - F_i \circ F_i^{-1}(y) \right|, \end{aligned}$$

or, F_i étant strictement croissante, nous avons

$$F_i^{-1}(y) - \frac{1}{n} \leq T_{j_i} \leq F_i^{-1}(y) + \frac{1}{n},$$

d'où, presque sûrement,

$$\overline{\phi \circ f^{-1}}_m(y) - \frac{1}{n} \leq \widehat{\phi \circ f^{-1}}_{mn}(y) \leq \overline{\phi \circ f^{-1}}_m(y) + \frac{1}{n} \tag{5.10}$$

Donc, pour montrer la convergence au sens de la définition 5.11, il suffit de montrer la convergence presque sûre des deux bornes de l'inégalité (5.10)*. Nous traitons uniquement ici la borne supérieure, le traitement étant identique pour la borne inférieure. Nous avons donc*

$$\left| \overline{\phi \circ f^{-1}}_m(y) + \frac{1}{n} - \phi \circ f^{-1}(y) \right| \leq \left| \overline{\phi \circ f^{-1}}_m(y) - \phi \circ f^{-1}(y) \right| + \frac{1}{n},$$

d'où

$$\left\| \overline{\phi \circ f^{-1}}_m + \frac{1}{n} - \phi \circ f^{-1} \right\|_\infty \leq \left\| \overline{\phi \circ f^{-1}}_m - \phi \circ f^{-1} \right\|_\infty + \frac{1}{n},$$

or, d'après le théorème 5.3,

$$\left\| \overline{\phi \circ f^{-1}}_m - \phi \circ f^{-1} \right\|_\infty \xrightarrow[m \to +\infty]{ps} 0,$$

d'où la convergence de la borne supérieure de l'inégalité (5.10)*,*

$$\left\| \overline{\phi \circ f^{-1}}_m(y) + \frac{1}{n} - \phi \circ f^{-1} \right\|_\infty \xrightarrow[m,n \to +\infty]{ps} 0.$$

■

Comme pour le théorème 5.5 et pour le corollaire 5.9, nous nous contentons de traiter ici la normalité asymptotique de l'estimateur $\widehat{\phi \circ f^{-1}}_{mn}$ de façon vectorielle.

Théorème 5.14 *Nous nous plaçons sous les hypothèses* (5.3) *et* (5.5)*. Soit le vecteur $y = (y_1, \ldots, y_k)'$ de $\mathcal{I}^k$, $k \in \mathbb{N}^*$. Nous supposons de plus que $n = m^{\frac{1}{2}+\alpha}$ avec $\alpha > 0$. Alors,*

$$\sqrt{m} \left(\begin{pmatrix} \widehat{\phi \circ f^{-1}}_{mn}(y_1) \\ \vdots \\ \widehat{\phi \circ f^{-1}}_{mn}(y_k) \end{pmatrix} - \begin{pmatrix} \phi \circ f^{-1}(y_1) \\ \vdots \\ \phi \circ f^{-1}(y_k) \end{pmatrix} \right) \xrightarrow[m \to +\infty]{\mathcal{L}} \mathcal{N}_k \left(0, \Gamma(y)\right),$$

où $\Gamma(y)$ est la matrice de variance-covariance définie dans le théorème 5.5.

Démonstration 5.15 *Pour tout $i \in \{1,\ldots,k\}$, l'inégalité* (5.10) *établie dans la démonstration du théorème 5.12 s'écrit*

$$\overline{\phi \circ f^{-1}}_m(y_i) - \frac{1}{n} \leq \widehat{\phi \circ f^{-1}}_{mn}(y_i) \leq \overline{\phi \circ f^{-1}}_m(y_i) + \frac{1}{n},$$

donc, pour $n = m^{\frac{1}{2}+\alpha}$ $(\alpha > 0)$, nous avons

$$\sqrt{m}\left(\overline{\phi \circ f^{-1}}_m(y_i) - \phi \circ f^{-1}(y_i)\right) - \frac{1}{m^\alpha} \leq \sqrt{m}\left(\widehat{\phi \circ f^{-1}}_{mn}(y_i) - \phi \circ f^{-1}(y_i)\right)$$

et

$$\sqrt{m}\left(\widehat{\phi \circ f^{-1}}_{mn}(y_i) - \phi \circ f^{-1}(y_i)\right) \leq \sqrt{m}\left(\overline{\phi \circ f^{-1}}_m(y_i) - \phi \circ f^{-1}(y_i)\right) + \frac{1}{m^\alpha},$$

donc

$$\begin{aligned}
&\sqrt{m}\left(\begin{pmatrix} \widehat{\phi \circ f^{-1}}_{mn}(y_1) \\ \vdots \\ \widehat{\phi \circ f^{-1}}_{mn}(y_k) \end{pmatrix} - \begin{pmatrix} \phi \circ f^{-1}(y_1) \\ \vdots \\ \phi \circ f^{-1}(y_k) \end{pmatrix}\right) \\
&= \sqrt{m}\left(\begin{pmatrix} \overline{\phi \circ f^{-1}}_m(y_1) \\ \vdots \\ \widehat{\phi \circ f^{-1}}_m(y_k) \end{pmatrix} - \begin{pmatrix} \phi \circ f^{-1}(y_1) \\ \vdots \\ \phi \circ f^{-1}(y_k) \end{pmatrix}\right) + O_{\mathbf{P}}\begin{pmatrix} \frac{1}{m^\alpha} \\ \vdots \\ \frac{1}{m^\alpha} \end{pmatrix},
\end{aligned}$$

d'où, d'après le théorème 5.5, la convergence en loi.

■

Pour construire un estimateur de la fonction de déformation $\phi \circ h_i^{-1}$, pour tout $i \in \{1,\ldots,m\}$, dans le modèle échantillonné (5.2), nous allons utiliser la même idée que précédemment. C'est-à-dire que nous allons remplacer les valeurs inconnues par les valeurs connues les plus proches. Soient $i_0 \in \{1,\ldots,m\}$ et $t \in [a,b]$, nous définissons l'estimateur empirique, conditionnellement à $F_{i_0} = f_{i_0} \in A$, par

$$\widehat{\phi \circ h_{i_0}^{-1}}_{mn}(t) = \frac{1}{m-1}\sum_{\substack{i=1 \\ i \neq i_0}}^{m} T_{j_i}, \tag{5.11}$$

avec, pour tout $i \in \{1,\ldots,m\} \setminus i_0$,

$$T_{j_i} = \arg\min_{t_j \in \{t_0,\ldots,t_n\}} \left|F_i(t_j) - F_{i_0}(t_{j_0})\right|,$$

avec

$$t_{j_0} = \arg\min_{t_j \in \{t_0,\ldots,t_n\}} |t_j - t|.$$

L'estimateur $\widehat{\phi \circ h_{i_0}^{-1}}_{mn}$ se déduit de l'estimateur $\overline{\phi \circ h_{i_0}^{-1}}_m$ de la même manière que l'estimateur $\widehat{\phi \circ f^{-1}}_{mn}$ se déduisait de l'estimateur $\overline{\phi \circ f^{-1}}_m$. La principale difficulté rencontrée, pour démontrer la consistance et la normalité asymptotique de $\widehat{\phi \circ h_{i_0}^{-1}}_{mn}$, est liée au fait que deux approximations successives sont utilisées dans la

définition (5.11). Plus précisément, la première approximation consiste à substituer la valeur de $t \in [a,b]$ sur laquelle nous voulons estimer $\phi \circ h_{i_0}^{-1}$, par la valeur la plus proche sur la grille $\{t_0, \ldots, t_n\}$. Pour mener à bien les démonstrations qui vont suivre, nous commençons par définir *le module d'uniforme continuité de déformation.*

Définition 5.16 *Soient H_1 et H_2 deux processus de déformation indépendants et de même loi que H défini par* (5.3)*. Nous définissons le processus aléatoire K_H, appelé module d'uniforme continuité de déformation, par*

$$K_H(\delta) = \sup_{\substack{(t_1,t_2)\in[a,b]^2 \\ |t_1-t_2|\leq\delta}} \left|H_1 \circ H_2^{-1}(t_1) - H_1 \circ H_2^{-1}(t_2)\right|, \delta \geq 0. \tag{5.12}$$

De plus, nous posons

$$L_H^1(H_2,\delta) = \mathbb{E}\left(K_H(\delta)|H_2\right), \delta \geq 0,$$

et

$$L_H^2(H_2,\delta) = \mathbb{E}\left(K_H^2(\delta)|H_2\right), \delta \geq 0.$$

Nous remarquons que presque sûrement, pour tout $\delta \geq 0$,

$$0 \leq K_H(\delta) \leq b - a,$$

donc $L_H^1(H_2,\cdot)$ et $L_H^2(H_2,\cdot)$, l'espérance conditionnelle et le moment conditionnel d'ordre 2 de K_H, existent bien. De plus, presque sûrement,

$$L_H^1(H_2,0) = L_H^2(H_2,0) = 0.$$

Lemme 5.17 *Les fonctions $L_H^1(H_2,\cdot)$ et $L_H^2(H_2,\cdot)$ sont presque sûrement continues à droite en 0.*

Preuve 5.18 *Ce lemme découle directement de la définition du module d'uniforme continuité de déformation $K_H(\cdot)$, du théorème de la convergence dominé et de la continuité presque sûre du processus H.*

■

Le corollaire suivant nous montre, pour tout $i_0 \in \{1, \ldots, m\}$ fixé, la consistance de l'estimateur $\widehat{\phi \circ h_{i_0}^{-1}}_{mn}$ dans $(\mathcal{C}([a,b]), \|\cdot\|_\infty)$.

Corollaire 5.19 *Nous nous plaçons sous les hypothèses* (5.3) *et* (5.5)*. Soit $i_0 \in \{1, \ldots, m\}$ fixé. Alors, conditionnellement à $F_{i_0} = f_{i_0} \in A$,*

$$\left\|\widehat{\phi \circ h_{i_0}^{-1}}_{mn} - \phi \circ h_{i_0}^{-1}\right\|_\infty \xrightarrow[m,n\to+\infty]{ps} 0.$$

Preuve 5.20 *Soit $i_0 \in \{1, \ldots, m\}$ fixé. Pour tout $t \in [a,b]$, conditionnellement à $F_{i_0} = f_{i_0} \in A$, nous avons*

$$\begin{aligned}\forall i \in \{1, \ldots, m\} \setminus i_0, T_{j_i} &= \arg \min_{t_j \in \{t_0,\ldots,t_n\}} |F_i(t_j) - f_{i_0}(t_{j_0})| \\ &= \arg \min_{t_j \in \{t_0,\ldots,t_n\}} \left|F_i(t_j) - F_i \circ F_i^{-1} \circ f_{i_0}(t_{j_0})\right|,\end{aligned}$$

or, F_i étant strictement croissante,

$$F_i^{-1} \circ f_{i_0}(t_{j_0}) - \frac{1}{n} \leq T_{j_i} \leq F_i^{-1} \circ f_{i_0}(t_{j_0}) + \frac{1}{n}. \tag{5.13}$$

De plus, nous avons

$$t_{j_0} = \arg \min_{t_j \in \{t_0, \ldots, t_n\}} |t_j - t|,$$

donc

$$t - \frac{1}{n} \leq t_{j_0} \leq t + \frac{1}{n},$$

et, puisque f_{i_0} est strictement croissante,

$$f_{i_0}\left(t - \frac{1}{n}\right) \leq f_{i_0}(t_{j_0}) \leq f_{i_0}\left(t + \frac{1}{n}\right).$$

L'inégalité (5.13) *implique donc l'inégalité*

$$F_i^{-1} \circ f_{i_0}\left(t - \frac{1}{n}\right) - \frac{1}{n} \leq T_{j_i} \leq F_i^{-1} \circ f_{i_0}\left(t + \frac{1}{n}\right) + \frac{1}{n},$$

d'où

$$\overline{\phi \circ h_{i_0}^{-1}}_m \left(t - \frac{1}{n}\right) - \frac{1}{n} \leq \widehat{\phi \circ h_{i_0}^{-1}}_{mn}(t) \leq \overline{\phi \circ h_{i_0}^{-1}}_m \left(t + \frac{1}{n}\right) + \frac{1}{n}. \tag{5.14}$$

Pour montrer la convergence au sens de la définition 5.11, il suffit de montrer la convergence presque sûre des deux bornes de l'inégalité (5.14)*. Le traitement de ces deux quantités étant identique, nous nous intéressons seulement ici à la borne supérieure. Nous avons*

$$\begin{aligned}
&\left|\overline{\phi \circ h_{i_0}^{-1}}_m \left(t + \frac{1}{n}\right) + \frac{1}{n} - \phi \circ h_{i_0}^{-1}(t)\right| \\
&= \left|\overline{\phi \circ h_{i_0}^{-1}}_m \left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}_m(t) + \frac{1}{n} + \overline{\phi \circ h_{i_0}^{-1}}_m(t) - \phi \circ h_{i_0}^{-1}(t)\right| \\
&\leq \left|\overline{\phi \circ h_{i_0}^{-1}}_m \left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}_m(t)\right| + \frac{1}{n} + \left|\overline{\phi \circ h_{i_0}^{-1}}_m(t) - \phi \circ h_{i_0}^{-1}(t)\right|.
\end{aligned}$$

Donc

$$\begin{aligned}
&\sup_{t \in [a,b]} \left|\overline{\phi \circ h_{i_0}^{-1}}_m \left(t + \frac{1}{n}\right) + \frac{1}{n} - \phi \circ h_{i_0}^{-1}(t)\right| \\
&\leq \sup_{t \in [a,b]} \left|\overline{\phi \circ h_{i_0}^{-1}}_m \left(t + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}_m(t)\right| + \frac{1}{n} + \left\|\overline{\phi \circ h_{i_0}^{-1}}_m - \phi \circ h_{i_0}^{-1}\right\|_\infty.
\end{aligned}$$

D'après le corollaire 5.7, nous avons

$$\left\|\overline{\phi \circ h_{i_0}^{-1}}_m - \phi \circ h_{i_0}^{-1}\right\|_\infty \xrightarrow[m \to +\infty]{ps} 0.$$

Il nous suffit donc de montrer que

$$\sup_{t\in[a,b]}\left|\overline{\phi\circ h_{i_0}^{-1}}_m\left(t+\frac{1}{n}\right)-\overline{\phi\circ h_{i_0}^{-1}}_m(t)\right|\xrightarrow[m,n\to+\infty]{ps}0.$$

Nous avons, pour tout $t\in[a,b]$,

$$\begin{aligned}
&\left|\overline{\phi\circ h_{i_0}^{-1}}_m\left(t+\frac{1}{n}\right)-\overline{\phi\circ h_{i_0}^{-1}}_m(t)\right|\\
&=\left|\frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m}F_i^{-1}\circ f_{i_0}\left(t+\frac{1}{n}\right)-\frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m}F_i^{-1}\circ f_{i_0}(t)\right|\\
&\leq\frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m}\left|H_i\circ h_{i_0}^{-1}\left(t+\frac{1}{n}\right)-H_i\circ h_{i_0}^{-1}(t)\right|\\
&\leq\frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m}\sup_{\substack{(t_1,t_2)\in[a,b]^2\\ |t_1-t_2|\leq\frac{1}{n}}}\left|H_i\circ h_{i_0}^{-1}(t_1)-H_i\circ h_{i_0}^{-1}(t_2)\right|,
\end{aligned}$$

d'où

$$\begin{aligned}
&\sup_{t\in[a,b]}\left|\overline{\phi\circ h_{i_0}^{-1}}_m\left(t+\frac{1}{n}\right)-\overline{\phi\circ h_{i_0}^{-1}}_m(t)\right|\\
&\leq\frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m}\sup_{\substack{(t_1,t_2)\in[a,b]^2\\ |t_1-t_2|\leq\frac{1}{n}}}\left|H_i\circ h_{i_0}^{-1}(t_1)-H_i\circ h_{i_0}^{-1}(t_2)\right|,
\end{aligned}\tag{5.15}$$

avec, pour tout $n\in\mathbb{N}^*$, *les variables aléatoires*

$$\left(\tilde{K}_H^i\left(\frac{1}{n}\right)=\sup_{\substack{(t_1,t_2)\in[a,b]^2\\ |t_1-t_2|\leq\frac{1}{n}}}\left|H_i\circ h_{i_0}^{-1}(t_1)-H_i\circ h_{i_0}^{-1}(t_2)\right|\right)_{i=1,\dots,m,\, i\neq i_0}$$

sont i.i.d. et de même loi que

$$K_H\left(\frac{1}{n}\right)\Big|\,H_2=h_{i_0}.$$

Soit $\varepsilon>0$. *Puisque par hypothèse* $f_{i_0}=f\circ h_{i_0}^{-1}\in A$, *alors* h_{i_0} *est continue. D'où, d'après le lemme 5.17,* $L_H^1(H_2=h_{i_0},\cdot)$ *est continue à droite en 0, et donc il existe une valeur* $\eta_\varepsilon>0$, *telle que*

$$L_H^1(H_2=h_{i_0},\eta_\varepsilon)<\varepsilon.$$

De plus, la monotonie du module de continuité K_H *nous permet d'assurer que pour un* $n\in\mathbb{N}$ *assez grand* ($n>\frac{1}{\eta_\varepsilon}$) *l'inégalité* (5.15) *devient*

$$\sup_{t\in[a,b]}\left|\overline{\phi\circ h_{i_0}^{-1}}_m\left(t+\frac{1}{n}\right)-\overline{\phi\circ h_{i_0}^{-1}}_m(t)\right|\leq\frac{1}{m-1}\sum_{\substack{i=1\\ i\neq i_0}}^{m}\tilde{K}_H^i(\eta_\varepsilon).$$

Par la loi forte des grands nombres, nous déduisons de cette inégalité que

$$\begin{aligned}0 \leq & \lim_{m,n\to+\infty} \sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}_m \left(t+\frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}_m(t) \right| \\ \leq & \lim_{m\to+\infty} \frac{1}{m-1} \sum_{\substack{i=1\\ i\neq i_0}}^{m} \tilde{K}_H^i(\eta_\varepsilon) = L_H^1(H_2 = h_{i_0}, \eta_\varepsilon) < \varepsilon.\end{aligned}$$

Ceci étant vrai pour tout $\varepsilon > 0$, alors, presque sûrement, nous avons

$$\lim_{m,n\to+\infty} \sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}_m \left(t+\frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}_m(t) \right| = 0.$$

D'où la convergence presque sûre de la borne supérieure de l'inégalité (5.14),

$$\sup_{t\in[a,b]} \left| \overline{\phi \circ h_{i_0}^{-1}}_m \left(t+\frac{1}{n}\right) + \frac{1}{n} - \phi \circ h_{i_0}^{-1}(t) \right| \xrightarrow[m,n\to+\infty]{ps} 0.$$

D'où le résultat.

■

Comme pour les résultats de convergence en loi précédents, théorème 5.5, corollaire 5.9 et théorème 5.14, nous montrons dans le corollaire suivant une version vectorielle de la normalité asymptotique de l'estimateur empirique $\widehat{\phi \circ h_{i_0}^{-1}}_{mn}$.

Corollaire 5.21 *Nous nous plaçons sous les hypothèses* (5.3) *et* (5.5). *Soit $i_0 \in \{1,\ldots,m\}$ fixé. Soit, de plus, $n = m^{\frac{1}{2}+\alpha}$ avec $\alpha > 0$. Alors, conditionnellement à $F_{i_0} = f_{i_0} \in A$, pour tout vecteur $t = (t_1,\ldots,t_k)'$ de $[a,b]^k$, $k \in \mathbb{N}^*$,*

$$\sqrt{m}\left(\begin{pmatrix} \widehat{\phi \circ h_{i_0}^{-1}}_{mn}(t_1) \\ \vdots \\ \widehat{\phi \circ h_{i_0}^{-1}}_{mn}(t_k) \end{pmatrix} - \begin{pmatrix} \phi \circ h_{i_0}^{-1}(t_1) \\ \vdots \\ \phi \circ h_{i_0}^{-1}(t_k) \end{pmatrix}\right) \xrightarrow[m\to+\infty]{\mathcal{L}} \mathcal{N}_k\left(0,\Gamma_0(t)\right),$$

où $\Gamma_0(t)$ est la matrice de variance-covariance définie dans le corollaire 5.9.

Preuve 5.22 *Soit $i_0 \in \{1,\ldots,m\}$ fixé. De plus, soit $n = m^{\frac{1}{2}+\alpha}$ avec $\alpha > 0$, alors, d'après l'inégalité* (5.14) *de la démonstration du corollaire 5.19, nous avons pour tout $t_i \in \{t_1,\ldots,t_k\}$,*

$$\begin{aligned}& \sqrt{m}\left(\overline{\phi \circ h_{i_0}^{-1}}_m \left(t_i - \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}_m(t_i)\right) \\ & + \sqrt{m}\left(\overline{\phi \circ h_{i_0}^{-1}}_m(t_i) - \phi \circ h_{i_0}^{-1}(t_i)\right) - \frac{1}{m^\alpha} \\ \leq & \sqrt{m}\left(\widehat{\phi \circ h_{i_0}^{-1}}_{mn}(t_i) - \phi \circ h_{i_0}^{-1}(t_i)\right) \\ \leq & \sqrt{m}\left(\overline{\phi \circ h_{i_0}^{-1}}_m \left(t_i + \frac{1}{n}\right) - \overline{\phi \circ h_{i_0}^{-1}}_m(t_i)\right) \\ & + \sqrt{m}\left(\overline{\phi \circ h_{i_0}^{-1}}_m(t_i) - \phi \circ h_{i_0}^{-1}(t_i)\right) + \frac{1}{m^\alpha}.\end{aligned} \tag{5.16}$$

Pour avoir la normalité asymptotique, d'après le corollaire 5.9, il nous suffit de montrer que

$$\sqrt{m}\left(\begin{pmatrix}\widehat{\phi\circ h_{i_0}^{-1}}_{mn}(t_1)\\ \vdots\\ \widehat{\phi\circ h_{i_0}^{-1}}_{mn}(t_k)\end{pmatrix}-\begin{pmatrix}\phi\circ h_{i_0}^{-1}(t_1)\\ \vdots\\ \phi\circ h_{i_0}^{-1}(t_k)\end{pmatrix}\right)$$
$$=\sqrt{m}\left(\begin{pmatrix}\overline{\phi\circ h_{i_0}^{-1}}_{m}(t_1)\\ \vdots\\ \widehat{\phi\circ h_{i_0}^{-1}}_{m}(t_k)\end{pmatrix}-\begin{pmatrix}\phi\circ h_{i_0}^{-1}(t_1)\\ \vdots\\ \phi\circ h_{i_0}^{-1}(t_k)\end{pmatrix}\right)+o_{\mathbf{P}}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix},$$

or, pour cela, il nous suffit de montrer que pour tout $i\in\{1,\dots,k\}$,

$$\sqrt{m}\left(\widehat{\phi\circ h_{i_0}^{-1}}_{mn}(t_i)-\phi\circ h_{i_0}^{-1}(t_i)\right)$$
$$=\sqrt{m}\left(\overline{\phi\circ h_{i_0}^{-1}}_{m}(t_i)-\phi\circ h_{i_0}^{-1}(t_i)\right)+o_{\mathbf{P}}(1),$$

et, d'après l'inégalité 5.16, cela revient à montrer que

$$\sqrt{m}\left(\overline{\phi\circ h_{i_0}^{-1}}_{m}\left(t_i-\frac{1}{n}\right)-\overline{\phi\circ h_{i_0}^{-1}}_{m}(t_i)\right)=o_{\mathbf{P}}(1)$$

et

$$\sqrt{m}\left(\overline{\phi\circ h_{i_0}^{-1}}_{m}\left(t_i+\frac{1}{n}\right)-\overline{\phi\circ h_{i_0}^{-1}}_{m}(t_i)\right)=o_{\mathbf{P}}(1).$$

Ces deux résultats se traitant de manière identique, nous nous contenterons de montrer ici le résultat pour la seconde quantité. D'après l'inégalité (5.15) *de la démonstration du corollaire 5.19, nous avons*

$$\left|\sqrt{m}\left(\overline{\phi\circ h_{i_0}^{-1}}_{m}\left(t_i+\frac{1}{n}\right)-\overline{\phi\circ h_{i_0}^{-1}}_{m}(t_i)\right)\right|\leq\sqrt{m}\frac{1}{m-1}\sum_{\substack{j=1\\ j\neq i_0}}^{m}\tilde{K}_H^j\left(\frac{1}{n}\right),$$

avec, pour tout $j\in\{1,\dots,m\}\setminus i_0$,

$$\tilde{K}_H^j\left(\frac{1}{n}\right)=K_H\left(\frac{1}{n}\right)\Big|H_2=h_{i_0}$$
$$=\sup_{\substack{(s_1,s_2)\in[a,b]^2\\ |s_1-s_2|\leq\frac{1}{n}}}\left|H_j\circ h_{i_0}^{-1}(s_1)-H_j\circ h_{i_0}^{-1}(s_2)\right|.$$

Comme nous l'avons vu dans la démonstration du corollaire 5.19, les variables aléatoires $\left(\tilde{K}_H^j\left(\frac{1}{n}\right)\right)_{j=1,\dots,m,\,j\neq i_0}$ *sont i.i.d.. De plus, en posant*

$$Z_m=\sqrt{m}\frac{1}{m-1}\sum_{\substack{j=1\\ j\neq i_0}}^{m}\tilde{K}_H^j\left(\frac{1}{n}\right),$$

nous avons

$$\begin{aligned}\mathrm{Var}\,(Z_m) &= \frac{m^2}{(m-1)^2}\mathrm{Var}\left(\tilde{K}_H^1\left(\frac{1}{n}\right)\right)\\ &= \frac{m^2}{(m-1)^2}\left(L_H^2\left(H_2 = h_{i_0},\frac{1}{n}\right) - \left(L_H^1\left(H_2 = h_{i_0},\frac{1}{n}\right)\right)^2\right).\end{aligned}$$

Or, d'après le lemme 5.17, puisque $n = m^{\frac{1}{2}+\alpha}$ *avec* $\alpha > 0$, *nous avons*

$$\mathrm{Var}\,(Z_m) \xrightarrow[m\to+\infty]{} 0.$$

D'où le résultat.

■

5.2.3 Estimation de l'espérance structurelle

Dans les deux parties précédentes (5.2.1 et 5.2.2) nous avons étudié des estimateurs de l'inverse de l'espérance structurelle $(\phi \circ f^{-1})$ et des fonctions de déformation individuelles $\left(\phi \circ h_i^{-1}, i = 1,\ldots,m\right)$. Dans cette partie, nous allons définir un estimateur de l'espérance structurelle elle-même.

L'estimateur de l'inverse de l'espérance structurelle est défini, pour tout $y \in \mathcal{I}$ fixé, par

$$\widehat{\phi \circ f^{-1}}_{mn}(y) = \frac{1}{m}\sum_{i=1}^{m} T_{j_i},$$

où

$$\forall i = 1,\ldots,m, T_{j_i} = \arg\min_{t_j \in \{t_0,\ldots,t_n\}} |F_i(t_j) - y|\,.$$

Par construction, la fonction $\widehat{\phi \circ f^{-1}}_{mn}$ est une fonction en escalier croissante, avec un nombre $K(m,n)$ de sauts en les points $(v_k)_{k=1,\ldots,K(m,n)}$ tels que

$$f(a) = v_0 < v_1 < \cdots < v_{K(m,n)} < v_{K(m,n)+1} = f(b).$$

Donc, pour tout $y \in \mathcal{I} = [f(a),f(b)] \setminus (v_k)_{k=1,\ldots,K(m,n)}$, nous pouvons écrire

$$\widehat{\phi \circ f^{-1}}_{mn}(y) = \sum_{k=0}^{K(m,n)} u_k \mathbb{1}_{]v_k,v_{k+1}[}(y),$$

où les valeurs $(u_k)_{k=0,\ldots,K(m,n)}$ sont telles que

$$a = u_0 < u_1 < \cdots < u_{K(m,n)-1} < u_{K(m,n)} = b.$$

Ainsi, de façon naturelle, nous définissons un estimateur de l'espérance structurelle $f \circ \phi^{-1}$ de la façon suivante. Pour tout $t \in [a,b]$, nous posons

$$\widehat{f \circ \phi^{-1}}_{mn}(t) = \sum_{k=0}^{K(m,n)-1}\left(\frac{v_{k+1}-v_k}{u_{k+1}-u_k}(t-u_k) + v_k\right)\mathbb{1}_{[u_k,u_{k+1}[}(t) + v_{K(m,n)}\mathbb{1}_{\{b\}}(t).$$

Par construction, notre estimateur $\widehat{f \circ \phi^{-1}}_{mn}$ est continu et strictement croissant sur l'intervalle $[a,b]$. De plus, le théorème suivant nous assure sa consistance.

Théorème 5.23 *Sous les hypothèses* (5.3) *et* (5.5)*, l'estimateur empirique* $\widehat{f \circ \phi^{-1}}_{mn}$ *converge presque sûrement vers l'espérance structurelle* $f \circ \phi^{-1}$*:*

$$\left\| \widehat{f \circ \phi^{-1}}_{mn} - f \circ \phi^{-1} \right\|_\infty \xrightarrow[m,n \to +\infty]{ps} 0.$$

Démonstration 5.24 *Nous montrons, dans un premier temps, que notre estimateur empirique* $\widehat{f \circ \phi^{-1}}_{mn}$ *converge ponctuellement vers* $f \circ \phi^{-1}$*:*

$$\widehat{f \circ \phi^{-1}}_{mn} \rightsquigarrow f \circ \phi^{-1}.$$

A une transformation linéaire près, notre estimateur $\widehat{\phi \circ f^{-1}}_{mn}$ *peut être considéré comme une fonction de répartition empirique. Ainsi, en utilisant le lemme 21.2 de [vdV98], nous avons l'équivalence suivante sur les convergences simples:*

$$\widehat{\phi \circ f^{-1}}_{mn} \rightsquigarrow \phi \circ f^{-1} \Leftrightarrow \left(\widehat{\phi \circ f^{-1}}_{mn} \right)^{-1} \rightsquigarrow \left(\phi \circ f^{-1} \right)^{-1},$$

où, pour toute fonction $F : \mathcal{I} \to [a,b]$*, l'opérateur* F^{-1} *est appelé inverse généralisée de* F*, et est défini par*

$$\begin{array}{rccl} F^{-1} : & [a,b] & \to & \mathcal{I} \\ & t & \mapsto & F^{-1}(t) = \inf \{ y \in \mathcal{I};\ F(y) \geq t \}. \end{array}$$

Donc, pour tout $t \in [a,b]$*, puisque* $\phi \circ f^{-1}$ *est continue et strictement croissante, nous avons*

$$\left(\widehat{\phi \circ f^{-1}}_{mn} \right)^{-1}(t) \xrightarrow[m,n \to +\infty]{ps} \left(\phi \circ f^{-1} \right)^{-1}(t) = f \circ \phi^{-1}(t).$$

De plus, pour tout $t \in [a,b]$*, nous avons*

$$\left(\widehat{\phi \circ f^{-1}}_{mn} \right)^{-1}(t) = v_0 \mathbb{1}_{\{a\}}(t) + \sum_{k=0}^{K(m,n)-1} v_{k+1} \mathbb{1}_{]u_k, u_{k+1}]}(t).$$

Il nous suffit maintenant, pour avoir la convergence simple de notre estimateur $\widehat{f \circ \phi^{-1}}_{mn}$*, de montrer que, pour tout* $t \in [a,b]$*,*

$$\left| \widehat{f \circ \phi^{-1}}_{mn}(t) - \left(\widehat{\phi \circ f^{-1}}_{mn} \right)^{-1}(t) \right| \xrightarrow[m,n \to +\infty]{ps} 0.$$

Or, soit $t \in]u_{k(m,n)}, u_{k(m,n)+1}] \subset [a,b]$*, avec* $k(m,n) \in \{0, \ldots, K(m,n) - 1\}$*, nous avons, par construction,*

$$v_{k(m,n)} \leq \widehat{f \circ \phi^{-1}}_{mn}(t) \leq v_{k(m,n)+1},$$

donc, nous avons

$$v_{k(m,n)} - v_{k(m,n)+1} \leq \widehat{f \circ \phi^{-1}}_{mn}(t) - \left(\widehat{\phi \circ f^{-1}}_{mn}\right)^{-1}(t) \leq 0.$$

De plus, nous avons une fonction limite $f \circ \phi^{-1}$ qui est continue, donc

$$v_{k(m,n)} - v_{k(m,n)+1} \xrightarrow[m,n\to+\infty]{} 0.$$

D'où la convergence simple.

Pour conclure, nous faisons appel au théorème de Dini (voir [Wag95]) qui nous dit qu'une suite de fonctions croissantes continues sur un intervalle $[a,b]$ converge simplement vers une fonction croissante continue si, et seulement si, elle converge uniformément. D'où notre convergence uniforme.

■

5.2.4 Réduction des hypothèses sur le processus de déformation

Les hypothèses (5.3.ii) faites sur le processus de déformation H dans la section 5.1 sont assez restrictives. En effet, ces hypothèses sont $H(a) \overset{ps}{=} a$ et $H(b) \overset{ps}{=} b$. Nous montrons ici que nous pouvons affaiblir ces hypothèses, en remplaçant (5.3.ii) par

ii′) $H^{-1}(\cdot,a)$ et $H^{-1}(\cdot,b)$ sont des variables aléatoires et leurs lois sont à support compact, avec de plus

$$\sup_{w\in\Omega} H^{-1}(w,a) \overset{ps}{<} \inf_{w\in\Omega} H^{-1}(w,b).$$

Nous noterons (5.3′) les hypothèses (5.3) dans lesquelles nous avons remplacé l'hypothèse (5.3.ii) par l'hypothèse (5.3.ii′).

Nous nous plaçons ici dans le cadre du modèle théorique (5.6) et nous nous contentons de montrer la consistance de l'estimateur empirique (5.7) sous les hypothèses (5.3′). En effet, pour la normalité asymptotique et pour l'extension au modèle d'échantillonnage (5.2), les calculs sont similaires.

L'estimateur (5.7) est

$$\overline{\phi \circ f^{-1}}_m = \frac{1}{m}\sum_{i=1}^{m} F_i^{-1}.$$

Ici, cet estimateur ne peut pas être calculé sur la totalité de l'intervalle $\mathcal{I}$. En effet, avec l'hypothèse (5.3.ii′), les fonctions $(f_i)_{i=1,\ldots,m}$ peuvent ne pas être toutes à valeurs dans $\mathcal{I}$. Par contre, puisque ces trajectoires sont croissantes, elles sont toutes à valeurs dans l'intervalle

$$\mathcal{I}_m = \left[\max_{i\in\{1,\ldots,m\}} f_i(a), \min_{i\in\{1,\ldots,m\}} f_i(b)\right].$$

En considérant les estimateurs associés aux bornes de cet intervalle, c'est-à-dire

$$\max_{i\in\{1,\dots,m\}} F_i(a) \text{ et } \min_{i\in\{1,\dots,m\}} F_i(b),$$

puisque $H^{-1}(a)$ et $H^{-1}(b)$ sont à valeurs dans un compact, nous avons les convergences presque sûres des extrêmes:

$$\max_{i\in\{1,\dots,m\}} F_i(a) \xrightarrow[m\to+\infty]{ps} \sup_{w\in\Omega} f\circ H^{-1}(w,a)$$

et

$$\min_{i\in\{1,\dots,m\}} F_i(b) \xrightarrow[m\to+\infty]{ps} \inf_{w\in\Omega} f\circ H^{-1}(w,b).$$

Comme la fonction f est continue et strictement croissante, cela se traduit par

$$\max_{i\in\{1,\dots,m\}} F_i(a) \xrightarrow[m\to+\infty]{ps} f\left(\sup_{w\in\Omega} H^{-1}(w,a)\right)$$

et

$$\min_{i\in\{1,\dots,m\}} F_i(b) \xrightarrow[m\to+\infty]{ps} f\left(\inf_{w\in\Omega} H^{-1}(w,b)\right).$$

De plus, l'hypothèse (5.3′) nous assure que l'intervalle

$$\tilde{\mathcal{I}} = \left[f\left(\sup_{w\in\Omega} H^{-1}(w,a)\right), f\left(\inf_{w\in\Omega} H^{-1}(w,b)\right)\right]$$

n'est pas vide. Dans le théorème suivant, nous montrons la consistance de l'estimateur empirique (5.7) dans $\left(\mathcal{C}\left(\tilde{\mathcal{I}}\right), \|\cdot\|_\infty\right)$.

Théorème 5.25 *Sous les hypothèses (5.3′) et* (5.5)*, l'estimateur* $\overline{\phi\circ f^{-1}}_m$ *converge presque sûrement vers* $\phi\circ f^{-1}$ *dans* $\left(\mathcal{C}\left(\tilde{\mathcal{I}}\right), \|\cdot\|_\infty\right)$*:*

$$\left\|\overline{\phi\circ f^{-1}}_m - \phi\circ f^{-1}\right\|_\infty \xrightarrow[m\to+\infty]{ps} 0.$$

Démonstration 5.26 *Cette démonstration est la même que la démonstration du théorème 5.3. Il s'agit d'une application directe du corollaire 7.10 page 189 de [LT91]. Notre estimateur*

$$\begin{aligned}\overline{\phi\circ f^{-1}}_m &= \frac{1}{m}\sum_{i=1}^{m} F_i^{-1}\\ &= \frac{1}{m}\sum_{i=1}^{m} H_i\circ f^{-1}\end{aligned}$$

est bien défini et strictement croissant sur l'intervalle $\tilde{\mathcal{I}}$*. De plus, en posant*

$$X_i = H_i\circ f^{-1} - \phi\circ f^{-1}, i\in\mathbb{N}^*$$

et

$$S_n = \sum_{i=1}^{n} X_i, n \in \mathbb{N}^*,$$

les variables aléatoires $(X_i)_{i=1,\ldots,m}$ *sont, par définition, boréliennes à valeurs dans* $B = \mathcal{C}(\tilde{\mathcal{I}})$ *où* $\tilde{\mathcal{I}}$ *est un compact de* $\mathbb{R}$*, et* B *est un espace de Banach séparable. De plus, le dual de* B *est l'ensemble des mesures bornées sur* $\tilde{\mathcal{I}}$*. Nous sommes donc bien dans le cadre du chapitre 7 de [LT91] et nous pouvons appliquer le corollaire 7.10. D'où le résultat.*

■

La figure 5.3 illustre l'estimation de la moyenne structurelle obtenue pour deux courbes. Ces deux courbes représentent les valeurs prises sur l'intervalle [0,1] par deux fonctions de répartition gaussienne, d'espérances différentes ($\mu_1 \neq \mu_2$) et de même variance. L'inverse de la moyenne structurelle (en pointillés) est ici calculée sur l'intervalle $\mathcal{I}_2$ représenté par les deux droites horizontales en tirés. Cette moyenne structurelle correspond à la fonction de répartition d'une gaussienne d'espérance $\frac{\mu_1+\mu_2}{2}$.

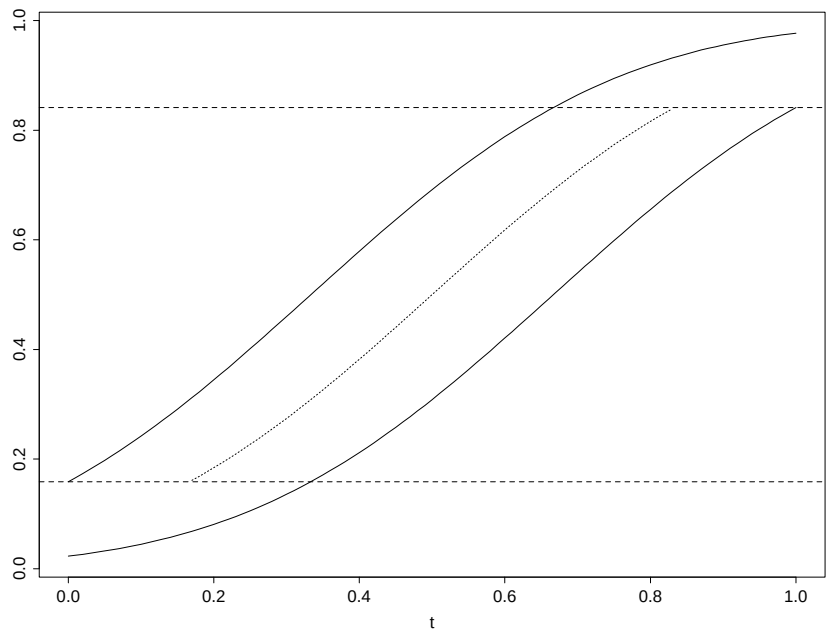

FIG. 5.3 – *Les courbes en trait plein correspondent à deux fonctions de répartition gaussienne, avec deux espérances différentes* ($\mu_1 \neq \mu_2$) *et une même variance. La courbe en pointillés correspond à la moyenne structurelle. Cette moyenne est la fonction de répartition gaussienne d'espérance* $\frac{\mu_1+\mu_2}{2}$.

Avec un raisonnement identique, il est possible d'étendre les résultats relatifs à l'estimation des fonctions de déformation $\phi \circ h_{i_0}^{-1}$ avec $i_0 \in \{1, \ldots, m\}$ fixé. Il suffit

dans ce cas de restreindre les estimateurs à des valeurs de t appartenant à l'intervalle

$$\left[f_{i_0}^{-1}\left(\max_{i\in\{1,\dots,m\}} f_i(a)\right), f_{i_0}^{-1}\left(\min_{i\in\{1,\dots,m\}} f_i(b)\right)\right].$$

Comme précédemment, en considérant les estimateurs associés aux bornes de cet intervalle, conditionnellement à $F_{i_0} = f_{i_0} \in A$, nous avons

$$f_{i_0}^{-1}\left(\max_{i\in\{1,\dots,m\}} F_i(a)\right) \xrightarrow[m\to+\infty]{ps} f_{i_0}^{-1}\left(\sup_{w\in\Omega} f\circ H^{-1}(w,a)\right)$$

et

$$f_{i_0}^{-1}\left(\min_{i\in\{1,\dots,m\}} F_i(b)\right) \xrightarrow[m\to+\infty]{ps} f_{i_0}^{-1}\left(\inf_{w\in\Omega} f\circ H^{-1}(w,b)\right).$$

De plus, puisque f_{i_0} est strictement croissante, et en écrivant $f_{i_0} = f\circ h_{i_0}^{-1}$, donc $f_{i_0}^{-1} = h_{i_0}\circ f^{-1}$, nous obtenons

$$f_{i_0}^{-1}\left(\max_{i\in\{1,\dots,m\}} F_i(a)\right) \xrightarrow[m\to+\infty]{ps} h_{i_0}\left(\sup_{w\in\Omega} H^{-1}(w,a)\right)$$

et

$$f_{i_0}^{-1}\left(\min_{i\in\{1,\dots,m\}} f_i(b)\right) \xrightarrow[m\to+\infty]{ps} h_{i_0}\left(\inf_{w\in\Omega} H^{-1}(w,b)\right).$$

5.3 Cas où la fonction f n'est pas croissante

Dans cette section, nous généralisons notre modèle à une fonction f qui n'est pas strictement croissante. En effet, lorsque la fonction f n'est pas strictement croissante, les estimateurs définis dans la section 5.2 ne sont plus applicables. L'idée que nous développons maintenant est donc de transformer une fonction f quelconque en une fonction g strictement croissante. Afin de définir une telle fonction g, nous supposerons dorénavant que f est telle qu'il existe des réels $(s_k)_{k=1,\dots,r} \in]a,b[^r$ avec $r\in\mathbb{N}^*$, tels que

$$a = s_0 < s_1 < \cdots < s_r < s_{r+1} = b$$

et, pour tout $k\in\{0,\dots,r-1\}$,

$$\begin{gathered}\forall (t_1,t_2)\in]s_k,s_{k+1}[,\ \forall (t_3,t_4)\in]s_{k+1},s_{k+2}[,\\ t_1<t_2 \text{ et } t_3<t_4 \Rightarrow (f(t_1)-f(t_2))\,(f(t_3)-f(t_4))<0.\end{gathered} \tag{5.17}$$

Notons,

$$\mathcal{F} = \{f : [a,b]\to\mathcal{I}\subset\mathbb{R} \text{ telle que } (5.17)\}$$

l'ensemble des fonctions strictement monotones sur un nombre fini d'intervalles contigus, dont le sens de variation est alterné. Il est donc naturel de construire une fonction croissante en utilisant la stricte monotonie de f sur chacun de ces intervalles.

5.3.1 Définition de notre transformation

Afin de simplifier les notations, nous définissons la fonction suivante.

Définition 5.27 *Soit $f \in \mathcal{F}$, soient $(s_k)_{k=1,\dots,r}$ définis par* (5.17)*, nous définissons la fonction Π par*

$$\begin{array}{rcl} \Pi : \ ]a,b[\setminus\{s_1,\dots,s_r\} \times \mathcal{F} & \to & \{-1,1\} \\ s_{l(t)} < t < s_{l(t)+1} & \mapsto & \Pi(t,f) = \left\{ \begin{array}{lll} -1 & si & s_{l(t)} - s_{l(t)+1} > 0, \\ 1 & si & s_{l(t)} - s_{l(t)+1} < 0, \end{array} \right. \end{array}$$

avec $l(t) \in \{0,\dots,r\}$. Nous noterons, à f fixé, $\pi = \Pi(\cdot,f)$.

Ainsi, la fonction π nous indique, pour tout $t \in]a,b[\setminus\{s_1,\dots,s_r\}$, le sens de variation de la fonction f sur l'intervalle $\left]s_{l(t)},s_{l(t)+1}\right[$. Nous utilisons cette fonction Π dans la définition de notre transformation.

Définition 5.28 *Soit $f \in \mathcal{F}$. La transformation $\mathcal{G}(\cdot,f)$ de f est définie, pour tout $t \in]a,b[\setminus\{s_0,\dots,s_{r+1}\}$, par*

$$\mathcal{G}(t,f) = f(t)\pi(t) - \sum_{k=0}^{r} \pi(t)f(s_k)\mathbb{1}_{]s_k,s_{k+1}[}(t) + f(s_0) + \sum_{k=1}^{r} |f(s_{k-1}) - f(s_k)|\mathbb{1}_{]s_k,b[}(t),$$

et, pour tout $k \in \{0,\dots,r+1\}$, $\mathcal{G}(\cdot,f)$ est définie par

$$\mathcal{G}(s_k,f) = f(a) + \sum_{l=1}^{k} |f(s_{l-1}) - f(s_l)|,$$

avec $\sum_{l=1}^{0} |f(s_{l-1}) - f(s_l)| = 0$. Nous noterons, à f fixée, $g = \mathcal{G}(\cdot,f)$.

5.3.2 Propriétés

Proposition 5.29 *Soit $f \in \mathcal{F}$. Alors, la fonction $\mathcal{G}(\cdot,f)$ est srictement croissante.*

Preuve 5.30 *Soit $f \in \mathcal{F}$. Soit $k \in \{0,1,\dots,r\}$, soit $t \in]s_k,s_{k+1}[$,*

$$g(t) = f(t)\pi(t) - \pi(t)f(s_k) + f(s_0) + \sum_{l=1}^{k} |f(s_{l-1}) - f(s_l)|,$$

en posant $\sum_{l=1}^{0} |f(s_{l-1}) - f(s_l)| = 0$. Ainsi,

$$g(t) - f(s_0) - \sum_{l=1}^{k} |f(s_{l-1}) - f(s_l)| = \pi(t)(f(t) - f(s_k)),$$

or, la fonction f est strictement monotone sur l'intervalle $]s_k,s_{k+1}[$, donc, si f est strictement croissante (resp. strictement décroissante), alors

$$f(\cdot) - f(s_k)$$

est strictement croissante (resp. strictement décroissante), de plus, pour tout $t \in]s_k,s_{k+1}[$,

$$\pi(t) = 1 \text{ (resp. } -1),$$

d'où

$$\pi(\cdot)(f(\cdot) - f(s_k))$$

est strictement croissante, d'où, g *est strictement croissante sur* $]s_k,s_{k+1}[$. *De plus, si* f *est strictement croissante sur l'intervalle* $]s_k,s_{k+1}[$,

$$g(t) = f(t) - f(s_k) + f(s_0) + \sum_{l=1}^{k} |f(s_{l-1}) - f(s_l)|,$$

or, d'après la définition,

$$g(s_k) = f(s_0) + \sum_{l=1}^{k} |f(s_{l-1}) - f(s_l)|$$

et

$$g(s_{k+1}) = f(s_0) + \sum_{l=1}^{k+1} |f(s_{l-1}) - f(s_l)|,$$

d'où, pour tout $t \in]s_k,s_{k+1}[$,

$$g(s_k) - g(t) = f(s_k) - f(t) < 0$$

et

$$g(t) - g(s_{k+1}) = f(t) - f(s_{k+1}) < 0,$$

donc

$$g(s_k) < g(t) < g(s_{k+1}).$$

Donc, g *est strictement croissante sur l'intervalle fermé* $[s_k,s_{k+1}]$. *Dans le cas où* f *est strictement décroissante sur l'intevalle* $]s_k,s_{k+1}[$ *, les calculs sont similaires. Donc, pour tout* $k = 0,\ldots,r$, *la fonction* g *est strictement croissante sur l'intervalle* $[s_k,s_{k+1}]$. *Donc,* g *est strictement croissante sur* $[a,b]$.

■

Notre opérateur $\mathcal{G}$ transforme donc une fonction $f \in \mathcal{F}$ en une fonction g strictement croissante. La figure 5.4 nous montre l'exemple de la transformation $\mathcal{G}$ appliquée à la fonction $f : [0,1] \to \mathcal{I} \subset \mathbb{R}$, définie par

$$t \mapsto f(t) = 1397t^6 - 3877t^5 + 4004t^4 - 1899t^3 + 409t^2 - 33t + 3. \tag{5.18}$$

La fonction f appartient bien à $\mathcal{F}$. Elle possède $r = 5$ optima locaux dans l'intervalle $]0,1[$, en les valeurs $s_1 \approx 0.06438$, $s_2 \approx 0.22752$, $s_3 \approx 0.44066$, $s_4 \approx 0.67062$ et $s_5 \approx 0.90951$. Ainsi, la figure 5.4 nous montre comment agit la transformation $\mathcal{G}$ sur la fonction f. Pour tout $k = 0,\ldots,r$, la fonction f est strictement monotone sur l'intervalle $[s_k,s_{k+1}]$ et la fonction g associée à cet intervalle n'est autre que

$$\begin{cases} \text{une translation positive de } f \text{ si } f \text{ est croissante,} \\ \text{une translation positive de } -f \text{ si } f \text{ est décroissante.} \end{cases}$$

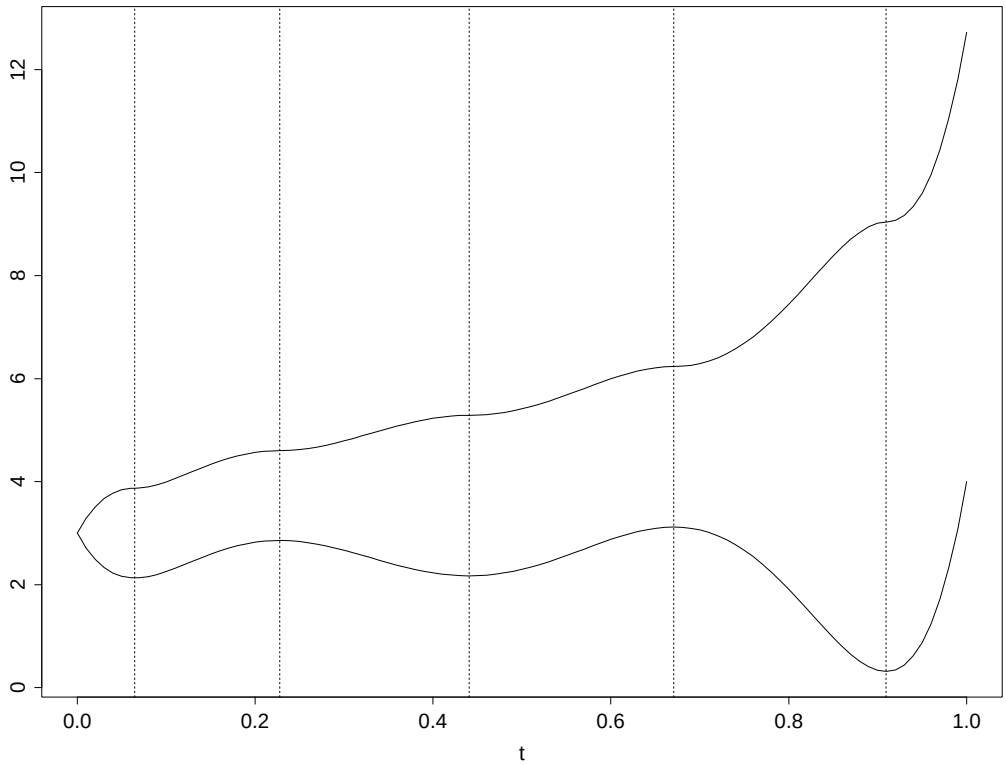

FIG. 5.4 – *Les courbes correspondent, d'une part à la fonction f définie par* (5.18), *et d'autre part, à la transformation associée g qui est strictement croissante. Les lignes en tirés verticales correspondent aux valeurs $(s_k)_{k=1,\ldots,r=5}$.*

Une autre propriété intéressante de l'opérateur $\mathcal{G}$ est que cette transformation appliquée sur les fonctions $(f_i)_{i=1,\ldots,m}$ nous replace dans le cadre du modèle (5.2) et donc dans le cadre des hypothèses des théorèmes 5.12 et 5.14 et des corollaires 5.19 et 5.21. La proposition suivante précise cette propriété.

Proposition 5.31 *Soient la fonction $f \in \mathcal{F}$ et les fonctions de déformation $(h_i)_{i=1,\ldots,m}$ définies par* (5.3). *Alors, les transformées par $\mathcal{G}$ des fonctions $(f_i)_{i=1,\ldots,m}$, notées $(g_i)_{i=1,\ldots,m}$, sont telles que*

$$g_i = g \circ h_i^{-1}.$$

Preuve 5.32 *Soit $i \in \{1,\ldots,m\}$ fixé. Il s'agit, dans un premier temps, de montrer que la fonction $f_i \in \mathcal{F}$. En effet, $f \in \mathcal{F}$, donc il existe des valeurs $(s_k)_{k=1,\ldots,r}$ définies par* (5.17). *Posons,*

$$\forall k \in \{1,\ldots,r\},\, s_k^i = h_i(s_k).$$

Ainsi, soit $k \in \{0,\ldots,r-1\}$, soient

$$(t_1,t_2) \in \left]s_k^i,s_{k+1}^i\right[,\, (t_3,t_4) \in \left]s_{k+1}^i,s_{k+2}^i\right[\;/\; t_1 < t_2 \text{ et } t_3 < t_4,$$

alors,

$$(f_i(t_1) - f_i(t_2))\,(f_i(t_3) - f_i(t_4)) = \left(f \circ h_i^{-1}(t_1) - f \circ h_i^{-1}(t_2)\right)\left(f \circ h_i^{-1}(t_3) - f \circ h_i^{-1}(t_4)\right),$$

or, la fonction de déformation h_i^{-1} est strictement croissante, donc

$$s_k < h_i^{-1}(t_1) < h_i^{-1}(t_2) < s_{k+1}$$

et

$$s_{k+1} < h_i^{-1}(t_3) < h_i^{-1}(t_3) < s_{k+2},$$

d'où, puisque $f \in \mathcal{F}$, nous avons

$$(f_i(t_1) - f_i(t_2))\,(f_i(t_3) - f_i(t_4)) < 0.$$

Donc $f_i \in \mathcal{F}$. Ainsi, nous pouvons calculer la transformée g_i de f_i. Pour tout $t \in]a,b[\setminus\{s_1^i, \ldots, s_r^i\}$, nous avons

$$g_i(t) = f_i(t)\Pi(t,f_i) - \sum_{k=0}^{r} \Pi(t,f_i) f_i(s_k^i) \mathbb{1}_{]s_k^i, s_{k+1}^i[}(t) + f_i(s_0) + \sum_{k=1}^{r} \left|f_i(s_{k-1}^i) - f_i(s_k^i)\right| \mathbb{1}_{]s_k^i, b[}(t),$$

or

$$\begin{aligned} f_i(t) &= f \circ h_i^{-1}(t), \\ \Pi(t,f_i) &= \Pi(t, f \circ h_i^{-1}) = \Pi(h_i^{-1}(t), f), \\ \mathbb{1}_{]s_k^i, s_{k+1}^i[}(t) &= \mathbb{1}_{]h_i(s_k), h_i(s_{k+1})[}(t) = \mathbb{1}_{]s_k, s_{k+1}[}\left(h_i^{-1}(t)\right). \end{aligned}$$

De plus, pour les valeurs $\left\{s_0^i, \ldots, s_{r+1}^i\right\}$, le calcul est direct. D'où

$$g_i = g \circ h_i^{-1}.$$

■

5.3.3 Application à notre modèle

Ainsi, d'après les propriétés du paragraphe précédent, la transformation $\mathcal{G}$ effectuée sur chacune des fonctions $(f_i)_{i=1,\ldots,m}$ nous permet de remplacer le modèle (5.2), à savoir

$$Y_{ij} = f \circ H_i^{-1}(t_j),\, i = 1, \ldots, m,\, j = 0, \cdots, n,$$

par le modèle

$$Z_{ij} = g \circ H_i^{-1}(t_j),\, i = 1, \ldots, m,\, j = 0, \cdots, n. \tag{5.19}$$

Le problème majeur, pour l'utilisation pratique de la transformation $\mathcal{G}$, est que nous connaissons chaque fonction f_i, $i = 1, \ldots, m$, sur la grille $(t_j)_{j=0,\ldots,n}$, mais que nous ne connaissons pas les valeurs $r \in \mathbb{N}$, $(s_k)_{k=1,\ldots,r}$ et $(s_k^i)_{i=1,\ldots,m,\, k=1,\ldots,r}$. Pour résoudre ce problème, c'est-à-dire pour pouvoir se placer dans le modèle (5.19), il est nécessaire de pouvoir approcher au mieux les valeurs $(Z_{ij})_{i=1,\ldots,m,\, j=0,\ldots,n}$, et ceci sans faire appel aux paramètres inconnus $r \in \mathbb{N}$, $(s_k)_{k=1,\ldots,r}$ et $(s_k^i)_{i=1,\ldots,m,\, k=1,\ldots,r}$.

D'après la structure de la transformation $\mathcal{G}$, il paraît naturel d'approcher les valeurs $(Z_{ij})_{i=1,\ldots,m,\, j=0,\ldots,n}$ par des accroissements successifs. C'est pourquoi nous définissons la suite de variables aléatoires (pour $i \in \{1,\ldots,m\}$ fixé)

$$\tilde{Z}_{i0} = Y_{i0}$$

et

$$\forall j \in \{1,\ldots,n\},\, \tilde{Z}_{ij} = \tilde{Z}_{ij-1} + |Y_{ij} - Y_{ij-1}|\,. \tag{5.20}$$

Proposition 5.33 *Soit $f \in \mathcal{F}$. Soient $i \in \{1,\ldots,m\}$ et $t \in [a,b]$ fixés. Soit une suite*

$$(j(n))_{n\in\mathbb{N}} \text{ avec } \frac{j(n)}{n} \xrightarrow[n\to+\infty]{} t.$$

Alors,

$$\tilde{Z}_{ij(n)} - Z_{ij(n)} \xrightarrow[n\to+\infty]{ps} 0.$$

Preuve 5.34 *Soit $i \in \{1,\ldots,m\}$ fixé. Soit $t \in]a,b[\setminus \{s_1^i,\ldots,s_r^i\}$ fixé. Soit $l \in \{0,1,\ldots,r\}$ tel que $t \in \left]s_l^i,s_{l+1}^i\right[$. Soit une suite*

$$(j(n))_{n\in\mathbb{N}} \text{ avec } \frac{j(n)}{n} \xrightarrow[n\to+\infty]{} t,$$

alors

$$\exists n_0 \in \mathbb{N} \,/\, \forall n \in \mathbb{N}, n \geq n_0 \Rightarrow \frac{j(n)}{n} \in \left]s_l^i,s_{l+1}^i\right[.$$

Soit $n \in \mathbb{N}$ avec $n \geq n_0$, nous avons

$$\begin{aligned}\tilde{Z}_{ij(n)} &= \tilde{Z}_{ij(n)-1} + \left|Y_{ij(n)} - Y_{ij(n)-1}\right| \\ &= Y_{i0} + \sum_{k=1}^{j(n)} |Y_{ik} - Y_{ik-1}|\,.\end{aligned}$$

De plus,

$$\begin{aligned}Z_{ij(n)} &= g \circ H_i^{-1}\left(t_{j(n)}\right) \\ &= F_i\left(t_{j(n)}\right)\Pi\left(t_{j(n)},F_i\right) + \Pi\left(t_{j(n)},F_i\right)F_i\left(s_l^i\right) + F_i(s_0) + \sum_{k=1}^{l}\left|F_i\left(s_{k-1}^i\right) - F_i\left(s_k^i\right)\right|,\end{aligned}$$

avec $\sum_{k=1}^{0}\left|F_i\left(s_{k-1}^i\right) - F_i\left(s_k^i\right)\right| = 0$, d'où,

$$\begin{aligned}Z_{ij(n)} &= \Pi\left(t_{j(n)},F_i\right)\left(Y_{ij(n)} - F_i\left(s_l^i\right)\right) + Y_{i0} + \sum_{k=1}^{l}\left|F_i\left(s_{k-1}^i\right) - F_i\left(s_k^i\right)\right| \\ &= \left|Y_{ij(n)} - F_i\left(s_l^i\right)\right| + Y_{i0} + \sum_{k=1}^{l}\left|F_i\left(s_{k-1}^i\right) - F_i\left(s_k^i\right)\right| \\ &= A + Y_{i0} + B.\end{aligned}$$

Notons, pour tout $k = 1, \ldots, l+1$,

$$s_{k-1}^i \leq t_{j_k - p_k} < \cdots < t_{j_k} \leq s_k^i.$$

Ainsi, pour tout $k = 1, \ldots, l$, *nous avons*

$$\begin{aligned}
&\left|F_i\left(s_{k-1}^i\right) - F_i\left(s_k^i\right)\right| \\
&= \left|F_i\left(s_{k-1}^i\right) - F_i\left(t_{j_k-p_k}\right) + \sum_{q=1}^{p_k}\left(F_i\left(t_{j_k-q}\right) - F_i\left(t_{j_k-q+1}\right)\right) + F_i\left(t_{j_k}\right) - F_i\left(s_k^i\right)\right| \\
&= \left|F_i\left(s_{k-1}^i\right) - F_i\left(t_{j_k-p_k}\right)\right| + \sum_{q=1}^{p_k}\left|F_i\left(t_{j_k-q}\right) - F_i\left(t_{j_k-q+1}\right)\right| + \left|F_i\left(t_{j_k}\right) - F_i\left(s_k^i\right)\right| \\
&= \left|F_i\left(s_{k-1}^i\right) - Y_{ij_k-p_k}\right| + \sum_{q=1}^{p_k}\left|Y_{ij_k-q+1} - Y_{ij_k-q}\right| + \left|Y_{ij_k} - F_i\left(s_k^i\right)\right|,
\end{aligned}$$

donc

$$\begin{aligned}
B &= \sum_{k=1}^{l}\left|F_i\left(s_{k-1}^i\right) - F_i\left(s_k^i\right)\right| \\
&= \sum_{k=1}^{l}\left|F_i\left(s_{k-1}^i\right) - Y_{ij_k-p_k}\right| + \sum_{k=1}^{l}\left|Y_{ij_k} - F_i\left(s_k^i\right)\right| + \sum_{k=1}^{l}\sum_{q=1}^{p_k}\left|Y_{ij_k-q+1} - Y_{ij_k-q}\right| \\
&= \sum_{k=1}^{l}\left|F_i\left(s_{k-1}^i\right) - Y_{ij_k-p_k}\right| + \sum_{k=1}^{l}\left|Y_{ij_k} - F_i\left(s_k^i\right)\right| + \sum_{k=1}^{j_l}\left|Y_{ik} - Y_{ik-1}\right| \\
&\quad - \sum_{k=1}^{l-1}\left|Y_{ij_{k+1}-p_{k+1}} - Y_{ij_{k+1}-p_{k+1}-1}\right| - \left|Y_{i1} - Y_{i0}\right|.
\end{aligned}$$

De même, nous avons

$$A = \left|Y_{ij(n)} - F_i\left(s_l^i\right)\right| =$$
$$\sum_{q=j_{l+1}-j(n)+1}^{p_{l+1}+1}\left|Y_{ij_{l+1}-q+1} - Y_{ij_{l+1}-q}\right| + \left|Y_{ij_{l+1}-p_{l+1}} - F_i\left(s_l^i\right)\right| - \left|Y_{ij_{l+1}-p_{l+1}} - Y_{ij_{l+1}-p_{l+1}-1}\right|.$$

Donc

$$\begin{aligned}
Z_{ij(n)} - \tilde{Z}_{ij(n)} &= \sum_{k=1}^{l+1}\left|F_i\left(s_{k-1}^i\right) - Y_{ij_k-p_k}\right| + \sum_{k=1}^{l}\left|Y_{ij_k} - F_i\left(s_k^i\right)\right| \\
&\quad - \sum_{k=1}^{l}\left|Y_{ij_{k+1}-p_{k+1}} - Y_{ij_{k+1}-p_{k+1}-1}\right| - \left|Y_{i1} - Y_{i0}\right|.
\end{aligned}$$

D'où, par continuité de la fonction f, F_i *est continue, donc*

$$Z_{ij(n)} - \tilde{Z}_{ij(n)} \xrightarrow[n\to+\infty]{ps} 0.$$

Pour les valeurs de $t \in \left\{s_0^i, \ldots, s_{r+1}^i\right\}$, *les calculs sont similaires.*

■

Comme pour les fonctions aléatoires $(F_i)_{i=1,\dots,m}$, mais cette fois-ci de façon déterministe, nous définissons la suite

$$(y_j = f(t_j))_{j=0,\dots,n}$$

et la suite croissante associée

$$(z_j = g(t_j))_{j=0,\dots,n}\,.$$

De la même façon que dans la définition (5.20), nous définissons une suite $(\tilde{z}_j)_{j=0,\dots,n}$ telle que

$$\tilde{z}_0 = y_0$$

et

$$\forall j \in \{1,\dots,n\},\, \tilde{z}_j = \tilde{z}_{j-1} + |y_j - y_{j-1}|\,. \tag{5.21}$$

Nous déduisons directement de la proposition 5.33 que pour une valeur de $t \in [a,b]$ fixée, et pour une suite

$$(j(n))_{n\in\mathbb{N}} \text{ avec } \frac{j(n)}{n} \xrightarrow[n\to+\infty]{} t,$$

nous avons

$$\tilde{z}_{j(n)} = z_{j(n)} + o(1).$$

La figure 5.5 nous montre le graphe de la fonction f définie en (5.18) et de sa transformée $\mathcal{G}(\cdot,f)$, ainsi que les valeurs des suites $(y_j)_{j=0,\dots,n}$ et $(\tilde{z}_j)_{j=0,\dots,n}$ pour différentes valeurs de n. L'approximation de la fonction $g = \mathcal{G}(\cdot,f)$ par les valeurs $(\tilde{z}_j)_{j=0,\dots,n}$ est satisfaisante dès que $n \geq 20$. De plus, ces graphes nous permettent de comprendre que ce qui est important pour une bonne approximation de g, ce n'est pas tant la valeur de n, mais plutôt le fait qu'il existe des valeurs sur la grille de discrétisation qui soient proches des valeurs $(s_k)_{k=1,\dots,r}$ où le sens de variation de f change. En effet, ces valeurs là servent de pivot et permettent de retrouver l'amplitude de g pour les valeurs de $(t_j)_{j=0,\dots,n}$ supérieures. Ainsi, il est évident que de manière asymptotique, comme le montre la propriété 5.33, l'ajustement est très satisfaisant, puisqu'il existe des valeurs de la grille de discrétisation qui sont de plus en plus proches des valeurs $(s_k)_{k=1,\dots,r}$.

Dans le modèle (5.19), l'intérêt principal n'est pas porté, en ce qui nous concerne, sur la moyenne structurelle, à savoir $g \circ \phi^{-1}$, mais plutôt sur les fonctions de déformation $(h_i)_{i=1,\dots,m}$ qui nous permettent, après une déformation inverse des fonctions $(f_i)_{i=1,\dots,m}$ de retrouver une estimation de l'espérance structurelle $f \circ \phi^{-1}$. Aussi, la proposition 5.33 nous permet de définir, de façon naturelle, un estimateur empirique des fonctions de déformation basé, non plus sur les valeurs $(Y_{ij})_{i=1,\dots,m,\, j=0,\dots,n}$, mais sur les valeurs $\left(\tilde{Z}_{ij}\right)_{i=1,\dots,m,\, j=0,\dots,n}$. Soit $i_0 \in \{1,\dots,m\}$ fixé et $t \in [a,b]$, nous définissons, conditionnellement à $F_{i_0} = f_{i_0}$, l'estimateur empirique de $\phi \circ h_{i_0}^{-1}(t)$ par

$$\widetilde{\phi \circ h_{i_0}^{-1}}_{mn}(t) = \frac{1}{m-1} \sum_{\substack{i=1\\ i\neq i_0}}^{m} T_{j_i}, \tag{5.22}$$

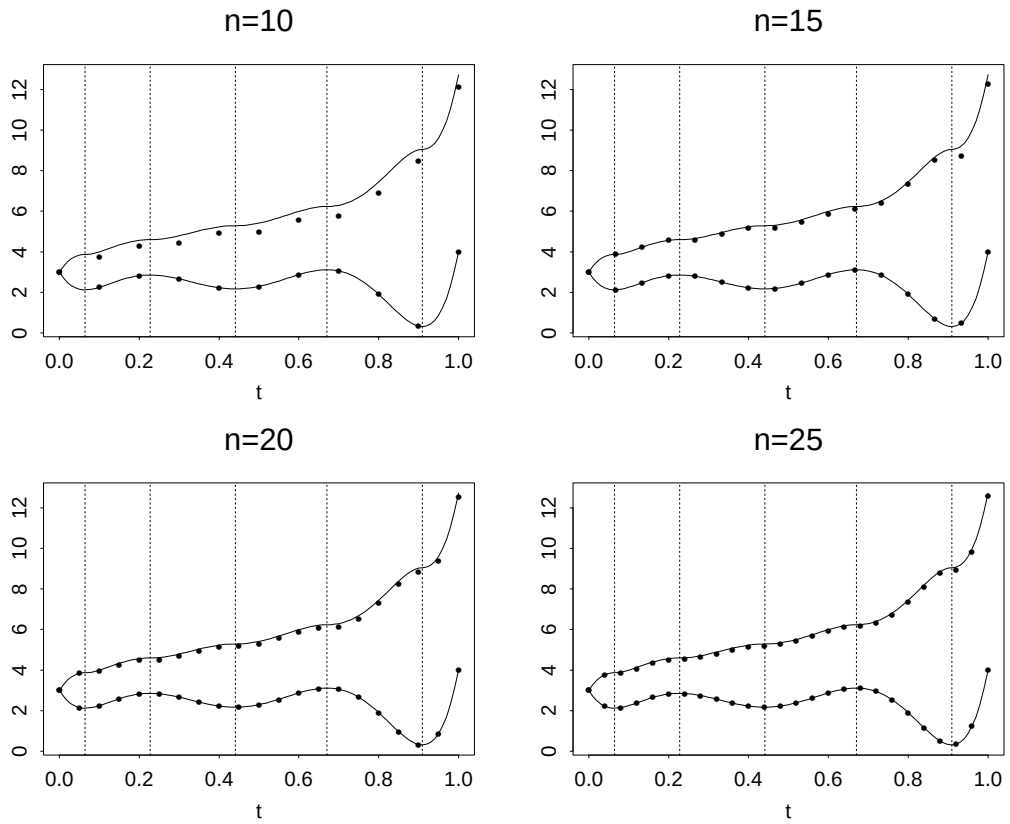

FIG. 5.5 – *Les courbes en trait plein correspondent, d'une part à la fonction f définie par* (5.18)*, et d'autre part, à la transformation associée g qui est strictement croissante. Les pointillés correspondent aux valeurs de $(y_j)_{j=0,\dots,n}$ et de $(\tilde{z}_j)_{j=0,\dots,n}$ définies par* (5.21) *et calculées sur $\left(t_j = \frac{j}{n}\right)_{j=0,\dots,n}$, avec $n = 10$, $n = 15$, $n = 20$ et $n = 25$. Les lignes en tirés verticales correspondent aux valeurs $(s_k)_{k=1,\dots,r=5}$. Ici, l'approximation de la fonction g par les valeurs $(\tilde{z}_j)_{j=0,\dots,n}$ est satisfaisante dès que $n \geq 20$.*

où, pour tout $i \in \{1,\dots,m\} \setminus i_0$,

$$T_{j_i} = \arg \min_{t_j \in \{t_0,\dots,t_n\}} \left| \tilde{Z}_{ij} - \tilde{Z}_{i_0 j_0} \right|,$$

avec

$$t_{j_0} = \arg \min_{t_j \in \{t_0,\dots,t_n\}} |t_j - t|.$$

Les estimations effectuées dans la section suivante seront faites à l'aide de cet estimateur. En effet, nous prendrons l'exemple de la fonction f définie par (5.18) qui n'est pas strictement croissante. Nous aurons donc besoin de faire appel à l'opérateur $\mathcal{G}$ pour estimer les fonctions de déformation.

5.4 Simulations

Les simulations sont effectuées avec la fonction f représentée sur la figure 5.4 et définie en (5.18). Cette fonction n'est pas strictement croissante mais appartient à l'ensemble $\mathcal{F}$ défini par (5.17) avec $r = 5$: $s_1 \approx 0.06438$, $s_2 \approx 0.22752$, $s_3 \approx 0.44066$, $s_4 \approx 0.67062$ et $s_5 \approx 0.90951$.

5.4.1 Simulation des fonctions de déformation

Les fonctions de déformation $(h_i)_{i=1,\ldots,m}$ sont simulées à partir d'une fonction que nous appelons *fonction élémentaire de déformation.* Une fonction élémentaire de déformation est définie par un couple $(c,d) \in]a,b[^2$ de la façon suivante:

$$u(c,d)(t) = \begin{cases} \frac{d-a}{c-a}(t-a)+a & \text{si} \quad t \in [a,c], \\ \frac{b-d}{b-c}(t-b)+b & \text{si} \quad t \in]c,b]. \end{cases} \tag{5.23}$$

La figure 5.6 nous montre un exemple de fonction élémentaire de déformation. Il s'agit d'une fonction strictement croissante, affine sur deux morceaux et passant par les points (a,a) et (b,b). Les graphes des fonctions de déformation portent en abscisse le label *temps absolu* et en ordonnée le label *temps relatif* lorsque la fonction $h_i, i = 1,\ldots,m$, est représentée, et réciproquement lorsque la fonction inverse $h_i^{-1}, i = 1,\ldots,m$, est représentée. Ce vocabulaire précise, naturellement, que le processus de déformation H déforme l'axe des abscisses sur lequel se définit la structure de la fonction f, et par là-même, le processus déforme la fonction f pour chaque individu $i \in \{1,\ldots,m\}$. Ainsi, pour retrouver l'espérance structurelle $f \circ \phi^{-1}$ à partir d'une fonction $f_i, i = 1,\ldots,m$, il suffit de représenter f_i dans le temps absolu, c'est-à-dire, sur l'axe des abscisses déformé par la fonction de déformation inverse $\phi \circ h_i^{-1}$.

La simulation d'une fonction de déformation h est définie par la composée de plusieurs fonctions élémentaires de déformation:

$$h = u(c_L,d_L) \circ \cdots \circ u(c_1,d_1),$$

avec $L \in \mathbb{N}$. Ainsi, la fonction h est, par composition, continue et strictement croissante, et les hypothèses (5.3) sont vérifiées. De plus, la proposition suivante nous montre que h est affine par morceaux.

Proposition 5.35 *Soit une suite de fonctions élémentaires de déformation, définie par $(u(c_l,d_l))_{l=1,\ldots,L}$ et notée de façon abrégée $(u_l)_{l=1,\ldots,L}$, avec $L \in \mathbb{N}$. Posons, pour tout $l \in \{0,1,\ldots,L\}$, $u_1^l = u_l \circ \cdots \circ u_1$, avec u_1^0 la fonction identité et $u_1^L = h$. Soit, pour tout $l \in \{1,\ldots,L\}$,*

$$\begin{cases} x_l &= \left(u_1^{l-1}\right)^{-1}(c_l), \\ y_l &= h(x_l). \end{cases}$$

Soient $(x_0,y_0) = (a,a)$ et $(x_{L+1},y_{L+1}) = (b,b)$. Soit la permutation σ de $\{0,\ldots,L+1\}$, telle que

$$x_0 = x_{\sigma(0)} < x_{\sigma(1)} \leq x_{\sigma(2)} \leq \cdots \leq x_{\sigma(L-1)} \leq x_{\sigma(L)} < x_{\sigma(L+1)} = x_{L+1}.$$

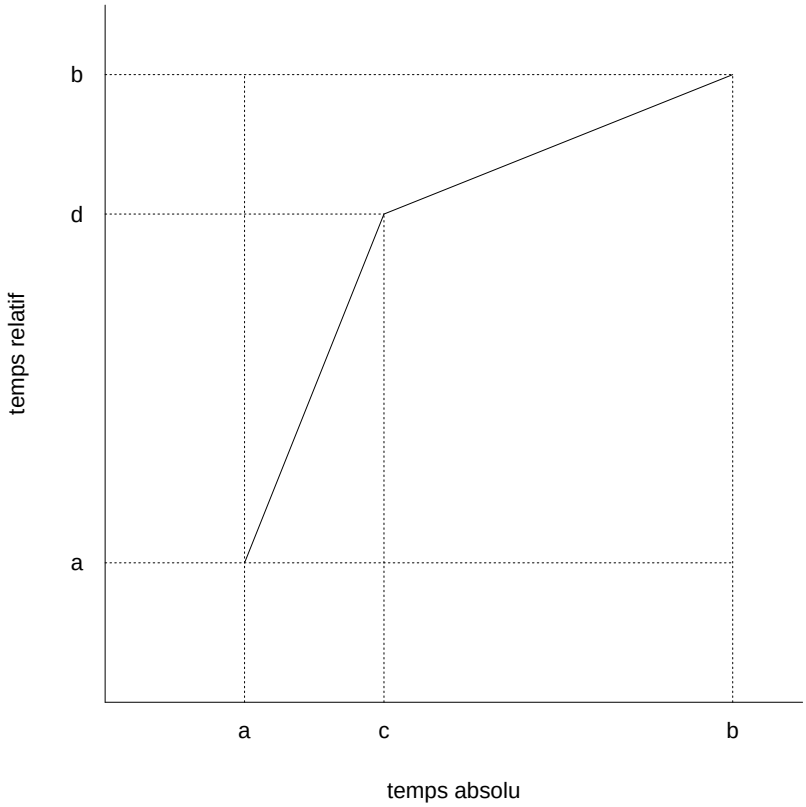

FIG. 5.6 – *Une fonction élémentaire de déformation* $u(c,d) : [a,b] \to [a,b]$ *définie par* (5.23) *est une fonction strictement croissante, affine sur deux morceaux et passant par* (a,a) *et* (b,b).

Alors, pour tout $k = 0,\ldots,L$, *la fonction* h *est affine sur* $\left[x_{\sigma(k)},x_{\sigma(k+1)}\right]$, *d'équation*

$$h(t) = \frac{y_{\sigma(k+1)} - y_{\sigma(k)}}{x_{\sigma(k+1)} - x_{\sigma(k)}} \left(t - x_{\sigma(k)}\right) + y_{\sigma(k)}.$$

Preuve 5.36 *Soit* $k \in \{0,\ldots,L\}$ *fixé et montrons que* h *est affine sur l'intervalle* $[x_{\sigma(k)},x_{\sigma(k+1)}]$. *Pour tout* $l \in \{1,\ldots,L\}$, *nous avons*

$$c_l \notin \left]u_1^{l-1}\left(x_{\sigma(k)}\right),u_1^{l-1}\left(x_{\sigma(k+1)}\right)\right[,$$

en effet, si

$$u_1^{l-1}\left(x_{\sigma(k)}\right) < c_l < u_1^{l-1}\left(x_{\sigma(k+1)}\right),$$

alors, puisque la composée $\left(u_1^{l-1}\right)^{-1}$ *est strictement croissante, nous avons*

$$x_{\sigma(k)} < \left(u_1^{l-1}\right)^{-1}(c_l) < x_{\sigma(k+1)},$$

or, ceci est impossible, au regard de la définition de la permutation σ. Donc, nous avons

$$c_l \leq u_1^{l-1}\left(x_{\sigma(k)}\right) \text{ ou } u_1^{l-1}\left(x_{\sigma(k+1)}\right) \leq c_l,$$

ainsi, u_l est affine sur $\left[u_1^{l-1}\left(x_{\sigma(k)}\right),u_1^{l-1}\left(x_{\sigma(k+1)}\right)\right]$. Ceci étant vrai pour tout $l \in \{1,\ldots,L\}$, et par composition de fonctions affines, la fonction h est affine sur $\left[x_{\sigma(k)},x_{\sigma(k+1)}\right]$.

■

La simulation de $m = 100$ fonctions de déformation $(h_i)_{i=1,\ldots,m}$ est effectuée comme suit. Pour tout $i \in \{1,\ldots,m\}$, nous posons

$$h_i = u(c_1,d_1^i) \circ u(c_2,d_2^i) \circ u(c_3,d_3^i) \circ u(c_4,d_4^i) \circ u(c_5,d_5^i), \tag{5.24}$$

avec

- $c_1 = 0.06438$,
- $c_2 = 0.22752$,
- $c_3 = 0.44066$,
- $c_4 = 0.67062$,
- $c_5 = 0.90951$,

et, pour tout $k \in \{1,\ldots,5\}$, les valeurs $(d_k^i)_{i=1,\ldots,m}$ sont des réalisations indépendantes de la variable aléatoire uniforme D_k définie par

- $D_1 \sim \mathcal{U}([c_1 - 0.03,c_1 + 0.03])$,
- $D_2 \sim \mathcal{U}([c_2 - 0.09,c_2 + 0.09])$,
- $D_3 \sim \mathcal{U}([c_3 - 0.17,c_3 + 0.17])$,
- $D_4 \sim \mathcal{U}([c_4 - 0.11,c_4 + 0.11])$,
- et $D_5 \sim \mathcal{U}([c_5 - 0.02,c_5 + 0.02])$.

Un calcul direct nous montre que chaque fonction élémentaire de déformation est d'espérance égale à l'identité. Donc, notre processus de déformation

$$H = u(c_1,D_1) \circ u(c_2,D_2) \circ u(c_3,D_3) \circ u(c_4,D_4) \circ u(c_5,D_5)$$

est d'espérance ϕ égale à l'identité. En effet, pour tout $t \in [0,1]$,

$$\begin{aligned}
\phi(t) =& \mathbb{E}[H(t)] \\
=& \mathbb{E}[u(c_1,D_1) \circ u(c_2,D_2) \circ u(c_3,D_3) \circ u(c_4,D_4) \circ u(c_5,D_5)(t)] \\
=& \mathbb{E}[\mathbb{E}(u(c_1,D_1) \circ u(c_2,D_2) \circ u(c_3,D_3) \circ u(c_4,D_4) \circ u(c_5,D_5)(t) \\
& \mid u(c_2,D_2) \circ u(c_3,D_3) \circ u(c_4,D_4) \circ u(c_5,D_5)(t))] \\
=& \mathbb{E}[u(c_2,D_2) \circ u(c_3,D_3) \circ u(c_4,D_4) \circ u(c_5,D_5)(t)],
\end{aligned}$$

et en répétant cette opération, c'est-à-dire en utilisant plusieurs fois l'espérance conditionnelle, nous arrivons à

$$\begin{aligned}
\phi(t) =& \mathbb{E}[u(c_4,D_4) \circ u(c_5,D_5)(t)] \\
=& \mathbb{E}[\mathbb{E}(u(c_4,D_4) \circ u(c_5,D_5)(t) \mid u(c_5,D_5)(t))] \\
=& \mathbb{E}[u(c_5,D_5)(t)] \\
=& t.
\end{aligned}$$

Donc, dans ce cas particulier, l'espérance structurelle $f \circ \phi^{-1}$ est égale à la fonction f. La figure 5.7 nous montre, d'une part, les fonctions de déformation obtenues, et d'autre part, la moyenne empirique de ces fonctions. La moyenne empirique est bien approximativement égale à l'identité.

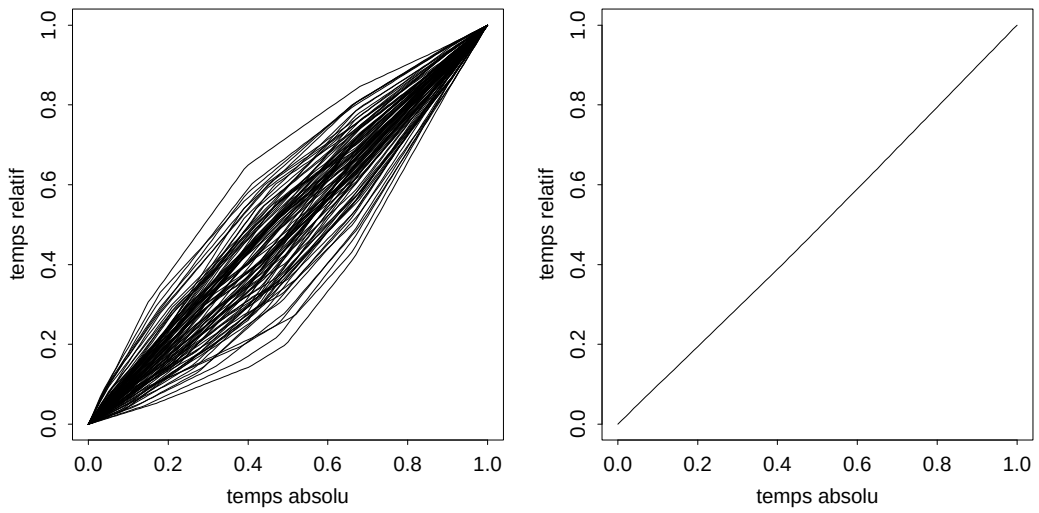

FIG. 5.7 – *A gauche, les fonctions de déformation simulées* $(h_i)_{i=1,\ldots,m=100}$. *A droite, la moyenne empirique de ces fonctions, soit une estimation de* ϕ. *Cette estimation est approximativement égale à la fonction identité.*

5.4.2 Simulation des données

Les données issues du modèle (5.2) sont donc simulées par

$$y_{ij} = f \circ h_i^{-1}(t_j), i = 1, \ldots, m, j = 0, \ldots, n, \tag{5.25}$$

où, $m = 100$, $n = 100$, la fonction f est définie en (5.18), les fonctions $(h_i)_{i=1,\ldots,m}$ sont simulées par (5.24) et $t_j = \frac{j}{n}$ pour tout $j \in \{0,1,\ldots,n\}$. La figure 5.8 nous montre, d'une part, les fonctions simulées, et d'autre part, la moyenne empirique des ces fonctions. Nous observons très bien que cette moyenne est une très mauvaise estimation de la fonction f. La fonction moyenne n'a plus la même amplitude que les fonctions $(f_i)_{i=1,\ldots,m}$ de notre échantillon.

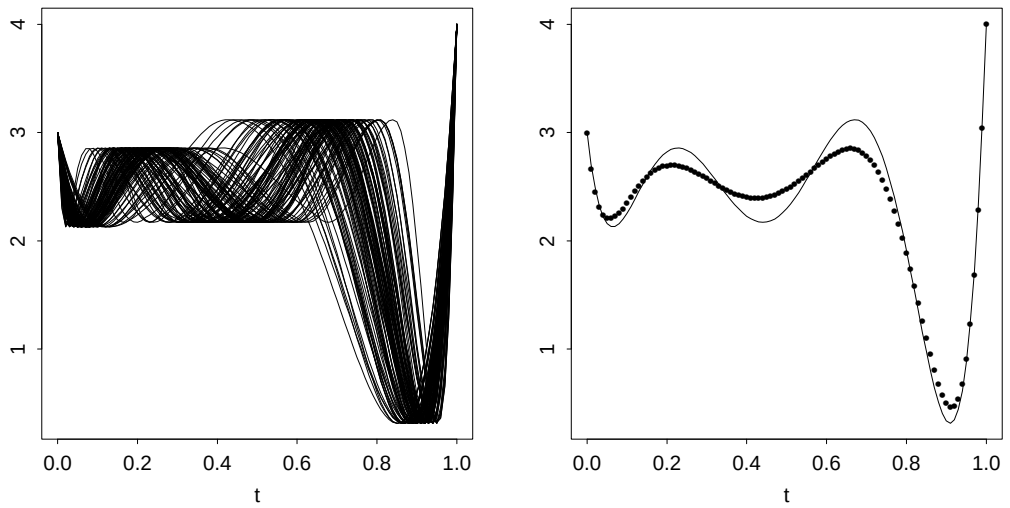

FIG. 5.8 – *A gauche, les fonctions simulées par* (5.25). *A droite, la moyenne de ces fonctions en les points* $(t_j)_{j=0,\ldots,n=100}$ *(en pointillés) et la fonction* f *(en trait plein).*

5.4.3 Estimation

L'estimation des fonctions de déformation $\left(\phi \circ h_i^{-1}\right)_{i=1,\ldots,m}$ est effectuée, en les points $(t_j)_{j=0,\ldots,n}$, par l'estimateur défini en (5.22). Ainsi, pour tout $t \in \{t_0, \ldots, t_n\}$, nous avons

$$\widetilde{\phi \circ h_{i_0}^{-1}}_{mn}(t) = \frac{1}{m-1} \sum_{\substack{i=1 \\ i \neq i_0}}^{m} T_{j_i},$$

où, pour tout $i \in \{1, \ldots, m\} \setminus i_0$,

$$T_{j_i} = \arg \min_{t_j \in \{t_0, \ldots, t_n\}} \left| \tilde{Z}_{ij} - \tilde{Z}_{i_0 j_0} \right|,$$

avec

$$t_{j_0} = \arg \min_{t_j \in \{t_0, \ldots, t_n\}} |t_j - t|.$$

La figure 5.9 nous présente le résultat du recalage des fonctions $(f_i)_{i=1,\ldots,m}$. Plus précisément, pout tout $i_0 \in \{1, \ldots, m\}$, cette figure trace les points

$$\left(\widetilde{\phi \circ h_{i_0}^{-1}}_{mn}(t_j), y_{i_0 j} \right)_{j=0,\ldots,n}. \tag{5.26}$$

Ces points correspondent à des estimations de l'espérance structurelle $f \circ \phi^{-1}$. En effet, pour tout $i_0 \in \{1, \ldots, m\}$, nous avons

$$f_{i_0} = f \circ h_{i_0}^{-1} = f \circ \phi^{-1} \circ \phi \circ h_{i_0}^{-1}.$$

La figure 5.9 montre que le recalage de chaque fonction f_i, $i = 1, \ldots, m$, sur la fonction f (f est en trait plein) est satisfaisant.

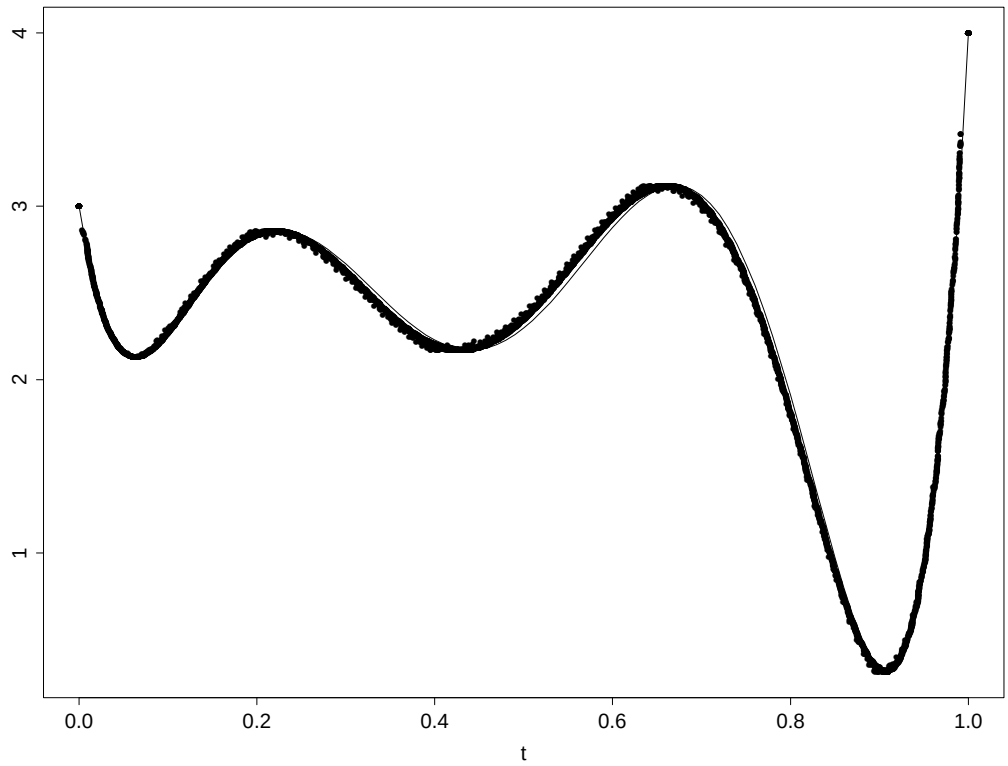

FIG. 5.9 – *Les fonctions recalées sont en pointillés, les coordonnées des points tracés sont donnés par* (5.26). *La fonction f est en trait plein.*

5.5 Cas où les mesures sont bruitées

Dans cette section, nous explicitons une méthode automatique pour résoudre le problème du recalage de courbes dans le cadre de données mesurées avec un bruit. Ainsi, les données sont issues du modèle

$$Y_{ij} = f \circ H_i^{-1}(t_j) + \varepsilon_{ij}, i = 1, \ldots, m, j = 0, \ldots, n, \tag{5.27}$$

avec les même hypothèses que pour le modèle (5.2). Seulement ici, nous avons un bruit sur nos mesures, modélisé par les variables aléatoires $(\varepsilon_{ij})_{i=1,\ldots,m,\, j=0,\ldots,n}$, supposées indépendantes et de même loi $\mathcal{N}(0,\sigma^2)$. Ainsi, les données simulées sont celles de la section 5.4 auxquelles nous avons ajouté les réalisations d'une variable aléatoire

normale centrée et d'écart-type $\sigma = 0.05$. Cet écart type correspond approximativement à 5% de l'amplitude entre les deux optima locaux successifs les plus proches de la fonction f.

La méthode d'estimation de la fonction $f \circ \phi^{-1}$ est basée sur l'estimation de chaque fonction f_i, $i = 1, \ldots, m$, par un estimateur à noyau (voir par exemple [WJ95]). Ensuite, l'estimation des fonctions de déformation $\left(\phi \circ h_i^{-1}\right)_{i=1,\ldots,m}$ nous permet de recaler les fonctions $(f_i)_{i=1,\ldots,m}$ et d'en calculer la moyenne, c'est-à-dire une estimation de $f \circ \phi^{-1}$. Plus précisément, la méthode que nous appellerons *processus de recalage*, est basée sur les cinq points suivants:

1. Pour tout $i \in \{1, \ldots, m\}$, nous estimons la fonction f_i par

$$\hat{f}_i(t_0) = \frac{1}{m} \sum_{i=1}^{m} Y_{i0} \xrightarrow[m \to +\infty]{ps} f(t_0),$$

$$\hat{f}_i(t_n) = \frac{1}{m} \sum_{i=1}^{m} Y_{in} \xrightarrow[m \to +\infty]{ps} f(t_n),$$

et, pour tout $j \in \{1, \ldots, n-1\}$, nous estimons $f_i(t_j)$ par

$$\hat{f}_i(t_j) = \frac{\sum_{k=0}^{n} Y_{ik} \Phi\left(\frac{t_k - t_j}{\nu_i}\right)}{\sum_{k=0}^{n} \Phi\left(\frac{t_k - t_j}{\nu_i}\right)}, \tag{5.28}$$

où, les valeurs $(Y_{ij})_{i=1,\ldots,m,\, j=0,\ldots,n}$ sont données par (5.27), la fonction Φ, appelée *le noyau*, est égale à la densité d'une loi $\mathcal{N}(0,1)$, et les valeurs $(\nu_i)_{i=1,\ldots,m}$, appelées *les largeurs de fenêtre*, sont dans $\mathbb{R}_+^*$.

2. Pour tout couple $(i,j) \in \{i = 1, \ldots, m\} \times \{0, \ldots, n\}$, nous calculons la valeur $\check{z}_{ij}$, en substituant, dans la définition (5.20), les variables $(Y_{ij})_{ij}$ par les estimations $\left(\hat{f}_i(t_j)\right)_{ij}$.

3. Pour tout couple $(i,j) \in \{i = 1, \ldots, m\} \times \{1, \ldots, n-1\}$, nous calculons une estimation de $\phi \circ h_i^{-1}(t_j)$, notée

$$\phi \check{\circ} h_i^{-1}(t_j), \tag{5.29}$$

en substituant dans la définition (5.22), les variables $\left(\tilde{Z}_{ij}\right)_{ij}$ par les estimations $(\check{z}_{ij})_{ij}$. Bien évidemment, pour tout $i \in \{1, \ldots, m\}$, nous posons $\phi \check{\circ} h_i^{-1}(t_0) = 0$ et $\phi \check{\circ} h_i^{-1}(t_n) = 1$.

4. Pour tout $i \in \{1, \ldots, m\}$, nous avons une estimation de la fonction $f \circ \phi^{-1}$, en les points $\{t_0, \ldots, t_n\}$, donnée par

$$\hat{f}^i(t_0) = \frac{1}{m} \sum_{i=1}^{m} Y_{i0},\ \hat{f}^i(t_n) = \frac{1}{m} \sum_{i=1}^{m} Y_{in},$$

et, pour tout $j \in \{1, \ldots, n-1\}$,

$$\hat{f}^i(t_j) = \frac{\sum_{k=0}^{n} Y_{ik} \Phi\left(\frac{\phi \check{\circ} h_i^{-1}(t_k) - t_j}{\tau_i}\right)}{\sum_{k=0}^{n} \Phi\left(\frac{\phi \check{\circ} h_i^{-1}(t_k) - t_j}{\tau_i}\right)}, \tag{5.30}$$

avec, pour tout $i \in \{1, \ldots, m\}$, $\tau_i \in \mathbb{R}_+^*$.

5. Enfin, l'estimation de $f \circ \phi^{-1}$ en les points $(t_j)_{j=0,\ldots,n}$ est donnée par

$$\hat{f}(t_j) = \frac{1}{m} \sum_{i=1}^{m} \hat{f}^i(t_j). \tag{5.31}$$

Pratiquement, la mise en place de cette méthode d'estimation nécessite de définir les valeurs des largeurs de fenêtre $(\nu_i)_{i=1,\ldots,m}$ utilisées en (5.28) et $(\tau_i)_{i=1,\ldots,m}$ utilisées en (5.30). Le choix des ces valeurs est bien évidemment très important. Plus précisément, le choix des valeurs $(\nu_i)_{i=1,\ldots,m}$ est plus important que celui des valeurs $(\tau_i)_{i=1,\ldots,m}$. En effet, les valeurs $(\nu_i)_{i=1,\ldots,m}$ interviennent dès le début du processus de recalage et l'erreur d'approximation risque de se propager, voire de s'amplifier, lors des approximations successives. Par contre, les valeurs $(\tau_i)_{i=1,\ldots,m}$ n'interviennent que dans le calcul de l'estimation $\hat{f}$ en fin de processus. Pour cette raison, nous choisissons d'estimer ces dernières simplement par validation croisée ([WJ95]). Quant au choix des valeurs $(\nu_i)_{i=1,\ldots,m}$, nous allons voir le résultat avec deux méthodes, d'une part la validation croisée, et d'autre part une méthode d'estimation basée sur la minimisation d'un critère empirique.

Ainsi, une méthode classique, en estimation non paramétrique, consiste à estimer chacune des largeurs de fenêtre par validation croisée. La figure 5.10 nous donne le résultat du processus de recalage et d'estimation de l'espérance structurelle $f \circ \phi^{-1}$ en utilisant la validation croisée. Le graphe des fonctions estimées $\left(\hat{f}_i\right)_{i=1,\ldots,m}$ (figure 5.10, en haut, à droite) nous montre que le choix de la validation croisée comme méthode d'estimation pour les valeurs $(\nu_i)_{i=1,\ldots,m}$, n'est pas optimal. En effet, il est reconnu que la validation croisée à tendance à estimer la largeur de fenêtre par une trop petite valeur, ce qui a pour conséquence d'avoir des estimations $\left(\hat{f}_i\right)_{i=1,\ldots,m}$ très proches de nos données, et donc pas assez débruitées. Ceci est manifeste dans la figure 5.10 (en haut, à droite). Dans notre contexte, cela détériore l'estimation de nos fonctions de déformation, et donc l'estimation de notre fonction $f \circ \phi^{-1}$ (figure 5.10, en bas). Ainsi, nous remarquons qu'il est plus intéressant de surévaluer les valeurs $(\nu_i)_{i=1,\ldots,m}$, plutôt que de les choisir trop petites. Cela se comprend aisément, puisque l'estimation des fonctions $(f_i)_{i=1,\ldots,m}$ est utilisée pour estimer les fonctions de déformation, et non pas pour avoir une estimation des fonctions $(f_i)_{i=1,\ldots,m}$ elles-mêmes. Ainsi, il s'agit de bien débruiter nos données, quite à avoir des estimations $\left(\hat{f}_i\right)_{i=1,\ldots,m}$ qui sous-estiment l'amplitude des fonctions $(f_i)_{i=1,\ldots,m}$. Nous pouvons ici conclure, que, pour cet exemple, la validation croisée ne donne pas de bon résultats.

En effet, la figure 5.10 nous montre qu'un grand nombre de fonctions de déformation est mal estimé (figure 5.10, en bas, à gauche) et que la fonction $f \circ \phi^{-1}$ n'est guère mieux estimée que par la moyenne (figure 5.10, en bas, à droite).

Une deuxième méthode d'estimation des valeurs $(\nu_i)_{i=1,\ldots,m}$ consiste à choisir les valeurs qui minimisent un critère approprié. Dans ce cas, il est évident que les calculs seront d'autant plus importants (en temps de calcul machine) que le nombre de paramètres à estimer sera important. Aussi, nous choisissons ici de n'utiliser qu'une seule largeur de fenêtre ν pour nos m fonctions $\left(\hat{f}_i\right)_{i=1,\ldots,m}$. Ainsi, nous remplaçons, dans le processus de recalage, l'estimation (5.28) par

$$\hat{f}_i(t_j) = \frac{\sum_{k=0}^{n} Y_{ik} \Phi\left(\frac{t_k - t_j}{\nu}\right)}{\sum_{k=0}^{n} \Phi\left(\frac{t_k - t_j}{\nu}\right)},$$

avec $\nu \in \mathbb{R}_+^*$. De plus, l'estimation de ν doit être optimale pour le recalage des courbes $(f_i)_{i=1,\ldots,m}$. Il est donc naturel de choisir la valeur de ν qui va minimiser le critère

$$\sum_{i=1}^{m} \sum_{j=0}^{n} \left| \hat{f}^i(t_j) - \hat{f}(t_j) \right|, \tag{5.32}$$

soit

$$\hat{\nu} = \arg\min_{\nu \in L} \sum_{i=1}^{m} \sum_{j=0}^{n} \left| \hat{f}^i(t_j) - \hat{f}(t_j) \right|,$$

avec L un ensemble convenablement choisi. Ici, nous avons choisi

$$L = \{\frac{5}{1000}, \frac{6}{1000}, \ldots, \frac{14}{1000}, \frac{15}{1000}\}.$$

Donc, la procédure d'estimation de $\hat{\nu}$ consiste à répéter les cinq étapes du processus de recalage pour les différentes valeurs de $\nu \in L$, et à choisir la valeur minimisant le critère (5.32). La figure 5.11 nous montre le résultat de ces calculs pour $\nu \in \{\frac{5}{1000}, \frac{6}{1000}, \ldots, \frac{14}{1000}, \frac{15}{1000}\}$. La valeur minimale est atteinte pour $\hat{\nu} = 0.011$.

La figure 5.12 nous donne le résultat du recalage des courbes $(f_i)_{i=1,\ldots,m}$ et de l'estimation de $f \circ \phi^{-1}$, avec l'estimation de ν: $\hat{\nu} = 0.011$. Le résultat du recalage des courbes et de l'estimation de $f \circ \phi^{-1}$ est ici assez satisfaisant.

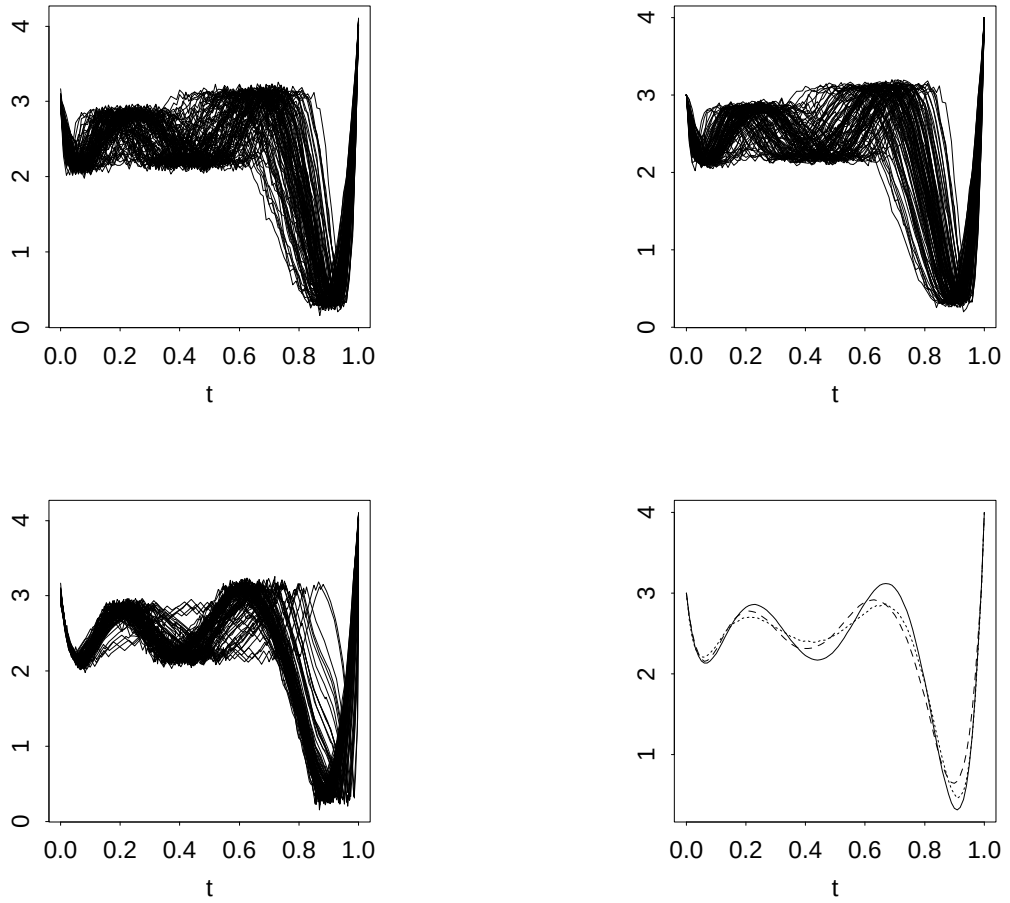

FIG. 5.10 – *En haut, à gauche, les données bruitées, simulées par* (5.27). *En haut, à droite, les fonctions* $\left(\hat{f}_i\right)_{i=1,\ldots,m}$ *calculées par* (5.28), *avec les largeurs de fenêtre* $(\nu_i)_{i=1,\ldots,m}$ *estimées par validation croisée. En bas, à gauche, les fonctions bruitées recalées, c'est-à-dire les points* $\left(\phi \check{\circ} h_i^{-1}(t_j), y_{ij}\right)_{i=1,\ldots,m,\, j=0,\ldots,n}$, *avec les valeurs* $\left(\phi \check{\circ} h_i^{-1}(t_j)\right)_{i=1,\ldots,m,\, j=0,\ldots,n}$ *estimées par* (5.29) *et les données* $(y_{ij})_{i=1,\ldots,m,\, j=0,\ldots,n}$ *simulées par* (5.27). *En bas, à droite, la fonction* f *(en trait plein), la moyenne des fonctions recalées (en tirés) estimée par* (5.31) *et la moyenne obtenue sans alignement (en pointillés).*

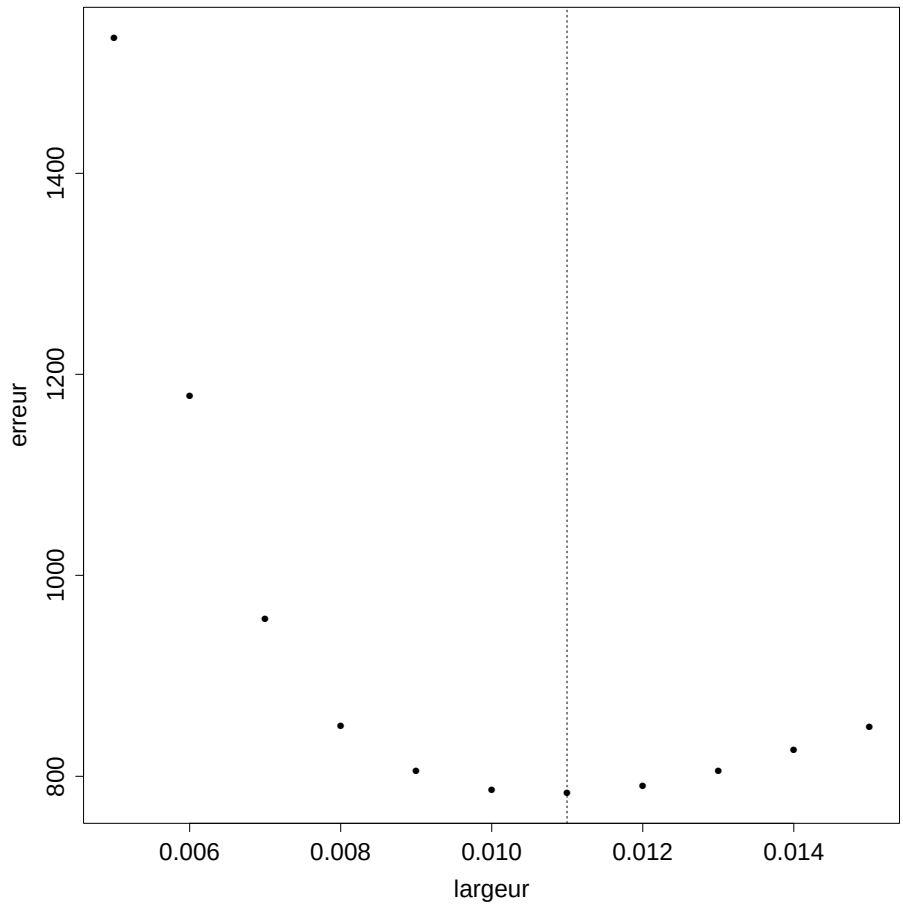

FIG. 5.11 – *Erreurs calculées selon le critère* (5.32) *pour les valeurs de* $\nu \in \{\frac{5}{1000}, \frac{6}{1000}, \ldots, \frac{14}{1000}, \frac{15}{1000}\}$. *Le minimum est atteint pour* $\hat{\nu} = 0.011$.

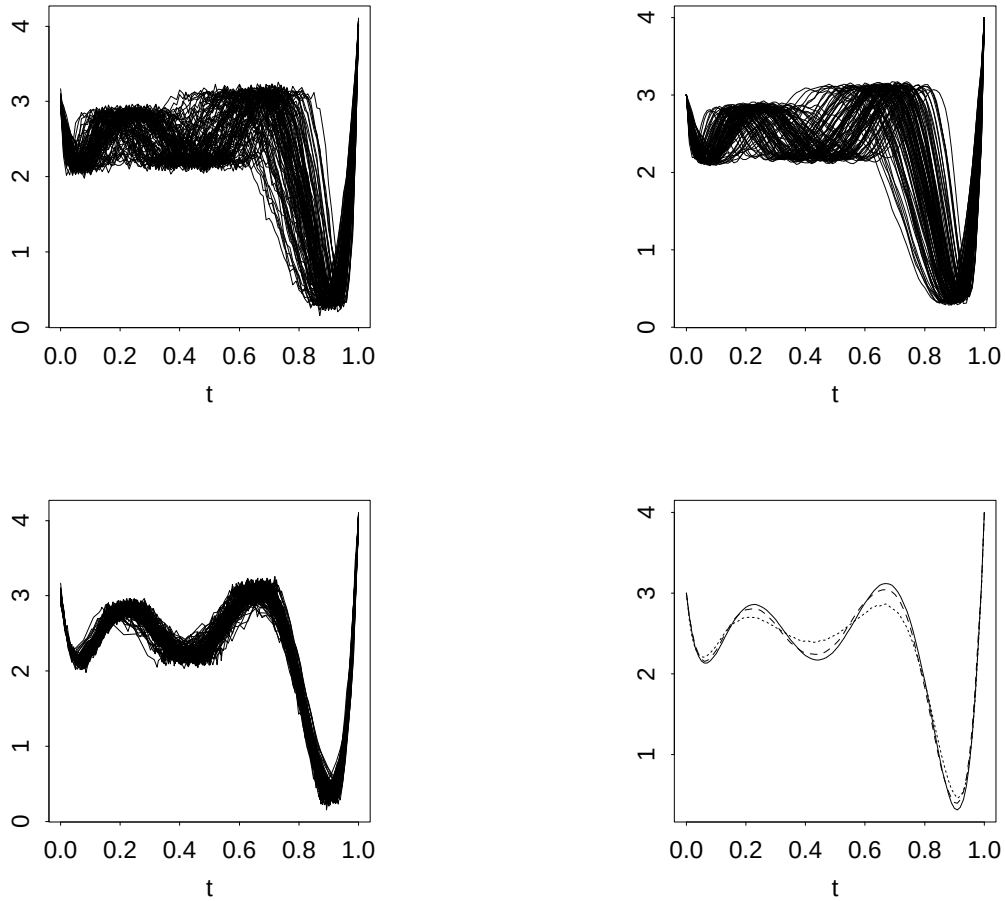

FIG. 5.12 – *En haut, à gauche, les données bruitées, simulées par* (5.27). *En haut, à droite, les estimations* $\left(\hat{f}_i\right)_{i=1,\ldots,m}$ *calculées par* (5.28), *avec les largeurs de fenêtre* $(\nu_i)_{i=1,\ldots,m}$ *toutes égales à* $\hat{\nu} = 0.011$. *En bas, à gauche, les fonctions bruitées recalées, c'est-à-dire les points* $\left(\phi \check{\circ} h_i^{-1}(t_j), y_{ij}\right)_{i=1,\ldots,m,\, j=0,\ldots,n}$, *avec les valeurs* $\left(\phi \check{\circ} h_i^{-1}(t_j)\right)_{i=1,\ldots,m,\, j=0,\ldots,n}$ *estimées par* (5.29) *et les données* $(y_{ij})_{i=1,\ldots,m,\, j=0,\ldots,n}$ *simulées par* (5.27). *En bas, à droite, la fonction* f *(en trait plein), la moyenne des fonctions recalées (en tirés) estimée par* (5.31) *et la moyenne obtenue sans alignement (en pointillés).*

Bibliographie

[Bar00] Y. Baraud. Model selection for regression on a fixed design. *Probab. Theory Related Fields*, 117(4):467–493, 2000.

[BBM99] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

[BCD83] M. Broniatowski, G. Celeux, and J. Diebolt. *Data Analysis and Informatics*. North Holland, 1983. Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabiliste.

[BFOS84] L. Breiman, J. Friedman, R. Olshen, and C.J. Stone. *Classification and regression trees*. Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA, 1984.

[BFV94] J. Boularan, L. Ferré, and P. Vieu. Growth curves: a two-stage nonparametric approach. *J. Statist. Plann. Inference*, 38(3):327–350, 1994.

[BJS03] D. Belomestny, V. Jentsch, and M. Schreckenberg. Completion and continuation of nonlinear traffic time series: a probabilistic approach. *J.Phys.A:Math. Gen.*, 36:11369–11383, 2003.

[BM85] K.E. Basford and G.J. McLachlan. Estimation of allocation rates in a cluster analysis context. *J. Amer. Statist. Assoc.*, 80(390):286–293, 1985.

[BM98] L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998.

[BM01] L. Birgé and P. Massart. Gaussian model selection. *J. Eur. Math. Soc. (JEMS)*, 3(3):203–268, 2001.

[Boo97] F.L. Bookstein. *Morphometric tools for landmark data*. Cambridge University Press, Cambridge, 1997. Geometry and biology, Reprint of the 1991 original.

[BtMvdBvdV03] R. J. Beerends, H. G. ter Morsche, J. C. van den Berg, and E. M. van de Vrie. *Fourier and Laplace transforms*. Cambridge University Press, Cambridge, 2003. Translated from the 1992 Dutch edition by Beerends.

[CCD95] G. Celeux, D. Chauveau, and G. Diebol. On stochastic version of the em algorithm. *Rapport de recherche INRIA*, 2514, 1995.

[CD92] G. Celeux and J. Diebolt. A stochastic approximation type em algorithm for the mixture problems. *Stochastics Reports*, 41:119–134, 1992.

[CDPB96] F. Couton, M. Danech-Pajouh, and M. Broniatowski. Application des mélanges de lois de probabilité à la reconnaissance de régimes de trafic routier. *Recherche Transports Sécurité*, 53:49–58, 1996.

[Cel88] G. Celeux. Classification et modèles. *Rev. Statist. Appl.*, 36(4):43–57, 1988.

[Che95] J. Chen. Optimal rate of convergence for finite mixture models. *Ann. Statist.*, 23(1):221–233, 1995.

[CL01] R.C.H. Cheng and W.B. Liu. The consistency of estimators in finite mixture models. *Scand. J. Statist.*, 28(4):603–616, 2001.

[Coh90] S. Cohen. *Ingénierie du trafic routier.* Presses de l'Ecole Nationale des Ponts et Chaussées, Institut National de Recherche sur les Transports et leur Sécurité - INRETS, France, 1990.

[DGT] A.S. Dalalyan, G.K. Golubev, and A.B. Tsybakov. Bayesian maximum likelihood and semiparametric second order efficiency. Prépublication numéro 842 du laboratoire de probabilités et modèles aléatoires, Université Paris 6 et Paris 7.

[DJ94] D.L. Donoho and I.M. Johnstone. Minimax risk over l_p-balls for l_q-error. *Probab. Theory Related Fields*, 99(2):277–303, 1994.

[DLB96] F. Dazy and J-F. Le Barzic. *L'analyse des données évolutives.* Éditions Technip, Paris, 1996. Méthodes et applications. [Methods and applications], With a preface by Françoise Lavallard and Gilbert Saporta.

[DLM99] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27(1):94–128, 1999.

[DLR77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–38, 1977. With discussion.

[Doc95] T. Dochy. Arbres de régression et réseaux de neurones appliqués à la prévision de trafic routier. *Thèse de l'université Paris Dauphine*, 1995.

[DPA94] M. Danech-Pajouh and M. Aron. Athena: Prévision à court terme du trafic sur une section de route. Technical report, Institut National de Recherche sur les Transports et leur Sécurité - INRETS, 1994.

[GCP93] V. Genon-Catalot and D. Picard. *Éléments de statistique asymptotique*, volume 11 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Paris, 1993.

[GK95] T. Gasser and A. Kneip. Searching for structure in curve samples. *J. Amer. Statist. Assoc.*, 90:1179–1188, 1995.

[GLL] E. Gassiat and C. Lévy-Leduc. Efficient semi-parametric estimation of the periods in a superposition of periodic functions with unknown shape. preprint.

[GLM03] F. Gamboa, J-M. Loubes, and E. Maza. Semi-parametric estimation of shifts. *preprint*, 2003.

[Gol88] G. Golubev. Estimation of the period of a signal with an unknown form against a white. *Problemy Peredachi Informatsii*, 24:38–52, 1988.

[Gor99] A.D. Gordon. *Classification – 2nd Edition.* CHAPMAN & HALL/CRC, University of St. Andrews, UK, 1999.

[GS58] U. Grenander and G. Szegö. *Toeplitz forms and their applications.* California Monographs in Mathematical Sciences. University of California Press, Berkeley, 1958.

[Jam78] M. Jambu. *Classification automatique pour l'analyse des données. I.* Dunod, Paris, 1978. Méthodes et algorithmes. [Methods and algorithms], Dunod Décision.

[KG92] A. Kneip and T. Gasser. Statistical tools to analyze data representing a sample of curves. *Ann. Statist.*, 20(3):1266–1305, 1992.

[KLMR00] A. Kneip, X. Li, K.B. MacGibbon, and J.O. Ramsay. Curve registration by local regression. *Canad. J. Statist.*, 28(1):19–29, 2000.

[Kne94] A. Kneip. Nonparametric estimation of common regressors for similar curve data. *Ann. Statist.*, 22(3):1386–1427, 1994.

[Lav02] M. Lavielle. On the use of penalized contrasts for solving inverse problems. *preprint*, 2002.

[LL95] B. Lindsay and M. Lesperance. A review of semiparametric mixture models. *J. Statist. Plann. Inference*, 47(1-2):29–39, 1995. Statistical modelling (Leuven, 1993).

[LML02] J-M. Loubes, E. Maza, and M. Lavielle. Road trafficking description and short term travel time forecasting, with a classification method. *Preprint*, 2002.

[LT91] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[McL82] G.J. McLachlan. On the bias and variance of some proportion estimators. *Comm. Statist. B—Simulation Comput.*, 11(6):715–726, 1982.

[MS86] D.L. McLeish and C.G. Small. Likelihood methods for the discrimination problem. *Biometrika*, 73(2):397–403, 1986.

[NARPV99] V. Núñez-Antón, J.M. Rodríguez-Póo, and P. Vieu. Longitudinal data with nonstationary errors: a nonparametric three-stage approach. *Test*, 8(1):201–231, 1999.

[PST98] M. Piccioni, S. Scarlatti, and A. Trouvé. A variational problem arising from speech recognition. *SIAM J. Appl. Math.*, 58(3):753–771 (electronic), 1998.

[Rao58] C. Rao. Some statistical methods for the comparison of growth curves. *Biometrics*, 14:1–17, 1958.

[RD91] J.O. Ramsay and C.J. Dalzell. Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B*, 53(3):539–572, 1991. With discussion and a reply by the authors.

[RL98] J.O. Ramsay and X. Li. Curve registration. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(2):351–363, 1998.

[RS91] J.A. Rice and B.W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, 53(1):233–243, 1991.

[RS97] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis.* Springer-Verlag, New York, 1997.

[RS02] J.O. Ramsay and B.W. Silverman. *Applied functional data analysis.* Springer Series in Statistics. Springer-Verlag, New York, 2002. Methods and case studies.

[Rud87] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, third edition, 1987.

[RWF95] J.O. Ramsay, X. Wang, and R. Flanagan. A functional data analysis of the pinch force of human fingers. *Appl. Statist.*, 44(1):17–30, 1995.

[SC78] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions, ASSP-26*, 1978.

[Sil95] B.W. Silverman. Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser. B*, 57(4):673–689, 1995.

[vdG00] S.A. van de Geer. *Applications of empirical process theory.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2000.

[vdV98] A.W. van der Vaart. *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

[VGDPMW98] H.J.M. Van Grol, M. Danech-Pajouh, S. Manfredi, and J. Whittaker. Daccord: on-line travel time prediction. In 8th *WCTR 1998*, volume 2, 1998.

[Vim04] M. Vimond. Estimation semi-paramétrique des paramètres de translations et de dilatations. Master's thesis, Laboratoire de Statistique et Probabilités - Université Paul Sabatier, 2004.

[Wag95] C. Wagschal. *Topologie et analyse fonctionnelle.* HERMANN, 1995.

[WG97] K. Wang and T. Gasser. Alignment of curves by dynamic time warping. *Ann. Statist.*, 25(3):1251–1276, 1997.

[WG99] K. Wang and T. Gasser. Synchronizing sample curves nonparametrically. *Ann. Statist.*, 27(2):439–460, 1999.

[WJ95] M.P. Wand and M.C. Jones. *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1995.