

Modélisation et estimation de la dépendance et de la régularité.
Synthèse des travaux de recherche en vue de l'obtention du
diplôme d'habilitation à diriger des recherches.

François Roueff

Soutenue le 19 juin 2007

à l'Université Paris X – Nanterre

devant le jury composé de

Jean-Marc Azaïs (Professeur à l'université Paul Sabatier)	Rapporteur
Gennady Samorodnitsky (Professeur à Cornell University)	Rapporteur
Rainer von Sachs (Professeur à l'université catholique de Louvain)	Rapporteur
Stéphane Jaffard (Professeur à l'université Paris XII)	Examineur
Eric Moulines (Professeur à Télécom Paris)	Examineur
Philippe Soulier (Professeur à l'université Paris X)	Examineur

Mots-clés

Dépendance longue ; Analyse en ondelettes ; Processus shot-noise ; Estimation semi-paramétrique ; Estimation non-paramétrique ; Modèles de mélange ; Stationnarité locale ; Estimation récursive ; Problèmes inverses ; Modèle feuilles mortes ; Drap stable.

Résumé

Nous considérons le problème de l'estimation statistique de paramètres de certains modèles rencontrés dans des contextes de mesures : le paramètre de Hurst d'un modèle de trafic réseau, la distribution de la marque d'un processus ponctuel de Poisson pour la spectrométrie γ , la fréquence d'un signal observé irrégulièrement (par exemple en astronomie). En parallèle à ces travaux, nous nous intéressons d'une part à l'étude théorique de certains modèles impliquant des *lois d'échelle* – essentiellement le drap stable fractionnaire et certains processus ARCH(∞) – et d'autre part à l'étude théorique de certaines méthodes d'estimation : estimation du paramètre de Hurst par une analyse en ondelettes, d'un modèle AR localement stationnaire par les moindres carrés récursifs et de la densité d'un mélange de loi discrètes par un estimateur de projection. Le modèle utilisé est dans ces derniers cas *choisi* pour éprouver la méthode.

key-words

Long dependence ; Wavelet analysis ; Shot-noise process ; semi-parametric estimation ; non-parametric estimation ; Mixture models ; Locally stationary processes ; Recursive estimation ; Inverse problems ; Dead leaves model ; Stable sheet.

Summary

We consider the problem of statistical estimation of the parameters of some models encountered in a context of measurements : the Hurst parameter of a network traffic model, the marks distribution of a point process for γ spectrometry, the frequency of an irregularly sampled signal (for instance in astronomy). In addition to these works, we have been interested on the one hand in the theoretical study of some models involving power laws – essentially linear fractional stable sheet and some ARCH(∞) processes – and on the other hand in the theoretical study of some estimation methods : wavelet methods for estimating the Hurst parameter, least mean square recursive estimation of a time varying AR process and a projection estimator for estimating the density of a mixture of discrete distributions. In those latter examples, the model is *chosen* for evaluating the method.

Remerciements

Cette habilitation n'aurait sans doute pas eu lieu sans la sollicitation, et le support constant depuis mon arrivée à Télécom Paris, du Professeur Eric Moulines. Je le remercie vivement pour le rôle tout particulier qu'il a eu sur l'ensemble de mes activités d'enseignant-chercheur au sein de son équipe.

Le Professeur Philippe Soulier a beaucoup contribué à la réalisation de l'habilitation en acceptant d'être mon "parrain" auprès de l'Université Paris X. Il a été en outre un collaborateur essentiel dans mon parcours scientifique de ces dernières années, en particulier depuis que nous avons co-organisé la conférence "Lois de puissance en probabilité et statistiques" à Luminy en 2004.

Je tiens à remercier les Professeurs Jean-Marc Azaïs, Gennady Samorodnitsky et Rainer von Sachs d'avoir accepté de servir de rapporteurs de cette habilitation et le Professeur Stéphane Jaffard d'avoir bien voulu participer au jury celle-ci. Je suis flatté que ces distingués chercheurs et mathématiciens aient bien voulu s'intéresser aux travaux que je présente dans cette synthèse. J'espère qu'ils y ont trouvé quelques satisfactions les récompensant du temps qu'ils y ont investi.

Tous les travaux présentés ici ont été accomplis en collaboration avec de nombreux co-auteurs, qui s'ajoutent aux sus-nommés Eric Moulines et Philippe Soulier : Antoine Ayache, Charles Bordenave, Olivier Cappé, Randal Douc, Gilles Faÿ, Yann Gousseau, Céline Lévy-Leduc, Pierre Priouret, Tobias Rydén, Antoine Souloumiac, Murad Taqqu, Thomas Trigano, Yimin Xiao. Chacun mériterait des remerciements particuliers tant ils ont contribué de manières diverses mais toujours significatives au chemin parcouru au long de ces 7 dernières années.

Table des matières

1	Introduction	1
2	Modèles considérés	3
2.1	Mémoire longue d'un processus à variance finie	3
2.1.1	Séries à accroissements stationnaires au second ordre	3
2.1.2	Transformée en ondelettes discrète	5
2.1.3	Propriétés du second ordre aux grandes échelles	7
2.1.4	Généralisations	9
2.2	Modèles d'observations directes	10
2.2.1	Modèles linéaires	10
2.2.2	Séries localement stationnaires	11
2.2.3	Modèle ARCH	12
2.2.4	Fonction périodique échantillonnée irrégulièrement	13
2.3	Problèmes inverses	14
2.3.1	Modèles de mélange	14
2.3.2	Processus de type "shot-noise"	15
2.3.3	Mesures pour la spectrométrie γ	16
2.3.4	Processus de flots à arrivées poissonniennes	17
2.3.5	Mosaïques aléatoires	20
2.4	Modélisation de la régularité	22
2.4.1	Quelques notions de régularité	22
2.4.2	Drap linéaire fractionnaire stable	23
2.4.3	Modèle feuilles mortes à loi d'échelle	25
3	Estimation semi ou non-paramétrique	31
3.1	Modèles d'observations directs	31
3.1.1	Identification d'un processus TVAR	31
3.1.2	Estimation d'une fréquence en échantillonnage irrégulier	32

3.1.3	Estimation du paramètre de mémoire longue d'un processus linéaire ou gaussien	33
3.2	Problèmes inverses	35
3.2.1	Estimation de la densité d'un mélange de distributions discrètes	35
3.2.2	Estimation de la densité de la marque d'un processus shot-noise	36
3.2.3	Estimation de l'indice de queue de la marque d'un processus shot-noise	37

Table des figures

2.1	(a) : mosaïque de Voronoï; (b) : mosaïque de Delaunay.	27
2.2	Simulation de modèles feuilles mortes avec P donné par : (a) support réduit à un disque fixé; (b) disque à rayon de loi uniforme; (c) rotation et homotétie uniformes d'une même "feuille"; (d) rotation uniforme d'un rectangle.	28
2.3	Convergences du modèle feuilles mortes quand Y est un disque fixé : (a) convergence vers un champ constant ($\alpha < 3$ et $r_1 \rightarrow \infty$); (b) convergence vers un bruit blanc ($\alpha > 3$ et $r_0 \rightarrow 0$); (c) : convergence vers un champ non-trivial $\alpha = 2.9$ et $r_0 \rightarrow 0$, puis zooms successifs d'un facteur 2 de ce champ pour (d), (e), (f)	29

Chapitre 1

Introduction

Dans cette synthèse, nous avons pris le parti de présenter succinctement la liste exhaustive des différents sujets de recherche abordés depuis la thèse de doctorat. Un traitement exhaustif d'un de ces sujets ne serait pas moins intéressant mais ne permettrait pas de mettre en valeur les passerelles existantes : les problèmes inverses, les processus ponctuels, les ondelettes, les lois de puissance pour la modélisation de la dépendance, de la régularité et des queues lourdes, pour citer les plus évidentes.

Les sujets abordés étant de nature probabiliste et statistique, nous avons séparé la présentation en deux chapitres, le premier exposant les modèles probabilistes étudiés et le second des résultats sur l'estimation de certains de ces modèles. La plupart des travaux abordés dans le second chapitre auront donc été déjà en partie introduits dans le premier.

Il nous a semblé intéressant de diviser les modèles en deux catégories : les modèles d'observations directs (paragraphe 2.2) et ceux qualifiés de problèmes inverses (paragraphe 2.3), séparation qui se retrouve naturellement dans le chapitre consacré à l'estimation. La mémoire longue et certains modèles de régularité sont plus spécifiquement considérés dans les paragraphes 2.1 et 2.4, car ces parties recouvrent les deux classes de modèles considérés.

On parlera ici de problème inverse quand le modèle des observations est obtenu par une transformation *connue*, plus ou moins complexe, d'un modèle plus simple.

Ces modèles sont en général motivés par une formation particulière des données due à un environnement physique, observées indirectement par le biais d'une mesure. Nous verrons ainsi plusieurs modèles issus de processus ponctuels inspirés d'une modélisation *physique* (voir paragraphes 2.3.3, 2.3.4, 2.3.5) mais que les mesures ne permettent pas d'observer directement.

Soit X un élément aléatoire défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans X et F une fonction connue, définie sur X . Notons \mathbf{P} la loi *inconnue* de X . Les problèmes inverses s'exprimeront ainsi :

- (Q-1) Comment les propriétés de \mathbf{P} se transmettent-elles à $\mathbf{P} \circ F^{-1}$?
- (Q-2) Quelle perte d'information a-t-il été subie au cours de la transformation F ?
- (Q-3) En particulier, comment élaborer des procédures d'estimation de la loi \mathbf{P} à partir de l'observation $F(X)$ et comment se compare-t-elles aux procédures classiques qui s'appliquent à l'observation X ?

Nous tenterons d'aborder ces questions pour les différents modèles que nous rencontrerons dans cette synthèse. Pour mieux comprendre cette classe des problèmes inverses, expliquons sommairement pourquoi les modèles du paragraphe 2.2 n'en font pas partie. L'opposition de ces modèles avec ceux du paragraphe 2.3 réside dans le fait que, dans le cadre d'un problème inverse, la transformation F est contingente à la formation des observations mais ne contient pas l'information qui nous intéresse. Au contraire, pour ce qui est des modèles *d'observations directes*, quand ils s'écrivent comme la transformation d'un modèle

simple, ce sont justement les paramètres de cette transformation qui donnent aux modèles leurs propriétés *structurelles* (celles que, dans le cadre de l'estimation, l'on cherche à déterminer) :

1. dans le cas des modèles linéaires (paragraphe 2.2.1), c'est le filtre $\{a_k, k \in \mathbb{Z}\}$, par exemple à travers le paramètre de mémoire longue (défini au paragraphe 2.1) qu'il définit, qui nous intéresse et non le bruit blanc $\{\epsilon_k, k \in \mathbb{Z}\}$,
2. dans le cas des modèles TVAR (paragraphe 2.2.2), les paramètres $\theta(t)$ et $\sigma(t)$,
3. dans le cas du modèle ARCH (paragraphe 2.2.3), les paramètres $\{a_k, k \geq 0\}$,
4. dans le cas de la fonction périodique échantillonnée irrégulièrement (paragraphe 2.2.4), la fonction s_* .

Le cas du drap linéaire fractionnaire stable abordé au paragraphe 2.4.2 est à mettre à part dans la mesure où les paramètres α et H interviennent tous deux dans les propriétés structurelles du modèle et que le premier intervient au niveau du modèle simple, le drap de Lévy α -stable Z_α et le second au niveau de la transformation (2.53) de celui-ci.

Chapitre 2

Modèles considérés

Sommaire du chapitre

2.1	Mémoire longue d'un processus à variance finie	3
2.1.1	Séries à accroissements stationnaires au second ordre	3
2.1.2	Transformée en ondelettes discrète	5
2.1.3	Propriétés du second ordre aux grandes échelles	7
2.1.4	Généralisations	9
2.2	Modèles d'observations directes	10
2.2.1	Modèles linéaires	10
2.2.2	Séries localement stationnaires	11
2.2.3	Modèle ARCH	12
2.2.4	Fonction périodique échantillonnée irrégulièrement	13
2.3	Problèmes inverses	14
2.3.1	Modèles de mélange	14
2.3.2	Processus de type "shot-noise"	15
2.3.3	Mesures pour la spectrométrie γ	16
2.3.4	Processus de flots à arrivées poissonniennes	17
2.3.5	Mosaïques aléatoires	20
2.4	Modélisation de la régularité	22
2.4.1	Quelques notions de régularité	22
2.4.2	Drap linéaire fractionnaire stable	23
2.4.3	Modèle feuilles mortes à loi d'échelle	25

2.1 Mémoire longue d'un processus à variance finie

2.1.1 Séries à accroissements stationnaires au second ordre

Nous allons définir le paramètre de mémoire longue de certaines série temporelle dont les accroissements d'ordre suffisamment élevé sont stationnaires. Cette présentation reprend, en la généralisant légèrement, l'approche de [49] en annexe, elle-même largement inspirée de [32].

La notion de mémoire longue que nous considérons ici repose sur un *comportement en loi de puissance* de la densité spectrale à la fréquence nulle. Il y a des sens plus ou moins forts que l'on peut donner à un tel comportement. Une fonction positive f définie dans un voisinage à droite de l'origine est dite avoir

un comportement en loi de puissance d'indice $a \in \mathbb{R}$ en $t = 0^+$

au sens faible si

$$\liminf_{t \downarrow 0} \frac{\log f(t)}{\log(t)} = \limsup_{t \downarrow 0} \frac{\log f(t)}{\log(t)} = a, \quad (2.1)$$

au sens de Karamata si

$$\liminf_{t \downarrow 0} \frac{f(tx)}{f(t)} = \limsup_{t \downarrow 0} \frac{f(tx)}{f(t)} = x^a \quad \text{pour tout } x > 0, \quad (2.2)$$

au sens fort si

$$0 < \liminf_{t \downarrow 0} t^{-a} f(t) = \limsup_{t \downarrow 0} t^{-a} f(t) < \infty. \quad (2.3)$$

La relation (2.1) est impliqué par (2.2) qui est elle-même impliquée par (2.3). Quelque soit le sens choisi, f a un comportement en loi de puissance d'indice $a \in \mathbb{R}$ si et seulement si $t \mapsto t^{-a} f(t)$ a un comportement en loi de puissance d'indice nul. En particulier, la relation (2.2) est équivalente à $f(t) = L(t) t^a$ avec L fonction à variation lente au sens de Karamata en 0^+ . La notion de loi de puissance introduite ci-dessus nous sera utile dans d'autres contextes que celui de la mémoire longue.

Considérons une série temporelle $X = \{X_t, t \in \mathbb{Z}\}$, stationnaire au second ordre, de mesure spectrale ν sur $[-\pi, \pi]$. La série X est dite avoir un *paramètre de mémoire longue* $d < 1/2$ au sens fort (resp. au sens de Karamata ou au sens faible) si ν admet une densité au voisinage de la fréquence nulle qui a un comportement en loi de puissance d'indice $-2d$ en l'origine au sens fort (resp. au sens de Karamata ou au sens faible). Quand $d > 0$, on dit de X qu'elle est à *mémoire longue positive* ou à *longue dépendance*.

On peut aisément généraliser cette définition à un paramètre de mémoire quelconque $d \in \mathbb{R}$ en restreignant l'hypothèse de stationnarité aux accroissements de X . Le processus X sera dit avoir un paramètre de mémoire longue $d \in \mathbb{R}$ si, pour tout entier positif $k > d - 1/2$, $(\mathbf{I} - \mathbf{B})^k X$ est stationnaire au second ordre et admet pour paramètre de mémoire longue $d - k$, où l'on a noté \mathbf{I} l'opérateur identité et \mathbf{B} l'opérateur de retard :

$$[\mathbf{B}X]_t = X_{t-1} \quad \text{pour tout } t \in \mathbb{Z}.$$

Quand un processus X est tel que $(\mathbf{I} - \mathbf{B})^k X$ est stationnaire au second ordre pour un entier positif k , on appellera ν *mesure spectrale généralisée* de X une mesure symétrique positive sur $[-\pi, \pi]$ (non-nécessairement finie) telle que $(\mathbf{I} - \mathbf{B})^k X$ ait pour mesure spectrale la mesure de densité $\lambda \rightarrow |1 - e^{-i\lambda}|^{2k}$ par rapport à ν . Réciproquement, on dira que X admet une mesure spectrale généralisée ν si il existe un entier positif k tel que

$$\int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{2k} \nu(d\lambda) < \infty,$$

et si, pour tout tel k , $(\mathbf{I} - \mathbf{B})^k X$ est stationnaire au second ordre de mesure spectrale la mesure de densité $\lambda \rightarrow |1 - e^{-i\lambda}|^{2k}$ par rapport à ν .

Comme $|\lambda|$ et $|1 - e^{-i\lambda}|$ sont équivalents quand $\lambda \rightarrow 0$, d'après les définitions précédentes, on voit que X a un paramètre de mémoire longue $d \in \mathbb{R}$ si et seulement si X admet une mesure spectrale généralisée

$$\nu(d\lambda) = |1 - e^{-i\lambda}|^{-2d} \nu^*(d\lambda), \quad \lambda \in [-\pi, \pi], \quad (2.4)$$

où ν^* est la mesure spectrale d'un processus stationnaire au second ordre de paramètre de mémoire longue nulle.

Pour tout $d \in \mathbb{R}$, on notera $(\mathbf{I} - \mathbf{B})^d$ l'opérateur de différentiation fractionnaire d'indice d , défini formellement par

$$(\mathbf{I} - \mathbf{B})^d = \mathbf{I} + \sum_{j=1}^{\infty} \frac{(-d)(1-d) \cdots (j-1-d)}{j!} \mathbf{B}^j = \sum_{j \geq 0} b_j(d) \mathbf{B}^j. \quad (2.5)$$

Cette définition formelle signifie que pour tout $d \in \mathbb{R}$, si X est un processus stationnaire de mesure spectrale ν telle que

$$\int_{-\pi}^{\pi} |1 - e^{-i\lambda}|^{2d} \nu(d\lambda) < \infty, \quad (2.6)$$

le processus $(\mathbf{I} - \mathbf{B})^d X$ est défini en posant $(\mathbf{I} - \mathbf{B})^d = \Phi[(1 - \exp(-i \cdot))^d]$, où Φ dénote l'isométrie spectrale de $L^2(\nu)$ dans l'ensemble des opérateurs s'appliquant à X définie sur la famille dense des exponentielles complexes par

$$\Phi[\exp(-ik \cdot)] = \mathbf{B}^k.$$

Si maintenant X est un processus de mesure spectrale généralisée ν qui vérifie (2.6), $(\mathbf{I} - \mathbf{B})^d X$ est défini par

$$(\mathbf{I} - \mathbf{B})^d X = (\mathbf{I} - \mathbf{B})^{d-k} (\mathbf{I} - \mathbf{B})^k X,$$

où k est un entier positif tel que $(\mathbf{I} - \mathbf{B})^k X$ est stationnaire. En particulier si X est un processus stationnaire de paramètre de mémoire longue $d \in \mathbb{R}$, $(\mathbf{I} - \mathbf{B})^d X$ est un processus stationnaire de mesure spectrale ν^* définie par (2.4) et donc, c'est un processus stationnaire au second ordre de paramètre de mémoire longue nulle.

Cette définition formelle peut être plus explicite dans certains cas. On peut montrer que, si $d \notin \mathbb{N}$,

$$|b_j(d)| \sim c j^{-d-1} \quad \text{quand } j \rightarrow \infty.$$

Il s'en suit que

1. si $d \geq 0$, le filtre $(\mathbf{I} - \mathbf{B})^d$ peut s'appliquer à tout processus stationnaire sous la forme

$$(\mathbf{I} - \mathbf{B})^d X = \lim_{k \rightarrow \infty} \sum_{j=0}^k b_j(d) \mathbf{B}^j X, \quad (2.7)$$

où la convergence doit être comprise au sens L^2 ;

2. si $d > -1/2$, il peut encore s'appliquer à tout processus stationnaire admettant une densité bornée au voisinage de l'origine.

En particulier si X est un processus stationnaire de paramètre de mémoire longue $d \in (-1/2, 1/2)$, comme, pour $d < 0$ il a par définition une densité bornée à l'origine, $(\mathbf{I} - \mathbf{B})^d X$ peut être défini par (2.7).

2.1.2 Transformée en ondelettes discrète

Ce paragraphe se veut une introduction rapide des méthodes d'ondelettes, sans doute inutile pour un lecteur qui en est déjà familier. Son objectif principal est d'introduire et de comprendre les hypothèses (W-1)–(W-4) qui accompagnent l'analyse en ondelette utilisée dans [49, 47, 48] et présentes dans [20] sous une forme différente.

Soit ψ une fonction $\mathbb{R} \rightarrow \mathbb{R}$ et $\psi_{j,k}$, $j, k \in \mathbb{Z}$ ses versions dilatées–translatées définies par

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k).$$

On appellera ψ une ondelette, si elle vérifie des hypothèses de localisation spectrale et temporelle et si elle admet un certain nombre de moments nuls :

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0 \quad \text{pour } m = 0, \dots, M-1. \quad (2.8)$$

Les ondelettes sont nées d'un outil d'analyse temps–échelle inventé par J. Morlet reposant uniquement sur une telle fonction, à laquelle il a donné ce nom d'*ondelette*. Elles ont connu un essor particulier au cours des années 90, quand, grâce à l'introduction de la notion d'analyse multi-résolution par S. Mallat et

Y. Meyer, une approche systématique a permis de construire des fonctions ψ pour lesquelles $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ forment une base hilbertienne de $L^2(\mathbb{R})$, généralisant ainsi le système de Haar à des fonctions ψ de régularité prescrite, éventuellement à support compact, cette dernière avancée étant dû à I. Daubechies. L'intérêt immédiat fut la caractérisation de certains espaces de régularité (en particulier les espaces de Besov, voir le paragraphe 2.4), mais aussi la décomposition de certains processus gaussien, initiée dans [43], comme alternative à la décomposition de Karhunen–Loève, l'introduction de nouvelles méthodes de compression d'image et de débruitage et, pour clore cette liste nullement exhaustive, l'analyse temps-échelle de certaines séries chronologiques.

Nous avons utilisé les ondelettes dans deux champs très distincts des travaux concernés par cet ouvrage de synthèse :

1. la décomposition de certains champs stables pour l'analyse de leur régularité, dont les résultats seront mentionnés au paragraphe 2.4.2 ;
2. l'analyse et l'estimation de la mémoire longue d'un processus (paragraphe 2.1.3, 2.3.4, 3.1.3 et 3.2.3).

Nous avons choisi, par souci de concision, de ne détailler l'utilisation des ondelettes que dans le cadre de ce deuxième point. Cette analyse repose tout d'abord sur le calcul des coefficients d'ondelette définis pour une fonction f par

$$c_{j,k}(f) = \int_{-\infty}^{\infty} f(t) \psi_{j,k}(t) dt, \quad j, k \in \mathbb{Z}, \quad (2.9)$$

puis sur l'analyse des coefficients $c_{j,k}(f)$ essentiellement en fonction de l'indice d'échelle j . La transformation qui à f associe ses coefficients $c_{j,k}(f)$ s'appelle *transformée en ondelette discrète* par opposition à la transformée en ondelette *continue*

$$C(t, a; f) = a^{-1/2} \int_{-\infty}^{\infty} f(t) \psi(a^{-1}(t-b)) dt, \quad a > 0, b \in \mathbb{R},$$

introduite initialement par J. Morlet. Pour mener notre analyse, la propriété d'orthogonalité, c'est-à-dire, le fait que pour $f \in L^2(\mathbb{R})$,

$$f = \sum_{j,k \in \mathbb{Z}} c_{j,k}(f) \psi_{j,k},$$

où la série converge au sens L^2 , ne sera pas indispensable ; aussi beaucoup de résultats d'analyse temps-échelle reposent-ils uniquement sur la localisation spectrale et temporelle et la propriété (2.8). D'un point de vue pratique cependant cet aspect, où du moins l'aspect analyse multi-résolution est crucial pour le calcul rapide des coefficients d'ondelette, puisqu'alors cette analyse permet de calculer les coefficients d'ondelette de l'échelle d'analyse la plus fine (en général $j = 0$) aux plus grossières (j croissant) par projections successives.

Pour une bonne compréhension de la suite, précisons les grandes lignes de cet algorithme. Une analyse multi-résolution repose sur une suite de sous-espaces vectoriels fermé $(V_j)_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$ telle que

- (i) $(V_j)_{j \in \mathbb{Z}}$ est une suite décroissante : $V_{j+1} \subset V_j$ pour tout $j \in \mathbb{Z}$;
- (ii) Il existe $\phi \in V_0$ tel que $\{\phi(\cdot - k), k \in \mathbb{Z}\}$ définisse une base hilbertienne de V_0 ;
- (iii) $\cup_{j \in \mathbb{Z}} V_j$ est dense dans $L^2(\mathbb{R}^n)$ et $\cap_{j \in \mathbb{Z}} V_j = \{0\}$;
- (iv) pour tout $j \in \mathbb{Z}$, $f \in V_j$ si et seulement si $f(2^{-j}\cdot) \in V_0$.

Toutes ces propriétés sont indispensables mais la dernière est la plus importante car elle implique que tous les espaces V_j se déduisent de V_0 (et donc de ϕ). La fonction ϕ s'appelle *fonction d'échelle* (parfois aussi *ondelette père*, par opposition à ψ , l'*ondelette mère*). L'ondelette ψ correspondant à cette analyse multi-résolution est telle que $\{\psi(\cdot - k), k \in \mathbb{Z}\}$ définisse une base hilbertienne de W_0 , espace défini par la relation

$$V_j = W_j \oplus_{\perp} V_{j+1}, \quad j \in \mathbb{Z}.$$

Les fonctions ϕ et ψ se trouvent intimement liés par cette relation, ce qui se manifeste par des propriétés de localisation temporelle et fréquentielle de même nature. De même, la propriété (2.8) se traduit-elle sur

ϕ par la propriété

$$\sum_{k \in \mathbb{Z}} k^m \phi(t - k) \text{ est polynomial en } t \text{ de degré } m \text{ pour } m = 0, 1, \dots, M - 1. \quad (2.10)$$

Nous avons pris pour convention que l'échelle d'indice $j = 0$ corresponde à l'échelle la plus fine de l'analyse en ondelette. Cette analyse est initiée en calculant la projection orthogonale de f sur V_0 , notée f_0 , puis obtenue récursivement, pour $j \geq 1$ en projetant f_{j-1} sur W_{j-1} et V_j , la première se décomposant sur la base $\{\psi_{j-1,k}, k \in \mathbb{Z}\}$ précisément avec les coefficients recherchés (2.9), la seconde, notée f_j , étant utilisée pour l'étape suivante. Hormis la première projection sur V_0 , toutes les autres projections sont des simples opérations de filtrage convolutif, qui à une suite de coefficients l^2 des coordonnées d'une fonction de V_0 dans sa base $\{\phi_{0,k}, k \in \mathbb{Z}\}$ associe les coefficients de ses projections sur les bases $\{\phi_{1,k}, k \in \mathbb{Z}\}$ et $\{\psi_{0,k}, k \in \mathbb{Z}\}$. Quand ϕ et ψ sont à support compact, ces filtres sont de plus à réponse impulsionnelle finie (filtre RIF). Dans ce cas, on appelle cet algorithme un *algorithme pyramidal*. Vu sous l'angle du filtrage, cet algorithme est un banc de filtre, tel qu'on l'entend classiquement en traitement du signal. La transformée en ondelette discrète est usuellement appliquée à des signaux à temps discret $x_k, k \in \mathbb{Z}$, en y appliquant le banc de filtre pyramidal décrit ci-dessus. Autrement dit, elle correspond au calcul de la transformée en ondelette discrète (2.9) de la fonction définie par

$$f(t) = \sum_{k \in \mathbb{Z}} x_k \phi(t - k). \quad (2.11)$$

Comme nous l'avons déjà mentionné, la notion d'analyse multi-résolution ne nous sera pas indispensable, mais par son importance pratique, nous nous devons de l'inclure dans notre description. Aussi, la transformée en ondelette discrète pour une série $\mathbf{x} = \{x_k, k \in \mathbb{Z}\}$ que nous utilisons est-elle simplement définie en appliquant (2.11) puis (2.9), sous les hypothèses suivantes pour ϕ et ψ :

- (W-1) ϕ et ψ ont des supports compacts, et vérifient $\hat{\phi}(0) = \int_{-\infty}^{\infty} \phi(t) dt = 1$ and $\int_{-\infty}^{\infty} \psi^2(t) dt = 1$;
- (W-2) Il existe $\alpha > 1$ tel que $\sup_{\xi \in \mathbb{R}} |\hat{\psi}(\xi)| (1 + |\xi|)^\alpha < \infty$, où $\hat{\psi}$ dénote la transformée de Fourier de ψ ;
- (W-3) La fonction ψ admet M moments nuls, *i.e.* vérifie (2.8) ;
- (W-4) La fonction ϕ interpole les polynômes jusqu'à l'ordre $M - 1$, *i.e.* vérifie (2.10).

On notera $W_{j,k}^{\mathbf{x}}$ les coefficients d'ondelettes obtenus suite aux deux étapes (2.11) et (2.9) :

$$W_{j,k}^{\mathbf{x}} = \int_{-\infty}^{\infty} \left(\sum_{l \in \mathbb{Z}} x_l \phi(t - l) \right) \psi_{j,k}(t) dt = \sum_{l \in \mathbb{Z}} \left\{ \int_{-\infty}^{\infty} \phi(t - l) \psi_{j,k}(t) dt \right\} x_l. \quad (2.12)$$

L'interversion des sommes et intégrales ici ne posent aucun problème sous l'hypothèse (W-1). On peut alors montrer, en manipulant le terme entre accolades sous les hypothèses (W-1)–(W-4) (voir le paragraphe 3 de [49] en annexe), que, pour tout $j \geq 0$,

$$W_{j,\cdot}^{\mathbf{x}} = \downarrow^j \circ \tilde{H}_j \circ (\mathbf{I} - \mathbf{B})^M(\mathbf{x}), \quad (2.13)$$

où $(\mathbf{I} - \mathbf{B})$ est l'opérateur de différentiation déjà considéré au paragraphe 2.1.1, \tilde{H}_j est un filtre passe-bas de réponse impulsionnelle finie et \downarrow est l'opérateur de décimation qui à $\{y_k, k \in \mathbb{Z}\}$ associe $\{y_{2k}, k \in \mathbb{Z}\}$. Pour j grand, les propriétés fréquentielles des filtres $H_j = \tilde{H}_j \circ (\mathbf{I} - \mathbf{B})^M$ héritées des hypothèses (W-1)–(W-4) sont données par la proposition 3 de [49] disponible en annexe.

2.1.3 Propriétés du second ordre aux grandes échelles

Nous nous sommes intéressés aux propriétés du second ordre du tableau

$$\{W_{j,k}, j \geq 0, k \in \mathbb{Z}\}$$

défini par (2.12) quand \mathbf{x} est remplacé par un processus à accroissements stationnaires de paramètre de mémoire longue donné $d \in \mathbb{R}$ tel que défini au paragraphe 2.1.1 et dans la limite des grandes échelles, *i.e.* $j \rightarrow \infty$. Nous avons montré que les propriétés du second ordre des coefficients d'ondelette sont asymptotiquement données par celles de la transformée en ondelette du *mouvement brownien fractionnaire généralisé* à temps continu défini par

$$W_{j,k}^{(d)} = B_{(d)}(\psi_{j,k}), \quad (2.14)$$

où $\{B_{(d)}(\theta)\}_\theta$ est le processus gaussien centré indexé par les fonctions θ vérifiant

$$\int_{-\infty}^{\infty} |\xi|^{-2d} |\hat{\theta}(\xi)|^2 d\xi < \infty,$$

de covariance

$$\text{Cov}(B_{(d)}(\theta_1), B_{(d)}(\theta_2)) = \int_{-\infty}^{\infty} |\xi|^{-2d} \hat{\theta}_1(\xi) \overline{\hat{\theta}_2(\xi)} d\xi.$$

La définition (2.14) est valide sous (W-1)–(W-4) si

$$M > d - 1/2 \quad \text{et} \quad \alpha > 1/2 - d, \quad (2.15)$$

car cela garantit alors que $\int_{-\infty}^{\infty} |\xi|^{-2d} |\widehat{\psi}_{j,k}(\xi)|^2 d\xi < \infty$ pour tout j, k . De plus pour $d \in (1/2, 3/2)$, les coefficients (2.14) ont même loi que

$$\int_{-\infty}^{\infty} \psi_{j,k}(t) B_H(s) ds,$$

où B_H est une version continue du mouvement brownien fractionnaire (MBF) de paramètre de Hurst $H = d - 1/2$; d'où le nom de *transformée en ondelette du mouvement brownien fractionnaire généralisé à temps continu*.

Précisons un peu la nature de la convergence de ces propriétés du second ordre aux grandes échelles. Pour décrire les propriétés du second ordre du tableau $\{W_{j,k}, j \geq 0, k \in \mathbb{Z}\}$, il n'est pas suffisant de considérer les séries $W_{j,\cdot}$ individuellement pour tout j . D'un autre côté, à cause de l'opérateur de décimation dans l'écriture (2.13), les séries $[W_{j,\cdot}, W_{j',\cdot}]^T$ ne sont pas conjointement stationnaires pour $j \neq j'$. En fait pour recouvrir une stationnarité jointe on peut considérer, par exemple pour $j > j'$, la série stationnaire $[W_{j,\cdot}, W_{j',2^{j-j'}\cdot}]^T$, dans laquelle la décimation relative $\downarrow^{j-j'}$ intervient. Mais pour que toutes les paires $\{W_{j,k}, W_{j',k'}\}$ soient considérées il faut prendre en compte *toutes* les séries $\{[W_{j,k}, W_{j',2^{j-j'}k+v}]^T\}_{k \in \mathbb{Z}}$, avec $v \in \{0, \dots, 2^{j-j'}\}$. On considère donc le processus à $2^{j-j'}$ composantes, appelé *processus inter-échelles*,

$$\{[W_{j,k}, \mathbf{W}_{j,k}(j-j')^T]^T\}_{k \in \mathbb{Z}}, \quad (2.16)$$

où, pour tout $u = 0, 1, \dots, j$,

$$\mathbf{W}_{j,k}(u) \stackrel{\text{def}}{=} [W_{j-u,2^u k}, W_{j-u,2^u k+1}, \dots, W_{j-u,2^u k+2^u-1}]^T. \quad (2.17)$$

Par convention, on a $\mathbf{W}_{j,k}(0) = W_{j,k}$.

Une partie importante de [49] consiste à étudier la co-densité spectrale $\mathbf{D}_{j,u} : [-\pi, \pi] \rightarrow \mathbb{C}$ du processus (2.16) définie par la relation

$$\text{Cov}(W_{j,k}, \mathbf{W}_{j,k'}(u)) = \int_{-\pi}^{\pi} e^{i\lambda(k-k')} \mathbf{D}_{j,u}(\lambda) d\lambda.$$

De la même façon, on peut définir la co-densité spectrale associée au tableau $\{W_{j,k}^{(d)}, j, k \in \mathbb{Z}\}$ défini par (2.14). Celle-ci prend la forme particulière

$$\text{Cov}(W_{j,k}, \mathbf{W}_{j,k'}(u)) = \int_{-\pi}^{\pi} e^{i\lambda(k-k')} 2^{2jd} \mathbf{D}_{\infty,u}(\lambda; d) d\lambda.$$

La notation $\mathbf{D}_{\infty,u}$ apparaîtra clairement ci-dessous pour des raisons déjà plus ou moins dévoilées.

Grâce à l'expression (2.13) de la transformée en ondelette discrète, la seule condition pour que ces co-densités existent consiste à avoir que $(\mathbf{I} - \mathbf{B})^M X$ admet une densité spectrale, ce qui est le cas pour un processus de paramètre de mémoire longue d dès que $M > d - 1/2$ (cf. (2.15)) et la mesure ν^* définie par (2.4) admet une densité spectrale sur $[-\pi, \pi]$. L'expression exacte de $\mathbf{D}_{j,u}$ valide sous cette hypothèse est donnée par [49, corollaire 1].

Enfin dans [49, théorème 1], est établi, sous des hypothèses classiques de régularité de la densité spectrale f^* de ν^* à la fréquence nulle (voir aussi (3.4) ci-dessous) et sous la condition (2.15) que

1. d'une part, $2^{-2jd} \text{Var}(W_{j,0}) \rightarrow \text{Var}(W_{0,0}^{(d)}) f^*(0)$ quand $j \rightarrow \infty$.
2. d'autre part, quelque soit $u \geq 0$, $2^{-2jd} \mathbf{D}_{j,u} \rightarrow \mathbf{D}_{\infty,u}(\cdot; d)$ au sens de la norme sup sur $[-\pi, \pi]$ quand $j \rightarrow \infty$.

Dans le cas gaussien ce type de résultat est suffisant pour analyser des estimateurs du paramètre de mémoire longue d comme nous le verrons au paragraphe 3.1.3. Dans le cas linéaire, des adaptations sont possibles, voir [48, théorème 1]; enfin des adaptations sont aussi possibles dans le cas où ν^* n'admet une densité qu'au voisinage de la fréquence nulle, voir [48, corollaire 2].

2.1.4 Généralisations

Il est bien connu que la régularité d'une fonction périodique, en particulier en 0, est liée au comportement de ses coefficients de Fourier à l'infini. Une définition communément utilisée d'absence de longue mémoire pour un processus stationnaire X se base sur la condition

$$\sum_{\tau \in \mathbb{Z}} |\gamma(\tau)| < \infty,$$

où l'on a noté γ la fonction d'auto-covariance de X . Cette condition implique en particulier, quand $n \rightarrow \infty$,

$$\text{Var} \left(\sum_{k=1}^n X_k \right) \sim \left(\sum_{\tau \in \mathbb{Z}} \gamma(\tau) \right) n$$

Examinons brièvement le comportement de cette variance à partir des conditions de loi de puissance de la mesure spectrale ν de X à la fréquence nulle. On a, pour tout $n \geq 1$,

$$\text{Var} \left(\sum_{k=1}^n X_k \right) = \int_{-\pi}^{\pi} n K_n(\lambda) \nu(d\lambda),$$

où K_n est le noyau de Fejér défini par

$$K_n(t) = \frac{1}{n} \left| \sum_{k=1}^n e^{it} \right|^2 = \frac{1}{n} \left| \frac{1 - e^{in\lambda}}{1 - e^{i\lambda}} \right|^2 = \frac{1}{n} \left(\frac{\sin(nt/2)}{\sin(t/2)} \right)^2.$$

Supposons que X admette un paramètre de mémoire longue $d \in (-1/2, 1/2)$ au sens de Karamata comme défini au paragraphe 2.1.1, *i.e.* il existe $\epsilon > 0$ tel que pour $|\lambda| \leq \epsilon$,

$$\nu(d\lambda) = |1 - e^{-i\lambda}|^{-2d} L(\lambda) d\lambda,$$

avec L fonction à variation lente en 0^+ . En utilisant que $nK_n(\lambda)$ est borné indépendamment de n pour $|\lambda| \in [\epsilon, \pi]$ et les théorèmes classiques de Karamata pour évaluer

$$\int_0^{\epsilon} nK_n(\lambda) |1 - e^{-i\lambda}|^{-2d} L(\lambda) d\lambda = \int_0^{\epsilon} |1 - e^{-in\lambda}|^2 |1 - e^{-i\lambda}|^{-2d-2} L(\lambda) d\lambda,$$

on obtient, en posant $c = \int_{-\infty}^{\infty} |1 - e^{-iu}|^2 |u|^{-2d-2} du$, quand $n \rightarrow \infty$,

$$\text{Var} \left(\sum_{k=1}^n X_k \right) \sim c L(n^{-1}) n^{2d+1}. \quad (2.18)$$

On retrouve que lorsque cette variance est de l'ordre de n , ce qui inclut le cas où l'auto-covariance est sommable, $\sum_{\tau} |\text{Cov}(X_0, X_{\tau})| < \infty$, on a nécessairement $d = 0$.

Dans le cas où les $X_k = B_H(k) - B_H(k-1)$ sont les accroissements d'un mouvement brownien fractionnaire (MBF) $\{B_H(t), t \in \mathbb{R}\}$ de paramètre de Hurst $H \in (0, 1)$, on a alors exactement

$$\text{Var} \left(\sum_{k=0}^n X_k \right) = \text{Var}(B_H(n)) = \sigma^2 n^{2H}.$$

C'est pourquoi on identifie souvent $d + 1/2$ à un *paramètre de Hurst* : c'est en effet le paramètre de Hurst du MBF dont les accroissements stationnaires ont pour paramètre de mémoire longue d .

Réciproquement, le processus $\{B_H(k+1) - B_H(k), k \in \mathbb{Z}\}$ des accroissements du MBF de paramètre de Hurst H ont pour paramètre de mémoire longue $d = H - 1/2$ et le mouvement brownien fractionnaire (MBF) échantillonné à temps discret $\{B_H(t), t \in \mathbb{Z}\}$ a donc quant à lui pour paramètre de mémoire longue $d = H + 1/2$.

La formule (2.18) est utile pour certaines généralisations de la notion de paramètre de mémoire longue que nous avons précédemment défini à partir des propriétés spectrales. En effet, pour certains processus, le comportement (2.18) s'obtient plus facilement que le comportement spectral. Un autre avantage est que cette approche se généralise plus facilement aux processus n'ayant pas d'accroissements stationnaires. Nous verrons un cas illustrant cette approche au paragraphe 2.3.4. Ainsi, pour un processus à temps continu $\{X(t), t \geq 0\}$, sans aucune hypothèse de stationnarité, on dira qu'il admet un paramètre de Hurst H si

$$\text{Var} \left(\int_0^T X(s) ds \right) = L(T) T^{2H} \text{ avec } L \text{ à variation lente en } +\infty.$$

2.2 Modèles d'observations directes

2.2.1 Modèles linéaires

Sous sa définition la plus générale, un processus linéaire $\{X_t, t \in \mathbb{Z}\}$ est obtenu comme le filtrage linéaire d'un bruit blanc $\{\epsilon_t, t \in \mathbb{Z}\}$,

$$X_t = \sum_{k \in \mathbb{Z}} a_k \epsilon_{t-k}, \quad (2.19)$$

où $\{a_k, k \in \mathbb{Z}\}$ est une suite réelle de $l^2(\mathbb{Z})$. Le processus $\{X_t, t \in \mathbb{Z}\}$ est donc stationnaire au second ordre de densité spectrale

$$f(\lambda) = \left| \sum_{k \in \mathbb{Z}} a_k e^{-ik\lambda} \right|^2, \quad \lambda \in (-\pi, \pi),$$

où la convergence de la série doit être comprise au sens $L^2(-\pi, \pi)$.

Tout processus gaussien stationnaire est un processus linéaire si et seulement si il admet une densité spectrale, et dans ce cas, le bruit blanc $\{\epsilon_t, t \in \mathbb{Z}\}$ peut lui-même être choisi gaussien sans perte de généralité, et donc blanc au sens fort.

Dans les problèmes d'estimation des propriétés du second ordre d'un processus stationnaire, ou d'un paramètre dépendant uniquement des propriétés du second ordre, l'hypothèse linéaire est souvent une extension de l'hypothèse gaussienne, par exemple à une hypothèse du type (2.19) avec

(i) $\{\epsilon_t, t \in \mathbb{Z}\}$ est un bruit blanc fort et $\sup_{t \in \mathbb{Z}} \mathbb{E}\epsilon_t^4 < \infty$.

Autrement dit, il ne sera pas attendu de l'hypothèse linéaire qu'elle change la nature d'un résultat qui s'applique au processus gaussien de même covariance.

Le cas de l'estimation du paramètre de mémoire longue, défini au paragraphe 2.1, sera un bon exemple :

1. nous montrerons des résultats d'estimation valides pour les modèles gaussiens et les généraliserons au moins partiellement pour les modèles linéaires (voir paragraphe 3.1.3);
2. puis nous étudierons un modèle non-linéaire, du type shot-noise, qui a les mêmes propriétés du second ordre mais pour lequel les propriétés du même estimateur sont différentes (voir paragraphe 3.2.3).

2.2.2 Séries localement stationnaires

La classe des processus localement stationnaires a été introduite par [12] puis étudiée dans [13] et [14]. L'objectif est de définir une densité spectrale locale $f(\cdot, t)$ en tout point t du temps. La première idée consiste à rendre le temps continu afin que le sens de "locale" corresponde au resserrement du temps autour d'un instant donné; la seconde est d'ajouter une dimension afin de pouvoir définir des processus stationnaires locaux à chaque instant. On obtient alors la définition suivante. Soit $\{X_{k,n}, k \in \mathbb{Z}, n \geq 1\}$ un tableau de v.a. réelles centrées et de variance finie telles que, pour tout $\tau \in \mathbb{Z}$,

$$\left| \text{Cov}(X_{k+\tau,n}, X_{k,n}) - \int_{-\pi}^{\pi} f(\lambda, t) e^{i\tau\lambda} d\lambda \right| \rightarrow 0, \quad (2.20)$$

quand $k/n \rightarrow t$ et $n \rightarrow \infty$, où, pour tout $t \in \mathbb{R}$, $f(\lambda, t)$ est appelée la *densité spectrale locale*. Dans le tableaux $\{X_{k,n}, k \in \mathbb{Z}, n \geq 1\}$ le temps correspond à k/n . Cette astuce de la seconde dimension permet de définir des densités spectrales locales très générales qui ne pourraient pas être obtenues si l'on avait considéré un processus à temps continu $\{X(t), t \in \mathbb{R}\}$ et remplacé $\text{Cov}(X_{k+\tau,n}, X_{k,n})$ par $\text{Cov}(X(t+\tau/n), X(t))$ dans la formule précédente.

Dans la définition originale de [12] des conditions particulières sont requises sur la convergence (2.20) ($O(n^{-1})$ et uniformité en $t \in [0, 1]$) afin d'obtenir des résultats d'estimation paramétrique quand la densité spectrale locale dépend d'un paramètre mais nous nous contenterons de cette définition moins contraignante dans cet exposé. De même, on peut par commodité indexer le temps par $t \in \mathbb{R}$, ou se contenter de le faire sur $t \in [0, 1]$, auquel cas on prend $k \in \{0, 1, \dots, n\}$, ou $t \in \mathbb{R}_+$, auquel cas on prend $k \in \mathbb{N}$. Dans la suite, nous choisirons ce dernier cas, car il autorise le choix des conditions initiales.

Un aspect important de cette classe de modèles est qu'elle permet une analyse asymptotique *consistante* pour l'estimation des paramètres de modèles très simples et très anciens qui inclut les processus AR à coefficients variants dans le temps (processus TVAR), voir notamment [56, 30, 27]. Ce sont les processus TVAR qui nous ont plus particulièrement intéressés et nous allons donc en détailler le cadre. Un processus TVAR $\{X_{k,n}, k \in \mathbb{N}, n \geq 1\}$ d'ordre $p \in \mathbb{N}$ est décrit par une équation de récurrence

$$X_{k+1,n} = \boldsymbol{\theta}_{k,n}^t \mathbf{X}_{k,n} + \sigma_{k+1,n} \epsilon_{k+1,n}, \quad (2.21)$$

où

- (i) on a noté $\mathbf{X}_{k,n} = [X_{k,n} \ \dots \ X_{k-p,n}]^T$,
- (ii) le tableau $\{\epsilon_{k,n}, k \in \mathbb{N}, n \geq 1\}$ est un bruit blanc fort centré unitaire,
- (iii) les coefficients variant dans le temps sont donnés par $\boldsymbol{\theta}_{k,n} = \boldsymbol{\theta}(k/n)$ avec $\boldsymbol{\theta} : \mathbb{R}_+ \rightarrow \mathbb{R}^p$ et $\sigma_{k,n} = \sigma(k/n)$ avec $\sigma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$.

Dans le cas du TVAR(p), comme indiqué dans [13], sous des hypothèses adéquates, la convergence (2.20) aura lieu en posant

$$f(\lambda, t) = \frac{\sigma^2}{2\pi} \left| 1 - \sum_{k=1}^p \boldsymbol{\theta}_k(t) e^{i\lambda kt} \right|^{-2},$$

qui est la densité spectrale du modèle AR(p) au point t , c'est-à-dire de coefficients $\theta_1(t), \dots, \theta_p(t)$ et d'innovation de variance $\sigma^2(t)$. Deux types d'hypothèses sont utilisées pour montrer la convergence (2.20) :

1. pour tout $t \geq 0$, $\theta(t)$ est le paramètre d'un AR causal :

$$1 - \sum_{k=1}^p \theta_k(t) z^k \neq 0 \quad \text{pour tout } z \in \mathbb{C} \text{ t.q. } |z| \leq 1. \quad (2.22)$$

2. Les fonctions $t \mapsto \theta(t)$ et $t \mapsto \sigma(t)$ sont régulières, par exemple lipschitziennes.

Une troisième condition est en fait utilisée, elle porte sur les conditions initiales $\mathbf{X}_{0,n}$. Le plus simple est de supposer que $\mathbf{X}_{0,n}$ suit la loi stationnaire du modèle AR(p) au point 0, autrement dit, d'étendre le modèle sur $k \leq 0$ en posant $\theta(t) = \theta(0)$ et $\sigma(t) = \sigma(0)$ pour tout $t \leq 0$.

La Proposition 5 de [45] en annexe précise la convergence (2.20) pour des conditions initiales générales :

$$\left| \text{Cov}(X_{k+\tau,n}, X_{k,n}) - \int_{-\pi}^{\pi} f(\lambda, k/n) e^{i\tau\lambda} d\lambda \right| \leq M (C \rho^k + n^{-\beta}), \quad (2.23)$$

où $M > 0$, $\rho \in (0, 1)$ et $C \geq 0$ sont des constantes indépendantes de (k, n) du moment que k/n appartient à un compact fixé. On peut poser $C = 0$ si la matrice d'auto-covariance de $\mathbf{X}_{0,n}$ est égale à celle d'un AR de paramètre $\theta(0)$ et $\sigma(0)$. Le paramètre $\beta > 0$ est un indice de régularité des fonctions $t \mapsto \theta(t)$ et $t \mapsto \sigma(t)$ qui impose en particulier la continuité de ces fonctions. Il s'en suit que la condition (2.22) peut être uniformisée sur $t \in K$ avec K compact de \mathbb{R}_+ :

$$\inf \left\{ |z| : z \in \mathbb{C}, \exists t \in K, 1 - \sum_{k=1}^p \theta_k(t) z^k = 0 \right\} > 1.$$

On remarque pour tout $t > 0$, si $n \rightarrow \infty$ et $k/n \rightarrow t$, on obtient bien (2.20) à la vitesse $O(n^{-\beta})$, et la convergence est uniforme sur tout compact inclus dans $(0, \infty)$. Nous renvoyons à [45] pour plus de détails sur ce résultat.

Sans pour l'instant parler d'un estimateur particulier des paramètres, un résultat important pour le modèle est l'établissement de vitesses minimax pour l'estimation, à partir du n -uplet $(X_{1,n}, X_{2,n}, \dots, X_{n,n})$ des fonctions $t \mapsto \theta(t)$ et $t \mapsto \sigma(t)$ sur $t \in [0, 1]$ dans des espaces de régularités usuels. Ces vitesses sont les vitesses habituellement obtenues dans les modèles de régression non-paramétrique, pour tout $u \in \mathbb{R}^p$ non-nul,

$$\inf_{n \geq 1} n^{2\beta/(1+2\beta)} \inf_{f: \mathbb{R}^n \rightarrow \mathbb{R}} \sup_{(\theta, \sigma) \in \mathcal{C}_\beta} \mathbb{E}_{\theta, \sigma} [\{f(X_{1,n}, \dots, X_{n,n}) - \theta^T u\}^2] > 0,$$

où $\mathbb{E}_{\theta, \sigma}$ représente l'espérance du modèle défini par les paramètres θ et σ et \mathcal{C}_β est un ensemble de fonctions dont la norme au sens de l'indice de régularité β est majoré. Dans ce résultat, comme seul les observations $X_{1,n}, \dots, X_{n,n}$ sont considérées, les paramètres θ et σ sont définis sur $t \in [0, 1]$. La définition précise de \mathcal{C}_β et les hypothèses nécessaires sur le tableau $\{\epsilon_{k,n}, k \in \mathbb{N}, n \geq 1\}$ sont détaillées au théorème 4 dans [45].

2.2.3 Modèle ARCH

Bien qu'il existe des modèles ARCH localement stationnaires, voir [15], c'est bien de modèles stationnaires dont nous allons discuter ici, et même de l'existence de solutions stationnaires à l'équation ARCH(∞) définie comme suit

$$X_n = \sigma_n z_n, \quad (2.24)$$

$$\sigma_n^2 = a_0 + \sum_{j=1}^{\infty} a_j X_{n-j}^2, \quad (2.25)$$

où $\{z_n, n \in \mathbb{Z}\}$ est un bruit blanc fort centré unitaire, $a_0 > 0$ et $\{a_j, j \geq 1\}$ est une suite à valeurs positives. Le cas où $\sum_{j=1}^{\infty} a_j < 1$ est par exemple abordé dans [64], où les solutions stationnaires du second ordre des équations (2.24) et (2.25) sont étudiées.

Nous nous sommes plus particulièrement intéressé au cas de l'équation IARCH(∞), cas particulier donné par la condition

$$\sum_{j=1}^{\infty} a_j = 1. \quad (2.26)$$

Une solution des équations (2.24) et (2.25) est dite causale si, pour tout $n \in \mathbb{Z}$, σ_n^2 est mesurable pour la tribu engendrée par la famille $\{z_{n-t}, t \geq 1\}$. Dans le cas donné par (2.26), il est clair qu'il n'existe pas de solution causale stationnaire au second ordre puisqu'une telle solution aurait pour variance un nombre V vérifiant

$$V = \mathbb{E}[X_n^2] = \mathbb{E}[\sigma_n^2] = a_0 + \sum_{j=1}^{\infty} a_j V = a_0 + V,$$

ce qui est exclu par le fait que l'on a supposé $a_0 > 0$. Remarquons que pour $a_0 = 0$ il y a une solution triviale qui consiste à prendre $\sigma_n = 0$ pour tout $n \in \mathbb{Z}$. Une conséquence du théorème 1 dans [18], en annexe, est que sous les hypothèses de ce théorème pour $\{a_j, j \geq 1\}$ et la distribution de z_0 , la solution nulle est l'unique solution causale pour $a_0 = 0$. Mais l'objectif de ce théorème est avant tout de donner un résultat d'existence et d'unicité d'une solution stationnaire causale des équations (2.24) et (2.25), qui est non-triviale dès que $a_0 > 0$, et qui a donc une variance infinie sous la condition (2.26).

Ce résultat étend un résultat d'existence du processus IARCH(∞) établi dans [33] par des techniques très différentes. Leurs hypothèses imposent une convergence géométrique de la série $\sum_{j=1}^{\infty} a_j$, ce qui exclut une classe importante de modèles IARCH(∞) : la classe des processus FIGARCH(d) définie pour $d \in (0, 1)$, qui consiste à prendre l'équation (2.25) de la forme

$$\sigma_n^2 = a_0 + \{\mathbf{I} - (\mathbf{I} - \mathbf{B})^d\} X_n^2,$$

où l'opérateur de différentiation fractionnaire d'indice d , noté $(\mathbf{I} - \mathbf{B})^d$, est défini par (2.5).

Dès leur introduction par [7], les processus FIGARCH(d) ont connu une certaine popularité, bien qu'à notre connaissance notre travail est le premier à montrer leur existence (soit 10 ans après leur introduction), voir le corollaire 3 dans [18], en annexe. Une des raisons de cette popularité est d'espérer obtenir un modèle à la fois conditionnellement hétéroscédastique et exhibant de la longue dépendance. Il faut cependant noter, que si notre travail répond à la question de l'existence du FIGARCH(d), la question de ses propriétés de dépendance restent largement ouvertes. En particulier, à notre connaissance, il n'est pas rigoureusement établi qu'il existe un modèle ARCH(∞) exhibant un comportement du type des processus à dépendance longue. Sous la condition (2.26), l'étude de tels comportements est compliquée par le fait que ces modèles ont une variance infinie et ne rentre donc pas dans le cadre décrit par le paragraphe 2.1.

2.2.4 Fonction périodique échantillonnée irrégulièrement

On considère une fonction périodique à valeurs réelles s_* observée par les mesures

$$Y_j = s_*(X_j) + \epsilon_j, \quad j = 1, 2, \dots, n, \quad (2.27)$$

où les instants d'observations sont donnés par un processus de renouvellement $(X_j)_{j \geq 1}$, $X_j = \sum_{k=1}^j V_k$ avec $\{V_j, j \geq 1\}$ suite de v.a. i.i.d. positives de moyenne finie, et $\{\epsilon_j, j \geq 1\}$ est un bruit blanc gaussien. Ce modèle est notamment utilisé pour des données de rayonnement astronomiques, pour lesquelles les instants de mesures X_j sont soumis aux aléas météorologiques.

Nous verrons au paragraphe 3.1.2 un estimateur de la période minimale de s_* à partir d'observations $(X_j, Y_j), j = 1, \dots, n$. Nous allons ici introduire quelques propriétés élémentaires de ce modèle. Examinons

plus précisément quelles sont les propriétés spectrales de la suite $\{s_*(X_j), j \in \mathbb{N}\}$. Grâce à la périodicité de s_* les propriétés de renouvellement de $(X_j)_{j \geq 1}$ se traduisent pour $s_*(X_j)$ par des propriétés de stabilité. Plus précisément, si V_1 admet une composante absolument continue non-nulle, alors $\{s_*(X_j), j \in \mathbb{N}\}$ est fonction d'une chaîne de Markov géométriquement ergodique. Sa version stationnaire est obtenue en remplaçant V_1 par une v.a. \tilde{V}_1 uniforme sur $(0, T_*)$ où l'on a noté T_* la période de s_* (toujours indépendante de $V_j, j \geq 2$), voir la preuve de [36, théorème 3] en annexe.

Quand V_1 admet une composante absolument continue non-nulle, on a en particulier que pour tout $\epsilon > 0$,

$$\sup_{|t| \geq \epsilon} |\Phi(t)| < 1 \text{ où } \Phi(t) = \mathbb{E}[e^{itV_1}]. \quad (2.28)$$

Examinons rapidement, sous cette hypothèse, les propriétés du second ordre de $s_*(X_j)$ dans le cas où les coefficients de Fourier de s_* sont absolument sommables, *i.e.*,

$$s_*(t) = \sum_{l \in \mathbb{Z}} c_l e^{2i\pi l t f_*} \text{ avec } \sum_{l \in \mathbb{Z}} |c_l| < \infty,$$

où $f_* = 1/T_*$ est la fréquence de s_* . On a alors, pour tous $j, k \geq 0$,

$$\text{Cov}(s_*(X_j), s_*(X_{j+k})) = \sum_{l, l'} c_l c_{l'} \Phi^j(2\pi(l+l')f_*) \Phi^k(2\pi l' f_*) - \sum_{l, l'} c_l c_{l'} \Phi^j(2\pi l f_*) \Phi^{j+k}(2\pi l' f_*).$$

Par convergence dominée et en utilisant (2.28), la première double somme se réduit à $l+l' = 0$ quand $j \rightarrow \infty$ et la seconde à $l = l' = 0$. Comme $c_l = \overline{c_{-l}}$, on obtient donc, pour tout $k \geq 0$, quand $j \rightarrow \infty$,

$$\text{Cov}(s_*(X_j), s_*(X_{j+k})) \rightarrow \sum_{l \in \mathbb{Z}, l \neq 0} |c_l|^2 \Phi^k(2\pi l f_*) = 2 \sum_{l=1}^{\infty} |c_l|^2 \text{Re}[\Phi^k(2\pi l f_*)].$$

Les propriétés du second ordre de $\{s_*(X_j), j \geq 1\}$ sont donc asymptotiquement celles d'une somme pondérée de processus ARMA(2,2) indépendants dont les pôles sont conjugués de phases $\pm \lambda_l, l = 1, 2, \dots$, où λ_l est défini par l'angle de la représentation en coordonnées polaires de $\Phi(2\pi l f_*)$,

$$\Phi(2\pi l f_*) = \rho_l e^{i\lambda_l}, \quad \rho_l \in \mathbb{R}_+, \lambda_l \in [-\pi, \pi].$$

Ainsi dans le cas d'un échantillonnage poissonnien, $\Phi(t) = (1-it)^{-1}$, d'où $\lambda_l = \arctan(2\pi l f_*)$. Le caractère oscillant de s_* se retrouve donc dans le contenu spectral de $\{s_*(X_j), j \geq 1\}$ mais sous forme amortie. C'est pourquoi une analyse de cette série ou de sa version bruitée $\{s_*(X_j) + \epsilon_j, j \geq 1\}$, indépendamment des temps d'observations $\{X_j, j \geq 1\}$ induit une perte d'information. Aussi nous étudierons au paragraphe 3.1.2 un estimateur qui utilise une analyse spectrale conjointe de Y_j et $X_j, j = 1, \dots, n$, basé sur le périodogramme de Lomb-Scargle, qui tient compte des instants d'échantillonnage,

$$I_n^{LS}(f) \propto \left| \sum_{j=1}^n Y_j e^{-2i\pi f X_j} \right|^2. \quad (2.29)$$

Du fait de la présence des X_j à la fois dans le signal Y_j et dans les exponentielles complexes, l'étude de ce périodogramme est plus difficile que celle du périodogramme classique.

2.3 Problèmes inverses

2.3.1 Modèles de mélange

Quand la transformation F consiste à n'observer qu'une partie d'un modèle simple,

$$F : X = Y \times Z \rightarrow Y, \quad (y, z) \mapsto y,$$

les composantes non-observées Z de $X = (Y, Z)$ sont appelées données latentes. Les modèles de mélanges sont des modèles à données latentes pour lesquels la loi de Y conditionnellement à Z est connue et la loi de Z , appelée *distribution mélangeante* est inconnue. La loi de Y conditionnellement à Z est en générale issue d'un modèle dominé paramétrique, auquel cas elle est donné par une famille de densités $(\pi_\theta)_{\theta \in \Theta}$. La distribution mélangeante est alors une distribution μ sur Θ , et la loi de l'observation Y est donné par la densité

$$\pi_\mu = \mu\pi. = \int_{\Theta} \pi_\theta \mu(d\theta) . \quad (2.30)$$

Le lecteur peut se référer à [37, 59] pour des introductions générales aux modèles de mélange. Le cas des mélanges finis correspond à supposer que μ est de support de cardinal fini. La question (Q-1) a ici relativement peu d'intérêt et ce sont surtout les questions (Q-2) et (Q-3) qui nous occuperont. Le cadre que nous avons choisi dans [53] est celui de l'estimation non-paramétrique de μ quand un n -échantillon i.i.d. de loi π_μ est observé. Une idée importante dans le cas des mélanges est d'utiliser l'opérateur qui à une fonction $h : Y \rightarrow \mathbb{R}$ associe une fonction $\Pi h : \Theta \rightarrow \mathbb{R}$ définie par

$$[\Pi h](\theta) = \pi_\theta h = \int_Y \pi_\theta h d\zeta , \quad (2.31)$$

où ζ est la mesure de domination de la famille de densités $(\pi_\theta)_{\theta \in \Theta}$. L'importance de cet opérateur est notamment mis en valeur par les travaux de P. Barbe, voir [8]. Par des arguments assez simples nous montrons comment obtenir une borne inférieure très générale du risque minimax pour estimer μ , voir la proposition 1 de [53] en annexe. Cette borne sera exploitée dans le cas de l'estimation de la densité de μ dans des classes de type Sobolev quand ζ est la mesure comptage sur $Y = \mathbb{Z}_+$ (autrement dit, π_μ est une loi discrète sur les entiers positifs). Ces espaces sont définis par la qualité d'approximation d'une densité par la suite de ces projections sur les espaces

$$V_m = \text{Vect}\{\Pi \mathbb{1}_k : k = 0, 1, \dots, m\}, \quad m = 0, 1, \dots , \quad (2.32)$$

où $\mathbb{1}_k$ est la fonction définie sur \mathbb{Z}_+ qui vaut 1 en k et s'annule ailleurs. La borne de la proposition 1 s'adapte directement à ce cadre (voir [53, théorème 1]). Elle est alors en étroite relation avec un estimateur de projection sur les espaces V_m . Aussi reviendrons-nous à ce cadre, et en particulier aux vitesses de convergence minimax de certains modèles de mélange discret au paragraphe 3.2.1. Pour une grande classe de modèles nous verrons que la vitesse minimax du modèle à données latentes est beaucoup plus lente que celle du modèle d'observation direct.

2.3.2 Processus de type “shot-noise”

Soit M un processus ponctuel marqué sur l'espace \mathbb{R} dont les marques appartiennent à $\mathbb{R}^{\mathbb{R}^+}$ muni d'une tribu pour laquelle la fonction $\Phi : \mathbb{R} \times \mathbb{R}^{\mathbb{R}^+} \rightarrow \mathbb{R}$ qui à (t, x) associe $x(t)$ est mesurable. On notera

$$M = \sum_{k \in \mathbb{Z}} \delta_{T_k, Z_k} ,$$

où T_k sont les instants d'arrivée, ordonnés dans l'ordre croissant et tels que $T_{-1} < 0 \leq T_0$ et Z_k sont les marques, on notera $\{Z_k(t), t \in \mathbb{R}_+\}$ le processus associé. Le processus shot-noise associé à M est défini par

$$\mathbf{X}(t) = \sum_{k \in \mathbb{Z}} Z_k(t - T_k) = \sum_{k \in \mathbb{Z}} \Phi(t - T_k, Z_k) = [F(M)](t) , \quad (2.33)$$

quand cette somme est bien définie. Ce ne sera pas toujours le cas sous les hypothèses que nous considérerons. En revanche, le processus

$$\mathbf{X}(t) = \sum_{k \in \mathbb{Z}_+} Z_k(t - T_k) = \sum_{k \in \mathbb{Z}_+} \Phi(t - T_k, Z_k) = [F(M)](t) , \quad (2.34)$$

sera quant à lui toujours correctement défini. On appellera un élément $\{Z_k(t - T_k), t \geq 0\}$ de ces sommes une *impulsion* ou un *flot* suivant le contexte. Les processus shot-noise sont utilisés dans de nombreux domaines, souvent quand l'on cherche à modéliser des phénomènes physiques, comme les activités sismiques, la mesure par spectrométrie ou par fluorescence, mais aussi dans certains modèles de télé-traffic ou de séries financières. Dans le cas très délicat d'activités sismiques, le processus des arrivées est souvent extrêmement complexe et, qui plus est, c'est celui-ci qui intéresse le traiteur de données. Au contraire, dans les applications que nous regarderons le processus d'arrivée sera très simple et c'est bien les propriétés des marques qui nous intéressent. Aussi dans la suite on supposera les hypothèses suivantes :

- (SN-1) le processus des arrivées $\sum_k \delta_{T_k}$ un processus de Poisson homogène sur la droite ou la demi-droite réelle d'intensité $\lambda > 0$ (inconnue) ;
 (SN-2) Les Z_k sont de copies i.i.d. d'un processus $Z = \{Z(t), t \geq 0\}$, indépendantes du processus des arrivées.

La première hypothèse est principalement simplificatrice, des généralisations de nos résultats sont sans doute à attendre dans un cadre de processus d'arrivées plus général, comme celui des processus régénératifs. Comme nous allons le voir, certaines propriétés de la loi de Z se transmettent au processus \mathbf{X} de façon plus ou moins inattendues, ce que nous exploiterons pour l'estimation de certains paramètres de cette loi.

Ces deux hypothèses seront vérifiées par les processus shot-noise que l'on rencontrera par la suite. Dans le paragraphe 2.3.3, nous verrons une décomposition classique du processus shot-noise stable en cycles de périodes actives et inactives. Dans le paragraphe 2.3.4, cette décomposition ne sera pas utilisée et d'ailleurs pas toujours valide sous les hypothèses que l'on utilisera.

2.3.3 Mesures pour la spectrométrie γ

Dans le cas de la spectrométrie γ les instants T_k sont les temps d'arrivée d'un photon envoyé par un instrument de mesure et Z_k est la réponse électrique générée par l'absorption de ce photon enregistrée par l'instrument de mesure. L'énergie correspondant à ce courant, est égale (à un coefficient multiplicatif près de changement d'unité physique) à

$$Y_k = \int_0^\infty Z_k(t) dt, \quad k \in \mathbb{Z}_+. \quad (2.35)$$

Les $\{Y_k, k \in \mathbb{Z}_+\}$ forment donc une suite de v.a. i.i.d.. Sa densité comporte la signature du composant mesuré, des pics étant présents aux énergies répertoriées pour les éléments chimiques d'intérêt (nous avons considéré dans [60] de mesures spectrométriques d'un mélange de composants radioactifs établies par le CEA). Pour identifier ces composants à partir de mesures $\{\mathbf{X}(t), t \in [0, T]\}$, il faut donc commencer par répondre à la question (Q-1), beaucoup moins simple que dans le cas des modèles de mélange, la transformation F étant moins triviale. Il est clair que la loi des Y_k n'est pas identifiable sans hypothèses supplémentaires sur Z . On supposera donc que

le support de $t \mapsto Z(t)$ est un intervalle $[0, X]$ avec $X > 0$ et $\mathbb{E}[X] < \infty$.

et on notera X_k la taille du support de Z_k pour tout $k \in \mathbb{Z}_+$. En utilisant les propriétés des processus de Poisson, on peut alors montrer que $\{\mathbf{X}(t), t \geq 0\}$ défini par (2.34) (le temps $t = 0$ correspond au lancement de la mesure) a pour support une réunion infinie d'intervalles disjoints

$$\mathcal{S} = \bigcup_{k=1,2,\dots} [T'_k, T'_k + X'_k], \quad (2.36)$$

où, d'une part, $T'_1 = T_1$ et pour tout $k \geq 2$,

$$T'_k = \min \left\{ T_i : i \geq 1 \text{ et } T_i > T'_{k-1} \vee \max_{j=0,\dots,i-1} (T_j + X_j) \right\},$$

et, d'autre part, pour tout $k \geq 1$,

$$X'_k = \max \{T_j + X_j - T'_k : j \in \mathbb{Z}_+ \text{ et } T_j < T'_{k+1}\} . \quad (2.37)$$

Autrement dit, T'_k est l'instant de la première arrivée suivant T'_{k-1} pour laquelle toutes les impulsions précédentes sont finies et X'_k est le temps qu'il faut attendre après cette arrivée pour que le processus $\mathbf{X}(t)$ se "vide" à nouveau. L'hypothèse $\mathbb{E}[X] < \infty$ garantit ce retour à zéro de $\mathbf{X}(t)$. On pose alors

$$Y'_k = \int_{T'_k}^{T'_k + X'_k} \mathbf{X}(t) dt , \quad (2.38)$$

qui n'est autre que la quantité totale d'énergie mesurée pendant la période d'activité $t \in [T'_k, T'_k + X'_k]$.

En utilisant les propriétés d'oubli des processus de Poisson, on a la propriété fondamentale suivante (qui est un cas particulier des résultats de [52]) :

- (i) La suite $\{(X'_k, Y'_k), k \geq 1\}$ est une suite de v.a. i.i.d., indépendante de la suite des durées des périodes inactives $\{T'_1, T'_k - (T'_{k-1} + X'_{k-1}), k \geq 2\}$ qui, quant à elle, est suite de v.a. i.i.d. de loi exponentielle de moyenne $1/\lambda$. On a de plus $\mathbb{E}[X'] = (\exp(\lambda\mathbb{E}[X]) - 1)/\lambda$.

Remarquons enfin que toutes ces variables se déduisent de la trajectoire $\{\mathbf{X}(t), t \geq 0\}$ grâce à la relation (2.36) qui donne les variables T'_k, X'_k uniquement à partir des instant de retour à zéro de cette trajectoire puis grâce à la relation (2.38). Notre approche pour l'identification de la loi des Y_k repose sur une relation entre la loi \mathbf{P} du couple (X_k, Y_k) et celle, notée \mathbf{P}' du couple (X'_k, Y'_k) , généralisant un résultat de [57] qui établissait ce résultat pour la file M/G/ ∞ , c'est-à-dire quand

$$Z(t) = \mathbb{1}_{[0, X)}(t) , \quad (2.39)$$

auquel cas on a simplement $Y_k = X_k$. Cette généralisation est établie dans le théorème 3.1 de [46], en annexe. Elle s'écrit ainsi : pour tous nombres complexes s, p de partie réelle positives,

$$\int_0^\infty e^{-(s+\lambda)u} \left\{ e^{\lambda \mathbb{E}[\exp(-pY)(u-X)_+]} - 1 \right\} du = \frac{\lambda \mathcal{L}\mathbf{P}'(s, p)}{(s + \lambda) \{s + \lambda - \lambda \mathcal{L}\mathbf{P}'(s, p)\}} , \quad (2.40)$$

où \mathcal{L} désigne l'opérateur de la transformée de Laplace définie par

$$\mathcal{L}\mathbf{P}'(s, p) = \int e^{-su - pv} d\mathbf{P}'(u, v) = \mathbb{E}[e^{-sX' - vY'}] ,$$

où (X', Y') est un couple de même loi que les (X'_k, Y'_k) .

Dans l'équation (2.40), le membre de gauche dépend uniquement de \mathbf{P} et celui de droite de \mathbf{P}' . Cette relation répond donc, au moins d'un point de vue calculatoire, à la question (Q-1). C'est aussi cette relation que nous avons utilisé pour répondre aux questions (Q-2) et (Q-3), en s'intéressant à la quantité qui nous intéresse plus particulièrement, la densité de Y , comme expliqué ci-dessus dans l'application de la spectrométrie γ .

Comme dans le cas du modèle de mélange, et malgré la complexité de la transformation reliant le modèle d'observation directe à celui des observations indirectes, nous verrons au paragraphe 3.2.2 que l'estimation de cette densité peut se faire à partir des observations indirectes. Nous verrons aussi que la perte de vitesse au sens minimax dépendra cette fois du comportement de la queue de distribution de la durée X .

2.3.4 Processus de flots à arrivées poissonniennes

On appelle ici *processus de flots à arrivées poissonniennes*, (*Infinite Source Poisson process* en anglais), le processus shot-noise vérifiant (SN-1) et (SN-2) tel que Z est donné par

$$Z(t) = U \mathbb{1}_{[0, \eta)}(t) , \quad (2.41)$$

où U est une v.a. à valeurs dans \mathbb{R} et η une v.a. à valeurs dans \mathbb{R}_+ . On peut supposer sans perte de généralité que $|U| > 0$ p.s., puisque dans le cas contraire, sous (SN-1) et (SN-2), le processus ponctuel des points T_k tels que $U_k \neq 0$ est aussi un processus de Poisson homogène (procédé d'amincissement). Ce modèle est un cas particulier de celui considéré dans le paragraphe 2.3.3, puisque son support est précisément $[0, \eta)$, qui inclus le modèle de la file M/G/ ∞ , obtenu pour $U \equiv 1$. Dans la suite l'intensité λ des arrivées ne joue pas de rôle significatif, aussi le prendra-t-on égal à 1. Remarquons enfin que sous les notations du paragraphe 2.3.3, on a

$$X = \eta \quad \text{et} \quad Y = \int_0^\infty Z(t) dt = U \times \eta.$$

Le processus \mathbf{X} défini par (2.33) ou (2.34) est utilisé pour modéliser le télé-traffic dans les réseaux IP au niveau des *flots*. Un flot est défini comme un flux de paquets IP entre une source et une destination du réseau. Ce flux est idéalisé sous la forme d'un taux constant U qui rend compte de la bande passante allouée à ce flot. Dans le contexte du réseau IP actuel, ce taux est en effet déterminé par la qualité de l'accès local au réseau (connections ADSL, câble ou LAN). Si le trafic est mesuré en un point du coeur du réseau, il est donc raisonnable de supposer l'indépendance de ces taux. Quant à l'hypothèse poissonnienne des arrivées de ces flots, nous allons voir qu'elle n'est nullement en contradiction avec l'observation empirique de mémoire longue positive sur les traces de télé-traffic. Cette remarque avait déjà été faite dans [40] dans le cas de la M/G/ ∞ : si la distribution des temps de service a une queue lourde, alors le processus du nombre de client exhibe de la mémoire longue positive. Cette propriété a des effets sur l'agrégation temporel de ce processus. Les propriétés asymptotiques de la version agrégée en temps du processus de flots à arrivées poissonniennes

$$\int_0^T \mathbf{X}(t) dt \tag{2.42}$$

quand λ et la durée d'agrégation T tendent conjointement vers l'infini, ont été étudiés dans [44]. Mentionnons enfin que l'observation directe des flots Z_k est très difficile en pratique car elle nécessite un traitement fin des entêtes des paquets IP afin d'appareiller tous les paquets d'un même flot, tâche rendue impossible par la quantité et la complexité du trafic en un point du réseau. D'où la nécessité de travailler sur des mesures indirectes des ces flots, à savoir le processus $\mathbf{X}(t)$, éventuellement échantillonné à temps discrets.

Nous allons dans un premier temps nous contenter d'examiner les propriétés du second ordre de $\{\mathbf{X}(t), t \in \mathbb{R}\}$ défini, en toute généralité, par (2.34). Le processus (2.33) n'est en effet correctement défini que pour $\mathbb{E}[\eta] < \infty$, auquel cas, c'est un processus stationnaire au sens strict. Or nous ne supposons pas nécessairement que $\mathbb{E}[\eta] < \infty$. De même la méthodologie proposée au paragraphe 2.3.3 de décomposer le processus en périodes actives et inactives n'est possible que dans le cas $\mathbb{E}[\eta] < \infty$. Pour prendre en compte une éventuelle dépendance de U et η (dépendance avérée dans le cas des flots du télé-traffic), les hypothèses de queue lourde de la distribution de η prendront la forme suivante

$$\mathbb{E}[|U|^p \mathbb{1}_{\eta > t}] = L_p(t) t^{-\alpha}, \tag{2.43}$$

où

1. $p \in \{0, 1, \dots, p^*\}$ avec p^* entier aussi élevé que nécessaire ($p^* = 2$ suffit pour les propriétés du second ordre),
2. $\alpha \in (0, 2)$ est l'indice de queue de η ,
3. L_p est une fonction à variation lente en ∞ .

La place du paramètre α par rapport à 1 est bien sûr liée au fait que $\mathbb{E}[\eta]$ est fini ou non, qui décide si le processus \mathbf{X} est stable (convergence vers le processus stationnaire défini par (2.33)) ou non. Un exemple assez général de loi jointe pour (U, η) modélisant un flot de type TCP est donné dans [20] en annexe, pour lequel l'hypothèse (2.43) est vérifiée pour tout entier $p \geq 0$.

Les propriétés du second ordre de $\{\mathbf{X}(t), t \in \mathbb{R}\}$ défini par (2.34) et de sa version stationnaire (2.33) quand elle existe sont données par les propositions 2.1 et 2.2 dans [20]. En particulier, on trouve que, dans le cas non-stationnaire, pour tout $t \geq s > 0$,

$$\text{Cov}(\mathbf{X}(Tt), \mathbf{X}(Ts)) \sim C(s, t) L_2(T) T^{2-\alpha} \quad \text{avec} \quad C(s, t) = \int_{t-s}^s v^{-\alpha} dv ,$$

où l'équivalence est obtenue quand $T \rightarrow \infty$, et dans le cas stationnaire, par exemple pour $\alpha > 1$, quand $t \rightarrow \infty$,

$$\text{Cov}(\mathbf{X}(0), \mathbf{X}(t)) \sim (\alpha - 1)^{-1} L_2(t) t^{2-\alpha} .$$

On peut de même calculer, en conditionnant par rapport au nombre d'arrivée sur $[0, T]$ avec l'hypothèse poissonnienne, quelque soit la valeur de $\alpha \in (0, 2)$, pour \mathbf{X} défini par (2.34),

$$\text{Var} \left(\int_0^T \mathbf{X}(s) ds \right) = T \mathbb{E} [U^2(\eta \wedge T)^2] - \frac{2}{3} \mathbb{E} [U^2(\eta \wedge T)^3] .$$

Comme, par ailleurs, on a par Fubini, pour $q = 2, 3$,

$$\mathbb{E} [U^2(\eta \wedge T)^q] = q \int_0^T t^{q-1} \mathbb{E} [U^2 \mathbf{1}(\eta > t)] dt ,$$

on obtient, en utilisant (2.43) avec $p = 2$ et le théorème de Karamata sur l'intégration des fonctions à variations lentes, quand $T \rightarrow \infty$,

$$\text{Var} \left(\int_0^T \mathbf{X}(s) ds \right) \sim \frac{2}{(2-\alpha)(3-\alpha)} L_2(T) T^{2H} \quad \text{avec} \quad H = \frac{3-\alpha}{2} .$$

Cette dernière équation signifie que le processus \mathbf{X} admet pour paramètre de Hurst $H = (3-\alpha)/2$ (voir paragraphe 2.1.4).

On voit que sous les hypothèses très générales (2.43) sur la loi jointe de (η, U) , les propriétés de queue de la distribution de η se traduisent par des propriétés de mémoire longue pour le processus \mathbf{X} , ce qui répond en partie à la question (Q-1). Nous reviendrons sur cette remarque pour le problème de l'identification de l'indice α à partir de l'observation du processus \mathbf{X} au paragraphe 3.2.3 mais nous pouvons dès à présent mentionner quelques résultats importants quant à la transformée en ondelette discrète du processus \mathbf{X} sous les hypothèses (W-1)-(W-4), définie comme (2.9) par

$$d_{j,k} = \int \mathbf{X}(t) \psi_{j,k}(t) dt, \quad j \geq 0, k \in \mathbb{Z} .$$

En particulier, pour $\mathbb{E}[\eta] < \infty$ et \mathbf{X} défini par (2.33) et (2.41), comme $\psi_{j,k}$ a un support compact et il y a localement un nombre fini de flots Z_k non-nuls p.s., on a, pour tout $j \geq 0, k \in \mathbb{Z}$,

$$d_{j,k} = \sum_{l \in \mathbb{Z}} U_l \int_{T_l}^{T_l + \eta_l} \psi_{j,k}(t) dt . \quad (2.44)$$

On a vu que si $\mathbb{E}[\eta] = \infty$ (notamment pour $\alpha < 1$), il faut utiliser la définition (2.34) pour \mathbf{X} , auquel cas il doit bien sûr adapter (2.44) en

$$d_{j,k} = \sum_{l \in \mathbb{Z}_+} U_l \int_{T_l}^{T_l + \eta_l} \psi_{j,k}(t) dt . \quad (2.45)$$

On peut en fait montrer que dès que ψ a un moment nul ($M \geq 1$ dans les hypothèses (W-1)-(W-4)), même dans le cas où $\mathbb{E}[\eta] = \infty$, les coefficients d'ondelettes (2.44) sont bien définis : la somme sur $l \in \mathbb{Z}$ est finie p.s., voir [20, lemme 5.1]. Autrement dit, même quand \mathbf{X} n'a pas de version stationnaire, sa

transformée en ondelette admet elle toujours une version stationnaire. Les propriétés du second ordre sous l'hypothèse (2.43) de la cette transformée en ondelette sont données en partie par [20, lemme 3.1]. Leur adaptation au cas où \mathbf{X} est échantillonné à des temps discrets est aisément obtenu. La partie la plus technique consiste en fait à adapter les résultats obtenus pour la version stationnaire (2.44) de la transformée en ondelettes à la *vraie* transformée en ondelettes (2.45) du processus non-stationnaire.

Les propriétés du second ordre des $\{d_{j,k}\}$ ont de fortes similitudes avec l'étude des propriétés du second ordre de la transformée en ondelette présentée au paragraphe 2.1.3 en prenant $d = 1 - \alpha/2$. Il serait intéressant de fournir un cadre relativement général incluant ces deux situations, ce qui est bien sûr possible. Pour le moment, contentons-nous de mentionner qu'en utilisant [20, lemme 4.3], certaines passerelles peuvent être établies.

2.3.5 Mosaiques aléatoires

Une mosaïque aléatoire de \mathbb{R}^d est un processus ponctuel défini sur l'ensemble \mathcal{F}' des fermés non-vides de \mathbb{R}^d , $M = \sum_k \delta_{F_k}$, qui vérifie

- (i) $\bigcup_k F_k = \mathbb{R}^d$;
- (ii) pour tout $k \neq l$, $F_k \cap \text{Int } F_l = \emptyset$.

Les points F_k de M sont appelés les cellules de la mosaïque. On appelle la frontière de M la réunion des frontières des F_k , qu'on notera

$$\partial M \stackrel{\text{def}}{=} \bigcup_k \partial F_k .$$

Les deux points (i) et (ii) reviennent alors à supposer que les ensembles $\text{Int } F_k$ et ∂M forment une partition de \mathbb{R}^d . Une autre propriétés des mosaïques découlera de la définition que l'on donne aux processus ponctuels sur \mathcal{F}' : la mesure M doit être σ -finie, c'est-à-dire finie sur tout compact inclus dans \mathcal{F}' . Nous devons donc munir \mathcal{F}' d'une topologie, par suite d'une tribu de boréliens, qui fournira à l'ensemble des mesures σ -finie la tribu habituelle qui permet de définir les processus ponctuels.

Précisons la structure topologique de l'espace \mathcal{F} des fermés de \mathbb{R}^d . Nous adoptons l'approche de G. Matheron ("hit and miss" topology). On munit \mathcal{F} de la topologie engendrée par la base des ouverts $\{\mathcal{F}^K, \mathcal{F}_G : K \in \mathcal{K}, G \in \mathcal{G}\}$ où \mathcal{K} est l'ensemble des fermés de \mathbb{R}^d , \mathcal{G} l'ensemble de ses ouverts et où on a posé

$$\mathcal{F}^A = \{F \in \mathcal{F} : F \cap A = \emptyset\} \quad \text{et} \quad \mathcal{F}_A = \{F \in \mathcal{F} : F \cap A \neq \emptyset\} .$$

Alors, d'après [42], \mathcal{F} est un espace compact. L'espace $\mathcal{F}' = \mathcal{F} \setminus \{\emptyset\}$ n'est pas compact mais ses compacts sont inclus dans des ensembles de la forme \mathcal{F}_K avec $K \in \mathcal{K}$ si bien que les mesures σ -finies correspondent aux mesures finies sur tout ensemble \mathcal{F}_K avec $K \in \mathcal{K}$. Aussi, une mosaïque aléatoire $M = \sum_k \delta_{F_k}$ vérifiera la condition

- (iii) pour tout $K \in \mathcal{K}$, $\#\{F_k : F_k \cap K \neq \emptyset\} < \infty$.

Les mosaïques aléatoires les plus simples sont les mosaïques de Voronoï et de Delaunay (voir¹ la figure 2.1), entièrement définies à partir d'un processus ponctuel $N = \sum_k \delta_{T_k}$ sur \mathbb{R}^d . Chaque cellule de la mosaïque de Voronoï est définie comme l'ensemble des points de \mathbb{R}^d les plus proches d'un des points T_k . Celles de la mosaïque de Delaunay sont alors obtenues en reliant les paires de points $\{T_k, T_l\}$ dont les cellules de Voronoï ont un côté commun. Dans ces deux exemples, les cellules peuvent être reconstituées uniquement à partir de la frontière de la mosaïque puisqu'elles sont données par les composantes connexes du complémentaire de cette frontière.

Le modèle feuilles mortes est un autre exemple de mosaïque aléatoire, introduit par G. Matheron dans [41]. Ce modèle est utilisé en traitement des images, notamment parce qu'il reproduit quelques propriétés statistiques fondamentales des images : non-gaussianité, discontinuités, frontières de discontinuité ayant des jonctions en T etc. Nous renvoyons aux références incluses dans [26, 25, 10] pour les

¹les figures sont placées en fin de chapitre

nombreuses applications de ce modèle et les justifications empiriques de celui-ci pour la modélisation des images naturelles.

La définition d'un modèle feuille morte, comme celui d'un processus shot-noise, repose sur la transformation d'un processus ponctuel marqué mais, à la différence du processus shot-noise, cette transformation n'est pas additive. Considérons un processus ponctuel $\Phi = \sum_k \delta_{x_k, t_k, X_k}$ défini sur $\mathbb{R}^d \times (-\infty, 0] \times \mathcal{F}$. On prendra pour Φ un processus de Poisson homogène sur $\mathbb{R}^d \times (-\infty, 0]$ avec des marques i.i.d. X_k indépendantes du processus $\sum_k \delta_{x_k, t_k}$ et de loi sur \mathcal{F} notée P . On considère alors la mosaïque aléatoire

$$M = \sum_k \delta_{V_k} \mathbb{1}\{\text{Int}V_k \neq \emptyset\} = F(\Phi) \quad (2.46)$$

des "parties visibles" V_k d'intérieur non-vidé définies par

$$V_k = (x_k + X_k) \setminus \bigcup_{t_j \in (t_k, 0]} (x_j + \text{Int}X_j). \quad (2.47)$$

Les t_j peuvent être compris comme des temps, et les x_j comme des "centres" de la "feuille" $x_j + X_j$. Dans la formule (2.47), l'ensemble V_k représente la partie de la feuille $x_k + X_k$, qui n'a pas été recouverte par les feuilles $x_j + \text{Int}X_j$ tombés après cette feuille ($t_j > t_k$). Cette construction rend compte du phénomène d'*occlusion* dans la formation d'une image : un objet est en partie caché par les objets placés entre cet objet et le point de prise du cliché. Des conditions suffisantes sur la loi P sont données dans [10] en annexe pour garantir que le processus ponctuel M défini ci-dessus soit bien une mosaïque aléatoire, c'est-à-dire vérifie les conditions (i), (ii) et (iii). Sous ces conditions, une méthode de simulation exacte d'un modèle feuille morte est possible sur tout compact de \mathbb{R}^d dès que la loi P peut être simulée.

On voit sur la figure 2.2 des simulations de modèles feuilles mortes sur un pavé de \mathbb{R}^2 pour différentes lois P des objets X_j . Contrairement aux mosaïques de la figure 2.1, les cellules V_j ne se déduisent pas nécessairement de la frontière de la mosaïque. Pour visualiser ces mosaïques, un niveau de gris tiré aléatoirement et uniformément est alloué à chaque feuille. On obtient le champ aléatoire

$$\mathbf{Y}(t) = \sum_k C_k \mathbb{1}_{\text{Int}V_k}(t), \quad (2.48)$$

où les C_k sont des v.a. i.i.d., indépendantes de M . C'est ce champ qui est utilisé pour la modélisation statistique des images dans [35]. On voit en particulier dans le cas (d) de la figure 2.2 que les éléments V_i peuvent être constitués de plusieurs composantes connexes du complémentaire de la frontière.

Dans [10] en annexe, nous nous sommes intéressés principalement à répondre partiellement à la question (Q-1), notamment, connaissant la loi P ,

1. Quelle est la loi du fermé ∂M ? Cette loi est déterminée par les propositions 5 et 6, la première donnant, pour tous compacts K_1, \dots, K_n la probabilité que ces compacts soient tous inclus dans l'intérieur d'une cellule de M , et la seconde montrant que cette probabilité caractérise la loi de la frontière ∂M ;
2. Quelle est la loi de la cellule typique non-recouverte? La proposition 7 donne sa densité par rapport à P .
3. Quelle est la loi du processus ponctuel défini comme l'intersection de ∂M avec une droite? Celle-ci est donnée par la proposition 8 dans le cas où P a son support sur les fermés convexes.

Ces questions avaient déjà été abordées dans [41], la première uniquement dans le cas $n = 2$. Nous en avons généralisé le cadre par l'utilisation systématique du calcul de Palm.

La réponse à la première question est particulièrement importante puisqu'elle permet de calculer les lois fini-dimensionnelles du modèle feuille morte colorié, celui où une couleur aléatoire est allouée à chaque partie visible. Il sera notamment utile pour définir le modèle feuilles mortes à loi d'échelle au paragraphe 2.4.3.

2.4 Modélisation de la régularité

2.4.1 Quelques notions de régularité

La plupart des modèles que nous avons abordés jusqu'ici, quand ils sont définis sur des temps ou des variables d'espace continues (*a fortiori* quand ils sont à temps discret) ont un comportement aux "échelles fines" qui présente peu d'intérêt :

1. Le modèle shot-noise (2.33) a les mêmes propriétés locales que $Z(t)$ puisqu'il localement égal à une somme finie de copies indépendantes de celui-ci. Il est par exemple constant par morceaux dans le cas où Z est donné par (2.41).
2. Le modèle feuilles mortes colorié (2.48) est aussi une somme localement finie (par la propriété (iii) des mosaïques aléatoires) de fonctions indicatrices pondérées. Si C_k est remplacé par un champ $C_k(t)$ dans (2.48), le modèle est potentiellement plus riche mais indépendamment de la "partie géométrique" du modèle qui est donné par le processus M défini par (2.46) et (2.47).

Nous verrons dans les paragraphes 2.4.3 et 2.4.2 deux modèles où les comportements aux échelles fines présentent plus d'intérêt. Ce comportement sera décrit par des quantités qui rendent compte de la régularité, par exemple le module de continuité, défini pour toute fonction bornée $f : I \rightarrow \mathbb{R}$ (avec I pavé inclus dans \mathbb{R}^d) et tout $u > 0$ par

$$\omega(f, u)_\infty = \sup_{x, y \in I, \|x-y\| < u} |f(y) - f(x)|. \quad (2.49)$$

Le module de continuité correspond au module de régularité

$$\omega(f, u)_p = \sup_{y : \|t\| \leq u} \left\| \{f(\cdot + t) - f(\cdot)\} \mathbb{1}_{I \cap (I-t)}(\cdot) \right\|_p \quad (2.50)$$

défini pour $f \in L^p(I)$ quand $p = \infty$. Une première approche de la régularité d'une fonction f consiste à se demander quel est le comportement du module de continuité (2.49) quand $u \downarrow 0$. Les comportements en loi de puissance, même au sens le plus faible (2.1), sont trop restrictifs et on se contente de définir un *indice de régularité* tout simplement par

$$a = \liminf_{u \downarrow 0} \frac{\log \omega(f, u)_\infty}{\log(u)},$$

c'est-à-dire identique à l'indice défini par (2.1) mais sans supposer que cette limite inf soit aussi une limite. Ce a appartient à $[0, 1]$, à moins que f soit constante auquel cas $a = \infty$, ce qu'on exclut ci-après. On appelle a l'*indice de régularité höldérienne*. En effet, on a

$$a = \sup \{s \in (0, 1) : f \in C^s(I)\},$$

où $C^s(I)$ est l'espace de Hölder contenant toutes les fonctions f telles que

$$\sup_{u > 0} u^{-s} \omega(f, u)_\infty < \infty$$

ou, de façon équivalente, telles que $\bullet^{-s} \omega(f, \bullet)_\infty \in L^\infty(\mathbb{R}_+)$. En fait, on définit

$$|f|_{C^s} \stackrel{\text{def}}{=} \left\| \bullet^{-s} \omega(f, \bullet)_\infty \right\|_\infty = \sup_{x \neq y} \frac{|f(y) - f(x)|}{\|y - x\|^s},$$

qui est une semi-norme (elle s'annule pour les fonctions constantes sur I qui ne sont pas nécessairement nulles) et

$$\|f\|_{C^s} \stackrel{\text{def}}{=} \|f\|_\infty + |f|_{C^s}$$

définit une norme. Les espaces de Besov sont définis de façon similaire en remplaçant le module de continuité par un module de régularité d'ordre p , ainsi l'espace $B_p^{s, \infty}(I)$ est l'ensemble des fonctions de

$L^p(I)$ telles que $\cdot^{-s}\omega(f, \cdot)_p \in L^\infty(\mathbb{R}_+)$ et, autre légère extension, $B_p^{s,q}(I)$ est l'ensemble des fonctions de $L^p(I)$ telles que $\cdot^{-s}\omega(f, \cdot)_p \in L^q(\mathbb{R}_+)$. On définit donc, comme ci-dessus,

$$|f|_{B_p^{s,q}} \stackrel{\text{def}}{=} \|\cdot^{-s}\omega(f, \cdot)_p\|_q \quad \text{et} \quad \|f\|_{B_p^{s,q}} \stackrel{\text{def}}{=} \|f\|_\infty + |f|_{B_p^{s,q}} \quad (2.51)$$

On peut montrer les points suivants (tous sont classiques, voir par exemple [11], mais tous ne sont cependant pas triviaux)

1. $C^s(I) = B_\infty^{s,\infty}(I)$;
2. on a l'inclusion $B_p^{s,q}(I) \subseteq B_{p'}^{s',q'}(I)$ si $p \geq p'$, $q \leq q'$ et $s' \leq s$ sauf si I n'est pas borné, auquel cas il faut prendre $p = p'$ pour que l'inclusion reste valide;
3. le sup des $s \in (0, 1)$ pour lesquels $f \in B_p^{s,q}(I)$ ne dépend pas de $q \in [1, \infty]$;
4. Ce sup est donné pour tout $q \in [1, \infty]$ par

$$\sup \{s \in (0, 1) : f \in B_p^{s,q}(I)\} = \liminf_{u \downarrow 0} \frac{\log \omega(f, u)_p}{\log(u)}.$$

Si les points 3 et 4 indiquent que le paramètre q n'influence pas l'indice de régularité Besov, il n'en est en général pas de même du paramètre p . On sait seulement, d'après les points 2 et 4 ci-dessus, dans le cas où I est compact, que si $p < p'$ et $\omega(f, \cdot)_p$ et $\omega(f, \cdot)_{p'}$ ont pour indices de régularité respectifs a et a' , alors nécessairement $a' < a$. Les espaces de Besov ont connu une certaine popularité dans la littérature des statistiques non-paramétriques en parallèle avec l'apparition des méthodes d'ondelettes promues par D. Donoho ([16, 17]). Ce n'est évidemment pas un hasard : les normes de Besov (2.51) sont équivalentes à des normes définies sur les coefficients d'ondelette de la fonction f pour une ondelette bien choisie (voir paragraphe 2.1.2). Cette propriété agréable des espaces de Besov n'est pas vérifiée pour des espaces de régularité très utilisés comme les espaces BV des fonctions à variations bornées mais certains encadrements permettent de conclure sur l'appartenance ou non à l'espace BV à partir des coefficients d'ondelettes. Par exemple si I est le pavé $[0, 1]^2$ de \mathbb{R}^2 , on a

$$B_1^{1,1} \subset BV \subset B_1^{1,\infty} \quad (2.52)$$

avec les inégalités correspondantes pour les normes associés à ces espaces.

L'introduction rapide que nous venons de faire de la régularité est uniquement basée sur des notions d'espaces fonctionnels. Il existe des notions alternatives, notamment l'approche plus géométrique de la dimension du graphe ou des ensembles images de f et une approche plus probabiliste qui est celle de l'étude du temps local. Nous ne détaillerons pas ici ces thèmes ; mentionnons simplement que régularité fonctionnelle, dimension du graphe, régularité du temps local et lois d'échelle (ou auto-similarité) sont bien sûr fortement liés. Pour le mouvement brownien fractionnaire (MBF) $\{\mathbf{X}(t), t \in \mathbb{R}\}$ de paramètre de Hurst $H \in (0, 1)$, tous les modules de régularités (définis avec $I = [0, 1]$) ont le même indice de régularité H , la dimension d de son graphe est aussi donné par la relation attendue $d = 2 - H$. Cette unification des points de vue ne se généralise pas à tous les processus à accroissements stationnaires H -auto-similaires comme ce fut remarqué par [34]. Le paragraphe 2.4.2 est consacré à l'étude du module de continuité et du temps local de certains champs stables. Cette étude a eu pour point de départ un travail ([3]) en collaboration avec A. Ayache sur des généralisations de mes travaux de thèses portant sur les relations entre régularité et dimension du graphe d'une fonction ou d'un processus.

2.4.2 Drap linéaire fractionnaire stable

Une bonne introduction aux processus stables se trouve dans [54]. Le drap linéaire fractionnaire stable (noté $(N, 1)$ -DLFS) de paramètres $\alpha \in (0, 2]$ et $H = (H_1, \dots, H_N) \in (0, 1)^N$ est un champ α -stable $\{X(t), t \in \mathbb{R}^N\}$ à valeurs réels défini par

$$X(t) = \int_{\mathbb{R}^N} \prod_{l=1}^N \left\{ (t_l - s_l)_+^{H_l - 1/\alpha} - (-s_l)_+^{H_l - 1/\alpha} \right\} dZ_\alpha(s), \quad (2.53)$$

où $\{Z_\alpha(s), s \in \mathbb{R}^N\}$ est un drap de Lévy α -stable dont l'intensité d'asymétrie (pour $\alpha < 2$) $\beta : \mathbb{R}^N \rightarrow [-1, 1]$ est donnée.

Deux cas particuliers sont plus connus :

1. pour $\alpha = 2$, on obtient le drap brownien fractionnaire (DBF), étudié par exemple dans [6, 63];
2. pour $N = 1$, on obtient le mouvement linéaire fractionnaire stable (MLFS) défini sur $t \in \mathbb{R}$, étudié par exemple dans [58, 34]. Ce dernier, si β est constant, est un processus H -auto-similaire à incréments stationnaires.

Dans la suite nous considérons uniquement le cas non-gaussien, $\alpha < 2$. Dans ce cas, le champ obtenu admet une version continue si et seulement si $\min(H_1, \dots, H_N) > \frac{1}{\alpha}$, ce que nous imposerons par la suite en supposant (sans perte de généralité) que

$$\alpha^{-1} < H_1 \leq \dots \leq H_N < 1. \quad (2.54)$$

Dans le cas multi-dimensionnel, $N > 1$, le $(N, 1)$ -DLFS est un champ anisotropique. En effet, la propriété d'auto-similarité du cas $N = 1$ est généralisée par la propriété suivante (toujours pour une intensité β constante) : pour toute matrice $N \times N$ diagonale A

$$\{X(At), t \in \mathbb{R}_+^N\} \stackrel{d}{=} \left\{ \prod_{j=1}^N a_j^{H_j} X(t), t \in \mathbb{R}_+^N \right\}, \quad (2.55)$$

où a_j est le j -ième élément diagonal de A et $\stackrel{d}{=}$ signifie égalité au sens des distributions fini-dimensionnelles. Ces propriétés anisotropiques peuvent être utiles en pratique pour la modélisation, comme remarqué dans [9] pour le cas gaussien.

Remarquons enfin que si on fixe toutes les composantes de $t \in \mathbb{R}^N$ sauf la n -ième,

$$t = (u_1, \dots, u_{n-1}, u, u_{n+1}, \dots, u_N) \stackrel{\text{def}}{=} (u, \hat{u}_n),$$

avec $\hat{u}_n = (u_1, \dots, u_n, u_{n+1}, \dots, u_N)$ fixé dans \mathbb{R}^{N-1} et u parcourant \mathbb{R} , le processus obtenu est un MLFS de paramètres α et H_n . La restriction du champ aux droites parallèles à l'axe de la n -ième coordonnée $\{X(u, \hat{u}_n), u \in \mathbb{R}\}$ héritera donc des propriétés établies par exemple par [58] : l'étude de son module de continuité indique une régularité höldérienne $H_n - 1/\alpha$ aux échelles fines. La question est donc de savoir ce que deviennent ces propriétés quand le champ est vu globalement. Les théorèmes 1.2 et 1.4 de [5] en annexe montrent par exemple que, pour tout $n \in \{1, \dots, N\}$, la régularité höldérienne de toutes les restrictions du processus aux droites parallèles à l'axe de la n -ième coordonnée est bien $H_n - 1/\alpha$. Un point important dû à l'anisotropie du modèle est que le comportement du module de continuité aux échelles fines n'est pas bien décrit par une loi de puissance de u^α , bien que nos résultats impliquent que l'indice de régularité höldérienne sur tout pavé borné I d'intérieur non-vidé est presque sûrement

$$\liminf_{u \downarrow 0} \frac{\log \omega(X, u)_\infty}{\log(u)} = \min(H_1, \dots, H_N) - \alpha^{-1} = H_1 - \alpha^{-1}.$$

En fait [5, théorèmes 1.2 et 1.4] impliquent le résultat plus précis : presque sûrement, pour tout pavé borné $I \subset \mathbb{R}^N$ non-réduit à un point,

$$\liminf_{u \downarrow 0} \frac{\log \tilde{\omega}^{(H+H_1)}(X, u)_\infty}{\log(u)} = H_1 - \alpha^{-1}, \quad (2.56)$$

où $H + H_1$ est le vecteur d'indices croissant $(0, H_2 - H_1, \dots, H_N - H_1)$ et où on a défini un module de continuité anisotropique sur I ,

$$\tilde{\omega}^{(H-H_1)}(X, u)_\infty = \sup_{t, s \in I, \|t-s\| \leq u} \frac{|X(t) - X(s)|}{\sum_{k=1}^N |t_k - s_k|^{H_k - H_1}}.$$

Ces résultats diffèrent du cas gaussien étudié dans [6] pour lequel le comportement du module de continuité anisotropique est donné par (2.56) avec le membre de droite remplacé par H_1 .

L'approche utilisée dans ces références (cas stable ou gaussien) repose sur l'utilisation d'une décomposition en base d'ondelette compliquée dans le cas stable (voir [5, théorème 1.1]) par le fait que la structure de dépendance des coefficients d'ondelette n'est plus simplement donnée par une covariance. Notons que cette décomposition fournit aussi des estimations du comportement du champ à l'infini (voir [5, théorèmes 1.3]).

L'étude menée dans [5] est complétée par un travail sur le temps local du (N, d) -DLFS, dont les d composantes sont des copies indépendantes du même $(N, 1)$ -DLFS. La bi-continuité du temps local du (N, d) -DLFS est montrée sous la condition de continuité des trajectoires (2.54) et la condition additionnelle

$$\sum_{k=1}^N H_k^{-1} > d,$$

condition, qui est montrée être nécessaire et suffisante pour l'existence du temps local, voir les théorèmes 2.2 et 2.3 de [4] en annexe.

2.4.3 Modèle feuilles mortes à loi d'échelle

Le modèle feuilles mortes, dont nous avons rappelé la définition au paragraphe 2.3.5, a été proposé pour rendre compte des statistiques empiriquement observées sur les images naturelles [35]. La transformation définie par (2.46) et (2.47) utilisée lors de la formation de ce modèle rend en effet à la fois compte du phénomène d'occlusion et de non-gaussianité, ou plus généralement d'une absence de stabilité additive des statistiques. Néanmoins ce modèle est en contradiction avec d'autres propriétés structurelles proposées par [50] pour la modélisation des images : en particulier l'invariance d'échelle. Cette hypothèse pousse à son paroxysme une observation communément admise pour les images naturelles : les structures géométriques présentes dans une images ne semblent pas liées à une échelle particulière. Pour preuve, on peut citer le réalisme bien connu des images synthétiques créées par des modèles très simples issus de la géométrie fractale. Comme montré dans [50], l'invariance d'échelle est une hypothèse très forte du point de vue de la modélisation ; aussi proposons-nous de s'attaquer à cette question sous l'angle "moins géométrique" de la régularité. De ce point de vue, pourtant moins restrictif, le modèle feuilles mortes usuel fait aussi défaut car, en tout point, les réalisations de sa version colorée (2.48) est p.s. localement constante et donc à variations bornées, pour peu que les ensembles X_k aient eux-même une frontière régulière. On peut de plus montrer que la norme BV de \mathbf{Y} sur un compact de \mathbb{R}^2 est même d'espérance finie, voir la proposition 6.1 dans [26] disponible en annexe dès que les ensembles X_k ont une frontière de longueur d'espérance finie, ce qui est contradiction avec les expériences menées dans [24] sur une base d'images naturelles.

Les modèles à structure additive, le plus souvent obtenus par modélisation des coefficients d'une image dans une base bien choisie, voir par exemple [55], ont l'avantage de pouvoir s'adapter facilement à des prescriptions de régularité. Par exemple la régularité Besov peut être prescrite par un modèle de série aléatoire d'ondelettes. En revanche ce type de modèle ne sont pas adaptés à la représentation de structures géométriques simples et des bases fortement redondantes accompagnées d'algorithmes complexes de codage sont par exemple nécessaires pour représenter de telles structures (voir par exemple [51]).

L'objectif principal de notre travail [26] est de définir un modèle feuille morte à loi d'échelle, obtenu comme la limite d'une suite de modèle feuilles mortes coloriés dont les propriétés de régularité sont non-triviales (en tout cas qui sorte du cadre BV) mais qui conservent cependant une structure géométrique héritées du modèle feuille morte usuel. On définit la famille $\{P_{r_0, r_1}, 0 < r_0 < r_1 < \infty\}$ des distributions des fermés aléatoires $X = RY$ où $Y \subset \mathbb{R}^2$ est un fermé aléatoire de distribution indépendante de r_0 et r_1 et R est facteur de dilatation aléatoire, indépendant de Y , de densité proportionnelle à $r^{-\alpha}$ pour $r_0 < r < r_1$. Le paramètre α est un indice de loi de puissance qui reproduit le comportement observée dans la distribution empirique de la taille des composantes connexes d'une image, [35, 2]. Les valeurs typiques de α sont dans l'intervalle $(2, 3)$. Quand le fermé aléatoire Y respecte quelques conditions de bases, (il a

une échelle bien déterminée et une frontière régulière) le modèle feuilles mortes défini par (2.46) et (2.47) associé à la distribution P_{r_0, r_1} , noté $M(r_0, r_1)$, est une mosaïque aléatoire et pour tout point du plan la probabilité d'appartenir à $\partial M(r_0, r_1)$ est nulle (voir proposition 4.1 dans [26]). Quelque soit $\alpha > 1$, la frontière $\partial M(r_0, r_1)$ converge en probabilité vers l'ensemble \mathbb{R}^2 , voir [26, proposition 4.2]. On peut donc espérer que la limite de ce modèle, si elle est correctement définie, a une norme BV infinie. Nous nous intéressons aux lois fini-dimensionnelles de la version colorisée (2.48) de $M(r_0, r_1)$, qui peuvent être obtenue à partir de la proposition 5 de [10] déjà mentionnée. La loi marginale de ce champ est donnée par la distribution des "couleurs" C_k apparaissant dans (2.48), qu'on prend uniforme par la suite. Les lois bi-dimensionnelles sont étudiées dans un premier temps pour éliminer les cas où celles-ci dégénèrent quand $r_0 \rightarrow 0$ ou $r_1 \rightarrow \infty$: ces cas sont donnés par [26, proposition 5.1], ils dépendent de la position de α par rapport à 3. Par exemple, si $\alpha > 3$ et $r_0 \rightarrow 0$, le champ tend vers un bruit blanc ; si $\alpha < 3$ et $r_1 \rightarrow \infty$, le champ tend vers un champ constant ; voir la figure 2.3(a) et (b). Le cas le plus intéressant nous semble celui où $\alpha < 3$ et $r_0 \rightarrow 0$ qui correspond aux valeurs empiriquement observées pour α et dont la limite a des propriétés de régularité non-triviales. Dans ce cas la convergence du champ est donnée par [26, théorème 5.4]. Une caractéristique des champs de la forme (2.48) est que leurs lois fini-dimensionnelles sont des mélanges finis dont seuls les poids dépendent de r_0 et r_1 . C'est donc la convergence de ces poids qu'il s'agit de vérifier ; de plus le champ limite obtenu, appelé *modèle feuilles mortes à loi d'échelle*, vérifie aussi cette propriété fondamentale : sa loi marginale est donnée par la distribution des couleurs C_k et toutes les lois fini-dimensionnelles sont des mélanges finis. C'est pourquoi la limite conserve des structures géométrique, d'où par exemple la présence visible de disques pour la figure 2.3(c). Cette structure ne serait pas conservée pour un champ gaussien de même auto-covariance que le modèle feuilles mortes à loi d'échelle. D'autre part les poids du mélange obtenus admettent des lois de puissance en fonction de l'échelle ; c'est ce qui est montré par la [26, proposition 5.2] pour la loi bi-dimensionnelle. Cette loi d'échelle se retrouve dans l'impression d'auto-similarité observée pour le champ limite de la figure 2.3(c)–(f).

La régularité Besov du modèle feuilles mortes à loi d'échelle est étudiée par [26, proposition 6.2] et de ce résultat et de l'inclusion (2.52), on trouve que la limite a une norme BV locale d'espérance infinie dès que $\alpha > 2$. Une discussion détaillée sur la comparaison de ces résultats avec d'autres approches pour modéliser la régularité des images peut être trouvée au paragraphe 7 de [26].

Mentionnons pour conclure que nous avons proposé dans [25] une méthode bayésienne de débruitage basée sur un *a priori* inspiré du modèle feuilles mortes à loi d'échelle.

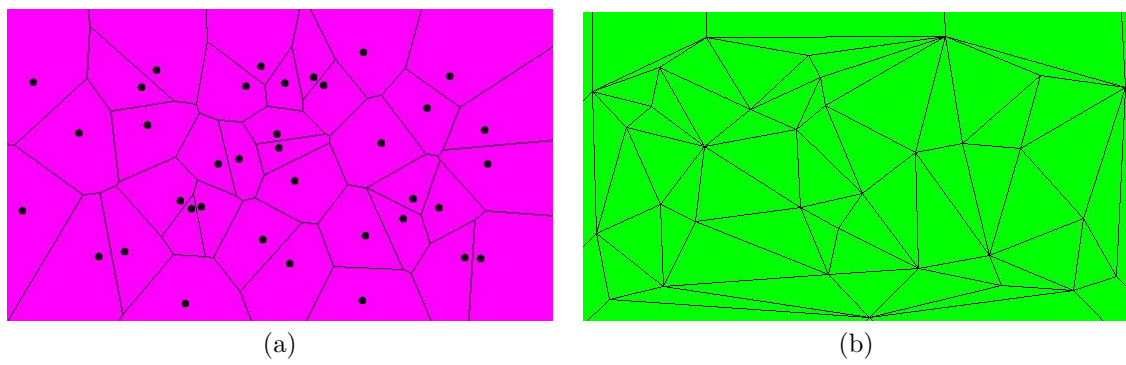


FIG. 2.1 – (a) : mosaïque de Voronoï; (b) : mosaïque de Delaunay.

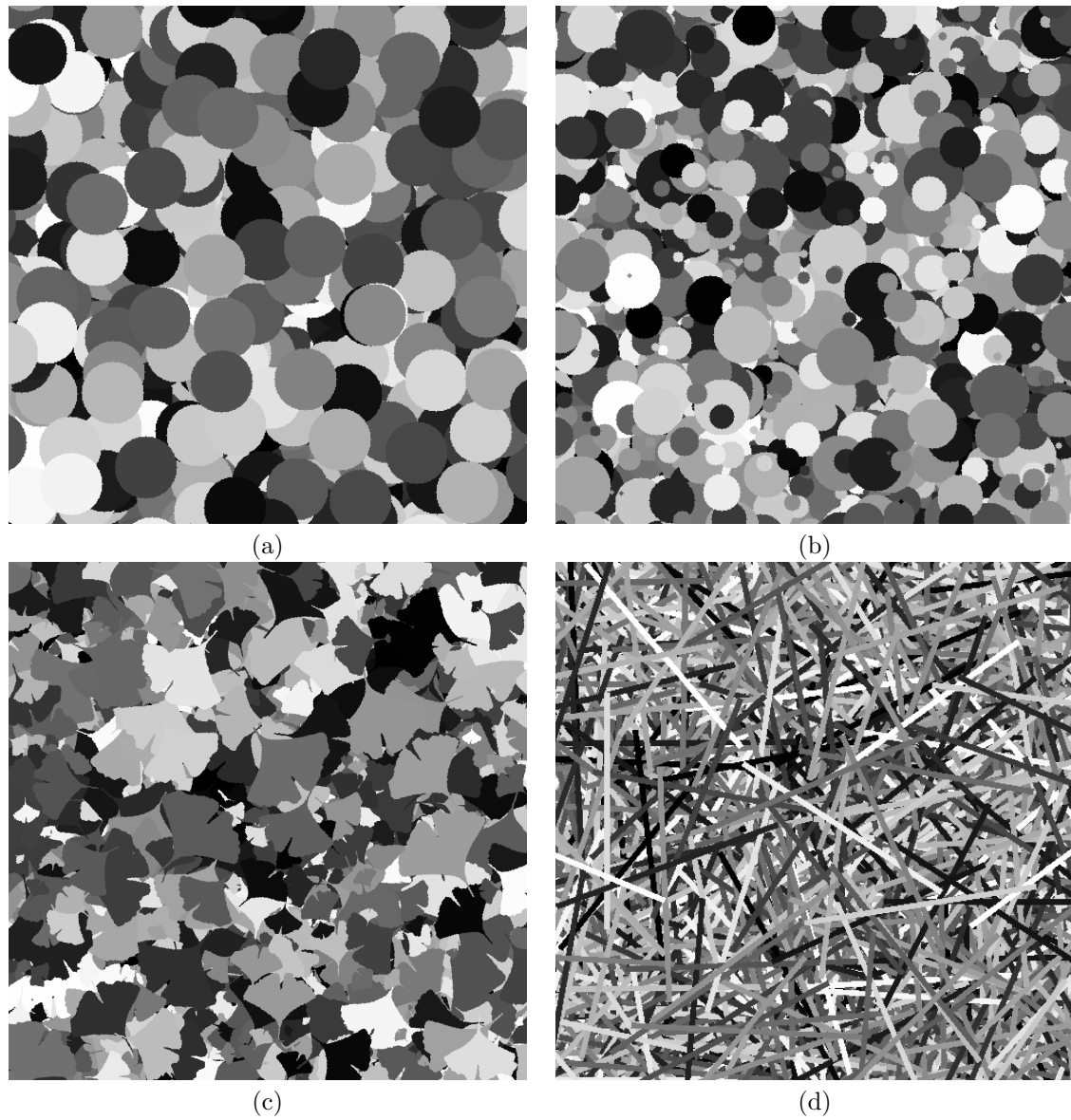


FIG. 2.2 – Simulation de modèles feuilles mortes avec P donné par : (a) support réduit à un disque fixé ; (b) disque à rayon de loi uniforme ; (c) rotation et homotétie uniformes d’une même “feuille” ; (d) rotation uniforme d’un rectangle.

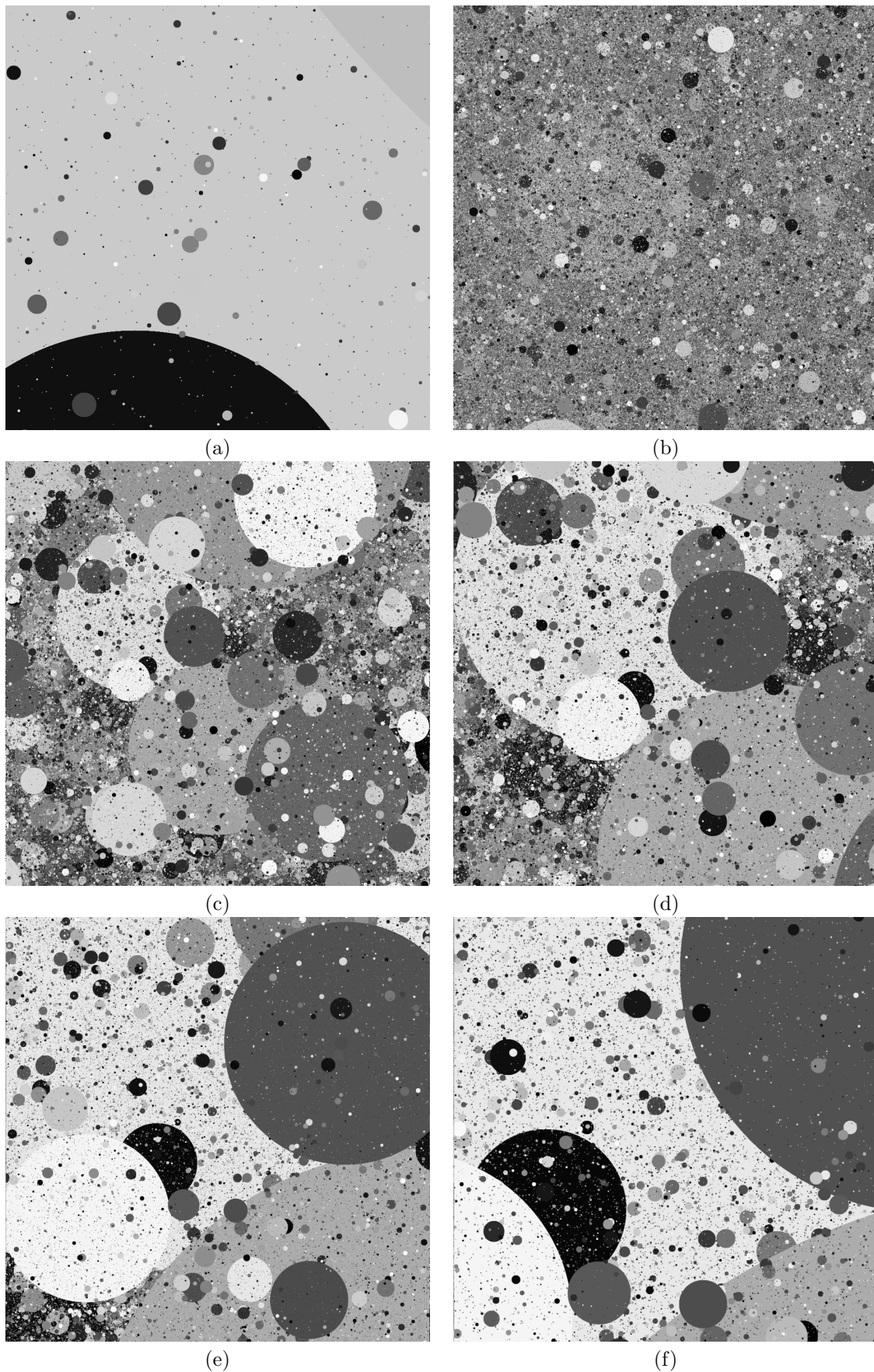


FIG. 2.3 – Convergences du modèle feuilles mortes quand Y est un disque fixé : (a) convergence vers un champ constant ($\alpha < 3$ et $r_1 \rightarrow \infty$); (b) convergence vers un bruit blanc ($\alpha > 3$ et $r_0 \rightarrow 0$); (c) : convergence vers un champ non-trivial $\alpha = 2.9$ et $r_0 \rightarrow 0$, puis zooms successifs d'un facteur 2 de ce champ pour (d), (e), (f)

Chapitre 3

Estimation semi ou non-paramétrique

Sommaire du chapitre

3.1 Modèles d'observations directs	31
3.1.1 Identification d'un processus TVAR	31
3.1.2 Estimation d'une fréquence en échantillonnage irrégulier	32
3.1.3 Estimation du paramètre de mémoire longue d'un processus linéaire ou gaussien	33
3.2 Problèmes inverses	35
3.2.1 Estimation de la densité d'un mélange de distributions discrètes	35
3.2.2 Estimation de la densité de la marque d'un processus shot-noise	36
3.2.3 Estimation de l'indice de queue de la marque d'un processus shot-noise	37

3.1 Modèles d'observations directs

3.1.1 Identification d'un processus TVAR

Considérons le processus TVAR $\{X_{k,n}, k \in \mathbb{N}, n \geq 1\}$ d'ordre p défini par (2.21) introduit au paragraphe 2.2.2. Nous avons déjà évoqué les vitesses d'estimations minimax de $\theta(t)$ pour $t \in [0, 1]$ pour $(\theta, \sigma) \in \mathcal{C}_\beta$ où \mathcal{C}_β est une classe fonctionnelle de régularité usuelle d'indice β , intersectée avec des conditions de stabilité du processus (voir [45] en annexe).

Nous avons étudié un estimateur récursif de $\theta(t)$ défini par une succession d'opérations simples qui, à chaque étape, prend en compte une nouvelle observation $X_{k,n}$, $k = 1, \dots, n$:

$$\hat{\theta}_{k+1,n} = \hat{\theta}_{k,n} + \mu \frac{(X_{k+1,n} - \hat{\theta}_{k,n}^T \mathbf{X}_{k,n}) \mathbf{X}_{k,n}}{1 + \mu |\mathbf{X}_{k,n}|^2}, \quad (3.1)$$

où

1. $\mu > 0$ est un paramètre de lissage réglé par l'utilisateur qui contrôle l'amplitude de la mise à jour de l'estimation ;
2. $|\cdot|$ dénote la norme euclidienne.

Cet estimateur est appelé estimateur récursif des moindres carrés normalisés. Son principe est de se déplacer dans la direction du gradient du critère des moindres carrés $(X_{k+1,n} - \theta^T \mathbf{X}_{k,n})^2$ en la valeur

courante du paramètre donné par $(X_{k+1,n} - \hat{\theta}_{k,n}^T \mathbf{X}_{k,n}) \mathbf{X}_{k,n}$ avec une amplitude donné par le pas μ . Ce type d'estimateurs a été étudié (voir par exemple [28]) dans un cadre de poursuite de paramètres θ_k évoluant aléatoirement. Le cadre des processus localement stationnaires permet d'affiner l'analyse asymptotique puisqu'il rend possible la consistance de $\hat{\theta}_{k,n}$ vers un paramètre déterministe $\theta(t)$ quand $n \rightarrow \infty$ avec $k/n \rightarrow t$. Pour obtenir la consistance, il faut prendre $\mu \rightarrow 0$ quand $n \rightarrow \infty$ et le choix de $\mu(n)$ détermine la vitesse de convergence. Ainsi, nous montrons dans [45, corollaire 3] que pour $\beta \in (0, 1]$ la vitesse minimax est atteinte dans la classe \mathcal{C}_β pour $\mu \asymp n^{-2\beta/(1+2\beta)}$.

Des équivalents asymptotiques du biais et de la variance sont fournis par [45, théorème 6], toujours pour $\beta \leq 1$. Ce résultat est adapté pour $\beta \in (1, 2]$, [45, théorème 7], montrant qu'un biais systématique de la forme $C(t) (\mu n)^{-1}$ entache l'estimation de $\theta(t)$, empêchant à l'estimateur d'atteindre la vitesse minimax. La méthode de Romberg est alors utilisée pour éliminer ce biais systématique; il s'agit tout simplement de définir, pour $\gamma \in (0, 1)$ fixé,

$$\tilde{\theta}_{k,n}(\mu) = (1 - \gamma)^{-1} \left(\hat{\theta}_{k,n}(\mu) - \gamma \hat{\theta}_{k,n}(\gamma\mu) \right), \quad (3.2)$$

qui permet d'éliminer le biais systématique. Enfin, il est montré que cet estimateur est bien minimax dans \mathcal{C}_β pour $\beta \in (0, 2]$ et $\mu \asymp n^{-2\beta/(1+2\beta)}$ (voir [45, corollaire 9]).

Ce travail et ces résultats illustrent un aspect très positif de l'approche introduite R. Dahlhaus pour modéliser la stationnarité locale. L'analyse classique de la poursuite de paramètres évoluant aléatoirement ne permet pas d'analyser aussi finement les estimateurs récursifs puisque μ y est constant. En particulier des équivalents asymptotique du biais et de la variance ne sont pas disponibles, ni l'adaptation simple (3.2) proposée pour améliorer la vitesse sous des conditions de régularité forte. Mentionnons au sujet de l'équivalent asymptotique de la variance qu'il est obtenu grâce à la normalisation particulière donnée par le dénominateur de l'équation (3.1) (voir [45, remarque 5]). La normalisation usuelle de l'estimateur des moindres carrés récursifs est en effet donné par $1 + |\mathbf{X}_{k,n}|^2$. Cette normalisation ne modifie pas la vitesse mais la variance asymptotique est bien plus complexe. En la remplaçant par $1 + \mu |\mathbf{X}_{k,n}|^2$, comme $\mu \rightarrow 0$, la variance asymptotique est la même que celle de l'estimateur récursif des moindres carrés sans normalisation, mais à la différence de celui-ci, s'applique sous des hypothèses plus générales.

Les résultats que nous avons obtenus découlent de deux étapes :

1. un contrôle uniforme des moments d'un produits de matrices dépendantes, voir [45, théorème 16], sous des conditions de Lyapounov et de persistance d'excitation;
2. une décomposition sous forme d'approximation polynomiale de l'erreur, voir [45, paragraphe 6] et des inégalités de Burkholder correspondant à chaque ordre de l'approximation ([45, lemme B.1, propositions B.2 et B.3]).

Le premier point est assez largement inspiré de résultats utilisés en poursuite mais, contrairement à ces derniers, il permet un contrôle uniforme, indispensable en estimation non-paramétrique pour l'obtention de vitesses minimax.

3.1.2 Estimation d'une fréquence en échantillonnage irrégulier

Nous avons étudié dans [36], disponible en annexe, pour le modèle (2.27) décrit au paragraphe 2.2.4 un estimateur de la fréquence f_* basé sur le périodogramme cumulé adapté au cas des échantillons irréguliers de la même façon que le périodogramme de Lomb-Scargle (voir (2.29) et défini par

$$\Lambda_n(f) = \frac{1}{n^2} \sum_{k=1}^{K_n} \left| \sum_{j=1}^n Y_j e^{-2i\pi k f X_j} \right|^2 \propto \sum_{k=1}^{K_n} I_n^{LS}(kf),$$

où K_n est le nombre d'harmoniques considéré à régler par l'utilisateur (en théorie $K_n \rightarrow \infty$ quand $n \rightarrow \infty$, par exemple à la vitesse $\log n$). Le cadre est semi-paramétrique car, en dehors d'une hypothèse

d'identifiabilité de la fréquence maximale, les seules hypothèses sur s_* sont des hypothèses de régularités décrites par des conditions de décroissances de ses coefficients de Fourier. Par souci de simplification et bien que ce ne soit pas indispensable, nous avons supposé que f_* est dans un intervalle fermé $[f_{\min}, f_{\max}] \subset (0, \infty)$ connu.

Le fait de cumuler les multiples d'une même fréquence dans la définition de Λ_n est une pratique classique pour l'estimation de la fréquence d'une fonction non-sinusoidale (voir [23]) qui permet d'obtenir un estimateur robuste même quand la première harmonique n'est pas dominante, ce qui est en général le cas. L'utilisation du périodogramme cumulé requiert néanmoins quelques précautions. En effet par construction, pour K_n grand, $\Lambda_n(f/q) > \Lambda_n(f)$ pour q entier. Néanmoins si f correspond bien à l'harmonique la plus basse $\Lambda_n(f/q) - \Lambda_n(f)$ reste contrôlable. Aussi l'estimateur est défini en deux temps :

1. on définit tout d'abord \tilde{f}_n par

$$\Lambda_n(\tilde{f}_n) = \tilde{\Lambda}_n \quad \text{où} \quad \tilde{\Lambda}_n = \sup_{f \in [f_{\min}, f_{\max}]} \Lambda_n(f),$$

2. puis on définit, pour $0 < r < f_{\min}/2$ et une suite (κ_n) telle que $\kappa_n \rightarrow 0$ quand $n \rightarrow \infty$,

$$\begin{aligned} \tilde{B}_n(j) &\stackrel{\text{def}}{=} \{f \in [f_{\min}, f_{\max}] : |f - j\tilde{f}_n| \leq r\}, \\ \hat{j}_n &= \max \left\{ j : \sup_{f \in \tilde{B}_n(j)} \Lambda_n(f) \geq (1 - \kappa_n)\tilde{\Lambda}_n \right\}, \\ \hat{f}_n &= \arg \sup_{f \in \tilde{B}_n(\hat{j}_n)} \Lambda_n(f). \end{aligned} \tag{3.3}$$

Nous montrons, pour des choix appropriés pour K_n et κ_n que l'estimateur \hat{f}_n est consistant, [36, théorème 1], sous les hypothèses très faibles pour s_* de sommabilité de ses coefficients de Fourier et sans aucune hypothèse supplémentaire pour les temps d'observation. La convergence $n(\hat{f}_n - f_*) \rightarrow 0$ p.s. est de plus obtenue si $\mathbb{E}[V_1^2] < \infty$.

Un théorème de la limite centrale est obtenu, [36, théorème 2], sous des hypothèses plus fortes (et plus classiques, cf. [23]) sur les coefficients de Fourier de s_* . La vitesse de convergence est $n^{-3/2}$. La variance de la limite est explicitée en fonction de s_* et Φ définie en (2.28). Cette variance n'égale pas la variance optimale, bien connue dans le cas du temps continu et retrouvée par [29] dans le cas d'échantillons irréguliers, mais elle s'en approche quand la variance du bruit de mesure est important. Néanmoins, à notre connaissance, nos résultats sont les seuls disponibles concernant l'établissement d'un estimateur de la période en échantillonnage irrégulier atteignant la vitesse optimale. En effet, dans [29], un estimateur de vitesse quasi-optimale est requis.

3.1.3 Estimation du paramètre de mémoire longue d'un processus linéaire ou gaussien

Le cadre semi-paramétrique classique pour l'estimation du paramètre de mémoire longue d'un processus X consiste à supposer que la mesure spectrale ν^* définie par (2.4) a une densité spectrale f^* pour laquelle, pour λ dans un voisinage de la fréquence nulle, on a

$$|f^*(\lambda) - f^*(0)| \leq \gamma f^*(0) |\lambda|^\beta, \tag{3.4}$$

où $\gamma > 0$, β est un indice de régularité höldérienne ponctuelle (au point $\lambda = 0$), $\beta \in (0, 2]$ et $f^*(0) \neq 0$. Cette condition implique donc que X admet d pour paramètre de longue mémoire au sens fort (voir paragraphe 2.1.1. La présence de $f^*(0)$ dans le membre de droite de (3.4) est une astuce simplificatrice permettant de supposer $f^*(0) = 1$ sans perte de généralité dans les preuves sans avoir à modifier γ . Ce cadre semi-paramétrique est héritée de la littérature étudiant les estimateurs de type Fourier du paramètre d , voir par exemple [22]. On sait notamment, par cette dernière référence, que, sous l'hypothèse (3.4), la

vitesse minimax de convergence pour l'estimation du paramètre d à partir de n observations consécutives de X est $n^{-\beta/(1+2\beta)}$.

Le comportement asymptotique de la transformée en ondelette discrète de X est décrit au paragraphe 2.1.3. Nous avons plus précisément étudié 2 estimateurs de type ondelette, tous deux basés sur le comportement asymptotique des formes quadratiques

$$\hat{\sigma}_j^2 = n_j^{-1} \sum_{k=0}^{n_j-1} W_{j,k}^2, \quad (3.5)$$

où n_j est le nombre de coefficients d'ondelette $W_{j,k}$ disponibles à l'échelle j à partir des n observations X_1, \dots, X_n . A j fixé, quand $n \rightarrow \infty$, on a bien sûr $\hat{\sigma}_j^2 \rightarrow \text{Var}(W_{j,0})$ et nous avons vu au paragraphe 2.1.3 que, lorsque $j \rightarrow \infty$,

$$\text{Var}(W_{j,0}) = \mathbb{E}[\hat{\sigma}_j^2] \sim f^*(0) \text{Var}(W_{0,0}^{(d)}) 2^{2jd}.$$

En fait ce comportement peut être précisé sous la condition (3.4) (voir [48, théorème 1]); on obtient

$$\text{Var}(W_{j,0}) = f^*(0) \text{Var}(W_{0,0}^{(d)}) 2^{2jd} (1 + O(2^{-\beta j})) \quad (3.6)$$

Cette formule permet de contrôler le biais des estimateurs de d de type ondelettes. Pour le reste, l'analyse de ces estimateurs se base sur des résultats de

- (i) contrôle L^2 de la variance empirique $\hat{\sigma}_j^2$ définie par (3.5) pour les processus à mémoire longue linéaire ou gaussien ([48, théorème 1]);
- (ii) contrôle L^2 du logarithme d'une forme quadratique de vecteurs gaussiens ([49, proposition 4]);
- (iii) théorème de la limite centrale pour de telles formes quadratiques [48, lemme 12].

Les deux estimateurs que nous avons étudiés sont adaptés d'estimateurs proposés au cours des années 90 dans des cadres paramétriques par [62] et [1] et sont respectivement définis par

1. la minimisation d'un critère de type Whittle dans le domaine des ondelettes,

$$\hat{d}_n^w \stackrel{\text{def}}{=} \arg \inf_{t \in \mathbb{R}} \left\{ \log \left[\sum_{j=L_n}^{U_n} n_j \hat{\sigma}_j^2 2^{-2tj} \right] - \frac{\sum_{j=L_n}^{U_n} n_j \log 2^{-2tj}}{\sum_{j=L_n}^{U_n} n_j} \right\}. \quad (3.7)$$

En référence aux approches de type Fourier, nous appelons cet estimateur l'*estimateur de Whittle en ondelette*.

2. la log-régression de $\hat{\sigma}_j^2$, $j = L_n, L_n + 1, \dots, U_n$,

$$\hat{d}_n^r \stackrel{\text{def}}{=} \sum_{j=L_n}^{U_n} w_j \log \hat{\sigma}_j^2 \text{ avec } \sum_{j=L_n}^{U_n} w_j = 0 \text{ et } 2 \log(2) \sum_{j=L_n}^{U_n} j w_j = 1.$$

Les échelles L_n et U_n sont respectivement les indices des échelles la plus fine et la plus grossière utilisées par les estimateurs. Le plus souvent, on considère $U_n - L_n$ fixé indépendamment de n , mais rien n'empêche de considérer le cas où $U_n - L_n \rightarrow \infty$. Le choix de L_n est un choix de fréquence de coupure similaire au paramètre m des estimateurs semi-paramétriques de type Fourier qui détermine la plus haute fréquence des coefficients de Fourier utilisés. Son choix détermine la vitesse de convergence car il influence à la fois le biais (contrôlé par l'hypothèse de régularité (3.4) et la variance (à travers le nombre de coefficients d'ondelettes qui est de l'ordre $n 2^{-L_n}$ à une constante multiplicative près quelque soit le choix de U_n). On a par exemple, dans le cadre des processus linéaires (voir [49, corollaire 4]), pourvu que $\liminf(U_n - L_n) \geq 2$,

$$2^{L_n} \asymp n^{(1+2\beta)^{-1}} \Rightarrow \hat{d}_n^w = d + O_p\left(n^{-\beta/(1+2\beta)}\right). \quad (3.8)$$

Le détail des résultats dans les références en annexe se présente comme suit. Dans [49, théorème 2] des résultats de contrôle de l'erreur quadratique moyenne de l'estimateur \hat{d}_n^r sous l'hypothèse gaussienne

quand $U_n - L_n$ est fixe et $L_n \rightarrow \infty$; Le théorème de la limite centrale de cet estimateur dans des hypothèses similaires est donné par [47, théorème 1]. L'estimateur \hat{d}_n^w est étudié dans [48] : le théorème 3 fournit les vitesses de convergences valides dans un cadre général incluant les modèles linéaires et le théorème 5 un théorème de la limite centrale dans le cas gaussien.

De ces travaux il ressort que les estimateurs de d par méthode d'ondelette atteignent la vitesse minimax de convergence $n^{-\beta/(1+2\beta)}$ et s'appliquent facilement au cas où $d \notin (-1/2, 1/2)$, sans requérir d'adaptations supplémentaires. Les constantes sont étudiées et plus généralement la comparaison avec les estimateurs de type Fourier sont discutées dans [48, paragraphe 4.4].

3.2 Problèmes inverses

3.2.1 Estimation de la densité d'un mélange de distributions discrètes

Nous avons étudié un estimateur de projection de la densité de la mesure mélangeante μ à partir d'un n -échantillon i.i.d. de loi π_μ donné par (2.30), avec $(\pi_\theta)_{\theta \in \Theta}$ famille de densités connues par rapport à une mesure dominante ζ . On note ν la mesure dominante pour μ et f la densité à estimer, $d\mu = f d\nu$, et par suite π_f la loi correspondante pour les observations. Nous avons déjà mentionné des bornes minimax de ce problème d'estimation non-paramétrique au paragraphe 2.3.1. Dans la suite nous considérerons uniquement le cas où ζ est la mesure de comptage sur \mathbb{Z}_+ . Notre analyse porte exclusivement sur le risque quadratique intégré définie en supposant que la densité f du mélange appartient à l'espace de Hilbert $\mathbb{H} = L^2(\nu)$ par

$$\mathbb{E}[\|\hat{f} - f\|_{\mathbb{H}}^2] = \mathbb{E} \int (f - \hat{f})^2 d\nu .$$

L'hypothèse $f \in L^2(\nu)$ ne semble pas naturelle à première vue mais elle est nécessaire pour l'utilisation du risque quadratique intégré, qui est classique en estimation non-paramétrique de densité. De plus cette hypothèse est de toute façon impliquée par des hypothèses de régularités minimales pour f , généralement considérées pour obtenir la consistance dans le cadre non-paramétrique. Nous allons voir que la structure hilbertienne joue un rôle particulier pour l'estimateur de projection que nous avons étudié.

L'idée de l'estimateur repose sur l'utilisation d'une base particulière de \mathbb{H} , construite à partir de la suite d'espaces emboîtés de dimension finie V_m , $m \geq 1$, définis par (2.32), déjà utilisée pour l'obtention de la borne inférieure de la vitesse minimax. Les résultats assez généraux obtenus proviennent de l'observation suivante. Si $h : \mathcal{Y} \rightarrow \mathbb{R}$ est telle que $\Pi h = g \in \mathbb{H}$ avec l'opérateur Π défini par (2.31), alors, pour tout $f \in \mathbb{H}$,

$$\pi_f h = (\Pi h, f)_{\mathbb{H}} = (g, f)_{\mathbb{H}} .$$

Or on remarque que

1. d'une part $\pi_f h$ s'estime naturellement par $P_n h = n^{-1} \sum_{k=1}^n h(Y_k)$;
2. d'autre part, la connaissance de $(g, f)_{\mathbb{H}}$ pour g parcourant un espace vectoriel V donné suffit au calcul du projeté orthogonal de f sur V .

D'où la définition de l'estimateur de projection, voir [53, définition 1]. L'expression la plus simple de cet estimateur est donnée par

$$\hat{f}_{m,n} = \sum_{k=0}^{d_m-1} [P_n \Pi^{-1} \phi_k] \phi_k ,$$

où $\{\phi_k, k = 0, \dots, d_m-1\}$ et une base orthonormée de V_m , par exemple obtenue par le procédé de Gram-Schmidt. Par cette approche générale, une borne supérieure du risque quadratique intégré minimax est obtenue, voir [53, théorème 2], qui complète la borne inférieure déjà mentionnée. Ces deux bornes ne coïncident pas en toute généralité et la suite de ce travail consiste à les expliciter dans des cas particuliers.

Trois classes de modèles de mélange sur \mathbb{Z}_+ sont considérés, pour lesquelles l'estimateur de projection proposé est étudié.

La classe la plus importante est celle des mélanges de séries entières sur \mathbb{Z}_+ définies par

$$\pi_\theta(k) = a_k \theta^k \tilde{Z}(\theta), \quad k \in \mathbb{Z}_+,$$

où $Z(t) = \sum_{k \geq 0} a_k t^k$ est une série entière à coefficients positifs. Pour ν égale à la mesure de Lebesgue restreinte à un intervalle compact, on montre que les classes fonctionnelles définies par les approximations dans les espaces V_m , sont dans ce cas des espaces d'approximations décrits par des modules de régularités particuliers (voir [53, propositions 7 et 8]). Des conditions générale sur $\{a_k\}$ sont ensuite données pour lesquels les bornes supérieures et inférieures du risque minimax précédemment obtenues montrent que l'estimateur de projection atteint la vitesse minimax pour un choix convenable de m en fonction de n . Ces vitesses sont logarithmiques, voir sous-logarithmiques (cf. [53, corollaires 1 et 2]), ce dernier cas étant celui des mélanges de lois de Poisson. Pour le mélange de lois de Poisson, des estimateurs de projections et des estimateurs à noyaux ([31, 38]) avaient déjà été proposés mais sans montrer qu'ils atteignaient une vitesse minimax, sauf pour des espaces de Sobolev d'indices entiers. Nos résultats fournissent un estimateur de projection asymptotiquement efficace au sens minimax, ce qui corrobore les conclusions de l'étude empirique menée par [39] indiquant que les estimateurs de projections ont de meilleurs performances que les estimateurs à noyaux. Qui plus est, cette efficacité est obtenue pour un choix universelle de m en fonction de n (c'est-à-dire ne dépendant pas des hypothèses de régularité pour f).

Les deux autres exemples de modèles de mélanges étudiés dans [53] illustrent le fait que l'estimateur de projection proposé atteint en général la vitesse minimax, du moins quand on la connaît. Pour un des exemples, celui de la déconvolution discrète, la borne inférieure générale du risque décrite au paragraphe 2.3.1 n'est pas optimale ; en fait cette borne ne semble optimale que lorsque les vitesses sont lentes, au mieux logarithmiques, ce qui n'est cependant pas rares pour l'estimation de densité mélangeante.

Mentionnons pour conclure qu'une approche relativement générale pour l'estimation d'une densité mélangeante est celle du maximum de vraisemblance non-paramétrique. Des vitesses de convergence de cet estimateur ont été obtenues par [61]. Il serait sans doute intéressant de comparer cet estimateur avec l'estimateur de projection, par exemple pour le mélange de lois de Poisson.

3.2.2 Estimation de la densité de la marque d'un processus shot-noise

Nous avons étudié dans [46, 60] deux estimateurs non-paramétriques de la densité des Y_k , $k \geq 1$ définis par (2.35) à partir des observations (X'_k, Y'_k) , $k = 1, \dots, n$ définies par (2.37) et (2.38) observables à partir de la trajectoire du processus shot-noise (2.33).

Ces estimateurs s'inspirent de la formule (2.40). On pose, pour x réel et p complexe,

$$a(x, p) \stackrel{\text{def}}{=} \exp(\lambda \mathbb{E}[e^{-pY}(x - X)_+]).$$

En interprétant la partie gauche de la formule (2.40) comme une transformée de Fourier, on obtient, par inversion, pour tout $c > 0$ fixé, pour tout $x \geq 0$ et tout p de partie réelle positive,

$$a(x, p) = 1 + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\lambda \mathcal{L}\mathbf{P}'(c + i\omega, p)}{(c + i\omega + \lambda) \{c + i\omega + \lambda - \lambda \mathcal{L}\mathbf{P}'(c + i\omega, p)\}} e^{(\lambda + c + i\omega)x} d\omega. \quad (3.9)$$

Le fait d'avoir pris $c > 0$ garantit que cette intégrale est correctement définie pour tout p de partie réelle positive puisqu'alors $|\mathcal{L}\mathbf{P}'(c + i\omega, p)| \leq 1$. Les estimateurs sont alors construits à partir des remarques suivantes :

- (E-1) λ s'estime facilement en utilisant les durées des périodes inactives $\{T'_1, T'_k - (T'_{k-1} + X'_{k-1}), k \geq 2\}$ (i.i.d. de loi exponentielle de paramètre λ) ;
- (E-2) la dérivée logarithmique de $a(\cdot, p)$ permet de retrouver la fonction caractéristique de Y , quand $x \rightarrow \infty$,

$$\partial_x \log a(x, p) = \lambda \mathbb{E}[e^{-pY} \mathbb{1}(X \leq x)] \rightarrow \lambda \mathbb{E}[e^{-pY}] ; \quad (3.10)$$

- (E-3) \mathcal{LP}' est facilement estimé à partir des observations (X'_k, Y'_k) , $k = 1, \dots, n$;
- (E-4) un estimateur de $\partial_x \log a(x, p)$ est obtenu à partir de la formule (3.9); la dérivée $\partial_x a(x, p)$ se fait en deux étapes en décomposant la fraction rationnelle sous l'intégrale en deux termes : le premier de ces deux termes est intégré explicitement en ω puis dérivé en x , l'intégral du second peut être dérivé sous le signe intégral, voir [46, Eq. (14)–(17)].
- (E-5) En utilisant (3.10) avec x grand et l'estimateur de $\partial_x \log a(x, p)$ du point précédent, un estimateur de $\mathbb{E}[e^{-pY}]$ est obtenu.

Les deux approches étudiées respectivement dans [46, 60] ont en commun cette succession d'opérations mais les étapes (E-3) et (E-5) diffèrent :

- (a) dans [46], \mathcal{LP}' est estimé par \mathcal{LP}'_n , où P'_n est la loi empirique associée aux observations (X'_k, Y'_k) , $k = 1, \dots, n$, puis la densité de Y est estimé par une inversion, lissée par un noyau de bande passante h , de l'estimateur de sa fonction caractéristique obtenu à l'étape (E-4) ci-dessus;
- (b) dans [60], \mathcal{LP}' est estimé en lissant la seconde composante par un noyau de bande passante h , puis la densité de Y est estimé par une inversion directe de l'estimateur de sa fonction caractéristique obtenu à l'étape (E-4).

La différence consiste donc uniquement à changer la place de l'étape de lissage : ou bien lors de l'inversion de l'estimateur de la fonction caractéristique de Y , ou bien lors de l'estimation de \mathcal{LP}' . Les deux estimateurs obtenus dépendent de deux paramètres : la taille de la bande passante h utilisée lors de l'étape de lissage et la valeur de x utilisée à l'étape 5. Le paramètre h est habituel en estimation non-paramétrique de la densité par des méthodes à noyau. Le paramètre x règle la quantité de biais due à l'approximation de $\mathbb{E}[e^{-pY}]$ par $\mathbb{E}[e^{-pY} \mathbb{1}(X \leq x)]$, cf. (3.10). Pour X borné, il suffit évidemment de prendre x telle que $\mathbb{P}(X \leq x) = 1$. Le comportement asymptotique de l'estimateur (a) est plus particulièrement étudié dans [46] où des bornes d'erreur d'estimation sont fournies. Nous montrons que si X est borné, l'estimateur (a) atteint la vitesse usuelle $n^{-\beta/(1+2\beta)}$ de convergence sous des hypothèses de régularité Sobolev d'indice β pour la densité, voir [46, théorème 4.1]. C'est donc la même vitesse que celle obtenue à partir d'observations directs de Y_1, \dots, Y_n . Le cas où X n'est pas borné est aussi examiné, voir [46, théorème 5.1]. En particulier, si la queue de distribution de X est sous-exponentielle les bornes obtenues prévoient une très faible perte de vitesse ($n^{-\beta/(1+2\beta)+\epsilon}$ pour $\epsilon > 0$ arbitrairement petit), voir [46, corollaire 5.1]. En revanche pour des queues exponentielles et sur-exponentielle, une perte significative de la vitesse est prévue par les bornes d'erreur données. En particulier, des queues de type Pareto donnent des vitesses logarithmiques. Il semble donc que le comportement de queue de distribution influe directement sur la qualité de l'estimation de la densité de Y à partir des observations indirectes (X'_k, Y'_k) , $k = 1, \dots, n$. La question de savoir (cf. 2) si cette conclusion correspond à une perte réelle d'information reste cependant ouverte.

Cette analyse peut s'adapter à l'estimateur (b) mais c'est surtout l'implémentation numérique, le choix des paramètres h et x d'après données et l'application à des données réelles qui sont traités dans [60].

3.2.3 Estimation de l'indice de queue de la marque d'un processus shot-noise

On considère le modèle de trafic du processus de flots à arrivées poissonniennes défini au paragraphes 2.3.2 et 2.3.4 par (2.33) et (2.41) sous une hypothèse de queue lourde d'indice $\alpha \in (0, 2)$ comme défini par (2.43).

Nous basant sur les similitudes entre les propriétés du second ordre du processus de flots à arrivées poissonniennes d'indice $\alpha \in (0, 2)$, avec les processus à paramètre de mémoire longue égal à $d = 1 - \alpha/2$ mises en valeur au paragraphe 2.3.4, nous avons proposé et étudié dans [20] l'estimateur (3.7) appliqué aux coefficients d'ondelettes de ce processus observé en temps continu, échantillonné à des temps discrets ou enfin observé sous forme de moyennes locales. Exprimé comme un estimateur de α et non de $d = 1 - \alpha/2$,

cet estimateur s'exprime

$$\hat{\alpha}_n \stackrel{\text{def}}{=} \arg \inf_{t \in \mathbb{R}} \left\{ \log \left[\sum_{j=L_n}^{U_n} n_j \hat{\sigma}_j^2 2^{(t-2)j} \right] - \frac{\sum_{j=L_n}^{U_n} n_j \log 2^{(t-2)j}}{\sum_{j=L_n}^{U_n} n_j} \right\}. \quad (3.11)$$

Dans cette formule, $\hat{\sigma}_j^2$ est définie comme en (3.5) en prenant pour $W_{j,k}$ les coefficients d'ondelettes de \mathbf{X} , en temps continu ou discret.

La consistance de $\hat{\alpha}_n$ est donnée par [20, théorème 4.1] sous la seule hypothèse que (2.43) soit valide pour $p = 0, 1, \dots, 4$. Mais cette consistance repose sur des conditions pour L_n et U_n différente de celles du cas gaussien ou linéaire. Ces conditions dépendent de α mais nous montrons (voir [20, remarque 4.1]) que l'on peut choisir L_n et U_n indépendamment de α et obtenir un estimateur consistant de α quelque soit le schéma d'observation (temps continu, temps discret ou moyennes locales). Ce point nous semble particulièrement important dans les applications de télé-traffic puisqu'il fournit un estimateur consistant de l'indice α qui détermine la stabilité du modèle de trafic.

Alors que dans le cas linéaire, la consistance était obtenue avec $2^{L_n} \sim n^\gamma$ pour n'importe quel $\gamma \in (0, 1)$, dans le cas du processus de flots, il faut nécessairement $\gamma < 1/\alpha$. Cette différence n'est pas due à une difficulté technique mais :

- (i) pour $\alpha \in [1, 2)$, à un comportement différent de la variance de $\hat{\sigma}_j^2$, en $n_j^{-1} 2^{4dj}$ dans le cas linéaire (voir [48, théorème 1]) et en $n_j^{-1} 2^{4dj} 2^{(\alpha-1)j}$ dans le cas processus de flots ;
- (ii) pour $\alpha \in (0, 1]$, au comportement non-stationnaire de X et de sa transformée en ondelette qui induit un biais supplémentaire dans le cas processus de flots.

Pour le processus de flots la coupure entre les cas $\alpha < 1$ et $\alpha > 2$ peut être précisé comme suit. On montre (voir [20, lemme 5.2]) que

$$\text{Var}(\hat{\sigma}_j^2) \sim n_j^{-1} 2^{\delta j} \text{ avec } \delta = (4 - 2\alpha) \vee (3 - \alpha) = \begin{cases} 4d & \text{si } \alpha \leq 1 \\ 4d + (\alpha - 1) & \text{sinon} \end{cases}$$

On peut penser que le problème du biais pour $\alpha \leq 1$ peut être corrigé afin de permettre $L_n \sim n^\gamma$ avec $\gamma \in [1/\alpha, 1)$ comme dans le cas linéaire mais cette question reste à être prouvée. Pour le cas $\alpha > 1$, la question est plus fondamentalement reliée à la différence de comportement du processus de flots poissonnien aux grandes échelles. On sait en effet par [44], que le processus agrégé en temps (2.42) bien renormalisé se comporte asymptotiquement comme un processus de Lévy α -stable alors que pour un processus linéaire, la renormalisation est différente, et la limite est un MBF.

Si nous discutons cette condition sur L_n pour la consistance, c'est surtout parce qu'elle a une forte incidence sur les vitesses de convergence établies pour $\alpha > 1$ dans [20, théorème 4.2]). Ces vitesses dépendent du paramètre β donné par la condition

$$\text{Var}(W_{j,0}) = C 2^{(2-\alpha)j} (1 + O(2^{-\beta j})), \quad (3.12)$$

où $\{W_{j,k}\}$ dénotent ici les coefficients d'ondelette de la version stationnaire du processus (qui existe puisque $\alpha > 1$). Cette condition est évidemment à rapprocher de (3.6) qui donnait déjà le biais (puis la vitesse) dans le cas des processus linéaires. Alors que dans le cas linéaire cette condition permettait d'obtenir la vitesse de convergence $n^{-\beta/(1+2\beta)}$, on obtient cette fois-ci la vitesse plus lente

$$\hat{\alpha}_n = \alpha + O_P \left(n^{-\beta/(\alpha+2\beta)} \right). \quad (3.13)$$

Remarquons enfin que pour des observations à temps discrets, le coefficient β de la condition (3.12) ne peut dépasser $2 - \alpha$ (c'est une conséquence de [20, lemme 5.6(iii)]).

Grâce à un résultat reliant les propriétés de queue de distribution de η ou de sa fonction caractéristique en 0 de la condition (3.12) (voir [20, lemme 4.3]), les vitesses de convergences sont obtenues dans deux

cas particuliers : η suit une loi de Pareto et η est la valeur absolue d'une v.a. α -stable symétrique (et, pour simplifier, U et η sont indépendantes).

Ainsi dans le cas d'une durée de flot de loi de Pareto, la vitesse obtenue est $n^{-(2-\alpha)/(4-\alpha)}$ alors que si les durées η_k étaient observées directement la vitesse serait $n^{-1/2}$.

Un résultat plus positif est celui où l'on considère une classe semi-paramétrique plus large : celle principalement définie par

$$L_2(t) = c + O(t^{-\beta}) \quad \text{quand } t \rightarrow \infty, \text{ avec } c > 0. \quad (3.14)$$

Dans ce cas en appliquant [20, lemme 4.3 et théorème 4.2], on obtient pour $\alpha > 1$ un estimateur de α à partir de $\mathbf{X}(t), t = 1, \dots, n$ de vitesse de convergence $n^{-\gamma/(2\gamma+\alpha)}$ avec $\gamma = \beta \wedge (2 - \alpha)$ si $\beta \neq 2 - \alpha$ et $\gamma < 2 - \alpha$ pour $\beta = 2 - \alpha$. En particulier, si $\beta < 2 - \alpha$, en définissant $\rho > 0$ par $\beta = \rho\alpha$, on trouve un estimateur qui converge à la vitesse $n^{-\rho/(1+2\rho)}$. Or on sait (voir par exemple [19]) que pour la sous-classe du modèle (3.14) définie par U indépendant de η et η de densité f telle que

$$f(t) = t^{-\alpha-1}(c + O(t^{-\rho\alpha})) \quad \text{quand } t \rightarrow \infty,$$

la vitesse d'estimation optimale au sens minimax de α à partir de m observations η_1, \dots, η_m est $m^{-\rho/(1+2\rho)}$. Comme le nombre de flots présents dans les observations $\mathbf{X}(t), t = 1, \dots, n$ est précisément de l'ordre $m \sim n$ quand $n \rightarrow \infty$, nous n'avons, pour le modèle défini par (3.14), aucune perte de vitesse par rapport au modèle d'observations directes.

Nous avons donc partiellement répondu à la question (Q-2) : pour le modèle semi-paramétrique défini par (3.14) avec $\beta < 2 - \alpha$, il n'y a pas de perte d'information au sens de la vitesse minimax pour l'estimation de α quand les marques η_1, \dots, η_m sont observées indirectement par $\mathbf{X}(t), t = 1, \dots, n$, avec $n \sim m$, par exemple $n = T_m$.

Pour $\beta \geq 2 - \alpha$, ou même pour η de loi de Pareto, nous observons une perte de vitesse par rapport au modèle d'observations directes. Il reste à savoir, comme au paragraphe 3.2.2, si cette perte de vitesse est due à la méthode d'estimation ou à une perte d'information due au problème d'inversion des mesures indirectes de ces durées à travers le processus \mathbf{X} .

Une autre question ouverte liée à notre travail est celui de l'extension du modèle de flots à taux constants (2.41) à des flots plus réalistes. Dans le cadre de la modélisation d'un trafic Internet au niveau IP, l'hypothèse la plus contestable du modèle de flots à arrivées poissonniennes que nous avons considéré est son idéalisation au "niveau paquets" : les flots correspondent souvent à une session TCP, lui-même constitué par des échanges de paquets. Un modèle prenant en compte la modélisation du trafic IP au niveau paquet a été proposé dans [21]; dans ce modèle, le processus $Z(t)$ est lui-même un processus ponctuel dont les points sont les instants de passage d'un paquet du flot Z . Dans [21], une étude empirique de l'estimateur que nous avons étudié pour les flots constants montre qu'il est robuste à ce modèle plus fin des flots mais les résultats théoriques restent à établir.

Bibliographie

- [1] P. Abry, D. Veitch, and P. Flandrin. Long-range dependence : revisiting aggregation with wavelets. *J. Time Ser. Anal.*, 19(3) :253–266, 1998. 3.1.3
- [2] L. Alvarez, Y. Gousseau, and J.-M. Morel. The size of objects in natural and artificial images. *Advances in Imaging and Electron Physics, Academic Press*, 111 :167–242, 1999. 2.4.3
- [3] A. Ayache and F. Roueff. A Fourier formulation of the Frostman criterion for random graphs and its applications to wavelet series. *Appl. Comput. Harmon. Anal.*, 14(1) :75–82, 2003. 2.4.1
- [4] A. Ayache, F. Roueff, and Xiao Y. Joint continuity of the local times of linear fractional stable sheets. *C. R. Acad. Sci. Paris Sér. I Math.*, 2007. To appear. 2.4.2
- [5] A. Ayache, F. Roueff, and Xiao Y. Local and asymptotic properties of linear fractional stable sheets. *C. R. Acad. Sci. Paris Sér. I Math.*, 2007. To appear. 2.4.2, 2.4.2
- [6] A. Ayache and Y. Xiao. Asymptotic properties and hausdorff dimensions of fractional brownian sheets. *J. Fourier Anal. Appl.*, 11 :407–439, 2005. 1, 2.4.2
- [7] Richard T. Baillie, Tim Bollerslev, and Hans Ole Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1) :3–30, 1996. 2.2.3
- [8] P. Barbe. Statistical analysis of mixtures and the empirical probability measure. *Acta Appl. Math.*, 50 :253–340, 1998. 2.3.1
- [9] Aline Bonami and Anne Estrade. Anisotropic analysis of some Gaussian models. *J. Fourier Anal. Appl.*, 9(3) :215–236, 2003. 2.4.2
- [10] C. Bordenave, Y. Gousseau, and F. Roueff. The dead leaves model : an example of a general tessellation. *Advances in Applied Probability*, 38(1) :31–46, mar 2006. 2.3.5, 2.3.5, 2.3.5, 2.4.3
- [11] A. Cohen. *Numerical analysis of wavelet methods*, volume 32 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 2003. 2.4.1
- [12] R. Dahlhaus. Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In *Athens Conference on Applied Probability and Time Series Analysis, Vol. II (1995)*, volume 115 of *Lecture Notes in Statist.*, pages 145–159. Springer, New York, 1996. 2.2.2, 2.2.2
- [13] R. Dahlhaus. On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.*, 62(1) :139–168, 1996. 2.2.2, 2.2.2
- [14] R. Dahlhaus. Fitting time series models to non-stationary processes. *Annals of Statistics*, 25 :1–37, 1997. 2.2.2
- [15] R. Dahlhaus and S. Subba Rao. A recursive online algorithm for the estimation of time-varying ARCH parameters. *Bernoulli*, 2007. To appear. 2.2.3
- [16] D. L. Donoho. De-noising via soft-thresholding. *IEEE Trans. Info. Theory*, 41 :613–627, 1995. 2.4.1
- [17] David L. Donoho and Iain M. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26(3) :879–921, 1998. 2.4.1
- [18] R. Douc, F. Roueff, and P. Soulier. On the existence of some ARCH(∞) processes. submitted, 2006. 2.2.3

- [19] H. Drees. Optimal rates of convergence for estimates of the extreme value index. *Ann. Statist.*, 26(1) :434–448, 1998. 3.2.3
- [20] G. Faÿ, F. Roueff, and P. Soulier. Estimation of the memory parameter of the infinite source Poisson process. *Bernoulli*, 2007. To appear. 2.1.2, 2.3.4, 2.3.4, 3.2.3, 3.2.3, 3.2.3, 3.2.3, 3.2.3
- [21] Faÿ G., B. González-Arévalo, T. Mikosch, and G. Samorodnitsky. Modeling teletraffic arrivals by a Poisson cluster process. *Queueing Systems*, pages 121–140, 2006. 3.2.3
- [22] L. Giraitis, P.M. Robinson, and A. Samarov. Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.*, 18 :49–61, 1997. 3.1.3
- [23] G. K. Golubev. Estimation of the period of a signal with an unknown form against a white noise background. *Problemy Peredachi Informatsii*, 24(4) :38–52, 1988. 3.1.2, 3.1.2
- [24] Y. Gousseau and J.-M. Morel. Are natural images of bounded variation? *SIAM J. Math. Analysis*, 33(3) :634–648, 2001. 2.4.3
- [25] Y. Gousseau and F. Roueff. A geometrical a priori for capturing the regularity of images. In *EUSIPCO 2005*, aug 2005. 2.3.5, 2.4.3
- [26] Y. Gousseau and F. Roueff. Modeling occlusion and scaling in natural images. *SIAM Multiscale Modeling and Simulation*, 2007. 2.3.5, 2.4.3, 2.4.3
- [27] Y. Grenier. Time-dependent ARMA modeling of nonstationary signals. *IEEE Transactions on ASSP*, 31(4) :899–911, 1983. 2.2.2
- [28] L. Guo. Stability of recursive stochastic tracking algorithms. *SIAM J. on Control and Optimization*, 32 :1195–1125, 1994. 3.1.1
- [29] Peter Hall, James Reimann, and John Rice. Nonparametric estimation of a periodic function. *Biometrika*, 87(3) :545–557, 2000. 3.1.2
- [30] M. Hallin. Mixed autoregressive-moving average multivariate processes with time-dependent coefficients. *J. Multivariate Anal.*, 8(4) :567–572, 1978. 2.2.2
- [31] N. Hengartner. Adaptive demixing in Poisson mixture models. *Ann. Statist.*, 25 :917–928, 1997. 3.2.1
- [32] C. M. Hurvich and B. K. Ray. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Ser. Anal.*, 16(1) :17–41, 1995. 2.1.1
- [33] Vytautas Kazakevičius and Remigijus Leipus. A new theorem on the existence of invariant distributions with applications to ARCH processes. *Journal of Applied Probability*, 40(1) :147–162, 2003. 2.2.3
- [34] N. Kôno and M. Maejima. Hölder continuity of sample paths of some self-similar stable processes. *Tokyo J. Math.*, 14 :93–100, 1991. 2.4.1, 2
- [35] A. Lee, D. Mumford, and J. Huang. Occlusion models for natural images : A statistical study of a scale invariant dead leaves model. *Int'l J. of Computer Vision*, 41 :35–59, 2001. 2.3.5, 2.4.3
- [36] C. Lévy-Leduc, E. Moulines, and F. Roueff. Semiparametric frequency estimation from irregularly sampled observations. 2007. Soumis. 2.2.4, 3.1.2, 3.1.2
- [37] B. G. Lindsay. *Mixture Models : Theory, Geometry and Applications*. Institute of Mathematical Statistics, Hayward, CA, 1995. 2.3.1
- [38] W.-H. Loh and C.-H. Zhang. Global properties of kernel estimators for mixing densities in discrete exponential family models. *Statist. Sinica*, 6 :561–578, 1996. 3.2.1
- [39] W.-H. Loh and C.-H. Zhang. Estimating mixing densities in exponential family models for discrete variables. *Scand. J. Statist.*, 24 :15–32, 1997. 3.2.1
- [40] Parulekar M. and A. M. Makowski. M/G/Infinity input processes : A versatile class of models for network traffic. In *Proceedings of the INFOCOM '97*, 1997. 2.3.4
- [41] G. Matheron. Modèle séquentiel de partition aléatoire. Technical report, CMM, 1968. 2.3.5, 2.3.5

- [42] G. Matheron. *Random Sets and Integral Geometry*. John Wiley & Sons, New York, 1975. 2.3.5
- [43] Y. Meyer, F. Sellan, and M. S. Taqqu. Wavelets, generalized white noise and fractional integration : the synthesis of fractional Brownian motion. *The Journal of Fourier Analysis and Applications*, 5 :466–494, 1999. 2.1.2
- [44] Thomas Mikosch, Sidney Resnick, Holger Rootzén, and Alwin Stegeman. Is network traffic approximated by stable Levy motion or fractional Brownian motion? *Annals of Applied Probability*, 12 :23–68, 2002. 2.3.4, 3.2.3
- [45] E. Moulines, P. Priouret, and F. Roueff. On recursive estimation for time varying autoregressive processes. *Ann. Statist.*, 33(6) :2610–2654, 2005. 2.2.2, 2.2.2, 3.1.1, 3.1.1, 3.1.1, 1, 2
- [46] E. Moulines, F. Roueff, A. Souloumiac, and T. Trigano. Nonparametric inference of photon energy distribution from indirect measurements. *Bernoulli*, 2007. To appear. 2.3.3, 3.2.2, 4, 3.2.2, a, 3.2.2
- [47] E. Moulines, F. Roueff, and M.S. Taqqu. Central Limit Theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context. submitted, 2006. 2.1.2, 3.1.3
- [48] E. Moulines, F. Roueff, and M.S. Taqqu. A wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series. 2006. submitted. 2.1.2, 2.1.3, 3.1.3, i, iii, 3.1.3, i
- [49] E. Moulines, F. Roueff, and M.S. Taqqu. On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. *J. Time Ser. Anal.*, 28(2), 2007. 2.1.1, 2.1.2, 2.1.2, 2.1.2, 2.1.3, ii, 3.1.3, 3.1.3
- [50] D. Mumford and B. Gidas. Stochastic models for generic images. *Quarterly Journal of Mathematics*, 59 :85–111, 2001. 2.4.3
- [51] E. Le Pennec and S. Mallat. Bandelet image approximation and compression. *SIAM Multiscale Modeling and Simulation*, 4(3) :992–1039, 2005. 2.4.3
- [52] R. Pyke. On renewal processes related to type I and type II counter models. *Ann. Math. Statist.*, 29 :737–754, 1958. 2.3.3
- [53] F. Roueff and T. Rydén. Nonparametric estimation of mixing densities for discrete distributions. *Ann. Statist.*, 33(5) :2066–2108, 2005. 2.3.1, 2.3.1, 2.3.1, 3.2.1
- [54] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian processes : stochastic models with infinite variance*. Chapman and Hall, 1994. 2.4.2
- [55] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18(1) :17–33, 2003. 2.4.3
- [56] T. Subba Rao. The fitting of non-stationary time-series models with time-dependent parameters. *J. Roy. Statist. Soc. Ser. B*, 32 :312–322, 1970. 2.2.2
- [57] L. Takacs. *Introduction to the Theory of Queues*. Oxford University Press, 1962. 2.3.3
- [58] K. Takashima. Sample path properties of ergodic self-similar processes. *Osaka J. Math.*, 26(1) :159–189, 1989. 2, 2.4.2
- [59] D. M. Titterton, A. F. M. Smith, and U. E Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York, 1985. 2.3.1
- [60] T. Trigano, T. Montagu, E. Moulines, F. Roueff, and A. Souloumiac. Statistical pileup correction method for HPGe detectors. *IEEE Trans. Signal Process.*, 2007. to appear. 2.3.3, 3.2.2, 3.2.2, b, 3.2.2
- [61] S. van de Geer. Rates of convergence for the maximum likelihood estimator in mixture models. *J. Nonparametr. Statist.*, 6 :293–310, 1996. 3.2.1
- [62] G.W. Wornell and A.V. Oppenheim. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Process.*, 40(3) :611 – 623, March 1992. 3.1.3
- [63] Y. Xiao and T. Zhang. Local times of fractional brownian sheets. *Probab. Th. Rel. Fields*, 124 :204–226, 2002. 1

- [64] P. Zaffaroni. Stationarity and memory of ARCH(∞) models. *Econometric Theory*, 20(1) :147–160, 2004. 2.2.3

N.B. Les articles décrits dans cette synthèse sont imprimés à la suite de cet ouvrage ou disponibles sur <http://www.tsi.enst.fr/~roueff/hdr/index.html>