

Numéro d'ordre : 017

HABILITATION À DIRIGER DES RECHERCHES
École Doctorale de Mathématique et Informatique Fondamentale

**Analyse temps-fréquence et modèles d'invariance d'échelle :
Contribution à l'étude de systèmes complexes**

Présenté par

Paulo GONÇALVES

Diplôme délivré par

L'École Normale Supérieure de Lyon
Laboratoire de l'Informatique du Parallélisme

Soutenue le 29 mars 2010 devant le jury constitué de :

Président	Monsieur Éric Fleury	Professeur ENS Lyon, France
Rapporteurs	Monsieur Stéphane Mallat	Professeur École Polytechnique, France
	Monsieur Ravi Mazumdar	Professeur Université de Waterloo, Canada
	Madame Valérie Perrier	Professeur INPG-ENSIMAG, France
Examineurs	Monsieur Gilles Celeux	Directeur de Recherche INRIA Saclay, France
	Monsieur Patrick Flandrin	Directeur de Recherche CNRS, ENS Lyon, France
	Monsieur Jean-Marc Vesin	Professeur EPFL, Suisse

Table des matières

1	Introduction	5
1.1	Parcours professionnel	5
1.2	Animation de la communauté scientifique et enseignement	6
1.3	Encadrement d'étudiants	7
1.3.1	Thèse de Patrick Loiseau	7
1.3.2	Thèse de Hugo Carrão	8
1.3.3	Thèse de Julien Gosme	8
1.3.4	Encadrements divers	9
1.4	Projets scientifiques	9
1.4.1	Laboratoire Commun INRIA - Alcatel Lucent Bell Labs	10
1.4.2	PetaFlow	10
1.4.3	DMASC - Propriétés d'Invariance d'Echelles de Signaux Cardiaques, Systèmes Dynamiques et Analyse Multifractale	11
1.5	Organisation du document	12
2	Représentations Temps-Fréquence	15
2.1	Introduction	15
2.1.1	Descriptions duales en temps et en fréquence d'un signal	15
2.1.2	Représentations atomiques temps-fréquence	17
2.1.3	Distributions bilinéaires temps-fréquence	20
2.2	Distributions pseudo-Wigner affines	25
2.2.1	Définition	25
2.2.2	Pseudo Affine Wigner Distributions and the Affine Class	29
2.3	Diffusion	31
2.3.1	Définitions	31
2.3.2	Isotropic diffusion	32
2.3.3	Anisotropic diffusion	38
2.4	Décomposition modale empirique	42
2.4.1	Principe	43
2.4.2	EMD vs. wavelets : a qualitative appreciation	45
2.5	Bilan et perspectives	47
3	Estimateurs et Modèles	49
3.1	Modèles d'invariances d'échelle	49
3.1.1	Auto-similarité.	50
3.1.2	Dépendance à longue portée	51
3.1.3	Régularité locale.	52
3.1.4	Mouvement Brownien fractionnaire : paradigme de l'invariance d'échelle ...	53
3.1.5	Invariances d'échelle non-stationnaires	54
3.2	Ondelettes et invariances d'échelle	55
3.2.1	Analyse en ondelettes de processus à invariance d'échelle	55
3.2.2	Estimation de lois d'échelle homogènes	58
3.2.3	Estimation de régularités locales non-stationnaires	61

3.3	EMD et invariances d'échelle	65
3.3.1	EMD analysis of broadband noise	65
3.3.2	EMD-based estimation of scaling laws	68
3.4	Un test d'existence des moments d'une v.a.	76
3.4.1	Principe	76
3.4.2	Detailed rationale	78
3.4.3	Application to the multifractal partition function estimation	86
3.5	Bilan et perspectives	90
4	Applications	93
4.1	Réseaux	93
4.1.1	Motivations	93
4.1.2	Une plate-forme de simulation et de métrologie	94
4.1.3	Validation expérimentale et extension du théorème de Taqqu	96
4.1.4	Nouvelles propriétés d'invariance d'échelle des flux TCP	101
4.2	Analyse du rythme cardiaque	103
4.2.1	Détermination de la nature du tonus cardiaque chez la souris	103
4.2.2	Analyse multifractale du rythme cardiaque	110
5	Conclusions, Bilan et Perspectives de Recherche	111
	Bibliographie Personnelle	114
	Bibliographie Générale	119

1

Introduction

1.1 Parcours professionnel

Ce document présente une synthèse des activités de recherche, d'animation et d'encadrement scientifique, que j'ai menées depuis l'obtention de mon doctorat en novembre 1993.

Depuis 1996, date à laquelle j'ai obtenu un poste de Chargé de Recherche à l'Institut National de Recherche en Informatique et Automatique (INRIA), j'ai successivement été membre permanent de trois Équipe-Projets INRIA (EPI), elles-mêmes rattachées à deux Centres de Recherche différents. L'organisation modulaire de l'INRIA repose sur une cohésion thématique forte des équipes-projets, chacune visant à l'excellence scientifique dans son domaine. En intégrant tour à tour trois EPI axées sur trois disciplines distinctes m'a conduit, non pas à une triple reconversion thématique, mais à adapter et à enrichir mes compétences en Traitement du Signal, pour servir au mieux de nos intérêts communs, les domaines de recherche spécifiques de mes équipes d'accueil. J'ai ainsi cherché à concevoir des outils – inspirés pour l'essentiel de la théorie des ondelettes – en réponse aux problèmes d'estimation, d'analyse ou de modélisation, qui parfois verrouillaient l'accès à des résultats originaux, importants pour l'activité du projet. Mais, véritables partenariats, chacune de ces collaborations m'a aussi permis d'acquérir ou d'approfondir certaines connaissances théoriques liées au domaine d'application de l'EPI. C'est cet enrichissement pluridisciplinaire que je me suis efforcé de mettre au profit des diverses études menées ces seize dernières années. Séduit par ce que je pense être la vocation même du Traitement du Signal – et plus largement celle des mathématiques appliquées – de se transposer à des champs disciplinaires aussi étendus que les réseaux de communication ou que le biomédical, c'est cette voie que j'ai choisie d'explorer, sachant qu'elle me conduirait à côtoyer des paysages scientifiques variés. Réel, et sans doute unique point commun à l'ensemble des activités de ces dernières années, la recherche dans les Sciences et Techniques de l'Information et de la Communication est aussi le fil conducteur autour duquel s'organise ce manuscrit.

Ci-dessous, je résume en quelques dates charnières les principales étapes de mon itinéraire professionnel.

1993 Thèse de doctorat « Représentations Temps-Fréquence et Temps-Echelle Bilinéaires : Synthèse et Contributions » effectuée au laboratoire de Physique de l'Ecole Normale Supérieure de Lyon (ENS Lyon) sous la direction de Patrick Flandrin (DR1 CNRS). Doctorat délivré par l'Ecole Doctorale « Signal - Image - Parole » de l'Institut National Polytechnique de Grenoble.

1994–96 Séjour post-doctoral avec Richard Baraniuk, professeur au Digital Signal Processing Department de Rice University, Houston (TX), USA.

1996 Recrutement CR2 à l'Institut National de Recherche en Informatique et Automatique (INRIA).

1996–99 Membre permanent de l'EPI FRACTALES dirigée par Jacques Lévy-Véhel (DR INRIA), INRIA Rocquencourt.

1999-2006 Membre permanent (CR1) de l'EPI IS2 dirigée par Gilles Celeux (DR1 INRIA), INRIA Rhône-Alpes. L'EPI IS2 s'est restructurée en 2004 et est devenue l'EPI MISTIS, dirigée par Florence Forbes (CR1 INRIA).

2003–05 Mise à disposition auprès de l'Institut Supérieur de Technologie de Lisbonne (Portugal), au département Systèmes et Robotique dirigé par João Sentieiro (Professeur).

depuis 2006 Membre permanent (CR1) de l'EPI RESO, dirigée par Pascale Vicat-Blanc Primet (DR2 INRIA). RESO est une équipe mixte INRIA Rhône-Alpes / LIP-ENS Lyon, et est physiquement localisée au LIP (Laboratoire d'Informatique).

Chacun de ces repères marque la transition entre deux défis scientifiques, transitions elles-mêmes systématiquement accompagnées d'un changement d'équipe, d'institut, voire de pays d'accueil... Cette mobilité thématique, autant que géographique, a été pour moi une source de renouveaux, parfois aussi de rupture avec une collaboration tarie, mais que je juge aujourd'hui formatrice et souhaitable à l'apprentissage, puis à la vivification d'une carrière de chercheur. J'ai donc vécu l'ensemble de ces coopérations comme une formidable opportunité de satisfaire une curiosité scientifique en diversifiant mon travail, ses enjeux, et à chaque fois, comme le besoin de remettre en question les objectifs de ma mission de chercheur.

1.2 Animation de la communauté scientifique et enseignement

En seize années d'activité professionnelle, j'ai eu le privilège d'être sollicité pour participer à différentes actions d'animation et de diffusion de la connaissance scientifique, dont voici la liste des événements les plus marquants.

1997-2002 Sur invitation de Jean-Marc Chassery (DR1 CNRS), alors Directeur du Groupement de Recherche « Information - Signal - Images - Vision » (ISIS), unité G 1047 du CNRS, j'assume avec Christian Doncarli (Professeur à l'Université de Nantes) la corresponsabilité du GT2, le Groupe de Travail « Analyse et Décision en Signal » (<http://gdr-isis.enst.fr/NEW/ISIS98/gt2.html>). Notre rôle est d'assurer l'animation scientifique et administrative d'un groupe d'une centaine de chercheurs et de doctorants de la communauté Traitement du Signal et des Images. Notre Groupe de Travail est structuré en 3 Opérations : « Décision et Interprétation Non-Stationnaire », « Séparation de Sources » et « Ondelettes et Fractales pour le Signal et les Images ».

1997-2002 Durant cette même période, je co-anime avec Patrice Abry (DR2 CNRS) l'une des 3 Opérations « Ondelettes et Fractales pour le Signal et les Images » du GT2, laquelle compte une quarantaine de membres inscrits. Dans ce cadre, nous organisons une série d'exposés trimestriels dont le but est de favoriser l'émergence de projets de recherche communs. Pour fertiliser les échanges entre équipes, le GdR ISIS nous alloue un budget de fonctionnement qui nous permet de financer diverses missions de doctorants.

2002,2009 En 1998, Bernard Dubuisson (DR CNRS), Directeur scientifique aux éditions Hermès Sciences Publications (Paris) présente au GdR ISIS son projet d'édition d'un traité français en 120 volumes sur le Traitement du Signal, intitulé « Information – Commande – Communication » (IC2). Les laboratoires partenaires du GdR ISIS étant à la fois la cible et les protagonistes privilégiés d'un tel projet, Bernard Dubuisson nous exhorte à participer à la

coordination des ouvrages collectifs et à la rédaction des chapitres. Dans ce contexte, pour y développer les thèmes de notre Opération, j'accepte la responsabilité de coordonner un des volumes du traité et je propose aussitôt à Patrice Abry de s'y associer. Par la suite, Jacques Lévy-Véhel, nous rejoindra et nous éditerons ensemble le tome « Lois d'Échelle, Fractales et Ondelettes » . Sa parution en 2002 comporte 14 chapitres écrits par les spécialistes du domaine, puis, toujours sous notre direction, une version actualisée et traduite en anglais, « Scaling, Fractals and Wavelets » , sera publiée en 2009 par les éditions ISTE - Wiley.

2004 En l'honneur du 20ème anniversaire de la parution de l'un des articles fondateurs de la théorie des ondelettes par Alex Grossmann (DR CNRS), nous avons organisé, Patrice Abry, Richard Baraniuk, Patrick Flandrin, Stéphane Jaffard et moi-même une école d'été sur le thème « Wavelets And Multifractal Analyses » (WAMA). Cette école, dont je suis directeur, se tient du 19 au 31 juillet 2004, à l'Institut d'Études Scientifiques de Cargèse, et compte 25 orateurs invités parmi les plus grands spécialistes internationaux de la discipline (<http://perso.ens-lyon.fr/paulo.goncalves/WAMA2004>). Pour soutenir la participation de jeunes chercheurs issus de pays moins favorisés, nous avons consacré une part importante de notre budget à l'attribution de bourses d'étude dont ont bénéficié plus d'un quart de la centaine de participants.

2002–04 À partir de 2002, et jusqu'en 2004, date à laquelle je décide d'abandonner cette charge, je suis éditeur associé de la revue scientifique IEEE, « Signal Processing letters » .

2007 Alors que ce tient à Lyon en 2007, la conférence internationale ICST-ACM GridNets 2007 pour laquelle je suis *local chair*, j'organise le workshop « MetroGrid » autour du thème métrologie pour les réseaux de grilles.

2009 Dans la nouvelle maquette du Master du Département d'Informatique de l'ENS Lyon, je suis responsable de la filière « Models and Optimization for Emerging Infrastructures » . Depuis 2007, je suis également chargé d'un cours de recherche sur les « Modèles de Trafic » .

Depuis mon recrutement à l'INRIA, en 1996, j'ai enseigné un cours sur « Ondelettes et Analyse fractale » dans diverses formations : École Centrale de Paris (1997-98), Universidad Nacional Autónoma de Mexico (1997-2001), INPG-ENSERG (2001-03).

1.3 Encadrement d'étudiants

1.3.1 Thèse de Patrick Loiseau

Dès mon arrivée à l'ENS Lyon en 2006, j'assume en co-tutelle avec Pascale Vicat-Blanc Primet, la direction du travail de thèse de Patrick Loiseau (ENS Lyon), dont la date de soutenance est prévue en décembre 2009. Nous nous sommes intéressés à l'identification, à l'estimation et à la modélisation de propriétés d'invariance d'échelle observées sur les signaux de trafic informatique. Du point de vue des réseaux, l'enjeu de cette étude est de mesurer l'impact des lois d'échelle sur la Qualité de Service, puis de caractériser les différents protocoles de transport vis à vis de ces mêmes propriétés.

Point fort de mon activité depuis mon intégration à l'EPI RESO, ce travail allie mes compétences en Traitement du Signal à celles en protocoles de transport et réseaux haut-débit de Pascale Vicat-Blanc Primet (co-directrice). La formation initiale de physicien de P. Loiseau, a quant à elle apporté toute la rigueur nécessaire à la démarche expérimentale que nous souhaitons adopter. De cette association est née une plate-forme de métrologie qui, couplée à la grille expérimentale Grid5000, nous permet de valider dans des conditions d'utilisation réaliste, les résultats théoriques avancés.

Parmi ceux-ci, sans doute que la contribution la plus marquante de la thèse de P. Loiseau porte sur l'étude de modèles théoriques capables de reproduire et d'expliquer certaines caractéristiques importantes du trafic Internet. C'est notamment le cas des invariances à petites échelles dont

l'origine, encore incertaine, peut être recherchée dans les protocoles (TCP notamment) et que les modèles de Markov réussissent, dans une certaine mesure, à reproduire. Le travail théorique qui est en cours vise justement à généraliser le cadre usuel de l'analyse multifractale à une mesure de variabilité (singularité) adaptée aux dynamiques de TCP. Ce travail couvre en fait deux difficultés : d'une part la définition mathématique de l'objet (mesure et spectre des grandes déviations associé), d'autre part l'estimation robuste de ces quantités sur des signaux réels. Ces deux aspects développés dans la thèse de P. Loiseau, et sur lesquels je reviendrai dans la partie technique de ce document, sont le fruit d'une collaboration initiée en 2008 avec Julien Barral de l'EPI SISYPHE de l'INRIA Rocquencourt.

P. Loiseau soutiendra sa thèse devant le jury constitué de Messieurs Philippe Robert (Rapporteur, DR1 INRIA), Jean-Yves Le Boudec (Rapporteur, Professeur EPFL), Rudolf Riedi (Rapporteur, Professeur Université de Fribourg) Madame Pascale Vicat-Blanc Primet (Co-directrice, DR2 INRIA) et moi-même (Directeur, CR1 INRIA), le 11 décembre 2009, à l'ENS Lyon.

1.3.2 Thèse de Hugo Carrão

Pendant ma mise à disposition à l'Institut Supérieur de Technologie de Lisbonne, de 2003 à 2005, j'ai entrepris avec Mário Caetano, Chercheur à l'Institut Portugais de Géographie (IGP), une collaboration sur la reconnaissance et la classification d'occupation de sol à partir de l'étude d'images satellite (*remote sensing*). M. Caetano m'a alors proposé de co-encadrer le travail de thèse de Hugo Carrão, doctorant à l'Institut Supérieur de Statistiques et de Gestion de l'Information (ISEGI) de l'Université Nouvelle de Lisbonne (UNL), sur ce même sujet. Nous avons commencé par élaborer une procédure de classification supervisée par Séparateurs à Vaste Marges (SVM pour la terminologie anglaise de *Support Vector Machine*) adaptée aux spécificités spectrales, spatiales et temporelles des mesures fournies par les capteurs embarqués Modis et Meris. Sur ces mêmes données, nous avons ensuite proposé un modèle harmonique non linéaire de séries temporelles, lequel se distingue des modèles usuels par ses bonnes propriétés d'extrapolation de l'évolution phénologique de différents types de végétations (agricultures saisonnières notamment). Nous utilisons alors ce modèle pour prédire sur un cycle annuel, certains des indices phénologiques les plus suivis.

Cette collaboration a été exceptionnellement fructueuse en production scientifique et humainement très enrichissante. Mais pour lui donner une suite plus consistante qui dépasserait le simple cadre marginal, il faudrait à l'avenir lui consacrer une part plus importante de mon activité, une part incompatible avec les choix que j'ai faits. C'est pourquoi je ne reviendrai pas dans ce document sur les aspects techniques de ce travail, et précise simplement que cette étude a partiellement été financée par le Programme d'Actions Intégrées Luso-Française (2007-2008). Entre autres missions, cela a permis que H. Carrão fasse deux séjours longs à l'INRIA : d'octobre 2005 à mars 2006 dans l'EPI MISTIS, et de juillet à décembre 2008 dans l'EPI RESO.

H. Carrão soutiendra sa thèse devant le jury constitué de Messieurs António Rendas (Recteur de l'UNL), Marco Painho (Professeur ISEGI), Fernando Bação (Professeur ISEGI), André Marçal (Professeur FCUP), Gilles Foody (Professeur University of Nottingham), Madame Ana Navarro (Professeur FCUL), Mário Caetano (Directeur, Chercheur IGP) et moi-même (Co-directeur, CR1 INRIA), en janvier 2010, à l'UNL.

1.3.3 Thèse de Julien Gosme

Dans la continuité thématique de son stage de DEA, effectué en 2001 sous ma responsabilité au sein de l'EPI is2, Julien Gosme a poursuivi à l'Université Technologique de Troyes, une thèse de Doctorat sous la Direction de Cédric Richard (Professeur UTT). Bien que n'ayant pas été officiellement co-encadrant de sa thèse, une part importante des publications d'alors ont été le résultat d'une collaboration que nous avons entretenue avec son Directeur jusqu'en décembre 2004, date de la soutenance de J. Gosme.

Nous lui avons donc proposé d'étendre son sujet de DEA sur la diffusion adaptative – EDP – des représentations temps-fréquence, en l'ouvrant à deux perspectives distinctes :

- Son application au problème de détection non paramétrique, l'idée étant de diminuer par lissage diffusif adapté la complexité du détecteur (complexité mesurée par la dimension de Vapnik-Chervonenkis) ;
- Proposer un schéma de diffusion équivalent à celui de la diffusion de la chaleur, mais qui soit covariant par les opérateurs de translation en temps et de changement d'échelle du groupe affine à deux paramètres.

C'est plus particulièrement sur ce deuxième axe qu'a porté notre collaboration, et c'est celui qui sera plus largement développé dans l'un des chapitres techniques de ce document.

1.3.4 Encadrements divers

Une composante essentielle de la mission de chercheur étant l'accompagnement d'étudiants séduits par une carrière scientifique, mais parfois encore indécis, j'ai naturellement supervisé, seul ou en tandem, les stages de quelques uns d'entre eux. Bien qu'elle ne soit remarquable ni par le nombre, ni par le prestige, si j'en dresse ici la liste, c'est à nouveau pour souligner le caractère pluridisciplinaire de mon activité scientifique.

- 2005 « Développement d'une boîte à outil d'analyse d'images satellite » , A. Pinheiro, IGP
Lisbonne ;
- 2003 « Décomposition Modale Empirique pour les Images » , B. Esterni. ENSIMAG et EPI
IS2 ;
- 2001 « Représentations temps-fréquence diffusées. Propriétés, applications et perspectives. Contrôle
de complexité de détecteurs temps-fréquence » , J. Gosme. Univ. Tech. de Troyes et EPI
IS2 ;
- 2000 « Optimisation semi-définie pour la synthèse d'ondelettes » , C. Triadou. ENSIMAG, EPI
NUMOPT et IS2 ;
- 2000 « Étude conjointe des variabilités du rythme cardiaque et de la repolarisation ventricu-
laire chez l'Homme insuffisant cardiaque modéré » , C. Froger. École Doctorale Physiologie
et Biologie de la Respiration et de la Circulation, Univ. Paris VII et Hôpital Lariboisière,
INSERM U572 et EPI IS2 ;
- 1998 « Segmentation de well-logs en zones homogènes » , R. Ramanantsiarovana, ESIEA et
EPI FRACTALES ;
- 97-98 « Développement de la boîte à outils Fraclab » , EPI FRACTALES ;
- 1995 « Hybrid Linea/Bilinear Time-Scale Analysis » , M. Pasquier. ENST Paris et Rice Univ.
(USA) ;
- 1993 « Signaux Browniens fractionnaires : Synthèse et estimation de la dimension fractale » ,
D. Janer, École Doctorale Signal Image Parole, (INPG) et ENS-Lyon.

1.4 Projets scientifiques

Ces dernières années, les guichets de financement de la recherche publique se sont multipliés avec bien sûr, la création d'agences de moyen, nationales et Européennes, mais aussi sous l'effet de politiques incitatives au développement de partenariats entre acteurs publiques et industriels. Soumis au principe des vases communicants, les opérateurs de recherche ont vu dans le même temps leur rôle de pilotage réduit, pour se concentrer sur leur mission de gestion du personnel. Ce mode de financement sur projets à trois ou quatre ans a sensiblement modifié les enjeux et raccourci les échelles de temps de la recherche – gagnant en réactivité ce qui peut-être s'est perdu en sérénité – cependant que le quotidien des chercheurs s'est alourdi en obligations administratives dopées par une inflation frénétique en rapports d'avancement et exercices d'évaluation.

J'ai participé à quelques unes de ces opérations contractuelles communes, et présente ci-dessous celles qui relèvent de mes thèmes de recherche actuels. Qu'ils soient appliqués aux communications dans les réseaux haut-débit, ou bien au rythme cardiaque, ces différents projets ont en commun l'étude d'invariances d'échelle dans des systèmes complexes, l'identification de leur origine, leur estimation et leur modélisation. Ils définissent également les grands axes de recherche que je souhaite

poursuivre et développer dans le cadre de mon activité à l'INRIA.

1.4.1 Laboratoire Commun INRIA - Alcatel Lucent Bell Labs

Afin justement de lisser l'effet ponctuel des opérations contractuelles, la création d'un laboratoire commun entre l'INRIA et les Bell Labs d'Alcatel-Lucent répond au besoin d'une structure stable pour abriter des objectifs de recherche à long terme et des travaux partagés entre les deux entités. Sous les traits d'une équipe-projet INRIA virtuelle et distribuée, le laboratoire commun donne à ce partenariat une meilleure visibilité externe, dans lequel peuvent se développer des synergies fortes entre les objectifs scientifiques de l'INRIA et les activités de recherche industrielle d'Alcatel-Lucent. Focalisé sur le thème des « réseaux auto-organisant », le laboratoire commun est sous-divisé en trois actions de recherche :

1. *High Manageability* (HiMa) ;
2. *Self-Networking* (SelfNet) ;
3. *Semantic Networking* (SemNet).

C'est sur ce troisième axe de recherche que se positionne l'EPI RESO, et que s'inscrit en partie mon activité.

Le traitement efficace et intelligent des flux - au sens de l'abstraction « flux = agrégat de paquets ayant une même proximité sémantique et gérés en tant que tels par le réseau » - est au centre de l'axe de recherche *Semantic Networking*. L'objectif est d'étudier comment prendre en compte la sémantique des flux pour améliorer d'une part les performances différenciées des flux hétérogènes circulant dans un réseau de télécommunications (dominé par Internet) et d'autre part de diminuer les coûts du réseau, en limitant le surdimensionnement requis aujourd'hui pour assurer la qualité de service (QoS). Nous proposons de mener un ensemble d'études complémentaires permettant :

- d'identifier et de caractériser efficacement et à très haut débit les flux,
- de les traiter et de les acheminer dans le réseau en améliorant simultanément leur fonction d'utilité propre en adéquation avec leur QoS et l'utilisation des différentes ressources dans le réseau,
- de faire en sorte que le réseau s'auto-gère, prenant de manière la plus autonome possible les décisions au niveau de la configuration et de la gestion de ces ressources.

Au final, les solutions développées dans le cadre de l'axe de recherche « Semantic Networking » devraient diminuer les coûts du réseau, que ce soit au niveau du CAPEX (infrastructure) par une meilleure utilisation des ressources des réseaux de très grande capacité (grâce à la notion de flux), ou au niveau de l'OPEX (opérations) en diminuant l'intervention humaine dans le contrôle et la gestion de ces réseaux (grâce à l'application de concepts de « self-management »).

S'agissant plus particulièrement de l'identification et de la classification des flux par type d'application, Marina Sokol, doctorante de l'EPI RESO a commencé en octobre 2009, une thèse sous ma direction, en co-tutelle avec Konstantin Avratchenkov, CR1 à l'EPI MAESTRO du centre de Sophia-Antipolis. Dans ce travail, nous recourons à l'analyse sémantique et comportementale des flux, des résultats antérieurs ayant montré que des statistiques simples calculées sur les tous premiers paquets de ceux-ci suffisent à caractériser les grandes catégories d'applications.

1.4.2 PetaFlow

Le projet ANR « PetaFlow » est une illustration saillante du type d'applications en prise directe avec la recherche amont effectuée dans le laboratoire commun INRIA - Alcatel Lucent. L'acronyme « PetaFlow » recouvre l'ensemble des trois concepts : Calculs intensifs, Visualisation scientifique et Transfert à haut débit et sur de longues distances de données d'échelle peta. Ces trois besoins naissent du problème de la modélisation de l'écoulement de l'air dans les voies supérieures de l'Homme, l'application médicale à l'origine de ce projet. Cette opération implique 2 EPI du centre Rhône-Alpes, RESO et MOAIS, une équipe CNRS de Grenoble, GIPSA-LAB, et une collaboration avec quatre équipes académiques japonaises, également partenaires du projet.

Le développement fulgurant des technologies de mesure et des ressources de calcul, fournissent et permettent le traitement de volumes de données en constante augmentation. Aujourd'hui par exemple, l'imagerie médicale doit stocker et analyser des matrices dont les tailles excèdent aisément les 10^{15} éléments ! pénétrant ainsi le domaine des peta-échelles. Mais cette explosion n'est possible qu'avec le développement simultané de ressources de calcul haute performance (petaflop). Il est toutefois déraisonnable – voire inutile – d'envisager d'équiper in-situ chaque installation peta-échelle d'un super-calculateur. En revanche, la factorisation des moyens de calcul et le recours au traitement délocalisé est sans doute une solution viable en réponse aux enjeux sociétaux et aux défis scientifiques qu'ils posent. Les « réseaux haute performance » permettent justement l'accès distant aux ressources de calcul haute performance et contribuent à un usage plus démocratique de celles-ci. L'application médicale visée, à savoir la visualisation de l'écoulement de l'air dans les voies supérieures de l'Homme, pose une difficulté supplémentaire liée à l'interactivité des représentations, lesquelles ne doivent pas souffrir de délais (incontrôlables) lorsque centre de calcul et système de visualisation sont distants de plusieurs milliers de kilomètres.

Le transfert, le traitement et la visualisation de données peta-échelle apparaît donc comme un vrai challenge pour les Sciences et Technologies de l'Information et de la Communication (STIC). Réciproquement, toute avancée dans ces disciplines, et en particulier les résultats de classification de flux en temps réel obtenus dans le cadre du Laboratoire Commun INRIA - Alcatel Lucent, contribuera à l'intégration de ce type de services dans la future génération d'Internet.

1.4.3 DMASC - Propriétés d'Invariance d'Echelles de Signaux Cardiaques, Systèmes Dynamiques et Analyse Multifractale

J'ai choisi de présenter ici le projet ANR « DMASC », car il est très représentatif d'un autre champ d'application, où une part importante des progrès accomplis repose sur des innovations liées aux mathématiques appliquées et en particulier au traitement du signal. Grâce au perfectionnement des instruments d'observation, le domaine biomédical est en constante évolution et les problèmes d'interprétation des signaux obtenus, sans cesse reposés.

Pour ma part, je me suis intéressé dès 1998, à l'étude du rythme cardiaque : D'abord chez la souris, en participant aux travaux de recherche de l'équipe du Professeur Bernard Swynghedauw de l'Unité INSERM U572 de l'hôpital Lariboisière (Paris) ; à présent chez l'Homme, d'une part, et chez le fœtus pendant la phase d'accouchement, d'autre part. Dans le premier cas, nous travaillons, avec l'EPI SISYPHE du centre de Rocquencourt, sur une base de données correspondant à des électrocardiogrammes de patients diabétiques, fournie par le Professeur Denis Chemla (de l'hôpital de Kremlin-Bicêtre, Université Paris 12) et ayant motivé la création du projet « DMASC ». La deuxième étude quant à elle, concerne l'aide à la prise de décision d'intervention par césarienne et porte sur une base de données remarquable (plusieurs milliers de cas) recueillie par le Docteur Muriel Doret, obstétricienne à l'Hôpital Femme-Mère-Enfant de Lyon. Ce travail fait l'objet d'une collaboration avec Patrice Abry (DR2 CNRS) de l'équipe SISYPHE du Laboratoire de Physique de l'ENS Lyon.

Des études numériques, empruntant certaines idées et concepts à la Physique Statistique, tels que la théorie des grandes déviations ou l'analyse fonctionnelle, ont révélé de surprenantes propriétés d'invariance d'échelle dans les longues séries temporelles d'intervalles R-R. Ces signaux, extraits d'électrocardiogrammes, correspondent aux intervalles de temps séparant deux battements cardiaques consécutifs. Les invariances d'échelle mesurées empiriquement sur ces données, sont l'expression de propriétés géométriques fractales que vérifient théoriquement certains objets mathématiques (mesures ou fonctions), communément désignés par multifractals (auto-similaires). Ces études numériques montrent également que ces lois d'échelle peuvent être de natures différentes, selon que les sujets observés sont sains ou souffrent de déficiences cardiaques. Ces observations confortent donc l'idée qu'une bonne compréhension des propriétés multifractales des signaux cardiaques puisse déboucher sur de nouveaux outils pertinents de diagnostic et de surveillance. Or, jusqu'à présent, aucune origine physiologique n'a su être associée à ces propriétés, pas plus qu'il n'existe de modèles mathématiques satisfaisants pour décrire ces signaux. Par conséquent, et avant

d'envisager une quelconque application médicale de cette piste, il est essentiel de dépasser le simple cadre de l'observation et de mener une étude approfondie des structures invariantes d'échelle, propres aux signaux cardiaques. Pour cela, il nous faut de nouveaux algorithmes, plus robustes, pour le traitement des signaux multifractals ; en particulier, nous souhaitons combiner l'approche statistique courante, avec une étude géométrique des invariances d'échelle. De plus, il est nécessaire de tester ces outils sur des bases de données représentatives de diverses cardio-pathologies, afin de proposer une classification basée sur les différentes expressions de lois d'échelle observées, et de pouvoir les interpréter en termes physiologiques. Tout ceci devrait contribuer au développement de nouveaux modèles mathématiques multifractals, précis et flexibles, dont les paramètres pourraient facilement s'ajuster aux différentes pathologies.

Enfin, il nous a semblé tout aussi intéressant d'inscrire au programme du projet « DMASC » l'analyse multifractale d'un autre signal fondamental en cardiologie : la pression artérielle. Plus précisément, voulons procéder à l'étude et à la modélisation conjointe du couple R-R / pression, en développant une approche multifractale vectorielle.

1.5 Organisation du document

Dans cette introduction, j'ai voulu montrer les nombreux changements de contexte institutionnel survenus dans mon cadre de travail. Et si j'ai cherché à mettre en avant la diversité des domaines d'application explorés, c'est pour élucider le manque apparent de cohésion thématique dans mon parcours professionnel, en soulignant le côté transverse de mes axes de recherche. Manifeste depuis plus longtemps dans les Sciences du Vivant, l'intérêt du Traitement du Signal, commence tout juste à se faire sentir dans la communauté réseaux. Alors curieusement, on découvre que deux systèmes complexes, aussi distants que le cœur et un réseau informatique, présentent du point de vue des lois d'échelle, des invariances très semblables. Sans dire qu'elles sont les signatures de structures comparables, ces propriétés font néanmoins appel aux mêmes outils d'analyse, d'estimation ou de modélisation, et c'est dans la conception de ceux-ci que se situe mon action. Sous cet angle, je pense donc qu'il existe une véritable « unité méthodologique » dans mes sujets de recherche, résistante aux changements successifs de mon cadre de travail. C'est ce fil conducteur qui m'a permis de décroiser les différents champs d'investigation vers lesquels j'ai été poussé par curiosité et par envie de mobilité.

La suite du document détaille les résultats les plus marquants de mon activité scientifique. La variété des applications qui ont été les cibles de ma recherche en Traitement du Signal, n'offre pas un cadre d'étude monolithique dans lequel j'aurais obtenu continûment un corpus de résultats. Une présentation chronologique de mon activité n'aurait pas plus de sens ici, dans la mesure où je me suis souvent détourné, puis recentré sur les mêmes thèmes de recherche. Il m'a finalement paru plus lisible de présenter mes contributions scientifiques en les regroupant par type de résultats obtenus. La suite du document s'organise donc en 3 chapitres techniques, sans implications directes entre eux, se rejoignant uniquement à travers les contextes de travail décrits dans l'Introduction.

Ainsi, les « outils » présentés aux chapitres 2 et 3 correspondent à des objets mathématiques originaux définis soit dans l'unique but d'enrichir un champ théorique – essentiellement celui des distributions temps-fréquence – soit pour répondre à un besoin spécifique d'analyse lié à une application – conduisant alors principalement à des estimateurs et à des modèles.

Enfin, le chapitre 4 nous mènera sur les terrains de deux applications distinctes, mais d'égal intérêt. Il sera cependant facile de voir que mon activité sur l'étude des réseaux étant plus aboutie que celle du rythme cardiaque, celle-ci cristallise d'avantage d'innovations théoriques et valorise mieux mes contributions scientifiques. C'est aussi le créneau qui, dans mon cas et à terme, me paraît le plus porteur.

Toutes ces zones d'activité sont des pépinières vivaces pour le développement du Traitement du Signal, et réciproquement, illustrent avec force les bénéfices de celui-ci en réponse aux problèmes sociétaux, industriels ou économiques. Sans avoir été pionnier dans aucun d'entre eux, ces sujets constituent le support actuel (parfois récurrent) de mon activité, et c'est sur tout ou partie de ceux-

ci que je souhaite ancrer ma recherche en Traitement du Signal, en l'inscrivant dans le contexte plus général d'une meilleure compréhension des « systèmes complexes ». Le chapitre 4 me donnera l'occasion de revenir sur cette idée, et de positionner mon périmètre de compétences dans cette perspective.

Chacun des points traités dans ces trois chapitres s'appuie sur un ou plusieurs articles parus dans des revues spécialisées. En reformuler le contenu scientifique n'apporterait pas d'éclairage radicalement nouveau, et ne révélerait d'autre dénominateur commun que le fond méthodologique – que par ailleurs j'espère inscrit en filigrane de tout le document. J'ai donc jugé cet exercice dispensable, et préféré reproduire, moyennant quelques adaptations au contexte, les publications originales. En contrepartie, je m'efforcerai de porter un regard critique sur l'intérêt que ces travaux ont pu susciter dans les différentes communautés concernées. Comment, si c'est le cas, ont-ils contribué à faire avancer la connaissance dans le domaine ? Quelle suite leur a-t-il été donnée, ou au contraire sur quelle impasse ont-ils débouché, et pourquoi ? Avec le recul, je tenterai de mesurer la résonance de mes travaux dans le milieu du Traitement du Signal. J'essaierai alors de relativiser la qualité des solutions proposées, en comparant leur impact à celui d'autres approches plus populaires, car soit plus efficaces, soit plus retentissantes, ou parfois les deux. . .

Cette mise en perspective de mes résultats m'aidera au chapitre 5, à dégager les orientations que je veux renforcer et celles que je souhaite voir émerger dans la suite de mon activité de recherche. J'y exposerai les enjeux scientifiques qui m'interpellent et qui couvrent les domaines de compétence auxquels je me suis formé.

2

Représentations Temps-Fréquence

2.1 Introduction

Une part importante de mes travaux a trait à la définition de nouveaux objets temps-fréquence et temps-échelle, ainsi qu'à leur rôle dans la résolution de problèmes d'estimation ou de détection. Pour situer le cadre et donner du relief aux résultats qui seront repris en détail par la suite, ce paragraphe introduit les concepts et pose quelques notations générales. Largement inspiré d'une présentation tutorielle faite aux deuxièmes ateliers sur l'analyse d'ondes gravitationnelles (Université d'Orsay, 1997), le modèle newtonien d'un type particulier de ces ondes sert de propos pour illustrer les différentes représentations introduites. La référence à un certain nombre d'ouvrages très complets consacrés à ces sujets [Flandrin, 1998, Mallat, 1998, Daubechies, 1992, Cohen, 1995, Hlawatsch and Auger, 2008] enrichira ce survol par un exposé holistique et approfondi de ces théories.

2.1.1 Descriptions duales en temps et en fréquence d'un signal

La représentation de Shannon. La représentation temporelle est certainement la façon la plus intuitive de décrire un signal x ¹

$$x(t) = \int x(u) \delta(u - t) du = \langle x, \delta_t \rangle, \quad (2.1)$$

où $\delta(t)$ désigne la distribution de Dirac, localisée en $t = 0$. Cette représentation, aussi appelée développement de Shannon, correspond à la décomposition de l'observation x sur une base continue générée par l'ensemble des versions translatées d'une fonction d'analyse parfaitement localisée en temps. Bien souvent donc, cette description « morphologique » du signal porte la signature du système physique à l'origine de l'observation. C'est par exemple le cas de la figure 2.1(a), qui montre la forme d'onde gravitationnelle qu'émettrait théoriquement, sous approximations newtoniennes, un système astrophysique massif dans le cas de la coalescence de deux binaires. Ce modèle correspond à la partie réelle du signal complexe suivant [Thorne, 1987] :

$$x(t; t_0, d) = A(t_0 - t)^{-\alpha} e^{-i2\pi d(t_0 - t)^\beta} U(t_0 - t), \quad (2.2)$$

où U est la fonction échelon de Heaviside, $\alpha = 1/4$, $\beta = 5/8$, t_0 l'instant de coalescence et d , A des constantes physiques. Cet exemple montre que l'amplitude instantanée de l'enveloppe du signal croît au cours du temps, alors que la période des oscillations tend vers zéro lorsqu'on approche l'instant t_0 de coalescence.

1. Par soucis de simplicité, on ne considérera dans cette introduction que les espaces d'observation (ou espaces directs) décrits par la variable « temps », étant entendu que les objets traités peuvent se transposer, avec plus ou moins de difficulté, à d'autres variables de description (variable d'« espace » par exemple) ainsi qu'aux signaux multi-dimensionnels. De plus, on se place dans le cadre des espaces vectoriels de Hilbert $L^2(T, dt)$ des signaux à énergie finie, i.e. $\int_0^T |x(u)|^2 du < +\infty$, où T est l'intervalle d'observation, et dt la mesure de Haar associée.

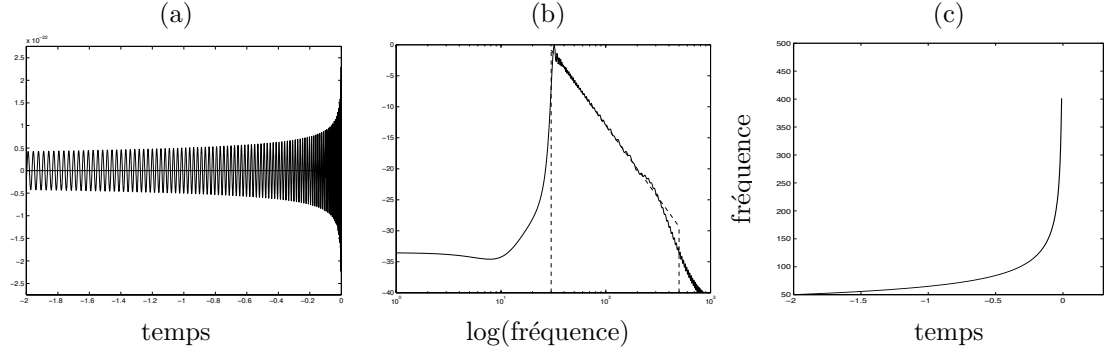


FIGURE 2.1 – Représentations d’une onde gravitationnelle (2.2). (a) Représentation temporelle (Shannon). (b) Densité spectrale d’énergie : module carré de la transformée de Fourier (axes en *Log-Log*). (c) Trajectoire de la fréquence instantanée dans le plan temps (abscisses) fréquence (ordonnées).

La représentation de Fourier. Une autre façon, moins triviale mathématiquement mais tout aussi incontournable de représenter le signal x est donnée par sa transformée de Fourier :

$$X(f) = \int x(u) e^{-i2\pi fu} du = \langle x, e_f \rangle. \quad (2.3)$$

La représentation ainsi obtenue, décompose le signal sur une base de fonctions harmoniques pures $\{e_f\}_{f \in \mathbb{R}}$ qui, contrairement aux impulsions de Dirac de la représentation de Shannon, sont de support temporel infini, mais parfaitement localisées (infiniment concentrées) dans le domaine des fréquences, $\langle e_{f_0}, e_f \rangle = \delta(f - f_0)$. La fonction (généralement complexe) $X(f)$ fournit une description harmonique permettant de livrer le contenu spectral du signal x : le module carré $|X(f)|^2$, appelé densité spectrale d’énergie (ou spectre en version courte), mesure la contribution énergétique de chaque composante harmonique e_f , alors que le spectre de phase $\text{Arg}(X(f))$, indique comment toutes ces composantes se synchronisent entre elles. Mais l’intérêt mathématique de la représentation de Fourier tient en très grande partie à l’existence d’une transformée de Fourier inverse

$$x(t) = \int X(f) e^{+i2\pi ft} df, \quad (2.4)$$

qui établit une correspondance biunivoque entre les représentations temporelle et fréquentielle d’un même signal. De plus, dans ce cadre de travail hilbertien, la relation de Parseval étend cette équivalence unitaire à une relation d’isométrie entre produits scalaires (et normes associées)

$$\langle x, y \rangle = \int x(t) y^*(t) dt = \int X(f) Y^*(f) df. \quad (2.5)$$

Reprenant l’exemple des ondes gravitationnelles, par un argument de phase stationnaire, on peut approcher l’intégrale de Fourier (2.3) et obtenir l’expression fréquentielle duale de (2.2)

$$X_{r,k}(f) = C f^{-(r+1)} e^{i\Psi_k(f)} U(f), \quad (2.6)$$

où $\Psi_k(f) = -2\pi (cf^k + t_0 f + \chi)$, $k = -5/3$ et $r = 1/6$. Le spectre correspondant, représenté à la figure 2.1(b), s’il montre bien le support et la décroissance en loi de puissance des contributions énergétiques spectrales, masque totalement les non-stationarités contenues dans le signal temporel 2.1(a), marquant en cela la complémentarité des deux représentations.

Fréquence instantanée. Afin de décrire fonctionnellement les non-stationarités du contenu fréquentiel d’un signal x , la fréquence instantanée est définie à partir du signal analytique associé²

2. Le signal analytique associé à un signal réel x , est un signal complexe $x_a(t) = x(t) + i\mathcal{H}x(t)$, où $\mathcal{H}$ est la transformation de Hilbert. En fréquence, cette relation s’écrit plus simplement $X_a(f) = 2X(f)U(f)$.

$x_a(t) = A(t)e^{i\varphi(t)}$ selon :

$$f_x(t) = \frac{1}{2\pi} \frac{d\varphi(t)}{dt}. \quad (2.7)$$

Dans le cas de signaux simples, c'est à dire de signaux ne présentant d'énergie qu'autour d'une seule fréquence à chaque instant, cette définition suffit à caractériser localement en temps l'évolution de leur contenu spectral. En ce sens, qui n'est pas celui des distributions d'énergie, le graphe de la fonction f_x suivant t , peut s'interpréter comme une signature conjointe en temps et en fréquence du signal x .

Retard de groupe. Inversement, le retard de groupe est une notion duale de celle de fréquence instantanée, qui vise à associer à chaque fréquence présente dans le bilan énergétique du signal, un temps caractéristique. On le définit à partir du spectre de phase de $X(f) = B(f)e^{i\Psi(f)}$, par la relation suivante :

$$t_x(f) = -\frac{1}{2\pi} \frac{d\Psi(f)}{df}. \quad (2.8)$$

D'interprétation moins intuitive, le retard de groupe n'est en général pas la simple fonction réciproque de la fréquence instantanée. Mais dans le cas des signaux à grand produit durée $\times$ bande passante, l'approximation $t_x(f_x(t)) \simeq t$ devient valable, ce qui permet de considérer indifféremment l'une ou l'autre de ces fonctions [Flandrin, 1998].

Appliqué à l'expression fréquentielle (2.6) des ondes gravitationnelles, le calcul du retard de groupe donne $t_x(f) = t_0 + ckf^{k-1}$, un monôme duquel est tiré le terme modulation (ou *chirp*) en loi de puissance pour désigner ce type de signal. Si alors on suppose que ces signaux sont à large produit durée $\times$ bande passante, on peut en déduire l'expression de la fréquence instantanée, tracée en figure 2.1(c). Cette trajectoire hyperbolique dans le plan temps-fréquence traduit de façon très perceptible la nature non-stationnaire des ondes gravitationnelles et permet de formaliser la loi d'accélération des oscillations au voisinage de l'instant de coalescence des binaires.

Il faut pourtant bien souligner les limites de ce type d'approches temps-fréquence, en rappelant que l'exploitation de la fréquence instantanée ou du retard de groupe n'est possible que dans le cas de signaux simples à composante unique. Or, cette condition est rarement satisfaite par les signaux issus de systèmes réels. Souvent plus complexes, les observations sont entachées de bruit additif large bande, rendant l'évaluation des définitions algébriques (2.7) et (2.8) numériquement très instables. Ce sont dans une large mesure, ce cadre d'analyse restreint et ce manque de robustesse qui ont amené une large communauté de chercheurs (du traitement du signal, des mathématiques, de la physique...) à élaborer des classes de représentations temps-fréquence plus lisibles et vérifiant un corpus de propriétés théoriques souhaitées.

2.1.2 Représentations atomiques temps-fréquence

Comme les représentations de Shannon (2.1) et de Fourier (2.3), les représentations atomiques reposent sur une décomposition linéaire du signal sur une famille de fonctions élémentaires d'analyse (appelées aussi « atomes »), via le produit scalaire suivant :

$$\Gamma_x(t, f; g) = \langle x, g_{t,f} \rangle, \quad \forall t \in T, \forall f \in \mathbb{R}. \quad (2.9)$$

Mais elles se distinguent des représentations à une seule variable, par la nature de la fonction d'analyse g qui, contrairement à la distribution de Dirac ou à l'harmonique complexe, doit posséder de bonnes propriétés de localisation à la fois en temps et en fréquence. L'ensemble des atomes d'analyse $\{g_{t,f}\}$ est alors obtenu par réplication du prototype g autour de chaque position $(t, f) \in (T \times \mathbb{R})$ du plan temps-fréquence³. Chacune des répliques $g_{t,f}$ résulte de l'action combinée d'un couple d'opérateurs de déplacement en temps et en fréquence sur la fonction g – en lieu et place des simples translations en temps ou en fréquence, définissant respectivement (2.1) et (2.3).

Bien que très peu contraintes dans leur principe, tout choix arbitraire des opérateurs de déplacement ou de la fonction d'analyse, ne conduit pas invariablement à la définition d'une

3. Il est possible sous certaines conditions, de restreindre l'action des opérateurs de déplacement à un maillage discret du plan temps-fréquence, et d'obtenir ainsi des décompositions discrètes.

décomposition atomique intéressante du point de vue de l'analyse des signaux. En outre, si l'on souhaite que la représentation $|\Gamma(t, f)|^2$ s'identifie à une distribution de l'énergie du signal analysé, il faut imposer une contrainte d'uni-modularité sur l'atome g . Il en est de même pour toute autre propriété de la décomposition, qui se traduit par une contrainte structurelle sur la fonction d'analyse ou sur le choix des opérateurs de déplacement [Cohen, 1995, Flandrin, 1998]. Mais pour donner tout son sens à la notion de décomposition temps-fréquence, il faut que celle-ci produise une image la plus fidèle possible de la façon dont l'énergie du signal s'organise et se déploie dans le plan temps-fréquence. En particulier, pour ne pas trop dénaturer la localisation des composantes du signal, il faut que l'atome g retenu ait des supports temporels et fréquentiels équivalents les plus étroits possibles, tout en étant conjointement soumis au principe d'incertitude de Gabor-Heisenberg, $\Delta t_g \Delta f_g \geq 1/4\pi$. Il en résulte alors une limitation importante des approches linéaires, puisque bornée inférieurement, l'aire occupée par g exclue toute décomposition atomique qui posséderait des résolutions en temps et en fréquence simultanément et arbitrairement grandes.

Détaillons maintenant les cas de deux décompositions offrant un intérêt particulier.

Transformée de Fourier à court-terme. Cette variante combine les opérateurs de translation en temps et de translation en fréquence, utilisés séparément dans les représentations de Shannon (2.1) et de Fourier (2.3) :

$$\Gamma_x(t, f; g) = \int x(u) g^*(u - t) e^{-i2\pi f u} du, \quad (t, f) \in (T \times \mathbb{R}). \quad (2.10)$$

Dans cette expression, l'atome g peut prendre n'importe quelle forme de fonction passe-bas (e.g. fonction gaussienne). Le choix particulier de ce couple d'opérateurs de déplacement, correspond à un pavage uniforme du plan temps-fréquence, schématisé dans sa version discrète à la figure 2.2(c).

Le spectrogramme, densité d'énergie égale au module carré de la transformée de Fourier à court-terme (2.10), est calculé sur notre exemple d'onde gravitationnelle et représenté à la figure 2.3(a). On voit clairement que l'énergie contenue dans le signal se distribue autour de la trajectoire de la fréquence instantanée (cf. figure 2.1(c)), sans toutefois atteindre une concentration entièrement satisfaisante dans les deux directions (en l'occurrence, l'atome choisi g privilégie d'avantage la résolution temporelle).

Transformée en ondelettes (continue). L'opérateur de translation en temps joue un rôle central dans l'étude des systèmes linéaires convolutifs. En le couplant, non plus aux translations en fréquence, mais à un opérateur de changement d'échelle (dilatation-compression), on obtient une décomposition atomique d'un tout autre type qui, singulièrement, a ouvert des perspectives remarquables en mathématiques (théorie des opérateurs, analyse vectorielle, fonctionnelle et fractale), en détection, estimation, compression ou encore codage de données... Proposée dans [Grossmann and Morlet, 1984], la transformée (continue) en ondelettes,

$$\Gamma_x(t, f; \psi) = \int x(u) \left(\frac{f}{f_0}\right)^{\frac{1}{2}} \psi^* \left(\frac{f}{f_0}(u - t)\right) du, \quad (t, f) \in (T \times \mathbb{R}_+^*), \quad (2.11)$$

projette le signal analysé sur les versions translatées en temps et changées d'échelle de l'« ondelette mère » ψ (ou atome), une fonction oscillante centrée sur la fréquence f_0 . Plus formellement, si l'on souhaite pouvoir reconstruire le signal initial à partir de sa transformée en ondelettes (sans perte d'information), il faut que celle-ci soit inversible, et pour cela l'ondelette ψ doit satisfaire la condition dite d'admissibilité :

$$\int \frac{|\Psi_0(\nu)|^2}{|\nu|} d\nu = 1, \quad (2.12)$$

où Ψ désigne la transformée de Fourier de ψ . La nature oscillante de l'ondelette $\Psi(0) = \int \psi(t) dt = 0$ est corollaire de cette condition d'admissibilité, et peut s'étendre à une propriété plus contraignante sur les monômes d'ordres supérieurs :

$$\int t^k \psi(t) dt = 0, \quad \forall 0 \leq k < n_\psi, \quad (2.13)$$

où n_ψ désigne le nombre de moments nuls (ou régularité) de l'ondelette mère.

Contrairement à l'opérateur de translation (en fréquence), l'opérateur de changement d'échelle n'est pas homomorphique, et son action ne laisse donc pas invariante la « forme » de l'ondelette mère. Il en résulte une géométrie particulière du pavage temps-fréquence associé aux décompositions en ondelettes ; dit à facteur de surtension constant, ce partitionnement du plan est symbolisé par la grille dyadique de la figure 2.2(d). Ainsi, en adaptant résolution temporelle et résolution spectrale à la fréquence analysée, la transformée en ondelettes permet de localiser avec précision la durée de composantes à variations rapides (caractéristiques de transitoires courts), et de mesurer avec finesse la bande passante des composantes plus basse fréquence (représentatives de phénomènes lents, tels que les dérives). Ce comportement est très bien illustré par la figure 2.3(b) représentant le module carré de la transformée en ondelettes (densité d'énergie usuellement appelée scalogramme) calculée sur notre exemple d'onde gravitationnelle. En la comparant à celle du spectrogramme 2.3(a), on remarquera la meilleure concentration énergétiques sur la partie basse fréquence de la modulation hyperbolique, alors que la résolution temporelle aux hautes fréquences reste sensiblement la même.

Avant de poursuivre sur des alternatives bilinéaires aux décompositions atomiques, il faut rendre justice à un argument théorique plus profond qui a valu aux décompositions en ondelettes de surpasser en efficacité et en popularité les transformées de Fourier à court-terme. Pour des raisons évidentes de dimensionnalité, les décompositions linéaires continues ne peuvent pas établir de correspondance biunivoque entre l'espace des signaux et l'espace à deux dimensions des représentations temps-fréquence⁴. Échantillonner le plan temps-fréquence pour obtenir un maillage plus « lâche » et donc moins redondant, se heurte dans le cas des transformées de Fourier à court-terme, à l'obstruction dite de Balian-Law, selon laquelle il n'existe pas de base orthogonale discrète construite sur le principe de la relation (2.10), et telle que l'atome g soit simultanément localisé en temps et en fréquence. Il en va tout autrement pour les ondelettes qui, reposant sur un groupe algébrique différent, permettent de lever ce verrou théorique. Ainsi, sous certaines conditions, les décompositions en ondelette peuvent non seulement engendrer des bases orthogonales à supports compacts [Daubechies, 1992], mais également être associées à des schémas d'analyse multirésolution particulièrement performants du point de vue de leur implémentation [Mallat, 1998].

Analyse multirésolution. Soit le couple $\{\phi, \psi\}$, engendrant une analyse multirésolution de l'espace $L^2(\mathbb{R})$ des signaux de carré intégrable. L'ensemble $\{\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), k = \dots, -1, 0, 1, \dots\}$ et l'ensemble $\{\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k), k = \dots, -1, 0, 1, \dots\}$ sont respectivement des bases orthogonales pour l'espace d'approximation $V_j \subset L^2(\mathbb{R})$, et pour l'espace des détails $W_j \subset L^2(\mathbb{R})$, de plus $V_{j+1} = V_j \oplus W_j$, et $W_k \perp W_j, \forall k \neq j$. Ainsi, tout signal $x \in L^2(\mathbb{R})$ peut se décomposer par projection orthogonale, sous forme d'une somme du type :

$$\begin{aligned} s(t) &= \sum_{k=-\infty}^{k=\infty} \langle s, \phi_{Jk} \rangle \phi_{Jk}(t) + \sum_{j=J}^{j=\infty} \sum_{k=-\infty}^{k=\infty} \langle s, \psi_{jk} \rangle \psi_{jk}(t) \\ &= A_J(t) + \sum_{j=J}^{j=\infty} D_j(t), \end{aligned} \quad (2.14)$$

où $A_J(t)$ est l'approximation de x à l'échelle $a_J = 2^{-J}$ (formellement $a \equiv f^{-1}$), et $D_j(t)$ le détail à l'échelle $a_j = 2^{-j}$.

Percée majeure en analyse harmonique, les ondelettes orthogonales lorsqu'elles sont adossées à une analyse multirésolution sont également un outil remarquable d'analyse numérique. Il est en effet possible de déduire du système $\{\phi, \psi\}$ un banc de filtres miroirs conjugués ($g[\cdot], h[\cdot]$) permettant une implémentation pyramidale extrêmement efficace (coût algorithmique en $O(N)$) des transformées en ondelettes discrètes ainsi que de leur inversion. Cette performance algorithmique, indissociable des propriétés mathématiques des ondelettes, a été déterminante dans l'importance et la notoriété que les analyses multirésolutions ont gagnées dans un grand nombre d'applications.

4. Pour être plus précis, il existe bien une correspondance entre l'espace des signaux et un sous-espace de Hilbert à noyau auto-reproduisant auquel appartiennent les décompositions atomiques.

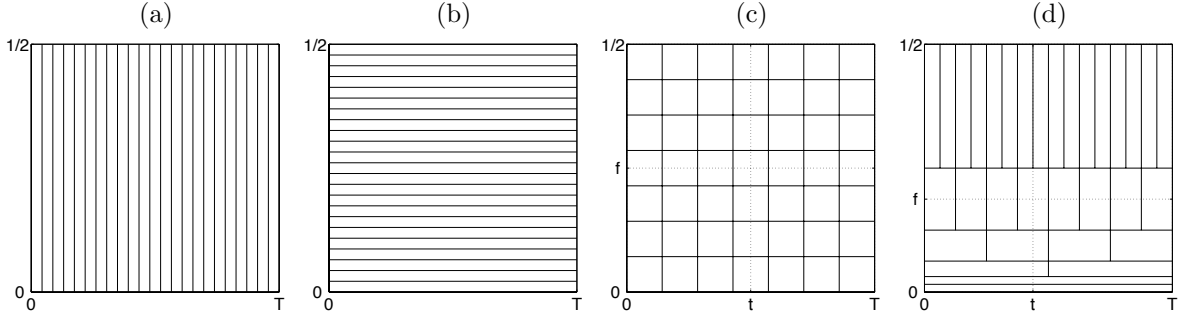


FIGURE 2.2 – Pavage Temps-Fréquence symbolisant différentes représentations d'un signal. L'axe des abscisses correspond au temps, celui des ordonnées à la fréquence. (a) Shannon. (b) Fourier. (c) Fourier à court-terme. (d) Ondelettes.

Toutefois, transformée de Fourier à terme ou décomposition en ondelettes (continues comme discrètes), quels que soient les opérateurs de déplacement retenus, et quel que soit la forme de l'atome g choisi, toutes les représentations construites sur le principe de projection linéaire (2.9) se rapportent à un type particulier de « pixélisation » du plan temp-fréquence. L'aire d'un pixel ne pouvant être arbitrairement petite (principe d'incertitude de Gabor-Heisenberg), accroître la résolution dans une direction ne peut se faire qu'au détriment de la résolution dans l'autre.

2.1.3 Distributions bilinéaires temps-fréquence

Les distributions bilinéaires (ou quadratiques) temps-fréquence sont une généralisation des décompositions atomiques, ou plus exactement du module carré de celles-ci :

$$|\Gamma_x(t, f; g)|^2 = \iint x(u) x^*(v) g_{t,f}(u) g_{t,f}^*(v) du dv, \quad (2.15)$$

où le produit $g_{t,f}(u) g_{t,f}^*(v)$ est remplacé par un noyau arbitraire plus général $K(u, v; t, f)$. La définition canonique qui s'en suit [Cohen, 1995, Flandrin, 1998] :

$$\rho_x(t, f; K) = \iint x(u) x^*(v) K(u, v; t, f) du dv, \quad (2.16)$$

offre une grande flexibilité dans le choix de K , choix que l'on sait analytiquement contraindre aux propriétés théoriques que l'on souhaite imposer à la distribution bilinéaire $\rho_x(t, f; K)$. Par exemple, pour que $\rho_x(t, f; K)$ vérifie la propriété de conservation d'énergie :

$$\iint \rho_x(t, f; K) dt df = \int |x(t)|^2 dt = \int |X(f)|^2 df = E_x, \quad (2.17)$$

et s'identifie ainsi à une distribution énergétique du signal, le noyau correspondant K doit respecter la condition marginale suivante (et que nous supposons implicite dans ce qui suit) :

$$\iint K(u, v; t, f) dt df = \delta(u - v). \quad (2.18)$$

Nous avons déjà eu l'occasion de souligner l'importance de l'opérateur de translation en temps, qui combiné à un autre opérateur de déplacement en fréquence (e.g., translation ou changement d'échelle) permet de définir la transformée de fourier à court-terme ou la décomposition en ondelettes. On peut donc tout aussi bien subordonner la définition (2.16) des distributions bilinéaires à la commutativité du diagramme suivant :

$$\begin{array}{ccc} x(t) & \longrightarrow & Tx(t) \\ \downarrow & & \downarrow \\ \rho_x(t, f; K) & \longrightarrow & \rho_{Tx}(t, f; K) = \tilde{T}\rho_x(t, f; K), \end{array} \quad (2.19)$$

où T est un opérateur de déplacement du signal dans le plan temps-fréquence, et $\tilde{T}$ son dual agissant sur la représentation. En fait, c'est cette contrainte qui est traditionnellement utilisée comme axiome pour réduire la classe très générale des distributions bilinéaires (2.16), à des sous-classes de représentations temps-fréquence dites covariantes. En précisant ce principe dans le cas de l'opérateur de déplacement temps-fréquence (des transformées de Fourier à court-terme), puis à celui de l'opérateur temps-échelle (des décompositions en ondelette), nous rappelons ci-dessous les deux seules classes irréductibles⁵ de distributions bilinéaires temps-fréquence : la classe de Cohen et la classe affine.

Classe de Cohen. Bien que ce ne soit pas la seule caractérisation possible, on peut définir la classe de Cohen des représentations temps-fréquence, comme l'ensemble des distributions bilinéaires covariantes par les opérateurs de translation en temps et de translation en fréquence :

$$\begin{array}{ccc} x(t) & \longrightarrow & x(t - t_0) e^{i2\pi f_0 t} \\ \downarrow & & \downarrow \\ \rho_x(t, f; K) & \longrightarrow & \rho_x(t - t_0, f - f_0; K). \end{array} \quad (2.20)$$

Traduite en termes de contraintes sur l'opérateur intégral à noyau (2.16), cette propriété impose alors que tout élément de cette classe puisse s'écrire sous la forme convolutive suivante :

$$C_x(t, f; \Pi) = \iint W_x(u, \theta) \Pi(u - t, \theta - f) du df, \quad (2.21)$$

entre la distribution de Wigner-Ville :

$$W_x(t, f) = \int x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-i2\pi f\tau} d\tau, \quad (2.22)$$

$$= \int X^*\left(f + \frac{\nu}{2}\right) X\left(f - \frac{\nu}{2}\right) e^{i2\pi t\nu} d\nu, \quad (2.23)$$

et Π , un noyau bi-dimensionnel arbitraire de paramétrisation. La distribution de Wigner-Ville est un pivot central (générateur) de la classe de Cohen, et vérifie un certain nombre de propriétés théoriques qui, selon le choix des noyaux Π , peuvent se propager à d'autres membres de la classe. Parmi ces propriétés, la localisation stricte sur des modulations linéaires (ou *chirps*) de la forme $x(t) = \exp\{i2\pi(\beta t^2/2 + f_0 t)\}$ est tout à fait remarquable et spécifique aux formes corrélatives définissant la distribution de Wigner (2.22)-(2.23). Comme un filtre adapté dans le plan temps-fréquence, la distribution de Wigner prend la forme diagonale suivante $W_x(t, f) = \delta(f - (f_0 + \beta t))$, réduisant ainsi le support de la densité à la trajectoire rectiligne de la fréquence instantanée du *chirp* linéaire. Une telle concentration énergétique contraste singulièrement avec les mosaïques temps-fréquence des décompositions atomiques, mais n'est pas sans un certain coût. Pour toute déviation au cas du *chirp* linéaire isolé, comme par exemple celui de composantes multiples, ou celui de la fréquence instantanée hyperbolique du modèle (2.2) d'onde gravitationnelle, la distribution de Wigner-Ville développe un réseau complexe d'interférences correspondantes aux termes croisés de la forme quadratique. La figure 2.3(c) illustre bien cette dualité : fine localisation de la composante propre sur sa trajectoire de fréquence instantanée, et coexistence de termes oscillants délocalisés, de contribution énergétique nulle. Un des grands défis de la classe de Cohen, a donc été de concevoir des noyaux de convolution Π dont la géométrie particulière soit capable de préserver la bonne localisation des composantes tout en lissant les zones interférentielles. Comme on le verra aux paragraphes suivants, une partie de mon activité a justement été consacrée à l'élaboration de telles solutions, mais avant cela, précisons que le spectrogramme est lui-même un élément de la classe de Cohen, traitant le problème des interférences sous l'angle extrême des noyaux passe-bas à variables non séparables. Pour s'en convaincre, il faut souligner une autre propriété intéressante de la distribution de Wigner-Ville, appelée « unitarité », elle traduit la préservation du produit

5. La covariance selon tout autre choix d'opérateur de déplacement du plan temps-fréquence conduit à une classe unitairement équivalente soit à la classe de Cohen, soit à la classe affine.

scalaire de $L^2(\mathbb{R})$ vers $L^2(\mathbb{R}^2)$, selon l'égalité de Moyal⁶ :

$$\left| \int x(t) y^*(t) dt \right|^2 = \iint W_x(t, f) W_y(t, f) dt df. \quad (2.24)$$

Combinée au principe de covariance (2.20), cette propriété d'unitarité transcrit en algèbre temps-fréquence, toute opération convolutive dans le domaine temporel ou fréquentiel. En particulier donc, il devient trivial de montrer que le spectrogramme (module carré de la transformée de Fourier à court-terme (2.10)) est un cas spécial de (2.21), pour lequel $\Pi(t, f) = W_g(t, f)$,

$$SP_x(t, f; g) := |\Gamma_x(t, f; g)|^2 = \iint W_x(u, \theta) W_g(u - t, \theta - f) du d\theta. \quad (2.25)$$

Classe affine. Revenons au principe général de covariance du schéma (2.19), et réécrivons le dans le cas de l'opérateur de translation en temps, associé cette fois-ci à celui de dilatation/compression. La classe affine (ou **classe des distributions temps-échelle**) désigne l'ensemble des distributions bilinéaires (2.16) covariantes par changement d'horloge, c'est à dire vérifiant la commutativité du diagramme suivant :

$$\begin{array}{ccc} x(t) & \longrightarrow & |a|^{-1/2} x\left(\frac{t-t_0}{a}\right) \\ \downarrow & & \downarrow \\ \rho_x(t, f; K) & \longrightarrow & \rho_x\left(\frac{t-t_0}{a}, af; K\right). \end{array} \quad (2.26)$$

Pour toute distribution de Wigner affine $P^{(k)}(t, f)$ covariantes selon ce principe, il est démontré dans [Bertrand and Bertrand, 1992] qu'il existe un réel k et une fonction de paramétrisation continue non nulle μ_k , tels que :

$$P_x^{(k)}(t, f) = f \int \mu_k(u) X(f\lambda_k(u)) X^*(f\lambda_k(-u)) e^{i2\pi t f \zeta_k(u)} du, \quad (2.27)$$

où

$$\begin{cases} \lambda_k(u) = (k(e^{-u} - 1)/(e^{-ku} - 1))^{1/(k-1)}, \forall k \neq 0, 1 \\ \lambda_0(u) = u/(1 - e^{-u}) \\ \lambda_1(u) = \exp(1 + (ue^{-u})/(e^{-u} - 1)) \\ \zeta_k(u) = \lambda_k(u) - \lambda_k(-u), \end{cases} \quad (2.28)$$

et où toutes les fonctions λ_k vérifient la propriété algébrique suivante

$$\lambda_k(-u) = e^{-u} \lambda_k(u). \quad (2.29)$$

La symétrie centrale $f \pm \nu/2$ qui est au cœur de la distribution de Wigner-Ville (2.23), est remplacée par une symétrie généralisée $f\lambda_k(\pm u)$ dans l'expression des distributions de Wigner affines (2.27). Ce ne sont alors plus les droites du plan temps-fréquence qui sont les lieux des points globalement invariants par ces symétries non-linéaires [20], mais un faisceau paramétré de courbes plus complexes d'expression $\{t^{(k)}(f) = t_0 + cf^{k-1}, k \in \mathbb{R}_-\}$. Chacune de ces courbes correspond à la loi de retard de groupe (2.8) d'un signal modulé en fréquence (ou *chirp* en loi de puissance) $X_k(f) = C f^{-\frac{1}{2}} e^{i\Psi_k(f)} U(f)$. Pour chaque valeur de l'exposant de modulation $k \leq 0$, il existe alors une distribution de Wigner affine $P^{(k)}(t, f)$ caractérisée par le choix⁷ :

$$\mu_k(u) = \mu_k^L(u) = \frac{d\zeta_k(u)}{du} (\lambda_k(u) \lambda_k(-u))^{\frac{1}{2}}, \quad (2.30)$$

qui, calculée sur le *chirp* $X_k(f)$, se localise parfaitement sur sa trajectoire $t^{(k)}(f)$, selon :

$$P_{x_k}^{(k)}(t, f) = C^2 f^{-1} \delta(t - t^{(k)}(f)). \quad (2.31)$$

6. Toute application $x \mapsto C_x$ n'est en général pas isométrique, l'égalité (2.24) n'étant vraie pour C_x que si le noyau associé vérifie la condition $|\Phi(\tau, \xi)| = 1$, où Φ désigne la transformée de Fourier de Π .

7. En toute rigueur, les cas $k < 0$ et $k = 0$ sont à traiter distinctement car correspondant à des groupes algébriques différents [Bertrand and Bertrand, 1992].

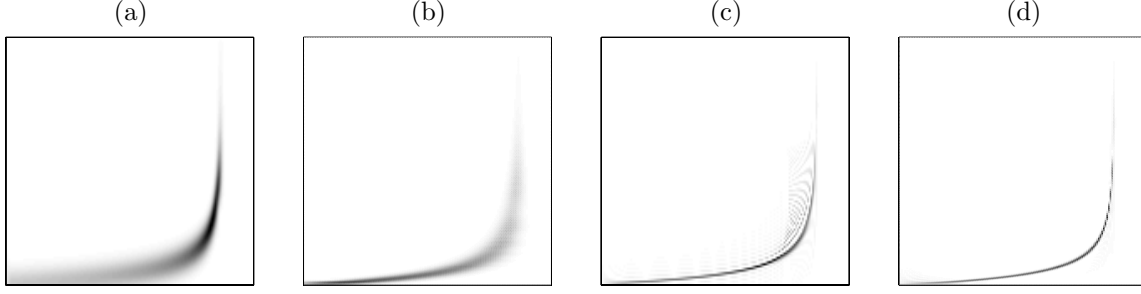


FIGURE 2.3 – Différents types de distributions énergétiques Temps-Fréquence calculées sur l'onde gravitationnelle d'expression (2.2). L'axe des abscisses correspond au temps, celui des ordonnées à la fréquence. (a) Spectrogramme, module carré de la transformée de Fourier à court-term (2.10). (b) Scalogramme, module carré de la décomposition en ondelettes (2.11). (c) Distribution de Wigner-Ville (2.22). (d) Distribution de Wigner affine (2.27) avec $k = -5/3$.

En première approximation, notre modèle d'onde gravitationnelle (2.6) correspond à un *chirp* en loi de puissance, avec un retard de groupe $t_x(f) = t^{(-5/3)}(f) = t_0 + cf^{-5/3-1}$. On vérifie donc bien que la distribution $P^{(-5/3)}(t, f)$ assortie au choix (2.30) de $\mu_{-5/3}(u)$, se localise parfaitement sur la trajectoire de $t^{(-5/3)}(f)$, mais contrairement à la distribution de Wigner-Ville, ne produit pas de termes interférentiels hors du support temps-fréquence de la composante propre du signal.

Toutes les distributions de Wigner affines (2.27) ne sont pas unitaires. À $k \leq 0$ fixé, seule la distribution associée au choix :

$$\mu_k(u) = \mu_k^U(u) = \left(\frac{d\zeta_k(u)}{du} \right)^{\frac{1}{2}} (\lambda_k(u) \lambda_k(-u))^{\frac{1}{2}}, \quad (2.32)$$

vérifie la formule de Moyal (2.24).

Les conditions de localisation (2.30) et d'unitarité (2.32) aboutissent en général à des solutions $\mu_k(u)$ distinctes. Seul le cas remarquable $k = 0$, conduit à l'égalité $\mu_0^U(u) = \mu_0^L(u)$, définissant ainsi l'unique distribution de Wigner affine (2.27), qui soit à la fois unitaire et localisée. Communément appelée distribution unitaire de Bertrand, $P^{(0)}(t, f)$ se localise sur les trajectoires de retards de groupe hyperboliques d'expression $t_x(f) = t^{(0)}(f) = t_0 + cf^{-1}$. Pour toutes les autres valeurs de $k < 0$, bien que les distributions localisées $P^{(k)}(t, f)$ ne soient pas unitaires, on peut toutefois leur adjoindre une forme dite passive,

$$\bar{P}^{(k)}(t, f) = \int \Psi_k(f(t - \tau)) P^{(k)}(\tau, f) d\tau \quad \text{où} \quad \Psi_k(s) = \int e^{i2\pi s \zeta_k(u)} du, \quad (2.33)$$

pour qu'ensemble elles vérifient une relation d'isométrie du type de (2.24)

$$\left| \int_0^{+\infty} X(f) Y^*(f) df \right|^2 = \int_0^{+\infty} \bar{P}_x^{(k)}(t, f) P_y^{(k)}(t, f) df dt. \quad (2.34)$$

La propriété d'unitarité jette un pont entre les décompositions linéaires et les représentations quadratiques d'un même groupe. Dans le cas de la décomposition en ondelettes (2.11) par exemple, combinée au principe de covariance par changement d'échelle (2.26), elle plonge directement le scalogramme dans la classe des représentations temps-fréquence bilinéaires, puisqu'alors :

$$|\Gamma_x(t, f; \psi)|^2 = \int \int_0^{+\infty} P_x^{(0)}(u, \theta) P_\psi^{(0)}\left(\frac{f}{f_0}(u - t), \theta \frac{f_0}{f}\right) d\theta du. \quad (2.35)$$

Cette expression est en effet un cas particulier d'une forme plus générale, équivalente à la définition (2.27), selon laquelle toute distribution de la classe affine peut s'écrire comme un produit de convolution affine entre la distribution unitaire de Bertrand et un noyau arbitraire de paramétrisation

Π , selon :

$$P_x^{(k)}(t, f) = \int \int_0^{+\infty} P_x^{(0)}(u, \theta) \Pi^{(k)}\left(\frac{f}{f_0}(u - t), \theta \frac{f_0}{f}\right) d\theta du. \quad (2.36)$$

Cette dernière expression est à rapprocher d'une autre formulation canonique de la classe affine, proposée dans [Rioul and Flandrin, 1992]. Centrée autour de la distribution de Wigner-Ville – que l'on retrouve calculée sur le signal analytique $2X(f)U(f)$, en posant $k = 2$ dans (2.28) et (2.27) – celle-ci passe par un noyau de paramétrisation $\Pi(t, f)$ différent du couple $(\lambda_k(u), \zeta_k(u))$, mais spécifique à chaque élément Ω de la classe affine selon

$$\Omega_x(t, f; \Pi) = \int \int_0^{+\infty} W_x(u, \theta) \Pi\left(\frac{f}{f_0}(u - t), \theta \frac{f_0}{f}\right) d\theta du. \quad (2.37)$$

De plus, par unitarité de la distribution de Wigner-Ville, cette convolution affine s'écrit dans le cas particulier du scalogramme :

$$SC_x(t, f; \psi) := |\Gamma_x(t, f; \psi)|^2 = \int \int_0^{+\infty} W_x(u, \theta) W_\psi\left(\frac{f}{f_0}(u - t), \theta \frac{f_0}{f}\right) d\theta du. \quad (2.38)$$

S'agissant de propriétés théoriques – telle que la localisation dans le plan temps-fréquence – et dès lors que l'on accepte d'abandonner la simplicité liée aux transformations linéaires, la liberté qu'offre le choix du noyau Π (ou de façon équivalente le couple $(\lambda_k(u), \zeta_k(u))$) ouvre sur un nombre illimité d'alternatives au scalogramme. Pourtant, la popularité des distributions bilinéaires n'a jamais atteint la mesure de leurs ressources. D'une part, parce que leur calcul repose sur la connaissance du signal à analyser dans son entier, elles se prêtent mal à des implémentations algorithmiques en ligne. Ensuite, parce qu'en dehors des structures propres de certains *chirps* modulés en fréquence, leur non linéarité crée un réseau complexe de termes interférentiels qui pénalisent sévèrement l'interprétation des représentations. Si bien qu'en dépit des multiples solutions proposées pour contourner ces obstacles, on constate que le potentiel théorique des distributions énergétiques temps-fréquence a encore aujourd'hui un coût trop élevé, que la plupart des applications en analyse du signal renâcle à payer. Cette « désaffection » pour les représentations quadratiques est néanmoins à dissocier du faux procès qui consisterait à les comparer au succès des transformations en ondelettes. Celles-ci, surtout dans leur version discrète associée à une structure de base, ont séduit la communauté du traitement du signal (et bien d'autres) par leur propriété d'inversibilité et de description parcimonieuse des signaux. Évidemment, ce cadre topologique échappe à la théorie des représentations bilinéaires dont l'objectif, rappelons le, est de fournir une carte temps-fréquence la plus fidèle et précise possible de l'occupation énergétique du signal.

Je pense avoir été avec différents collaborateurs, de ceux qui ont tenté de vulgariser l'usage des représentations temps-fréquence (c'est cet objectif pédagogique de synthèse que nous avons visé dans le chapitre tutoriel [44]). La suite de ce chapitre reproduit, adaptés au contexte, quelques uns de nos travaux ayant débouché sur des solutions originales qui concilient faible complexité algorithmique, propriétés de localisation et suppression des interférences.

2.2 Distributions pseudo-Wigner affines

Ce paragraphe est un large extrait adapté de l'article :

Pseudo Affine Wigner Distributions : Definition and Kernel Formulation

P. GONÇALVES and R. BARANIUK

IEEE Transaction on Signal Processing Volume 46, no. 6, pp. 2305-2314, June 1998

The affine Wigner distributions (2.27) have great potential as flexible tools for time-varying spectral analysis. They possess a number of desirable theoretical properties, including the ability to match a large class of different signal types. Unfortunately, their promise is offset by two major practical limitations. First, the entire signal enters into the calculation of these time-scale distributions (TSD) at every point (t, f) in the time-frequency plane, precluding their online operation with long signals. Second, due to their nonlinearity, interference components arise between each pair of signal components, complicating their interpretation (compare Figures 2.4(a)-(b)) [19, 20]. As a result, few affine Wigner distributions have been applied in practice (aside from in [Bertrand and Bertrand, 1991]).

In this section, we attack these limitations by introducing a set of (smoothed) *pseudo affine Wigner distributions*. These new TSDs offer not only asymptotically the same properties as the affine Wigner distributions, but also support efficient online operation and suppress troublesome cross-components. Our derivation relies on the strong analogy between time-frequency and time-scale analysis and is inspired by the pseudo Wigner distribution.

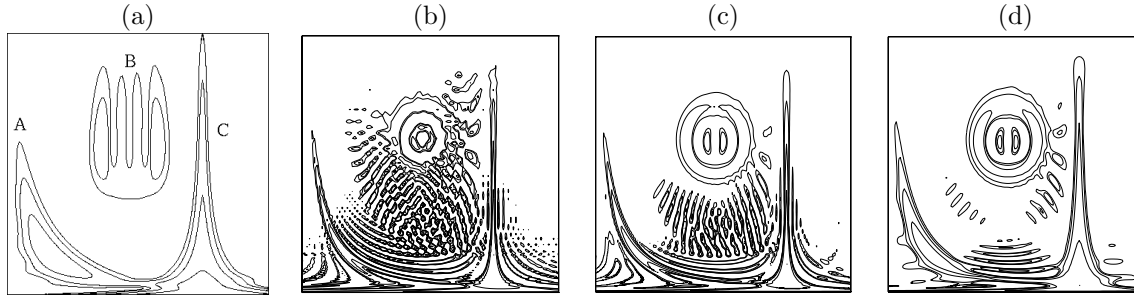


FIGURE 2.4 – Time-scale distributions (TSDs) of a test signal composed of a hyperbolic chirp $X_A(f) = e^{i2\pi\alpha \ln f}$ (component A), a third-order Hermite function (component B), and a Lipschitz singularity $x_C(t) = |t - t_0|^{-0.1}$ (component C). TSDs are plotted in equal-energy contours with the horizontal axis corresponding to time and the vertical axis corresponding to frequency. (a) The scalogram (squared magnitude of the wavelet transform $SC_x(t, f)$), computed using a Morlet wavelet ψ of quality factor $Q = 2$. (b) Unitary Bertrand distribution $P_x^{(0)}(t, f)$. (c) Pseudo Bertrand distribution $\tilde{P}_x^{(0)}(t, f)$ computed using a Morlet wavelet of $Q = 8$. (d) Smoothed pseudo Bertrand distribution computed using the same wavelet and a Gaussian frequency window G of $Q = 1$. The pseudo and smoothed pseudo Bertrand distributions permit a continuous transition between the (low-resolution but interference-free) scalogram and the (high-resolution but interference-ridden) Bertrand distribution.

2.2.1 Definition

Derivation. The well-known pseudo Wigner distribution [Claassen and Mecklenbräuker, 1980] of the Cohen class is obtained by introducing a window function into the Wigner distribution (2.22). An analogous windowing procedure leads to the pseudo affine Wigner distributions. In contrast to the pseudo Wigner case, however, this windowing must be frequency-dependent, to ensure that the resulting TSD remains affine covariant.⁸ As a result, the smoothing in frequency direction is proportional bandwidth, rather than constant-bandwidth as in the pseudo Wigner distribution.

⁸ Rioul and Flandrin consider the same covariance requirements in their definition of the affine pseudo Wigner distribution [Rioul and Flandrin, 1992]. In this paper, we generalize their definition to the entire class of affine Wigner distributions.

We first rewrite the general form (2.27) in the time domain

$$P_x^{(k)}(t, f) = f \int \mu_k(u) \left[\int x(\tau) e^{-i2\pi\lambda_k(u)f(\tau-t)} d\tau \right] \left[\int x(\tau') e^{-i2\pi\lambda_k(-u)f(\tau'-t)} d\tau' \right]^* du. \quad (2.39)$$

It is clear from this expression that at every point (t, f) in the time-frequency plane the affine Wigner distribution depends on the entire signal x . Since online operation requires that we consider the signal only in a sliding interval, we introduce a window function h in (2.39) to obtain

$$\begin{aligned} \tilde{P}_x^{(k)}(t, f) &\equiv f \int \mu_k(u) \left[\int x(\tau) h[f\lambda_k(u)(\tau-t)] e^{-i2\pi\lambda_k(u)f(\tau-t)} d\tau \right] \\ &\times \left[\int x(\tau') h[f\lambda_k(-u)(\tau'-t)] e^{-i2\pi\lambda_k(-u)f(\tau'-t)} d\tau' \right]^* du. \end{aligned} \quad (2.40)$$

The dependence of h on the analysis frequency f guarantees $\tilde{P}^{(k)}$ affine covariance to time shifts and scale changes.⁹ By analogy to the pseudo Wigner distribution, we call these new TSDs *pseudo affine Wigner distributions*.

The pseudo affine Wigner distributions can be formulated in terms of the wavelet transform. Introducing the bandpass wavelet function $\psi(\tau) = h(\tau) e^{i2\pi\tau}$, we can reorder (2.40) as

$$\begin{aligned} \tilde{P}_x^{(k)}(t, f) &= f \int \mu_k(u) \left[\int x(\tau) \psi^*[f\lambda_k(u)(\tau-t)] d\tau \right] \left[\int x(\tau') \psi^*[f\lambda_k(-u)(\tau'-t)] d\tau' \right]^* du \\ &= \int \frac{\mu_k(u)}{\sqrt{\lambda_k(u)\lambda_k(-u)}} \Gamma_x(t, \lambda_k(u)f) \Gamma_x^*(t, \lambda_k(-u)f) du, \end{aligned} \quad (2.41)$$

with Γ_x the wavelet transform (2.11) computed with wavelet ψ . This generalized “matched filter” correlation of the wavelet transform with itself echos the structure of (2.27) with the Fourier transform of the signal X replaced by its wavelet transform Γ_x . Thus, we can identify the pseudo affine Wigner distribution as one of the two fundamental bilinear TSDs derived from the wavelet transform :

The scalogram results from squaring the wavelet transform ; the pseudo affine Wigner distribution results from a generalized self-correlation of the wavelet transform across frequency.

Figure 2.5 illustrates the focusing effect of the generalized self-correlation (2.41) on a time slice of the wavelet transform. To compute the self-correlation at frequency f , we scale and warp $\Gamma_x(t, f)$ to the function $\Gamma_x(t, \lambda_k(u)f)$ and then compute the inner product over u between this function and its reversed twin $\Gamma_x^*(t, \lambda_k(-u)f)$. In contrast with a simple affine correlation, the function is not only scaled, but also reshaped before computing the inner product.

Time-frequency smoothing interpretation. The time windowing introduced in (2.40) acts as a proportional bandwidth frequency smoothing that suppresses interference components oscillating in the frequency direction. Compare, for example, the pseudo Bertrand distribution $\tilde{P}^{(0)}$ of Figure 2.4(c) with the unitary Bertrand distribution $P^{(0)}$ of Figure 2.4(b).

To suppress interference terms oscillating in the time direction, we must smooth in that direction or window the dual variable. The introduction of a low-pass function G in (2.41) limits the integration with respect to u (loosely speaking, the dual variable of the product tf), and thus performs proportional-bandwidth time smoothing of the TSD. We call the resulting time-scale distributions

$$\tilde{P}_x^{(k)}(t, f) = \int G(u) \frac{\mu_k(u)}{\sqrt{\lambda_k(u)\lambda_k(-u)}} \Gamma_x(t, \lambda_k(u)f) \Gamma_x^*(t, \lambda_k(-u)f) du, \quad (2.42)$$

the *smoothed pseudo affine Wigner distributions* (see Figure 2.4(d)).

⁹. Suppressing the $\lambda_k(\pm u)$ in h in (2.40) yields a different distribution with similar covariance properties. However this formulation does not appear to admit an efficient implementation.

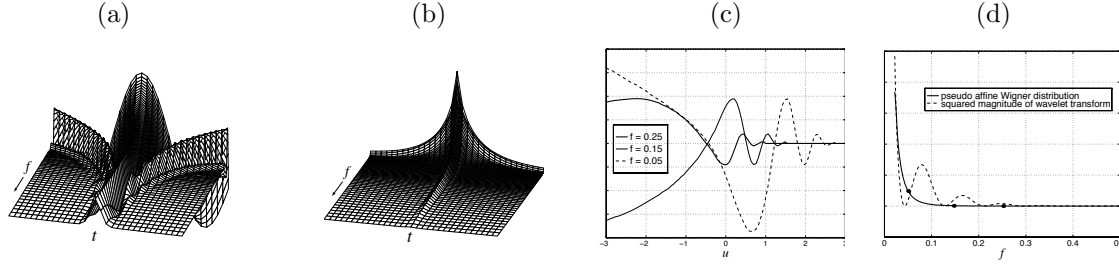


FIGURE 2.5 – Computation of the pseudo affine Wigner distribution from the wavelet transform, emphasizing the focusing effect of the generalized matched correlation underlying (2.41). (a) Wavelet transform (real part) $\Gamma_x(t, f)$ of the singularity (component C) from Figure 2.4. (Note that the frequency axis runs from top to bottom in this mesh plot.) (b) Pseudo affine Wigner distribution $\tilde{P}_x^{(0)}(t, f)$ of the same signal. Consider the computation (2.41) of a time slice $\tilde{P}_x^{(0)}(t_1, f)$ of this distribution in the vicinity of the singularity. (c) Scaled and warped wavelet transform $\Gamma_x(t_1, \lambda_0(u)f)$ plotted versus u , for $f = 0.25$, $f = 0.15$, and $f = 0.05$. (d) Result of the generalized matched correlation $\tilde{P}_x^{(0)}(t_1, f)$ compared to the corresponding time slice of the scalogram $|\Gamma_x(t_1, f)|^2$.

Even though the pseudo and smoothed pseudo affine Wigner distributions are smoothed versions of the affine Wigner distributions, they can still have resolution exceeding that of the scalogram (recall Figure 2.4(a)). This resolution enhancement compared to the scalogram is due precisely to the action of the generalized self-correlation in (2.42) : rather than simply squaring the wavelet transform, we match-filter it.

Implementation. The pseudo affine Wigner distributions can be interpreted as sliding versions of the original affine Wigner distributions, and as a result, they are naturally suited for online operation with long signals. To construct a pseudo affine Wigner distribution, we simply compute the wavelet transform of the signal and then at each time perform the generalized frequency correlation (2.41) or (2.42). The fast Mellin transform is a convenient tool for implementing this correlation efficiently [34], [Ovarlez et al., 1992]. The algorithm runs as follows :

1. Compute the wavelet transform $\Gamma_x(t, f)$ of the signal using wavelet $\psi(\tau) = h(\tau) e^{i2\pi\tau}$. Samples should be spaced uniformly in time and exponentially in frequency.
2. At each time t , for a range of u , rescale $\Gamma_x(t, f)$ to $\Gamma_x(t, \lambda(\pm u)f)$ using the Mellin transform [Ovarlez et al., 1992], which maps scale changes to simple phase shifts. Since the Mellin transform of a function $z(v)$ equals the Fourier transform of $e^{v/2}z(e^v)$, a fast Fourier transform (FFT) applied to an exponentially spaced set of frequency samples of $\Gamma_x(t, f)$ implements a fast Mellin transform.
3. At each time t , compute the inner product (2.41) or (2.42) with respect to u .

Using a fast algorithm for the computation of the wavelet transform [Ovarlez et al., 1992, Jones and Baraniuk, 1991], the computational cost of this procedure is $O(MN \log M)$ for N time and M frequency samples,¹⁰ which is on the same order as the cost for the spectrogram, pseudo Wigner distribution, and scalogram.

Examples.

$k = 2$. Affine pseudo Wigner distribution

In this case, (2.27) reduces to the ordinary Wigner distribution and (2.42) becomes the “affine smoothed pseudo Wigner distribution” of Rioul and Flandrin [Rioul and Flandrin, 1992].

$k = 0$. Pseudo Bertrand distribution

In the particular case of the unitary $P^{(0)}$ distribution, the special form for $\mu_0(u)$ cancels the $\sqrt{\lambda_0(u)\lambda_0(-u)}$ factor in (2.41), leaving us with a much simpler expression for $\tilde{P}^{(0)}$. The result is the pseudo Bertrand distribution of [34].

10. Assuming that the length of the wavelet at maximum dilation is of $O(M)$.

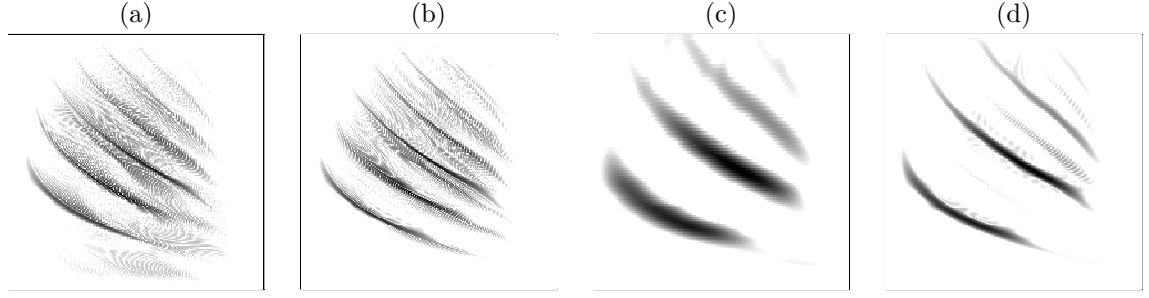


FIGURE 2.6 – TSDs of the echo-location chirp of the large brown bat, *Eptesicus Fuscus*. Horizontal axis corresponds to time and vertical axis to frequency. (a) Wigner distribution $P_x^{(2)}$. (b) Unitary Bertrand distribution $P_x^{(0)}$. (c) Scalogram $|\Gamma_x|^2$. (d) Smoothed pseudo Bertrand distribution $\tilde{P}_x^{(0)}$.

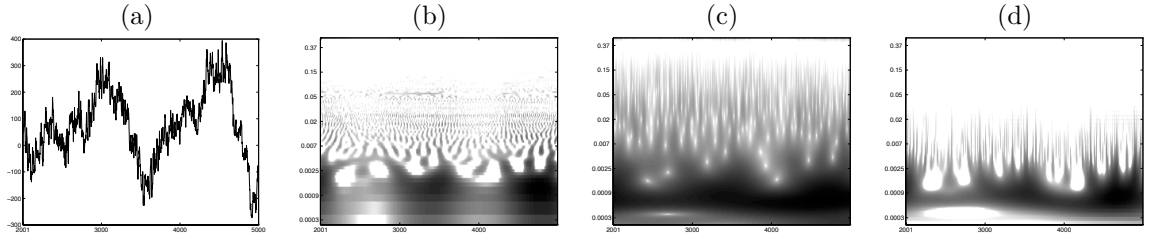


FIGURE 2.7 – Time-varying spectral representations in a machine fault monitoring application. (a) 5000 samples of a time series acquired from an accelerometer axially mounted on a Main Condensate Pump rotating at 892 rpm. The sampling rate was 50 kHz. (b) Pseudo Wigner TFD $\tilde{W}_x$ computed using a 1024-point sliding window. This Cohen's class TFD is not scale covariant, hence it cannot reveal both the low frequency rotational components and the high frequency transients. (Horizontal axis corresponds to time, vertical axis to normalized frequency in a logarithmic scale.) (c) Scalogram $|\Gamma_x|^2$ computed using a Mexican hat wavelet. This affine class TSD is scale covariant, but suffers from low resolution. (d) Pseudo Unterberger distribution $\tilde{P}_x^{(-1)}$ computed using the same Mexican hat wavelet (and $G \equiv 1$). This TSD localizes both the fundamental rotation frequency and the transient events. (Data taken from the “NRad” test data set, courtesy of Jeff Allen of NCCOSC and Douglas Lake of the Office of Naval Research.)

In Figure 2.4, we demonstrated the performance of this new TSD on a synthetic test signal. In Figure 2.6, we plot the Wigner, Bertrand, scalogram, and smoothed pseudo Bertrand TSDs of the echo-location chirp of the large brown bat, *Eptesicus Fuscus*. The approximate hyperbolic localization of the smoothed pseudo Bertrand distribution matches the chirping nature of the signal, while the proportional-bandwidth time-frequency smoothing suppresses the interference components that swamp both the Wigner and Bertrand distributions.

$k = -1$. Pseudo Unterberger distribution

Due to their affine covariance properties, TSDs have a unique ability to analyze low frequencies with good frequency resolution and high frequencies with good time resolution. The active Unterberger distribution, furthermore, is time-localized and preserves the scaling properties of signal components [18], making it ideal for the study of transients. The primary drawback of this bilinear TSD is the existence of interference components between transient events.

The pseudo Unterberger distribution smooths interference components in the frequency direction while preserving the correct scaling structure across frequency. In Figure 2.7, we compare the performance of this new TSD to the pseudo Wigner TFD and scalogram on a machine fault signature.

$k = \pm 5$. Approximate affine Wigner distributions

Up to this point, we have emphasized the ability of the smoothed pseudo affine Wigner distributions to control interference components through time-frequency smoothing. More generally, however, our approach allows us to efficiently approximate all (unsmoothed) affine Wigner distri-

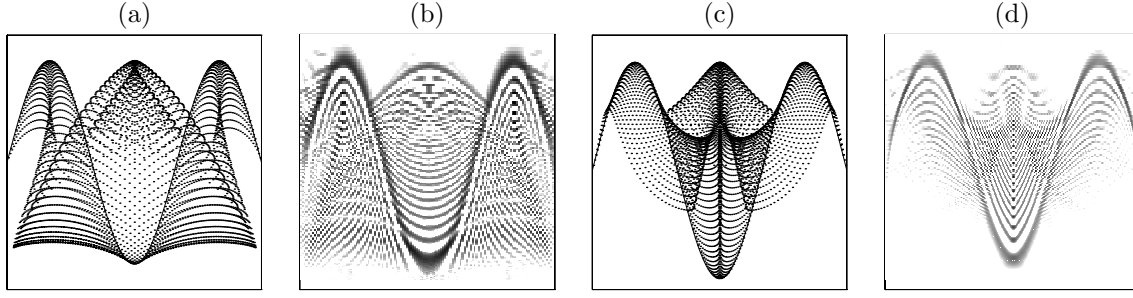


FIGURE 2.8 – Accurate approximation of affine Wigner distributions using pseudo affine Wigner distributions. Horizontal axis corresponds to time and vertical axis to frequency. The test signal x is a sinusoidal frequency modulation $x(t) = \exp[-i f_m^{-1} \cos(2\pi f_m t)]$. (a) Predicted locus of the true distribution $P_x^{(-5)}$ as determined using the geometric construction rules of [19, 20]. (b) Approximate affine Wigner distribution $\tilde{P}_x^{(-5)}$ computed as in (2.41) using a narrow-band wavelet of $Q = 130$. (c) Predicted locus of the true distribution $P_x^{(5)}$. (d) Approximate affine Wigner distribution $\tilde{P}_x^{(5)}$.

butions, even for unusual values of the index k for which the algorithm proposed in [Ovarlez et al., 1992] does not apply directly. In the limit as the bandwidth of the wavelet ψ falls to zero, we have $\tilde{P}_x^{(k)} \rightarrow P_x^{(k)}$. Therefore, a pseudo affine Wigner distribution (2.41) computed using a narrow-band wavelet will closely approximate its corresponding affine Wigner distribution, including its marginal, extended covariance and chirp localization properties. (In general, a (smoothed) pseudo affine Wigner distribution will not possess all possible theoretical properties of the affine Wigner TSDs; however, one could constrain the choice of the pre-processing wavelet ψ to preserve certain of them.)

In approximating an affine Wigner distribution, our goal is not to suppress interference components, but rather to *preserve* them. Figure 2.8 illustrates the close agreement between the approximate affine Wigner distributions $\tilde{P}_x^{(-5)}$ and $\tilde{P}_x^{(5)}$ and the theoretical loci of the true distributions as determined by the geometric construction rules of [19, 20]. To obtain these TSDs, we employed a narrow-band wavelet of $Q = 130$ in (2.41).

2.2.2 Pseudo Affine Wigner Distributions and the Affine Class

Like all affine Wigner distributions, the pseudo affine Wigner distributions belong to the affine class of TSDs, which, as defined in (2.36), revolves around the unitary Bertrand distribution $P^{(0)}$. Unfortunately, this formalism becomes awkward when one tries to derive an analytic form for the kernel Π corresponding to a pseudo affine Wigner distribution whose $k \neq 0$. In this section, we investigate an alternative canonical formulation for the affine class in which the kernels corresponding to the pseudo affine Wigner distributions have an easily identifiable, closed form.

For each $k \in \mathbb{R}$, we will replace $P^{(0)}$ in (2.36) with an alternative, matched generator TSD $W^{(k)}$ that provides a natural framework for the affine Wigner distributions $P^{(k)}$ and the pseudo affine Wigner distributions $\tilde{P}^{(k)}$. With this new formalism, all affine class TSDs can be written as

$$\rho_x(t, f) = \int_0^\infty \int_0^\infty W_x^{(k)}(\tau, \nu) \Pi\left(f(\tau - t), \frac{\nu}{f}\right) d\nu d\tau, \quad (2.43)$$

with a different kernel Π for each choice of generator $W^{(k)}$.

Affine Wigner generators. Any TSD that is continuously invertible (*regular* in the terminology of [Hlawatsch et al., 1993]) can play the rôle of $W^{(k)}$ in (2.43). In particular, $W^{(k)}$ does not have to be unitary. For the affine Wigner distributions, Fourier transformation of (2.27) followed by the reparameterization $\xi_k(u)$ of (2.28) and the symmetry property (2.29) leads to the inversion formula (for

X as a function of u)

$$X(e^u f_0) = \frac{1}{X^*(f_0)} \frac{\frac{d}{du}\xi(u)}{\mu(u)} \int P_x^{(k)}\left(t, \frac{f_0}{\lambda(-u)}\right) e^{-i2\pi t f_0(e^u - 1)} dt, \quad (2.44)$$

with f_0 a constant such that $X(f_0) \neq 0$. This inverse is continuous provided the term $C(u)$ in front of the integral remains bounded from above and below, with $0 < \epsilon \leq C(u) < \infty$. The affine Wigner distributions thus provide a family of generators for the affine class.

In Appendix A of [35], we singled out the affine Wigner generators most natural for studying the pseudo affine Wigner distributions. These generators take the form

$$W_x^{(k)}(t, f) \equiv f \int X(f\lambda_k(u)) X^*(f\lambda_k(-u)) e^{i2\pi t f \xi_k(u)} \frac{d\xi_k(u)}{du} du, \quad (2.45)$$

and correspond to the special choice $\mu_k(u) = \frac{d}{du}\xi_k(u)$ in (2.27).¹¹ With this generator installed in (2.43), the kernel corresponding to the (smoothed) pseudo affine Wigner distribution $\tilde{P}^{(k)}$ can be written in closed form as

$$\Pi^{(k)}(t, f) = f \iint G(u) \mu(u) \frac{\lambda_k(v)\lambda_k(-v)}{\lambda_k(u)\lambda_k(-u)} \Psi\left(f \frac{\lambda_k(v)}{\lambda_k(u)}\right) \Psi^*\left(f \frac{\lambda_k(-v)}{\lambda_k(-u)}\right) e^{i2\pi t f \xi_k(v)} du dv. \quad (2.46)$$

Examples. Each different $k \in \mathbb{R}$ yields a different generator $W^{(k)}$ matched to the specific geometry of the affine Wigner distributions $P^{(k)}$ and the pseudo affine Wigner distributions $\tilde{P}^{(k)}$. Examples include :

$k = 2$. *Wigner-Ville distribution*

Using $\mu_2(u) = \frac{d}{du}\xi_2(u) = 1 - \tanh^2 \frac{u}{2}$ in (2.27), we retrieve the Wigner distribution (2.23) (which also belongs to the affine class of time-frequency representations!) and the usual affine class definition proposed in [Rioul and Flandrin, 1992]. The Wigner distribution is the sole unitary generator of the form (2.45) (see Appendix B in [35] for the proof).

$k = 0$. *Non-unitary Bertrand distribution*

Using $\mu_0(u) = \frac{d}{du}\xi_0(u) = 1$ in (2.27), we obtain a (non-unitary) Bertrand distribution.

$k = -1$. *Active Unterberger distribution*

Using $\mu_{-1}(u) = \frac{d}{du}\xi_{-1}(u) = \cosh \frac{u}{2}$ in (2.27), we encounter the active form of the Unterberger distribution. This TSD is the sole generator of the form (2.45) that is time localized according to the definition of [Bertrand and Bertrand, 1992] (see Appendix C in [35] for the proof).

The scalogram has a different representation in terms of each generator. Recall that the unitarity of the affine Wigner distribution $P^{(0)}$ permitted us to write the scalogram as an affine correlation of two unitary Bertrand distributions (see Eq. (2.35)). This formula generalizes to the generators $W^{(k)}$ of (2.45). To see this, note that setting $G(u) = \delta(u)$ reduces (2.42) to the scalogram and (2.46) to

$$\begin{aligned} \Pi^{(k)}(t, f) &= \overline{W}_\psi^{(k)}(t, f) \\ &\equiv f \int \lambda_k(v) \lambda_k(-v) \Psi(f\lambda_k(v)) \Psi^*(f\lambda_k(-v)) e^{i2\pi t f \xi_k(v)} dv. \end{aligned} \quad (2.47)$$

Here, $\overline{W}^{(k)}$ is the *passive form* (2.33) of the generator $W^{(k)}$. While the *active form* $W^{(k)}$ is non-unitary in general, it cooperates with its passive form to produce the isometry-like relation (2.34). In this case, (2.43) simplifies to

$$|\Gamma_x(t, f)|^2 = \int \int_0^\infty W_x^{(k)}(\tau, \nu) \overline{W}_\psi^{(k)}\left(f(\tau - t), \frac{\nu}{f}\right) d\nu d\tau, \quad (2.48)$$

a natural generalization of (2.35).

11. With $\mu_k(u) = \frac{d}{du}\xi_k(u)$, the term $C(u)$ in front of the integral in (2.44) reduces to a constant.

2.3 Diffusion des représentations temps-fréquence et temps-échelle

Comme nous l'avons vu au paragraphe 2.1.3, les classes de Cohen et affines peuvent être vues comme des lissages respectivement convolutifs et affines, de la distribution de Wigner-Ville. La contribution ci-dessous revisite ces deux formulations canoniques sous l'angle des équations aux dérivées partielles. Partant du constat que dans le cas d'un noyau $\Pi(t, f)$ gaussien, l'expression (2.21) coïncide avec une solution à temps fini de l'équation de diffusion de la chaleur, nous avons initialement proposé dans [45], puis développé avec le travail de thèse de Julien Gosme, le prolongement de ce parallèle au-delà des cas isotropes homogènes [50–53]. Le gain obtenu en terme d'adaptativité et de concentration énergétique des représentations temps-fréquence est tout à fait surprenant, et certainement encore très sous-exploité. C'est cette ouverture de la classe de Cohen – et de la classe affine – au potentiel des EDP que nous détaillons ci-dessous.

Ce paragraphe s'inspire en grande partie de l'article :

Adaptive diffusion as a versatile tool for time-frequency and time-scale representations processing : a review

J. GOSME, C. RICHARD and P. GONÇALVES

IEEE Transactions on Signal Processing

Volume 53, No. 11, pp. 4136–4146, 2005

2.3.1 Definitions

Diffusion is the process by which matter is transported from areas of high concentration to areas of lower concentration as a result of the movement of an ensemble of molecules inside a region. This is mathematically formulated by a basic equation referred to as Fick's law [Crank, 1957] :

$$J = -D \nabla U, \quad (2.49)$$

where J is the flux of molecules, U their concentration, D the diffusion tensor, and ∇ the gradient operator. The negative sign indicates that the diffusing mass flows in the direction of decreasing concentration. If D is a scalar-valued conductance function, which implies that J and ∇U are collinear, the diffusion process is called *isotropic*. Otherwise it is called *anisotropic*. It is said that the diffusion process is *homogeneous* if the diffusion tensor D is constant over the region of interest. Location-dependent diffusion is called *non-homogeneous* or *inhomogeneous*. The law of conservation of mass is expressed as :

$$\frac{\partial U}{\partial \tau} = -\text{div}(J), \quad (2.50)$$

where τ is the diffusion time, and div the divergence operator. Combining this relationship with Fick's law produces the law of diffusion, which states :

$$\frac{\partial U}{\partial \tau} = -\text{div}(-D \nabla U). \quad (2.51)$$

In the case where D is a positive constant, the resulting isotropic and homogeneous process (2.51) is often referred to as *heat diffusion equation* since it describes the evolution of temperature within a finite homogeneous continuum with no internal sources of heat. In this paper, this diffusion is said to be *linear* because the tensor D does not vary with τ . Otherwise, it would be called *nonlinear*¹². Let us now restrict our discussion to the partial differential equations (PDE's) in two-plus-one dimension

$$\begin{cases} U(v_1, v_2; \tau = 0) = U_0(v_1, v_2) \\ \frac{\partial U}{\partial \tau} = \text{div}(\nabla U), \end{cases} \quad (2.52)$$

12. In the theory of partial differential equations, a diffusion process is called *nonlinear* if D is a function of U or its derivatives.

where $U_0 \in \mathcal{L}^1(\mathbb{R}^2)$ denotes the initial spatial condition. It is well-known that the solution to (2.52) is

$$U(v_1, v_2; \tau) = \begin{cases} (G * U_0)(v_1, v_2) & (\tau > 0) \\ U_0(v_1, v_2) & (\tau = 0), \end{cases} \quad (2.53)$$

with $*$ the usual 2-D convolution, and $G(v_1, v_2; \tau) = (4\pi\tau)^{-1} \exp(-[v_1^2 + v_2^2]/4\tau)$ an isotropic Gaussian kernel which is referred to as Green's function of the PDE given above. This means that the solution $U(v_1, v_2; \tau)$ of the heat diffusion equation (2.52) at each time instant τ can be simply obtained by convolution of the initial spatial condition $U_0(v_1, v_2)$ with Green's function $G(v_1, v_2; \tau)$. This result led Koenderink to use the basic PDE framework (2.52) for image processing, bringing a new insight into image restauration techniques viewed as homogeneous diffusion processes [Koenderink, 1984]. Note that earlier work was done by Iijima and published in Japanese [Iijima, 1971], as pointed out in [Weickert et al., 1999]. Thereafter Perona and Malik proposed a nonlinear inhomogeneous generalization of this approach¹³ based on a real-valued conductance function $\eta(\cdot)$ of spatial coordinates (v_1, v_2) and time τ [Perona and Malik, 1990] :

$$\frac{\partial U(v_1, v_2; \tau)}{\partial \tau} = \text{div}(\eta(|\nabla U|^2) \nabla U). \quad (2.54)$$

They chose to make the conductance function a decreasing function of the gradient magnitude in order to inhibit the diffusion process in areas of high contrast whereas areas of low contrast undergo faster diffusion. Systems such as (2.54) are intended to smooth out noise in images while sharp edges and contours are preserved. Significant advancements have been made since then, e.g., with diffusion processes acting differently along the gradient direction ∇U and the normal direction to ∇U . This last principle has been successfully applied by authors such as Weickert [Weickert, 1997] and Alvarez *et al.* [Alvarez et al., 1993]. It has also been used by Rudin *et al.* [Rudin et al., 1992] in the form of a backward diffusion along the gradient direction to sharpen edges. More recently, Gilboa *et al.* [Gilboa et al., 2002] and Smolka [Smolka, 2003] have proposed forward-and-backward diffusion processes where different regions of the image are either forward or backward diffused according to the geometry within a given neighborhood.

In the light of these considerations, we shall now investigate new signal-depend smoothing techniques of bilinear TFRs and TSRs that adapt over time, frequency and scale to remove interference terms and noise while emphasizing signal components. Special efforts will be made to preserve the covariance properties of distributions with respect to time shifts, frequency shifts or scale changes.

2.3.2 Isotropic diffusion

Distributions in the Cohen class (2.21) involves convolution of the Wigner-Ville distribution (WD) with a 2-D kernel Π . This interpretation has led many authors to investigate design aspects of Π for reducing noise and interference terms in TFRs. Unfortunately, reduced interference distributions with fixed kernel necessarily spread out and delocalize signal components as well. In order to overcome this incompatibility, we first revisited in [45], standard 2-D smoothing related to the Cohen class in terms of diffusion processes of the form (2.52) and (2.54). Here we start by briefly sketching the different steps of this approach and we check that resulting distributions are covariant with respect to time-frequency shifts. Next we extend this framework for tackling the problem of adaptive smoothing of TSRs of the affine class.

Isotropic diffusion in the Cohen class

The Gaussian-windowed spectrogram (2.25) with variance σ_w^2 illustrates the connection between smoothed TFRs and PDEs. Indeed, its kernel $\Pi(t, f) = W_g(t, f)$ is the Gaussian function $\Pi(t, f) = 2 \exp(-[t^2 + 4\pi^2\sigma_g^4 f^2]/\sigma_g^2)$, and happens to correspond to the Green function $G(t, f; \tau) =$

13. In [Perona and Malik, 1990, Voci et al., 2004], this diffusion process is referred to as *anisotropic*. However, according to the convention given in [Weickert, 1998], this scheme is *isotropic* since its flux J defined in (2.49) is collinear to ∇U .

$(4\pi\tau)^{-1} \exp(-[t^2 + f^2]/4\tau)$ of the heat equation

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(\nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f). \end{cases} \quad (2.55)$$

When the instant of diffusion $\tau = (4\pi)^{-1}$, the spectrogram $SP_x(t, f)$ mentioned above is then the solution of (2.55). In a more general way, $C_x(t, f; \tau)$ is a member of the Cohen class for all positive τ since it can be expressed as the 2-D convolution of $W_x(t, f)$ with $G(t, f; \tau)$.

Following the lines of [Perona and Malik, 1990, Voci et al., 2004] in an image processing context, a possible refinement of (2.55) consists in introducing inhomogeneity into the heat equation to perform time and frequency dependent smoothing [45] :

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(\eta_x(t, f) \nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f), \end{cases} \quad (2.56)$$

where $\eta_x(t, f)$ is a conductance function depending both on the analyzed signal x and on time-frequency location (t, f) . In their primary work, Perona and Malik chose to use a decreasing conductance function of the gradient magnitude. However, this solution is not adequate in a time-frequency context since the largest variations of the WD coincide with oscillating interference terms that we want to remove. The conductance function η_x can be chosen in many different ways, depending on the information we want to emphasize in TFRs. Several solutions have been proposed in [45, 51] for improving the readability of the WD. They involve an auxiliary TFR of the Cohen class, denoted by ρ_x , in order to combine the localization properties of the WD with some desired characteristics of ρ_x . For instance, with E. Payot, we suggested in [45] to use

$$\eta_{W_x}(t, f) = \mathbb{1}_{]-\infty, 0]}(W_x(t, f)), \quad (2.57)$$

where $\mathbb{1}_{]-\infty, 0]}(\cdot)$ equals 1 if its argument is negative, and 0 otherwise. This diffusion function is intended to smooth out interference terms, which are mostly located in the neighborhood of time-frequency areas where $W_x(t, f) < 0$ due to their oscillating nature. Note that (2.57) can be made dependent on the diffusion time τ by using the evolving distribution $C_x(t, f; \tau)$ instead of $W_x(t, f)$. Such a feedback leads to nonlinear diffusion filters. Another intuitive diffusion-based approach for obtaining interference-free distributions relies on the spectrogram. Although this distribution spreads out signal components in the time-frequency domain, it is approximately equal to zero over non-energetic areas where the interference terms of the WD are likely to be situated. Thus, as illustrated in Figure 2.9, making the conductance function a decreasing function of SP_x such as

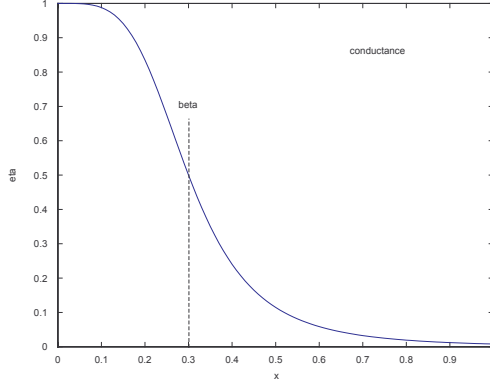
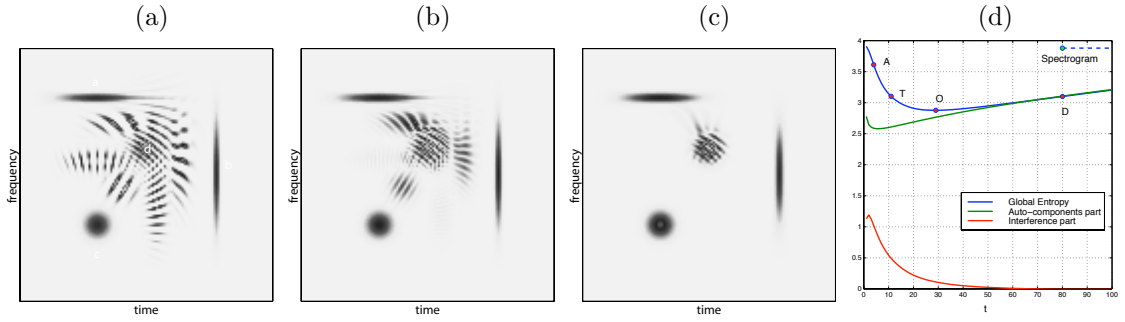
$$\eta_{SP_x}(t, f) = \left[1 + \left(\frac{SP_x(t, f)}{\beta} \right)^\alpha \right]^{-1} \quad (\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \quad (2.58)$$

allows for tuning the diffusion rate over the time-frequency domain. As a last example, the following conductance function

$$\eta_{C_x}(t, f; \tau) = \left[1 + \left(\frac{C_x(t, f; \tau)}{\beta} \right)^\alpha \right]^{-1} \quad (\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \quad (2.59)$$

assumes that signal components are more *structured* than interference terms. Hence, it inhibits the diffusion process in the neighborhood of denser signal components whereas sharp components undergo faster diffusion. Figure 2.10 illustrates the diffusion process (2.56) associated with the spectrogram-based diffusion function (2.58). Compared to the WD, only the signal components are left unchanged while the interference terms, exhibiting sharper patterns, are progressively smoothed out. As shown in Figure 2.11, this smoothing process still performs fairly well in a noisy environment. Because the WD maps a white noise to a 2-D random field in the time-frequency domain, positive and negative coefficients are locally balanced and partly cancelled out by the diffusion process. The robustness of the spectrogram with respect to noise also explains this interesting result.

We shall now prove that $C_x(t, f; \tau)$, under mild conditions on the initial distribution and the conductance function, is covariant with respect to time-frequency shifts for all positive τ . Let us first rewrite

FIGURE 2.9 – Conductance function (2.58) with $\alpha = 4$ and $\beta = 0.3$.FIGURE 2.10 – Adaptive diffusion process with spectrogram-based conductance function applied to the WD of a 4 tones signal : (a) WD, (b) diffusion until time instant τ , (c) diffusion until the optimum stopping time instant τ^* , (d) evolution of the entropy measure of $C_x(t, f; \tau)$.

the diffusion process (2.56) in the form : $\tilde{\mathcal{D}}_\tau[C_x(t, f; \tau = 0)] = C_x(t, f; \tau)$, where $\tilde{\mathcal{D}}_\tau$ is the operator describing the evolution of $C_x(t, f; \tau)$. Let us also define the translation operator $\tilde{\mathcal{W}}_{t_0, f_0}$ which, applied to any distribution $C_x(t, f)$, gives : $\tilde{\mathcal{W}}_{t_0, f_0}[C_x(t, f)] = C_x(t - t_0, f - f_0)$. The simple change of variables $t \leftarrow (t - t_0)$ and $f \leftarrow (f - f_0)$ in (2.56) allows us to show that :

$$\tilde{\mathcal{W}}_{t_0, f_0} \tilde{\mathcal{D}}_\tau[C_x(t, f; \tau)] = \tilde{\mathcal{D}}_\tau \tilde{\mathcal{W}}_{t_0, f_0}[C_x(t, f; \tau)], \quad (2.60)$$

meaning that diffusing a shifted TFR comes to shifting the diffused TFR, proving so the property of time-shift covariance. In the same way, we verify that $\tilde{\mathcal{W}}_{t_0, f_0} \tilde{\mathcal{D}}_\tau[C_x(t, f; \tau = 0)] = C_x(t - t_0, f - f_0; \tau)$ and that $\tilde{\mathcal{D}}_\tau \tilde{\mathcal{W}}_{t_0, f_0}[C_x(t, f; \tau = 0)] = \tilde{\mathcal{D}}_\tau[C_{\mathcal{W}x}(t, f; \tau = 0)]$ if the initial distribution $C_x(t, f; \tau = 0)$ is covariant with respect to time-frequency shifts. Moreover, $\tilde{\mathcal{D}}_\tau[C_{\mathcal{W}x}(t, f; \tau = 0)] = C_{\mathcal{W}x}(t, f; \tau)$, if the conductance function $\eta_x(t, f)$ itself verifies this covariance property, i.e., $\eta_{\mathcal{W}x}(t, f) = \eta_x(t - t_0, f - f_0)$. Under these two conditions, we can conclude with (2.60) that :

$$C_{\mathcal{W}x}(t, f; \tau) = C_x(t - t_0, f - f_0; \tau). \quad (2.61)$$

Also note that (2.56) conserves energy, i.e., $\iint C_x(t, f; \tau) dt df = \int |x(t)|^2 dt$, since (2.56) is a particular case of (2.51), which follows directly from the law of mass conservation (2.50).

Let us now present a discrete framework for the diffusion process (2.56). We suppose that TFRs are sampled on a discrete grid $(n \cdot \delta t, m \cdot \delta f)$, $(n, m) \in [0, N] \times [0, M]$, with δt and δf the sampling period in time and frequency [45], [Perona and Malik, 1990]. Using a simple first order Euler approximation gives the following approximate discrete evolution equation

$$C_x[n, m; \tau_{k+1}] = C_x[n, m; \tau_k] + \frac{1}{4} \sum_{i \in I} \eta_x^{(d)}[n, m] \nabla^{(d)} C_x[n, m; \tau_k]. \quad (2.62)$$

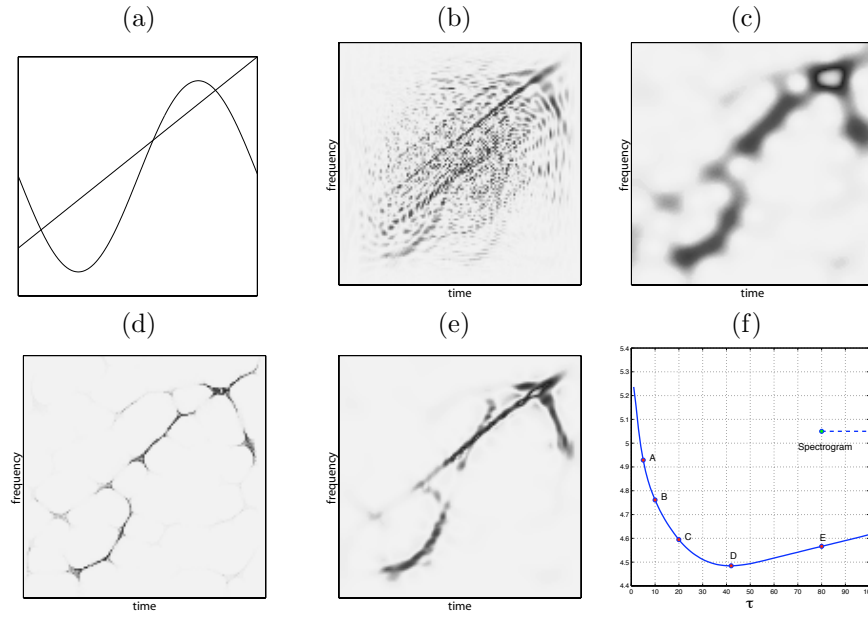


FIGURE 2.11 – Different time-frequency representations on a test signal composed of a linear chirp and a sinusoidal FM component. SNR is fixed to 3 dB with additive stationary Gaussian white noise. (a) Theoretical distribution of the signal without noise. (b) Wigner distribution. (c) Spectrogram using a Gaussian window. (d) Reassigned spectrogram. (e) Non uniform adaptive time-frequency distribution with the conductance function (2.58) where $\beta = 2$ and SP_x is the spectrogram displayed in Fig. 2.11(c). (f) Evolution of the Shannon entropy measured on the non stationary adaptive time-frequency distribution for each step of the diffusion algorithm (2.56). Point D marks the "optimal" stopping point ($\tau^* = 44$) corresponding to the non-uniform adaptive TFR of Fig. 2.11(e).

where $d \in I$ refers to the cardinal directions of the four neighbor points of $C_x[n, m; \tau_k]$, and $C_x[n, m; \tau_0] = W_x[n, m]$. In (2.62), $\eta_x^{(d)}$ denotes the average value of the conductance function along the direction d , meaning, for the upper neighbor $d = \text{north}$:

$$\eta_x^{(\text{north})}[n, m] = \frac{\eta_x[n, m] + \eta_x[n, m + 1]}{2}, \quad (2.63)$$

and

$$\nabla^{(\text{north})} C_x[n, m; \tau_k] = C_x[n, m + 1; \tau_k] - C_x[n, m; \tau_k]. \quad (2.64)$$

In the remainder, we either use straightforward adaptations of this numerical scheme, or the one proposed in [Weickert, 1998, Weickert and Schar, 2002].

Without constraint on the diffusion time, C_x would converge to a uniform distribution over the time-frequency domain regardless of the analyzed signal. Hence, the diffusion process has to be stopped when some criterion is achieved. The Shannon entropy, defined in the case of TFRs in [Baraniuk et al., 2001], as

$$H_C(\tau_k) = - \sum_n \sum_m |C_x[n, m; \tau_k]| \log |C_x[n, m; \tau_k]|, \quad (2.65)$$

is a natural candidate for measuring the complexity of TFRs¹⁴. Indeed, the TFRs of signals consisting of small numbers of components would lead to small entropy values while the TFRs of information-rich signals would lead to large ones. A simple way to justify the above stopping criterion is to consider an example, where the overall integrals can be decomposed into two distinct region integrals : the first one over the domain of interference terms and the second one over the domain of signal components. The first integral tends to zero as the interference terms are smoothed out, while the second one increases at a slower rate due to the action of the conductance function. When these antagonistic effects balance, the Shannon entropy reaches a minimum that causes the diffusion process to stop. This behavior is illustrated in Figures 2.10(d) and 2.11(f), where H_C rapidly decreases toward a minimum and then slowly increases beyond this optimum stopping time τ^* . In an image processing context, note that it was shown that the Shannon entropy monotonically increases with respect to the diffusion time for most commonly-used diffusion processes [Weickert, 1998, Ferraro et al., 1999, Sporring and Weickert, 1999]. This property is based on the convexity of the function $v \log v$, $v \in \mathbb{R}_+^*$, in the definition of the Shannon entropy, but does not systematically hold true for (2.65) which deals with non-positive distributions. As a final remark, it should be noticed that we could also consider the Rényi's entropies instead of the Shannon entropy, as suggested in [Baraniuk et al., 2001]. This would lead to different TFRs optimizing different criteria related to the trade-off "resolution *versus* interference terms removal".

In this section, we have investigated several forms of variable-conductance diffusion processes in order to locally adjust diffusion strength, and then to perform adaptive smoothing matching local characteristics of TFRs. We shall now extend this approach to tackle the problem of adaptive smoothing of TFRs of the affine class.

Isotropic diffusion in the affine class

Every bilinear affine TFR (thereafter termed time-scale representations and noted TSR in short) of the affine class is characterized by an *affine convolution* of the WD, as written in (2.37). Just like the spectrogram within the Cohen class, the scalogram (2.38) is a prominent member of the affine class. By analogy with the relations (2.55) that bonds smoothed TFRs of the Cohen class with the heat equation, we proposed to adapt the diffusion strength to the scale a as follows [51] :

$$\begin{cases} \frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = \text{div}(a^2 \nabla \Omega_x(t, a; \tau)) \\ \Omega_x(t, a; \tau = 0) = W_x(t, a), \end{cases} \quad (2.66)$$

14. Refer to [Baraniuk et al., 2001] for a detailed study of Rényi's entropies and time-frequency information measures. We also recommend [Ferraro et al., 1999, Sporring and Weickert, 1999] for a thorough study on Rényi's entropies as means for extracting information from images during diffusion.

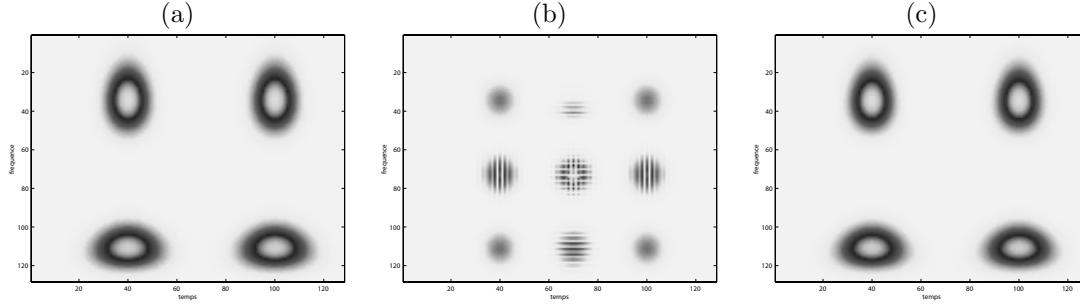


FIGURE 2.12 – (a) Affine isotropic diffusion applied to (b) the WD of a 4 tones signal. (c) Gaussian-windowed scalogram. These distributions are represented in a time-frequency form to clearly illustrate the effect of affine smoothing.

where $\Omega_x(t, a; \tau = 0)$ is the TSR to be processed. Note that (2.66) differs from (2.55) by the presence of the factor a^2 in the divergence term. As it can straightforwardly be proved with a proper change of variables in (2.66), this modification is crucial to obtain the following property :¹⁵

$$\tilde{\mathcal{A}}_{t_0, a_0} \tilde{\mathcal{D}}_\tau [\Omega_x(t, a; \tau)] = \tilde{\mathcal{D}}_\tau \tilde{\mathcal{A}}_{t_0, a_0} [\Omega_x(t, a; \tau)], \quad (2.67)$$

where we recognize the clock change covariance property (2.26) of the affine class. Here $\tilde{\mathcal{A}}_{t_0, a_0}$ is the operator defined by $\tilde{\mathcal{A}}_{t_0, a_0} [\Omega_x(t, a)] = \Omega_x\left(\frac{t-t_0}{a_0}, \frac{a}{a_0}\right)$, and $\tilde{\mathcal{D}}_\tau$ is the diffusion operator associated with (2.66). We can also check that : $\tilde{\mathcal{A}}_{t_0, a_0} \tilde{\mathcal{D}}_\tau [\Omega_x(t, a; \tau = 0)] = \Omega_x\left(\frac{t-t_0}{a_0}, \frac{a}{a_0}; \tau\right)$. In addition, if $\Omega_x(t, a; \tau = 0)$ is covariant with respect to the affine operator $\mathcal{A}$ acting in (2.26), we have : $\tilde{\mathcal{D}}_\tau \tilde{\mathcal{A}}_{t_0, a_0} [\Omega_x(t, a; \tau = 0)] = \Omega_{\mathcal{A}x}(t, a; \tau)$. Combining these two results via (2.67), we then conclude that :

$$\Omega_{\mathcal{A}x}(t, a, \tau) = \Omega_x\left(\frac{t-t_0}{a_0}, \frac{a}{a_0}; \tau\right), \quad \forall \tau \geq 0. \quad (2.68)$$

Notice that, as $\Omega_x(t, a; \tau)$ remains bilinear with respect to x (by linearity of the differential operators in (2.66)), it naturally belongs to the affine class¹⁶. Then, according to the canonical definition (2.37) of this class, some kernel $\Pi(t, a; \tau)$ associated with $\Omega_x(t, a; \tau)$ may well be found. However, its identification is a non-trivial issue because it deals with an inhomogeneous diffusion scheme. To the best of our knowledge, this problem remains an open question. Finally, let us remark that (2.66) is not the only diffusion process which produces TSRs of the affine class. For instance, it shares this property with :

$$\frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = a^2 \operatorname{div}(\nabla \Omega_x(t, a; \tau)) \quad (2.69)$$

$$\frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = a \operatorname{div}(a \nabla \Omega_x(t, a; \tau)). \quad (2.70)$$

Throughout the remainder of this section, we shall restrict ourselves to the diffusion process (2.66) which, like (2.51), preserves energy. For tackling the problem of adaptive smoothing of TSRs, we now propose to extend (2.66) using the following isotropic inhomogeneous scheme :

$$\begin{cases} \frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = \operatorname{div}(a^2 \eta_x(t, a) \nabla \Omega_x(t, a; \tau)) \\ \Omega_x(t, a; \tau = 0) = W_x(t, a). \end{cases} \quad (2.71)$$

A similar proof to the one sketched above, shows that the covariance of $\eta_x(t, a)$ with respect to clock changes guarantees this diffusion process to generate affine-covariant distributions $\Omega_x(t, a; \tau)$, for all

¹⁵. Replace variables t and a in (2.66) by $(t - t_0)/a_0$ and a/a_0 , respectively, and calculate the divergence term to get (2.67).

¹⁶. Another type of affine diffusion processes was proposed in an image processing context [Alvarez et al., 1993], but it did not preserve bilinearity.

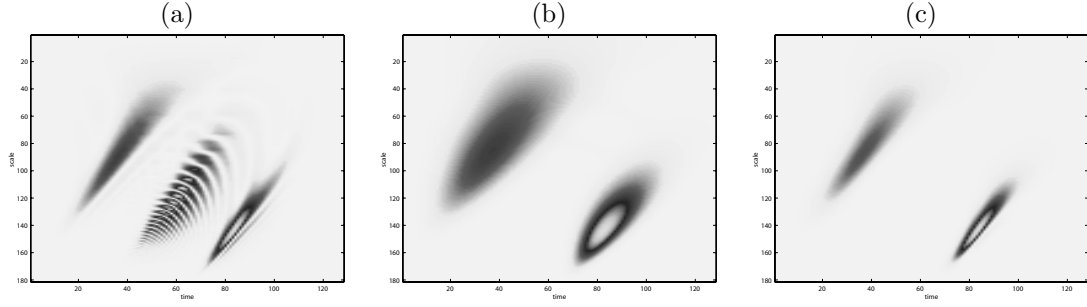


FIGURE 2.13 – *Affine diffusion applied to two Altes chirps. (a) WD, (b) isotropic diffusion, (c) adaptive isotropic diffusion with scalogram-based conductance function.*

positive τ . As already pointed out in the previous section, many strategies may be developed to design specific conductance functions. For instance, following the intuition that led to (2.58), a way to obtain interference-free TSRs relies on the conductance function

$$\eta_{SC_x}(t, a) = \left[1 + \left(\frac{SC_x(t, a)}{\beta} \right)^\alpha \right]^{-1} \quad (\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \quad (2.72)$$

derived from a scalogram. Figure 2.13 portrays the benefit of using this inhomogeneous model (2.71), rather than the homogeneous model (2.66) that tends to remove the interference terms but also to smooth out the signal components in the WD.

2.3.3 Anisotropic diffusion

So far, we have investigated diffusion-based models with scalar-valued conductance functions to locally adjust the diffusion strength. Here we propose to control both diffusion strength and orientation with the tensor D involved in the general diffusion model (2.51). In the spirit of the work by [Jones and Baraniuk, 1995] on signal-dependent time-frequency analysis, the anisotropic diffusion schemes presented below allow for the direction and for the amount of smoothing to be adjusted. The cases of TFRs of the Cohen class and TSRs of the affine class are considered.

Anisotropic diffusion in the Cohen class

In contrast to the isotropic process (2.55), where the diffusion is solely governed by the geometry of the distribution to be processed, we propose a more flexible scheme where the diffusion model is also controlled by a tensor D_{tf} built into the diffusion equation :

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(D_{tf} \nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f), \end{cases} \quad (2.73)$$

where D_{tf} is a (2×2) symmetric positive definite matrix in order to ensure the stability of the system [Weickert, 1998]. Note that (2.55) can be recovered using the identity matrix for D_{tf} . For a fixed matrix D_{tf} , denoted below by D , it is a well-known result that the solution $C_x(t, f; \tau)$ of (2.73) at each diffusion instant τ can be simply obtained by convoluting the initial distribution $W_x(t, f)$ with the following 2-D Gaussian kernel [Weickert, 1998] :

$$\Pi(t, f; \tau) = \frac{1}{4\pi\tau|D|^{\frac{1}{2}}} \exp\left(-\frac{(t \ f) D^{-1} (t \ f)^t}{4\tau}\right). \quad (2.74)$$

The eigenvectors and the eigenvalues of the diffusion tensor then define the two orthogonal directions of smoothing as well as the smoothing strength in these directions.

Some signals, such as linearly frequency modulated tones, may consist of time-frequency components with a dominant orientation. Incorporating such information as a prior knowledge in the diffusion process

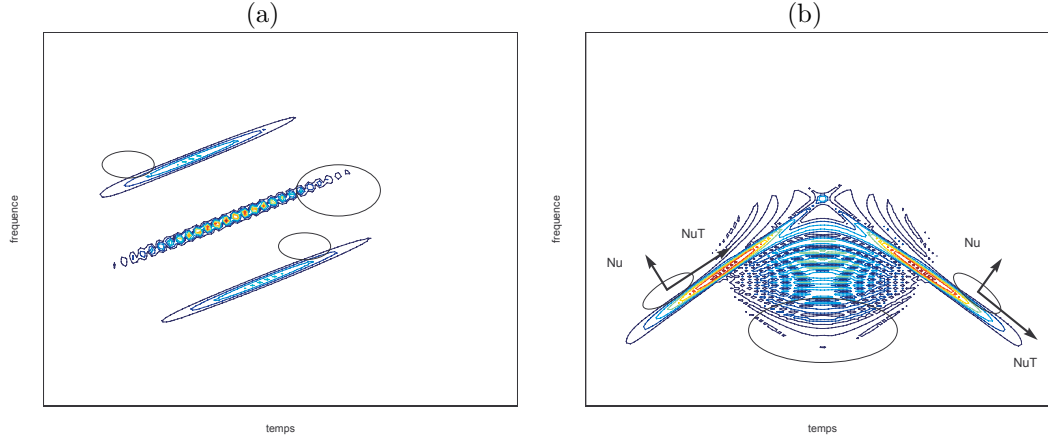


FIGURE 2.14 – Illustration of (a) isotropic and (b) anisotropic diffusion processes, with local kernels schematized by ellipses. Anisotropy allows the kernel orientation to be usefully adapted to the distribution.

should improve its performance, in particular if the diffusion along the signal components is enhanced as compared to the perpendicular direction. This can be achieved by adapting the Gaussian kernel orientation to the distribution to be processed. Here we suggest to choose D_{tf} in (2.73) as a function of the gradient ∇C_x at the vicinity of any location (t, f) of interest, as shown in Figure 2.14. To this end, the orthonormal system formed by the eigenvectors $\{v_1, v_2\}$ of D_{tf} is collinear to $(\nabla C_x, \nabla C_x^\perp)$. That is

$$D_{tf} = (v_1 \ v_2) \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} (v_1 \ v_2)^t \quad (2.75)$$

where the eigenvalue λ has to be chosen greater than 1 to produce a more aggressive smoothing along the signal components than in the perpendicular direction. A convenient way to estimate local orientation is to introduce the gradient-square tensor, defined as [Förstner and Gülch, 1987] :

$$J_x(t, f) = \nabla C_x (\nabla C_x)^t = \begin{pmatrix} \left[\frac{\partial C_x(t, f; \tau)}{\partial t} \right]^2 & \frac{\partial C_x(t, f; \tau)}{\partial t} \frac{\partial C_x(t, f; \tau)}{\partial f} \\ \frac{\partial C_x(t, f; \tau)}{\partial t} \frac{\partial C_x(t, f; \tau)}{\partial f} & \left[\frac{\partial C_x(t, f; \tau)}{\partial f} \right]^2 \end{pmatrix}. \quad (2.76)$$

The eigenvectors of $J_x(t, f)$ are collinear to ∇C_x and $\nabla C_x^\perp$. In the presence of noise, it is well-known that the gradient does not give reliable information about local orientation. In practice, a smoothing filter is used to produce a gradient-square tensor with mean orientation information related to a weighted neighborhood [Weickert, 1998, Förstner and Gülch, 1987]. This operation can be performed componentwise by any low-pass filter, which is usually a Gaussian window or a rectangular window filter L_σ of width σ :

$$\bar{J}_x(t, f; \sigma) = \begin{pmatrix} L_\sigma * \left[\frac{\partial C_x(t, f; \tau)}{\partial t} \right]^2 & L_\sigma * \left[\frac{\partial C_x(t, f; \tau)}{\partial t} \frac{\partial C_x(t, f; \tau)}{\partial f} \right] \\ L_\sigma * \left[\frac{\partial C_x(t, f; \tau)}{\partial t} \frac{\partial C_x(t, f; \tau)}{\partial f} \right] & L_\sigma * \left[\frac{\partial C_x(t, f; \tau)}{\partial f} \right]^2 \end{pmatrix}. \quad (2.77)$$

We can now propose an inhomogeneous anisotropic diffusion scheme to adapt the orientation and the strength of the smoothing operator to the TFR :

$$\begin{cases} \frac{\partial C_x(t, f; \tau)}{\partial \tau} = \text{div}(\eta_x(t, f) D_{tf} \nabla C_x(t, f; \tau)) \\ C_x(t, f; \tau = 0) = W_x(t, f), \end{cases} \quad (2.78)$$

where $\eta_x(t, f)$ is a conductance function that can be chosen among those listed in Section 2.3.2. Just like the isotropic diffusion process (2.56), and because D_{tf} only involves derivatives, this model provides shift-covariant distributions as long as the conductance function itself is covariant with respect to time-frequency shifts. In addition, it preserves energy as it is of the same form of (2.51). Figure 2.15

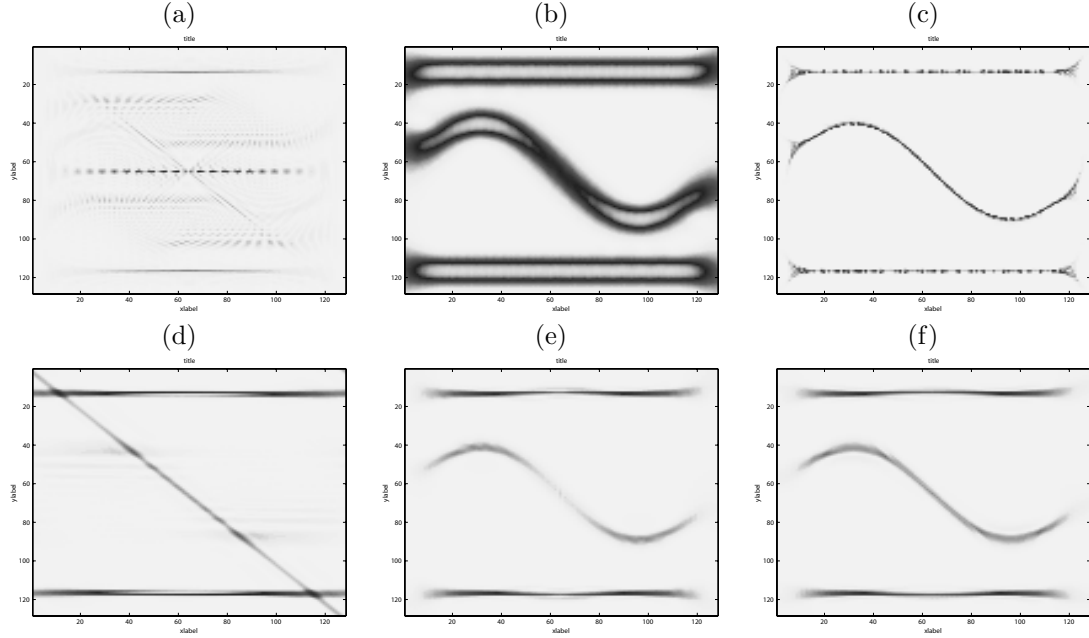


FIGURE 2.15 – Comparison of different TFRs : (a) WD, (b) homogeneous diffusion, (c) spectrogram modified by the reassignment method, (d) TFR with signal-dependent radially Gaussian kernel, (e) adaptive diffusion, and (f) anisotropic adaptive diffusion.

compares on a synthetic signal, anisotropic adaptive diffusion with other standard approaches used to improve the readability of TFRs. In Figure 2.16, the same comparison is carried out on a natural signal of bat echolocation.

Anisotropic diffusion in the affine class

The anisotropic process (2.73) can also be used to process distributions of the affine class. To be so, we simply need to replace the scaling variable a by the frequency variable f , in accordance with the formal equivalence $f \equiv a^{-1}$. Then, it can be shown that :

$$\begin{cases} \frac{\partial \Omega_x(t, f; \tau)}{\partial \tau} = \text{div}(D_f \nabla \Omega_x(t, f; \tau)) \\ \Omega_x(t, f; \tau = 0) = W_x(t, f), \end{cases} \quad \text{with} \quad D_f = \begin{pmatrix} f^{-2} & 0 \\ 0 & f^2 \end{pmatrix} \quad (2.79)$$

where $\Omega_x(t, f; \tau)$ satisfies to the clock change covariance. A straightforward signal-dependent extension of (2.79) could be obtained by adjusting the smoothing strength to a conductance function $\eta_x(t, f)$ embedded into the diffusion equation. However, this approach is not satisfactory as the smoothing orientation, given by D_f , does not depend on the analyzed signal. Intuitively, we could circumvent this, by substituting D_f with a gradient-square tensor based on an affine covariant gradient. Unfortunately, such a measure of local orientation does not exist, as angles in the time-frequency domain are not affine covariant¹⁷. Nonetheless, a closer view at the definition of the gradient-square tensor applied to $\Omega_{Ax}(t, a; \tau)$ (instead of $\Omega_{Ax}(t, f; \tau)$) shows that :

$$J_{Ax}(t, a) = \begin{pmatrix} \left[\frac{\partial \Omega_{Ax}(t, a; \tau)}{\partial t} \right]^2 & \frac{\partial \Omega_{Ax}(t, a; \tau)}{\partial t} \frac{\partial \Omega_{Ax}(t, a; \tau)}{\partial a} \\ \frac{\partial \Omega_{Ax}(t, a; \tau)}{\partial t} \frac{\partial \Omega_{Ax}(t, a; \tau)}{\partial a} & \left[\frac{\partial \Omega_{Ax}(t, a; \tau)}{\partial a} \right]^2 \end{pmatrix} = \frac{1}{a_0^2} J_x \left(\frac{t - t_0}{a_0}, \frac{a}{a_0} \right). \quad (2.80)$$

As a result, information on local orientation conveyed by eigenvectors of $J_x(t, a)$ or $\bar{J}_x(t, a; \sigma)$ is covariant with respect to clock changes. It can thus be used in (2.75) to determine the diffusion

17. See [Olver et al., 1999] and references therein for a discussion on first and second order affine invariant/covariant descriptors.

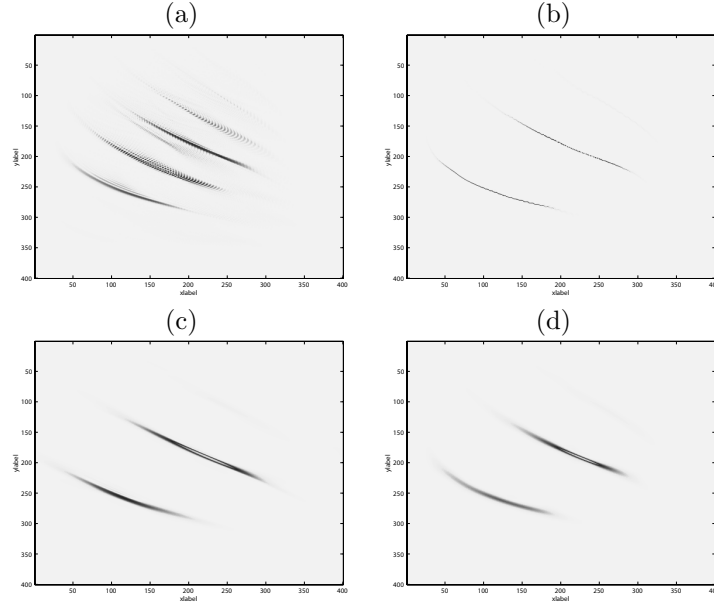


FIGURE 2.16 – Comparison of different TFRs applied to a 500-sample bat echolocation signal. (a) WD, (b) spectrogram, (c) spectrogram modified by the reassignment method, (d) TFR with signal-dependent radially Gaussian kernel, and (e) anisotropic adaptive diffusion. Signal kindly provided by C. Condon, K. White, and A. Feng from Beckman Institute of the University of Illinois.

tensor D_{ta} and to propose the following anisotropic inhomogeneous diffusion process :

$$\begin{cases} \frac{\partial \Omega_x(t, a; \tau)}{\partial \tau} = \text{div}(a^2 \eta_x(t, a) D_{ta} \nabla \Omega_x(t, a; \tau)) \\ \Omega_x(t, a; \tau = 0) = W_x(t, a). \end{cases} \quad (2.81)$$

Like (2.71), affine covariance of this process is subject to affine covariance of the conductance function $\eta_x(t, a)$. In addition, as a particular case of (2.51), this scheme preserves the energy of the analyzed signal.

2.4 Décomposition modale empirique

En marge des décompositions atomiques linéaires (2.9), la décomposition Modale Empirique (désignée dans la suite du document par son abréviation anglaise EMD, pour *Empirical Mode Decomposition*) partage avec celles-ci, l'idée que tout signal peut s'expliquer par une somme de contributions individuelles de briques « élémentaires » localisées dans le plan temps-fréquence. Elle s'en distingue, en considérant les signaux à l'échelle de leurs oscillations locales, sans que celles-ci soient nécessairement harmoniques au sens de Fourier.

Dans son principe, l'EMD vise à isoler de façon récursive et locale, l'oscillation la plus rapide, définie comme l'onde interpolant trois extrema consécutifs du signal. Une fois isolée cette composante « haute fréquence », appelée première IMF pour *Intrinsic Mode Function*, celle-ci est soustraite du signal original, et la même procédure d'identification est appliquée sur le résidu. Ainsi, toutes les IMFs d'ordres supérieurs sont récursivement identifiées et l'algorithme s'arrête lorsque le résidu après extraction de la $K^{\text{ème}}$ composante ne présente au plus que trois extrema. Par construction, on retrouve de façon rigoureusement exacte le signal original en sommant le résidu et l'ensemble des IMFs obtenues, selon :

$$x(t) = m_K[x](t) + \sum_{k=1}^K d_k[x](t), \quad (2.82)$$

où, $d_k[x](t)$ représente l'IMF d'ordre k , et $m_K[x](t)$, le résidu de fin de décomposition. Plus précisément, la procédure d'extraction des IMFs, dite de tamisage (*sifting process* en anglais) est elle-même itérative et procède en cinq étapes :

1. Identifier les indices des points extrema du signal : les minima notés $\{t_-\}$ et les maxima notés $\{t_+\}$;
2. Interpoler la série des $\{t_-\}$ (respectivement celle des $\{t_+\}$) par une enveloppe inférieure $e_-(t)$ (respectivement supérieure $e_+(t)$) ;
3. Calculer l'enveloppe moyenne $m(t) = (e_-(t) + e_+(t))/2$;
4. Extraire le détail $d(t) = x(t) - m(t)$;
5. Réitérer les étapes 1 à 4 sur le détail $d(t)$.

Le critère d'arrêt est atteint lorsque l'enveloppe moyenne $|m(t)|$ est en tous points inférieure à un seuil fixé.

La décomposition modale empirique a pour principaux atouts d'être adaptative (entièrement pilotée par le signal analysé) et locale en temps. Elle se prête donc naturellement bien à l'analyse de signaux non stationnaires, et à l'inverse des décompositions linéaires usuelles (type ondelettes (2.11)), elle ne requière le choix d'aucun filtre a priori. À son désavantage, l'EMD souffre de n'être définie qu'au travers d'un algorithme, et se prête mal à une formalisation mathématique qui permettrait de mieux en comprendre le mécanisme, d'en évaluer finement les performances et les limites, et aussi de caractériser plus rigoureusement les IMFs. Cette méthode de décomposition des signaux n'en reste pas moins très attractive et complémentaire de techniques plus classiques.

C'est ce potentiel que nous avons cherché à montrer dans une série de publications [21, 22, 25–27, 32, 73–76], dont beaucoup sont le fruit du travail de thèse de Gabriel Rilling [Rilling, 2007], sous la direction de Patrick Flandrin. L'article partiellement reproduit ci-après reprend dans le détail les principes de l'EMD. Puis, nous reviendrons sur ce travail dans le chapitre 3 consacré aux estimateurs, pour dresser dans le cas de l'analyse de bruits colorés, une étude comparée avec les décompositions en ondelettes.

Ce paragraphe est extrait de l'article :

Empirical Mode Decomposition as a Data-Driven Wavelet-Like Expansions

P. FLANDRIN and P. GONÇALVES

International Journal of Wavelets, Multiresolution and Information Processing

Volume 2, No. 4, pp. 477–496, 2004

2.4.1 Principle

EMD basics The simplest model for an oscillatory waveform $x(t)$ is given by circular functions of the type $a \cos 2\pi f_0 t$, or by a combinations of them. Such “Fourier modes” are of particular interest in the case of stationary signals and linear systems, since they are eigenfunctions of linear time-invariant operators. However, many physical situations are known to undergo nonstationary and/or nonlinear behaviors, thus calling for more elaborate and more meaningful representations. In this respect, one can, e.g., think of representing signals in terms of amplitude and frequency modulated (AM-FM) components such that

$$x(t) = \sum_{k=1}^K a_k(t) \cos \varphi_k(t). \quad (2.83)$$

The rationale for such a modelling is to compactly encode possible nonstationarities in a time variation of the amplitudes and frequencies of Fourier-like modes. More generally, signals may also be generated by nonlinear systems for which oscillations are not necessarily associated with circular functions, thus suggesting decompositions of the form

$$x(t) = \sum_{k=1}^K x_k(t), \quad (2.84)$$

where the components $x_k(t)$ may present both amplitude variations and non-harmonic, time-varying, oscillations, while being “independent” of each other in some way.

EMD [Huang et al., 1998] is a technique which has been designed primarily for obtaining representations of this type in the case of signals which are oscillatory (possibly nonstationary and/or generated by a nonlinear system), in some automatic, fully data-driven, way. In a nutshell, the starting point of EMD is to consider oscillatory signals at the level of their local oscillations and to formalize the idea that :

« signal = fast oscillations superimposed to slow oscillations »

and to iterate on the slow oscillations component considered as a new signal.

More precisely (see Figure 2.17), if we look at the evolution of a signal $x(t)$ between two consecutive local extrema (say, two minima occurring at times t_- and t'_-), we can heuristically define a (local) “high-frequency” part $\{d(t), t_- \leq t \leq t'_-\}$. This *detail* $d(t)$ corresponds to the oscillation terminating at the two minima and passing through the maximum which necessarily exists in between them. For the picture to be complete, we also identify the corresponding (local) “low-frequency” part $m(t)$, or local *trend*, so that we have $x(t) = m(t) + d(t)$ for $t_- \leq t \leq t'_-$. Assuming that this is done in some proper way for all the oscillations composing the entire signal, we get what is referred to as an *Intrinsic Mode Function* (IMF) as well as a *residual* consisting of all local trends. The procedure can then be applied to this residual, considered as a new signal to decompose, and successive constitutive components of a signal can therefore be iteratively extracted. The only definition of such a so-extracted “component” is that it is locally (*i.e.*, at the scale of one single oscillation) in the highest frequency band.

Given a signal $x(t)$, the effective algorithm of EMD can therefore be summarized as the following main loop [Huang et al., 1998] :

1. identify all extrema of $x(t)$

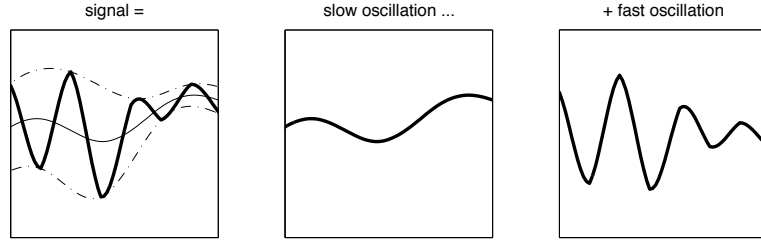


FIGURE 2.17 – A schematic illustration of the EMD idea. The original signal (thick line in the left diagram) is viewed as the superposition of a slow oscillation (middle diagram) and a fast oscillation (right diagram), the slow oscillation being obtained as the mean of two envelopes passing through the signal extrema. Once the slow oscillation has been identified, it can be considered as a new signal onto which the same procedure can be applied.

2. interpolate between minima (resp. maxima), ending up with some “envelope” $e_{\min}(t)$ (resp. $e_{\max}(t)$)
3. compute the average $m(t) = (e_{\min}(t) + e_{\max}(t))/2$
4. extract the detail $d(t) = x(t) - m(t)$
5. iterate upon the residual $m(t)$

In practice, the above procedure has to be refined by a *sifting* process, an inner loop that iterates steps (1) to (4) upon the detail signal $d(t)$, until this latter can be considered as zero-mean according to some stopping criterion¹⁸. Once this is achieved, the detail is considered as the effective IMF, the corresponding residual is computed and only then, step (5) applies. Eventually, the original signal $x(t)$ is first decomposed through the main loop as

$$x(t) = d_1(t) + m_1(t), \quad (2.85)$$

and the first residual $m_1(t)$ is itself decomposed as

$$m_1(t) = d_2(t) + m_2(t), \quad (2.86)$$

so that

$$\begin{aligned} x(t) &= d_1(t) + m_1(t) \\ &= d_1(t) + d_2(t) + m_2(t) \\ &= \dots \\ &= \sum_{k=1}^K d_k(t) + m_K(t). \end{aligned} \quad (2.87)$$

By construction, the number of extrema decreases when going from one residual to the next, thus guaranteeing that the complete decomposition is achieved in a finite number of steps (typically, K is at most $O(\log_2 N)$ for N data points). Moreover, the whole decomposition being only based on elementary subtractions, it obviously allows for a perfect reconstruction of the initial signal $x(t)$, given the collection of details $\{d_k(t), k = 1, \dots, K\}$ and the residual $m_K(t)$.

Modes and residuals have been heuristically introduced on “spectral” arguments, but this must not be considered from a too narrow perspective. First, it is worth stressing the fact that, even in the case of harmonic oscillations, the high vs. low frequency discrimination mentioned above applies only *locally* and corresponds by no way to a pre-determined sub-band filtering. Indeed, selection of modes rather corresponds to an automatic and adaptive (data-driven) time-variant filtering. In this direction, Figure 2.18 shows an example, where a signal composed of two AM-FM components significantly overlapping in time and frequency is successfully decomposed by the method.

¹⁸. It is not the purpose of this introduction to address algorithmic issues which have been considered in some detail elsewhere [Huang et al., 1998], [73]. Let us just mention that the main reason for which a proper IMF has to be zero-mean is that this is a pre-requisite for its AM-FM demodulation with Hilbert transform techniques, a post-processing aspect of EMD that will not be considered here.

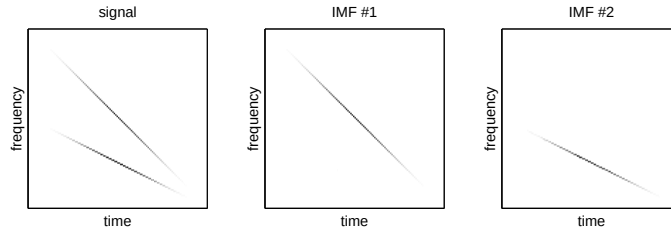


FIGURE 2.18 – An example of EMD-based time-variant filtering. A signal consisting in the superposition of two amplitude modulated linear chirps is represented in the time-frequency plane (left diagram), revealing an overlap of the two components in both time and frequency. The EMD of this signal ends up with essentially two IMF's whose time-frequency signatures (middle and right diagrams) directly correspond to each of the components. (All time-frequency transforms are reassigned spectrograms.)

Another example that puts emphasis on the potentially “non-harmonic” nature of EMD is given in Figure 2.19. In this case, the analyzed signal is composed of a “low frequency” triangular waveform to which is superimposed a “middle frequency” sine wave whose amplitude is quickly (linearly) decaying and a “high frequency” triangular waveform whose amplitude is slowly (linearly) growing. Both linear and nonlinear oscillations are effectively identified and separated by EMD, whereas any “harmonic” analysis (Fourier, wavelets, ...) would end up in the same context with a much less compact and physically less meaningful decomposition.

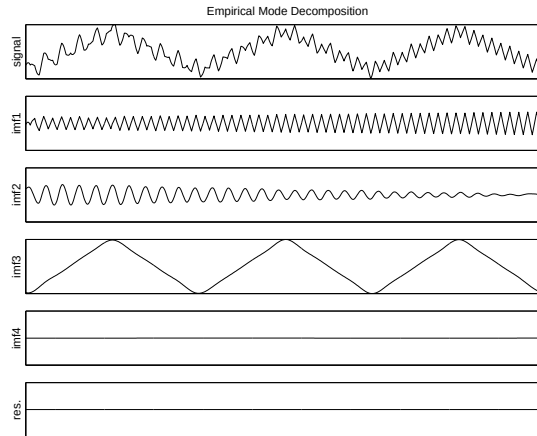


FIGURE 2.19 – An example of an EMD-based “non-harmonic” decomposition. The original signal (top diagram) is decomposed into three main components revealing both linear and nonlinear oscillations (one sine wave (IMF 2) and two triangular waveforms (IMFs 1 and 3)), together with possible amplitude modulations.

2.4.2 EMD vs. wavelets : a qualitative appreciation

The brief EMD description sketched above is somehow reminiscent of wavelet analysis [Mallat, 1998], at least in the sense that it corresponds to an iterative decomposition scheme aimed at progressively scrutinizing coarser and coarser scales in a signal. In parallel with what has been said on the basic idea underlying EMD, we can say that the starting point of a wavelet decomposition is to formalize the idea that :

$$\ll \text{signal} = \text{high-frequency detail superimposed to low-frequency approximation} \gg$$

and to iterate on the approximation considered as a new signal.

The way this decomposition is conventionally achieved consists in a repeated application of two filtering operations. Firstly, a high-pass (resp. low-pass) filter extracts the detail (resp. approximation)

part of the signal by identifying it as the one which lives in the highest (resp. lowest) frequency band. Secondly, the very same filters are applied to the approximation considered as a new, full-band, signal after a decimation by a factor of two. Compared to EMD, the two main differences are (i) that splitting the signal is achieved at each step of the decomposition, on a pre-determined spectral basis, and (ii) that the use of linear time-invariant filters precludes the possibility of adapting to local variations of the oscillations.

However, wavelet decompositions present the advantage of being based on solid and well-understood theoretical foundations, and to be equipped with extremely efficient fast algorithms. This of course contrasts with the present situation of EMD, whose definition is only given as the output of an algorithm, and which clearly lacks from a well-established theory.

2.5 Bilan et perspectives : un point de vue personnel

Historiquement les représentations temps-fréquence continues, en offrant une lecture séquentielle du contenu spectral des signaux, ont joué un rôle pionnier déterminant dans l'analyse des processus non stationnaires. Les contributions détaillées dans ce chapitre ont eu pour cible :

- l'amélioration de cette lisibilité en termes de résolution accrue des composantes ;
- le maintien de propriétés structurelles permettant de transposer simplement au plan temps-fréquence, certains opérateurs usuels du traitement du signal ;
- un redéploiement de l'information contenue dans le signal pour mettre en évidence certaines organisations ou propriétés difficiles d'identifier autrement.

Une grande part de ces avancées repose sur la redondance des représentations temps-fréquence continues. Mais derrière cette redondance se cache un cadre mathématique rigide (plus contraint encore dans le cas des distributions bilinéaires) difficilement compatible avec les traitements nécessitant d'inverser les représentations pour revenir à une estimation du signal initial (comme par exemple le débruitage ou plus généralement tous les problèmes inverses). Pour faire face, les représentations temps-fréquence, et plus exactement les représentations temps-échelle, se sont alors orientées vers les schémas discrets, et conduit aux décompositions sur des bases orthonormées d'ondelettes. Dans le confort mathématique de ce nouveau contexte, il est devenu « facile » de reconsidérer quelques grands défis du traitement du signal et dans cet élan, le nombre considérable de progrès décisifs accomplis a détourné l'intérêt des chercheurs des représentations continues, et en particulier des distributions bilinéaires.

De surcroît, avantageées par des propriétés bien particulières, les bases d'ondelettes ont creusé le sillon des représentations parcimonieuses des signaux, auxquelles les travaux récents de Candès sur le *compressive sensing* ont donnée une dimension supplémentaire [Candès et al., 2006]. Et même si certaines approches astucieuses réussissent à combiner les deux [Flandrin and Borgnat, 2008] (ou sous un tout autre angle ceux proposés dans [69, 70]), représentations quadratiques et décompositions parcimonieuses suivent des chemins et des objectifs qui se sont radicalement éloignés.

Cela étant, sur beaucoup de fronts, les distributions temps-fréquence font valoir la redondance comme une source d'information statistique capable d'accroître les performances des estimateurs et des classifieurs. Nous en avons une excellente illustration dans les travaux sur la reconnaissance de forme [Honeiné et al., 2005], la classification [Honeiné and Richard, 2009, Honeiné et al., 2007] [50], la détection [Chassande-Mottin and Flandrin, 1999] ou encore sur les tests de stationnarité [Borgnat et al., 2009]. Or, s'agissant plus particulièrement de la diffusion des représentations (affine) temps-fréquence (Paragraphe 2.3), je pense qu'il y a là un potentiel latent qui mériterait une plus grande attention. Hormis les applications de classification et de contrôle de complexité [50], la richesse offerte par le choix de la fonction de conductance (notamment dans le cas affine), n'a jamais été vraiment explorée. Dans cette veine, on pourrait par exemple imaginer d'étendre le principe de rétro-diffusion aux lignes de maxima locaux des décompositions en ondelette continues [Muzy et al., 1993] ; ou bien, dans une toute autre direction, s'inspirer des algorithmes de *matching pursuit* pour identifier des familles d'analyse adaptées (et non fixées a priori) minimisant un critère entropique global. Ce pourrait être une piste permettant d'hybrider de façon plus souple *local cosine transforms* et transformées en ondelettes comme cela est par exemple fait en codage audio dans [Molla and Torrèsani, 2005].

C'est un enthousiasme plus tempéré que m'inspire les distributions pseudo-Wigner affines. D'abord parce que ce sont des objets relativement bien circonscrits, les perspectives d'évolution sont moins ouvertes. Ensuite, ces versions à temps glissant offrent un gain essentiellement algorithmique par rapport aux distributions de Wigner affine, si bien que leur retentissement est fortement conditionné à l'intérêt qui est porté à ces dernières.

Dans ce paysage temps-fréquence mathématiquement bien formalisé, la décomposition modale empirique est apparue comme une curiosité pouvant renouveler un certain nombre de points de vue sur l'analyse harmonique. Le premier réflexe a donc été de l'appliquer, parfois sans discernement

ni précautions, à toutes sortes de situations où les ondelettes avaient pu se révéler décevantes... À côté de cet engouement, des études plus circonspectes ont démonté le mécanisme de l'EMD, ce qui a permis de mieux comprendre son fonctionnement, de cerner plus rigoureusement ce qu'est un « mode intrinsèque », d'en évaluer les performances et aussi d'en éprouver certaines limites. Mais, en dépit de ces efforts, l'EMD continue de ne reposer que sur un principe et de n'exister qu'à travers un algorithme, demeurant rétive à tout formalisme mathématique satisfaisant.

Pour autant, se demander « l'EMD est-elle une fausse bonne idée ? » me paraît réducteur, car certains de ses avantages sont bien réels. Mais sans doute faut-il continuer à s'interroger sur le fond théorique de l'intuition qui est à la source de cette décomposition, et sur le crédit que l'on peut accorder à un résultat qu'on ne sait ni formellement caractériser ni physiquement toujours bien interpréter.

3

Estimateurs et Modèles

3.1 Modèles d'invariances d'échelle

L'analyse des propriétés d'invariance d'échelle, a toujours été une source récurrente de motivation de mon activité de recherche. Qu'il s'agisse de mes travaux sur les distributions affines, sur les ondelettes ou sur l'EMD, tous convergent peu ou prou, vers un même objectif : identifier des lois d'échelle qui structurent une grande variété de signaux physiques. L'ouvrage collectif [4] offre à cet égard, un panorama très diversifié de l'actualité internationale sur ce domaine.

Avant donc d'aborder les différents outils élaborés, le paragraphe qui suit pose quelques définitions et repères terminologiques relatifs aux lois d'échelle.

Ce paragraphe et le suivant sont repris et adaptés du chapitre :

Scale Invariance and Wavelets

P. FLANDRIN, P. GONÇALVES and P. ABRY

Scaling, Fractals and Wavelets, Chap. 2, pp. 71-102, 2009.

Les processus présentant des spectres en loi de puissance (souvent regroupés sous le vocable réducteur, mais générique, de processus “en $1/f$ ”) apparaissent dans des domaines très variés, allant de l'hydrologie [Beran, 1994] à la finance [Mandelbrot, 1997], en passant par les télécommunications [Leland et al., 1994, Park and Willinger. W., 2000], la turbulence [Frisch, 1995, Castaing, 1996] ou la biologie [Teich et al., 2001], et bien d'autres domaines encore [Wornell, 1996]. Les caractéristiques de ces processus sont liées à des concepts tels que fractalité, auto-similarité ou longue dépendance et, bien que ces différentes notions ne soient pas équivalentes, elles partagent toutes une caractéristique commune : celle de gommer l'idée d'une structure liée à une échelle de temps privilégiée pour la remplacer par celle d'une mise en relation, invariante, entre différentes échelles.

L'étude des processus invariants en échelle pose de nombreux problèmes de modélisation, d'analyse et de traitement pour lesquels peu d'outils étaient disponibles jusqu'à un passé récent [Beran, 1994]. La possibilité concrète de manipuler commodément ces processus s'est récemment trouvée renforcée par l'apparition des techniques multirésolution.

D'un point de vue qualitatif, l'idée de spectre en $1/f$ recouvre les situations où, étant donné un signal temporel observé, sa densité spectrale de puissance¹ empirique se comporte en $\mathcal{S}(f) = C|f|^{-\alpha}$, avec $\alpha > 0$. D'un point de vue pratique, c'est évidemment la forme équivalente

$$\log \mathcal{S}(f) = \log C - \alpha \log |f| \quad (3.1)$$

qui est la plus parlante, le caractère “ $1/f$ ” se traduisant par une droite en coordonnées log-log.

1. Adaptation stochastique du module carré de la transformée de Fourier (2.3) appliquée aux processus stationnaires.

Dans le cas d'observations physiques, parler de comportement spectral en $1/f$ n'a généralement de sens que relativement à une bande d'analyse. Si l'on convient d'introduire sur la demi-droite des fréquences positives deux fréquences (réglables) f_{bf} et f_{hf} telles que $0 < f_{bf} < f_{hf} < +\infty$, on est conduit à une classification en trois régimes selon que l'on observe le comportement en $1/f$ dans l'une (au moins) des trois régions que cette partition définit.

Considérons chacun des cas :

1. $f_{bf} \leq f \leq f_{hf}$: on est là dans une situation où, dans un domaine passe-bande, on observe simplement une décroissance algébrique de la densité spectrale, sans fréquence prédominante, on parle alors de processus en $1/f$;
2. $f_{hf} \leq f \leq \infty$: le caractère $1/f$ est prépondérant dans la limite des hautes fréquences et rend compte de la régularité locale des trajectoires, de leur variabilité, ou de leur dimension fractale ;
3. $0 \leq f \leq f_{bf}$: la loi de puissance du spectre intervient ici dans la limite des basses fréquences et il en résulte une divergence à l'origine de la densité spectrale. Si $0 < \alpha < 1$, cette divergence correspond à une décroissance algébrique de la fonction de corrélation suffisamment lente pour que cette dernière ne soit pas sommable : il y a dépendance à longue portée ou longue mémoire.

Ces trois régimes recouvrent, en fait, des réalités de nature différente et n'ont aucune raison de se manifester de façon simultanée. Leur dénominateur commun est cependant d'être tous trois liés à une idée commune d'*invariance d'échelle*, selon laquelle — dans une gamme d'échelles et à une renormalisation près — les propriétés du tout sont semblables à celles des parties (auto-similarité). En effet, un spectre en loi de puissance appartient à la classe des fonctions *homogènes*. Sa *forme* est par suite insensible aux changements d'échelle, au sens où, pour tout $c \in \mathbb{R}^+$, on a :

$$\mathcal{S}(f) = C |f|^{-\alpha} \Rightarrow \mathcal{S}(cf) = C |cf|^{-\alpha} = c^{-\alpha} \mathcal{S}(f). \quad (3.2)$$

Sachant que, par transformation de Fourier, une dilatation ou compression dans l'espace des fréquences se traduit par une compression/dilatation dans l'espace temporel, on est ainsi en droit de s'attendre à ce que les processus "en $1/f$ " soient étroitement couplés aux processus auto-similaires.

3.1.1 Auto-similarité.

Définition 3.1.1 *Un processus $x = \{x(t), t \in \mathbb{R}\}$ est dit auto-similaire d'index $H > 0$ si, pour tout $c \in \mathbb{R}$, on a l'égalité en loi des trajectoires [Beran, 1994, Samorodnitsky and Taqqu, 1994] :*

$$\{x(t), t \in \mathbb{R}\} \stackrel{\mathcal{L}}{=} \{c^H x(t/c), t \in \mathbb{R}\}, \forall c > 0. \quad (3.3)$$

Ainsi défini, un processus auto-similaire ne possède aucune échelle caractéristique, dans la mesure où il reste (statistiquement) identique à lui-même après un changement d'échelle quelconque. Si, d'un point de vue théorique, l'auto-similarité est susceptible de s'étendre des plus grandes aux plus petites échelles, la définition ci-dessus doit généralement s'accorder d'un domaine d'échelles pour lequel l'invariance a un sens : Par exemple, la durée toujours finie d'une observation fixe l'échelle maximale atteignable, alors que la résolution toujours finie d'un capteur limite l'échelle la plus fine.

On notera que si un processus est auto-similaire, il est nécessairement non stationnaire, et tous ses moments finis vérifient la relation :

$$\mathbb{E} |x(t)|^q = \mathbb{E} |x(1)|^q |t|^{qH}. \quad (3.4)$$

On ne peut donc, *stricto sensu*, associer un spectre à un processus auto-similaire. Il existe néanmoins une sous-classe intéressante de la classe des processus auto-similaires, s'approchant en un sens de celle des processus stationnaires : c'est celle des processus à *accroissements stationnaires*, définie par [Beran, 1994, Samorodnitsky and Taqqu, 1994].

Définition 3.1.2 *Un processus $x = \{x(t), t \in \mathbb{R}\}$ est dit à accroissements stationnaires si et seulement si, pour tout $\theta \in \mathbb{R}$, la loi du processus des accroissements :*

$$x^{(\theta)} := \left\{ x^{(\theta)}(t) := x(t + \theta) - x(t), t \in \mathbb{R} \right\}$$

ne dépend pas de t .

Cette définition est bien sûr extensible à des accroissements d'ordre plus élevé ("accroissements d'accroissements"). Dans cette définition, le paramètre θ joue le rôle d'échelle de temps relativement à laquelle on étudie le processus x . L'auto-similarité de ce dernier se traduit sur ses accroissements par :

$$\{x^{(\theta)}(t), t \in \mathbb{R}\} \stackrel{\mathcal{L}}{=} \{c^H x^{(\theta/c)}(t/c), t \in \mathbb{R}\}, \forall c > 0. \quad (3.5)$$

Coupler auto-similarité d'index H et accroissements stationnaires contraint le paramètre H à exister dans la gamme $0 < H < 1$ et la fonction de covariance d'un processus (centré et nul à l'origine) à être de la forme :

$$\mathbb{E}x(t)x(s) = \frac{\sigma^2}{2} (|t|^{2H} + |s|^{2H} - |t - s|^{2H}), \quad (3.6)$$

où $\sigma^2 := \mathbb{E}|x(1)|^2$.

De plus, la fonction de corrélation du processus "accroissements" $x^{(\theta)}$ s'écrit :

$$\mathbb{E}x^{(\theta)}(t)x^{(\theta)}(t + \tau) = \frac{\sigma^2}{2} |\tau|^{2H} \left(\left| 1 + \frac{\theta}{\tau} \right|^{2H} + \left| 1 - \frac{\theta}{\tau} \right|^{2H} - 2 \right).$$

On peut alors en explorer les comportements asymptotiques dans les deux limites, des grands et des petits τ . On montre, par exemple, que, dans la limite $\tau \rightarrow +\infty$ (i.e., $\tau \gg \theta$), la fonction d'autocorrélation décroît asymptotiquement en $\tau^{2(H-1)}$:

$$\mathbb{E}x^{(\theta)}(t)x^{(\theta)}(t + \tau) \sim \frac{\sigma^2}{2} 2H(2H - 1) \theta^2 \tau^{2(H-1)}. \quad (3.7)$$

Par dualité de Fourier, ce comportement induit une divergence spectrale algébrique d'exposant $1 - 2H$ à l'origine. Les processus auto-similaires à accroissements stationnaires se trouvent ainsi étroitement liés aux processus à longue portée. Dans l'autre limite, $\tau \rightarrow 0$ (i.e., $\tau \ll \theta$), on montre que l'on a, pour $H > 1/2$:

$$\mathbb{E}x^{(\theta)}(t)x^{(\theta)}(t + \tau) \sim \sigma^2 \theta^{2H} (1 - \theta^{-2H} |\tau|^{2H}). \quad (3.8)$$

Ce comportement caractérise la régularité locale de chacune des trajectoires du processus x . Les paragraphes suivants explicitent les notions associées à chacune de ces limites : dépendance à longue portée d'une part, régularité locale d'autre part.

3.1.2 Dépendance à longue portée

Définition 3.1.3 *Un processus $x = \{x(t), t \in \mathbb{R}\}$, stationnaire au second-ordre, est dit « à longue portée » (ou « à longue dépendance », ou encore « à longue mémoire ») si sa fonction d'autocorrélation $c_x(\tau) := \mathbb{E}x(t)x(t + \tau)$ est telle que [Beran, 1994, Samorodnitsky and Taqqu, 1994] :*

$$c_x(\tau) \sim c_r |\tau|^{-\beta}, \tau \rightarrow +\infty, \quad (3.9)$$

avec $0 < \beta < 1$. De façon équivalente, la densité spectrale de puissance d'un processus à longue portée est telle que :

$$\mathcal{S}_x(f) := \int c_x(\tau) e^{-i2\pi f\tau} d\tau \sim c_f |f|^{-\gamma}, f \rightarrow 0, \quad (3.10)$$

avec $0 < \gamma = 1 - \beta < 1$ et c_f une constante.

Sous sa forme (3.9), la longue portée est liée au fait que la décroissance (algébrique) de la fonction de corrélation aux temps longs est lente, au point de ne pas être sommable : il y a ainsi un effet longue mémoire, au sens où des liens statistiques significatifs se maintiennent entre échantillons très éloignés. Cette situation contraste évidemment avec celle des processus markoviens à courte mémoire, caractérisés par une décroissance asymptotique exponentielle des corrélations. Par définition, l'argument de l'exponentielle implique une échelle de temps caractéristique, alors que la décroissance algébrique n'en requiert pas : ils s'agit donc bien d'un comportement en loi d'échelle. Par transformation de Fourier, la longue portée implique une divergence spectrale en l'origine $\mathcal{S}_x(0) = \infty$, que l'on retrouve également avec la loi de puissance (3.10). Enfin, même si la propriété de dépendance à longue portée existe et se définit indépendamment de celle d'auto-similarité, la relation (3.7) exprime un lien fort entre ces deux notions, puisqu'elle indique que le processus « accroissement » d'un processus autosimilaire à accroissements stationnaires présente, si $H > 1/2$, de la dépendance à longue portée.

3.1.3 Régularité locale.

Plutôt qu'au comportement de l'autocovariance aux temps longs, intéressons-nous maintenant à celui des temps courts. Soit x un processus aléatoire, stationnaire au second ordre, dont la fonction d'autocorrélation se comporte à l'origine en

$$\mathbb{E} x(t)x(t+\tau) \sim \sigma^2(1 - C|\tau|^{2h}), \tau \rightarrow 0, 0 < h < 1. \quad (3.11)$$

Il est alors possible de montrer que cette signature de la covariance à l'origine est équivalente à un comportement algébrique de la variance des accroissements calculés pour des pas de temps courts :

$$\mathbb{E} |x(t+\tau) - x(t)|^2 \sim C|\tau|^{2h}, \tau \rightarrow 0, 0 < h < 1,$$

Cette relation nous renseigne sur la régularité locale de chaque réalisation particulière de trajectoire du processus x . Pour des processus gaussiens par exemple, elle indique que celles-ci sont continues d'ordre $h' < h$. Lorsque $0 < h < 1$, cela signifie que les trajectoires de x sont partout continues mais nulle part dérivables. Plus précisément, pour décrire cette régularité locale on recourt à la notion d'exposant de Hölder :

Définition 3.1.4 *Un signal $x(t)$ est de régularité höldérienne $h \geq 0$ en t_0 s'il existe un polynôme local $P_{t_0}(t)$ de degré $n = \lfloor h \rfloor$ et une constante C tels que :*

$$|x(t) - P_{t_0}(t)| \leq C|t - t_0|^h. \quad (3.12)$$

Dans le cas où $0 \leq h < 1$, la partie régulière de $x(t)$ se réduit à $P_{t_0}(t) = x(t_0)$, conduisant à la caractérisation de la régularité höldérienne de $x(t)$ en t_0 par la relation :

$$|x(t_0 + \theta) - x(t_0)| \leq C|\theta|^h. \quad (3.13)$$

Moralement, l'exposant de Hölder offre une mesure de la rugosité de la trajectoire de x : plus l'exposant est proche de 1, plus la trajectoire est douce et régulière ; plus il est proche de 0, plus celle-ci paraît rugueuse sous l'effet d'une forte variabilité.

Le comportement asymptotique algébrique de la variance des accroissements traduit donc une régularité höldérienne $h' < h$ des trajectoires du processus x . Cette correspondance entre comportement asymptotique algébrique des accroissements et régularité locale reste valable même si le processus x n'est plus stationnaire, mais seulement à accroissements stationnaires. Les processus dont les trajectoires ont une régularité locale homogène et constante sont dits *monofractals*. Pour les processus autosimilaires à accroissements stationnaires, il est facile de voir qu'en partant de (3.7), les accroissements ont un comportement algébrique, pour tout θ en général et donc en particulier dans la limite $\theta \rightarrow 0$:

$$\mathbb{E} |x(t+\theta) - x(t)|^2 = \mathbb{E} |x(\theta) - \underbrace{x(0)}_{=0}|^2 = \sigma^2 |\theta|^{2H}. \quad (3.14)$$

Tandis que la relation (3.8) correspond bien à une autocovariance de type (3.11) pour le processus des accroissements. Un processus autosimilaire à accroissements stationnaires, tout comme le processus accroissement associé, a donc pour régularité locale uniforme (i.e., en moyenne et partout) $h' < H$.

3.1.4 Mouvement Brownien fractionnaire : paradigme de l'invariance d'échelle ...

Le mouvement Brownien fractionnaire (mBf) est un modèle simple et parmi les plus répandus, de processus auto-similaire [Mandelbrot and Van Ness, 1968]. Il est caractérisé par un exposant réel $0 < H < 1$, appelé exposant de Hurst.

Définition 3.1.5 *Le mBf $B_H = \{B_H(t), t \in \mathbb{R}; B_H(0) = 0\}$, est le seul processus gaussien centré qui soit auto-similaire d'indice H et à accroissements stationnaires.*

L'auto-similarité et la stationnarité des accroissements garantissent que la fonction de covariance du mBf vérifie bien la forme (3.6), le caractère gaussien assurant quant à lui, que la loi du mBf est entièrement déterminée par cette structure de covariance. La fonction d'auto-corrélation des accroissements qui s'en déduit :

$$\mathbb{E} (B_H(t + \theta) - B_H(t)) (B_H(t) - B_H(t - \theta)) = \sigma^2 (2^{2H-1} - 1) |\theta|^{2H},$$

démontre que le mBf est une généralisation du mouvement Brownien ordinaire (ou processus de Wiener) – dont on retrouve l'indépendance des accroissements lorsque $H = 1/2$ – mais qui induit une corrélation positive (persistance) ou négative (anti-persistance) selon que $H > 1/2$ ou $H < 1/2$.

Définition 3.1.6 *On appelle bruit Gaussien fractionnaire (bGf) le processus d'accroissements $G_{H;\theta} := \{G_{H;\theta}(t), t \in \mathbb{R}\}$, défini par :*

$$G_{H;\theta}(t) := \frac{1}{\theta} B_H^{(\theta)}(t), \quad \theta > 0, \quad (3.15)$$

où B_H est un mBf.

C'est, par construction, un processus stationnaire, partout continu mais nulle part dérivable, que l'on peut voir comme une extension du bruit blanc gaussien. Bien qu'on ne puisse pas envisager sans certaines précautions le passage à la limite $\theta \rightarrow 0$ dans la définition (3.15), on voit d'après (3.7) et (3.15) que le bGf se comporte pour les « petits » pas d'accroissement comme :

$$c_{G_{H;\theta}}(\tau) := \mathbb{E} G_{H;\theta}(t) G_{H;\theta}(t + \tau) \sim \frac{\sigma^2}{2} 2H(2H - 1) \tau^{2(H-1)}, \tau \gg \theta.$$

Ce comportement indique, d'une part que le bGf présente de la mémoire longue, d'autre part que la densité spectrale de puissance du bGf est proportionnelle à $|f|^{-(2H-1)}$. On peut ainsi justifier par divers arguments (type intégration/dérivation) [Flandrin, 1992] que le mBf lui-même a un « spectre moyen » de la forme $\mathcal{S}_{B_H}(f) \propto |f|^{-(2H+1)}$.

Parallèlement à son rôle d'exposant spectral, le paramètre H contrôle également la régularité höldérienne des trajectoires du mBf et du bGf, qui s'avère valoir $h' < H$ en tout point. À cette (ir)régularité s'associe naturellement pour les processus gaussiens une notion de fractalité, la dimension de Hausdorff des trajectoires valant $\dim_H \text{graph}(B_H) = 2 - H$ (pour une définition précise de la dimension de Hausdorff, on pourra consulter les Chapitres 1 et 4 de l'ouvrage [4]).

Il se trouve ainsi que le mBf présente l'avantage (ou l'inconvénient) d'être globalement auto-similaire sur tout l'axe des fréquences, le seul paramètre H rendant compte, suivant les besoins, de l'un ou l'autre des trois régimes pré-cités : autosimilarité, longue mémoire, régularité locale. Du point de vue de la modélisation, le mBf apparaît donc comme un point de départ particulièrement intéressant (comme cela peut être le cas du bruit blanc gaussien dans des contextes stationnaires). Cette simplicité se paie évidemment en terme de réalisme, chaque fois que l'on cherche à modéliser des données réelles par ce type de processus, et de nombreuses variations et extensions ont alors été construites autour des processus auto-similaires. Les deux grandes classes de modèles évoquées ci-dessous, et sur lesquelles nous reviendrons au paragraphe 3.2.3, font l'objet détaillé de l'ouvrage collectif [4] et consistent à remplacer l'unique exposant H par une collection d'exposants.

3.1.5 Invariances d'échelle non-stationnaires

On peut d'abord envisager de modifier la relation (3.14) en introduisant une dépendance temporelle sur l'exposant de régularité hölderienne :

$$\mathbb{E} |x(t + \theta) - x(t)|^2 \sim C(t) |\theta|^{2h(t)}, \theta \rightarrow 0.$$

Lorsque $0 < h(t) < 1$ est une fonction déterministe suffisamment régulière, on dit que le processus x est *multifractionnaire* ou *localement autosimilaire* (si x est gaussien), pour exprimer l'idée qu'autour de t , $x(t)$ « ressemble » à un mBf de paramètre $H = h(t)$. La régularité locale n'est plus une quantité homogène le long de la trajectoire mais au contraire, variable selon $h(t)$, permettant ainsi de décrire l'évolution (temporelle ou spatiale) de la rugosité. Ouvrons ici une parenthèse chronologique, pour rappeler qu'une première généralisation de H à une fonction douce $h(t)$ a été proposée dans [18, 38, 39], dans le cas des fonctions de Weierstrass, puis dans le cas de mBf continus par morceaux. Notre démarche d'alors, totalement intuitive, anticipait sur les résultats théoriques ultérieurs [Peltier and Levy-Vehel, 1995, Daoudi et al., 1998, Benassi et al., 1998a, Benassi et al., 1998b, Benassi et al., 2000] (voir aussi les chapitres 5 et 6 dans [4]), qui devaient préciser les conditions de régularité sur $h(t)$ garantissant l'existence des processus multifractionnaires, et démontrer des relations de densité entre espaces selon lesquelles la trajectoire d'un mouvement brownien multifractionnaire admet en chacun de ses points un mouvement Brownien fractionnaire tangent.

Lorsque $h(t)$ est une fonction fortement irrégulière (discontinue), voire elle-même un processus aléatoire tel qu'à t fixé, l'observation $h(t)$ dépend de la réalisation de x , le processus x est dit *multifractal*. Les variations de régularité ne sont alors plus décrites par la connaissance de $h(t)$ (impossible à identifier) mais par un *spectre multifractal* $D(h)$ qui caractérise la dimension de Hausdorff de l'ensemble des points t où $h(t) = h$ (voir par exemple le Chapitre 4 dans [4]). L'une des conséquences majeures de la multifractalité des processus réside dans le fait que des quantités, usuellement appelées *les fonctions de partition*, se comportent dans la limite des petites échelles comme des lois de puissance :

$$\frac{1}{n} \sum_{k=1}^n |x^{(\tau)}(t + k\tau)|^q \simeq c_q |\tau|^{\zeta(q)}, |\tau| \rightarrow 0. \quad (3.16)$$

Pour des processus à accroissements stationnaires, ces moyennes temporelles peuvent être vues comme des estimateurs empiriques des moyennes d'ensemble $\mathbb{E} |x^{(\tau)}(t)|^q$. La relation (3.16) ci-dessus est donc réminiscente de l'équation (3.4), conséquence fondamentale de l'autosimilarité. Une différence essentielle existe cependant, les exposants $\zeta(q)$ ci-dessus n'ont a priori aucune de raison de présenter le même comportement linéaire en qH . En d'autres termes, la description des lois d'échelles dans les données se complique et passe d'un unique exposant à toute une collection d'exposants. La transformée de Legendre-Fenchel calculée sur les exposants $\zeta(q)$ mesurés, constitue alors un estimateur possible du spectre multifractal d'un processus.

Encore une fois, les visites guidées proposées aux Chapitres 1, 3 et 4 de [4] sont une excellente porte d'entrée vers l'abondante littérature consacrée aux processus multifractals, leur identification et à leurs fondements mathématiques.

3.2 Ondelettes et invariances d'échelle

Nous allons à présent voir comment exploiter les principes de base (opérateur de changement d'échelle) et les propriétés essentielles (nombre de moments nuls, localisation) des décompositions en ondelettes, pour caractériser et mesurer simplement les phénomènes d'invariance d'échelle décrits ci-dessus. Suivant le contexte, il sera tantôt plus simple d'exprimer les résultats obtenus sur les versions continues (2.11) des décompositions en ondelette, tantôt à partir de leur forme discrète associée le cas échéant à une base orthogonale (2.15). Dans les deux cas, nous adopterons la reparamétrisation en échelle via l'équivalence formelle $a \equiv 1/f$ (où la fréquence centrale f_0 de l'ondelette mère est posée sans perte de généralité égale à 1), et noterons $\Gamma_x[j, k]$ le champ des coefficients d'ondelettes orthogonales calculés sur la discrétisation ($t \mapsto 2^{-j}k$, $a \mapsto 2^{-j}$, $(k, j) \in \mathbb{Z} \times \mathbb{Z}$) du maillage temps-fréquence de la figure 2.2(d).

3.2.1 Analyse en ondelettes de processus à invariance d'échelle

Auto-similarité.

Proposition 3.2.1 *Les coefficients d'ondelette issus de la décomposition d'un processus auto-similaire d'index H vérifient l'égalité*

$$(\Gamma_x[j, 0], \dots, \Gamma_x[j, N_j - 1]) \stackrel{\mathcal{L}}{=} 2^{-j(H+1/2)} (\Gamma_x[0, 0], \dots, \Gamma_x[0, N_j - 1]) \quad (3.17)$$

Ce résultat initialement établi pour le mouvement Brownien fractionnaire [Flandrin, 1992] et généralisé ensuite à l'ensemble des processus auto-similaires [Averkamp and Houdré, 1998], repose sur l'adéquation entre opérateur de dilatation/compression (2.26) et invariance d'échelle (3.3) propre aux processus auto-similaires. La conséquence principale de cela est que, lorsqu'ils existent, les moments d'ordre q des coefficients d'ondelette vérifient l'égalité

$$\mathbb{E} |\Gamma_x[j, k]|^q = 2^{-jq(H+1/2)} \mathbb{E} |\Gamma_x[0, k]|^q.$$

Proposition 3.2.2 *Les coefficients d'ondelettes issus de la décomposition d'un processus à accroissements stationnaires, sont stationnaires à chaque échelle 2^j .*

Pour comprendre l'origine de ce résultat, remarquons que le processus accroissement de la définition 3.1.2 s'identifie à une décomposition en ondelettes (2.11) où l'ondelette d'analyse prendrait la forme élémentaire $\psi(t) = \delta(t - 1) - \delta(t)$ (aussi appelée « ondelette du pauvre »). En fait, la structure naturellement oscillante des ondelettes admissibles ($n_\psi \geq 1$) et la covariance par translation en temps des formes convolutives (2.11), sont les deux arguments qui garantissent cette « stationnarisation » des processus à accroissements stationnaires. On s'en convainc aisément par le calcul (avec $j = 0$ pour alléger l'écriture) :

$$\begin{aligned} \Gamma_x[0, k + k_0] &= \int x(u) \psi(u - k - k_0) du = \int x(u + k_0) \psi(u - k) du \\ &= \int [x(u + k_0) - x(k_0)] \psi(u - k) du \\ &\stackrel{\mathcal{L}}{=} \int [x(u) - x(0)] \psi(u - k) du = \int x(u) \psi(u - k) du = \Gamma_x[0, k]. \end{aligned}$$

Ce résultat s'étend aux processus à accroissements d'ordre p stationnaires, mais il faut alors choisir une ondelette possédant $n_\psi \geq p$ moments nuls pour que la série de coefficients $\Gamma_x[j, k]$ obtenue soit stationnaire à chaque échelle [Averkamp and Houdré, 1998]. Cette condition sur n_ψ met en lumière le rôle d'opérateur pseudo-différentiel joué par les ondelettes : par dualité temps-fréquence, le paramètre n_ψ contrôle le comportement en zéro du gabarit $|\Psi(f)| \sim |f|^{n_\psi}$, $f \rightarrow 0$, comportement que l'on peut en première approximation, identifier à l'opérateur de dérivation d'ordre n_ψ .

Proposition 3.2.3 *Les coefficients d'ondelettes issus de la décomposition d'un processus x centré, auto-similaire d'index H , de variance finie et à accroissements stationnaires ($H - ASAS$) ont, lorsqu'ils existent, des moments d'ordre q vérifiant la loi d'échelle suivante*

$$\mathbb{E} |\Gamma_x[j, k]|^q = \mathbb{E} |\Gamma_x[0, 0]|^q 2^{-jq(H+1/2)}.$$

Ce dernier résultat découle directement du couplage des deux propositions précédentes. Pour les processus de variance finie, comme par exemple le mouvement brownien fractionnaire, cette relation prend la forme spécifique suivante :

$$\mathbb{E} |d_X(j, k)|^2 = \mathbb{E} |d_X(0, 0)|^2 2^{-j(2H+1)}. \quad (3.18)$$

S'agissant des statistiques d'ordre deux, la forme particulière (3.6) que prend la structure de covariance d'un processus $H - ASAS$ permet de déduire le comportement asymptotique de la structure de dépendance des coefficients d'ondelettes [Flandrin, 1992, Tewfik and Kim, 1992].

Proposition 3.2.4 *La structure asymptotique de covariance des coefficients d'ondelette d'un processus x centré, auto-similaire d'index H , de variance finie et à accroissements stationnaires ($H - ASAS$) prend la forme :*

$$\mathbb{E} \Gamma_x[j, k] \Gamma_x[j', k'] \approx |2^{-j}k - 2^{-j'}k'|^{2(H-n_\psi)}, \quad |2^{-j}k - 2^{-j'}k'| \rightarrow \infty,$$

Ceci montre que le nombre de moments nuls réduit d'autant la portée de la corrélation, à tel point que si $n_\psi > H - 1/2$, la dépendance à longue portée, existant dans les accroissements du processus si $H > 1/2$ est transformée en dépendance à court terme [Flandrin, 1992, Tewfik and Kim, 1992] [2].

Dépendance à longue portée. Comme on l'a vu au paragraphe 3.1.2, les processus stationnaires à "dépendance longue" se caractérisent par une décroissance très lente de leur fonction d'auto-corrélation (3.9). Ainsi, les liens statistiques forts maintenus entre des échantillons même très éloignés, $x(t)$ et $x(t + \tau)$, compliquent notablement l'étude et l'analyse de ces processus, en pénalisant, par exemple, la convergence des algorithmes reposant sur les estimateurs empiriques de moments. Nous allons voir que la décomposition en ondelettes d'un processus à longue portée permet de contourner cette difficulté puisque sous certaines conditions la série des coefficients $\Gamma_x[j, k]$ présente une dépendance à court terme. La fonction de covariance des coefficients d'ondelettes, réécrite pour faire apparaître $\Psi(f)$,

$$\mathbb{E} \Gamma_x[j, k] \Gamma_x[j', k'] = 2^{-\frac{j+j'}{2}} c_f \int \frac{\Psi(2^{-j}f) \Psi(2^{-j'}f)}{|f|^\gamma} e^{-i2\pi f(2^{-j}k - 2^{-j'}k')} df, \quad (3.19)$$

montre que son comportement asymptotique, i.e., pour les grandes valeurs de l'intervalle $|2^{-j}k - 2^{-j'}k'|$, est équivalent à celui de sa transformée de Fourier à l'origine et donc à celui du rapport :

$$\frac{|\Psi(2^{-j}f) \Psi(2^{-j'}f)|}{|f|^{-\gamma}} \underset{f \rightarrow 0}{\sim} \frac{2^{(-j-j')n_\psi} |f|^{2n_\psi}}{|f|^\gamma} = \frac{2^{-(j+j')n_\psi}}{|f|^{\gamma-2n_\psi}}.$$

Le nombre de moments nuls n_ψ de l'ondelette a donc pour effet de compenser la divergence à l'origine de la densité spectrale du processus. En choisissant une ondelette d'analyse telle que $n_\psi \geq \gamma/2$, la dépendance longue du processus est gommée dans les séquences de coefficients de la décomposition, la corrélation résiduelle (similaire à celle de la proposition 3.2.4, avec $\gamma = 2H + 1$) décroît alors d'autant plus vite que n_ψ est grand.

De l'équation (3.19) ci-dessus, on montre également que la variance des coefficients ondelettes se comporte en loi de puissance avec l'échelle d'analyse :

$$\mathbb{E} |d_X(j, k)|^2 = 2^{-j(1-\beta)} c_f \int \frac{|\Psi(f)|^2}{|f|^\gamma} df = c_0 2^{-j\gamma}. \quad (3.20)$$

Cette relation sera au cœur de la procédure d'estimation du paramètre γ .

Régularité locale. Les propriétés de régularité locale des trajectoires de processus ont été introduites dans la section 3.1.3. Leur empreinte sur les « ondelettes » ont été le plus souvent formulées le champ des coefficients discrets, étant entendu qu'elles peuvent avec plus ou moins de difficulté se décliner sur les variétés de coefficients continus.

Théorème 3.2.1 *Soient x un signal de régularité höldérienne $h \geq 0$ en t_0 , et ψ une ondelette suffisamment régulière ($n_\psi \geq \lceil h \rceil$), alors il existe une constante $c > 0$, telle que $\forall j, k \in \mathbb{Z} \times \mathbb{Z}$:*

$$|\Gamma_x[j, k]| \leq c 2^{-(\frac{1}{2}+h)j} (1 + |2^j t_0 - k|^h).$$

Réciproquement, si $\forall j, k \in \mathbb{Z} \times \mathbb{Z}$, on a

$$|\Gamma_x[j, k]| \leq c 2^{-(\frac{1}{2}+h)j} (1 + |2^j t_0 - k|^{h'}),$$

pour $h' < h$, alors x est de régularité höldérienne h en t_0 .

La démonstration de ce théorème a été proposée indépendamment par S. Jaffard [Jaffard, 1989] et M. Holschneider & P. Tchamitchian [Holschneider and Tchamitchian, 1990]. A la lumière de celui-ci, on voit, une fois encore, que c'est la dynamique de décroissance des coefficients d'ondelette à travers les échelles qui rend compte de la régularité locale de la trajectoire de x . La deuxième partie du théorème 3.2.1 montre aussi que la seule connaissance des coefficients se situant à la verticale du point singulier ($|2^j t_0 - k| = 0$) ne suffisent pas pour déterminer la régularité locale de x en t_0 . En toute rigueur, il faudrait considérer la décomposition dans son ensemble, laissant entendre qu'une singularité isolée peut affecter tous les coefficients $\Gamma_x[j, k]$ à l'intérieur du cône dit d'*influence*.

Du point de vue de l'estimation, l'implication directe du théorème 3.2.1 révèle les limites pratiques de la transformée en ondelettes orthogonales, étant en effet très peu probable que l'abscisse t_0 de la singularité coïncide avec une ligne de coefficients sur la grille dyadique. En pratique donc, c'est plus souvent un schéma d'analyse continue qui est préféré, pour lequel on a la version approchée et incomplète (implication directe) du théorème 3.2.1 :

Proposition 3.2.5 *Si x est de régularité Höldérienne $n < h < n + 1$ en t_0 , pour une ondelette d'analyse ψ ayant $n_\psi \geq h$ moments nuls, alors on a le comportement asymptotique suivant :*

$$\Gamma_x(t_0, a) \sim O(a^{h+1/2}), \quad a \rightarrow 0^+ \quad (3.21)$$

Précisons que si l'ondelette ψ n'est pas de régularité suffisante, le comportement de $\Gamma_x(t_0, a)$ est tronqué et c'est alors la valeur de n_ψ qui impose la décroissance des coefficients d'ondelette aux petites échelles.

Il ne faut toutefois pas se méprendre sur l'interprétation de la proposition 3.2.5. C'est uniquement parce que l'on s'intéresse au cas limite des résolutions infiniment grandes que l'influence de la singularité semble se localiser parfaitement en $t = t_0$. En pratique, l'intérêt de cette propriété est limité par la résolution nécessairement finie de l'échantillonnage des données, qui ne permet pas de scruter les données en-deçà d'une échelle minimale, par convention égale à 1.

Enfin, notons que dans le cas des processus stochastiques auto-similaires à accroissements stationnaires dont l'exposant de Hölder est constant le long des trajectoires, la variance (3.14) des accroissements se traduit directement sur celle des coefficients d'ondelette continus ou discrets :

$$\begin{aligned} \mathbb{E} |\Gamma_x(a, t)|^2 &\sim a^{2h+1}, & a \rightarrow 0, \\ \mathbb{E} |\Gamma_x[j, k]|^2 &\sim 2^{-j(2h+1)}, & j \rightarrow +\infty. \end{aligned} \quad (3.22)$$

Ces relations sont à mettre en regard de celles obtenues dans les cas de l'autosimilarité (3.18) et de la mémoire longue (3.20) et serviront de point de départ dans la construction des estimateurs du paragraphe suivant.

3.2.2 Estimation de lois d'échelle homogènes

Cette section, consacrée à la mise en œuvre des analyses en ondelettes pour réaliser l'étude des phénomènes d'invariance d'échelle, reprend les résultats initialement publiés dans :

Wavelets, spectrum analysis and $1/f$ processes

P. ABRY, P. GONÇALVES and P. FLANDRIN

Lecture Notes in Statistics (A. Antoniadis, ed.), 1995.

Les sections qui précèdent ont mis en relief le comportement en loi de puissance de la variance des coefficients d'ondelettes. Qu'il s'agisse de l'auto-similarité avec l'équation (3.18), de la longue mémoire avec l'équation (3.20), ou de la fractalité des trajectoires avec l'équation (3.22), ces relations cristallisent une partie du potentiel des "ondelettes" comme outil d'analyse des lois d'échelles.

Nous commencerons donc par une situation idéalisée où on suppose qu'il existe une gamme d'octaves $j_1 \leq j \leq j_2$, déjà identifiée, pour laquelle la relation :

$$j_1 \leq j \leq j_2, \quad \mathbb{E} |\Gamma_x[j, k]|^2 = C 2^{j\alpha},$$

est satisfaite, et nous allons nous concentrer sur l'estimation du paramètre α .

Estimation des paramètres d'invariance d'échelle. Pour estimer l'exposant α , une idée simple consiste à mesurer la pente d'un diagramme $\log_2 \mathbb{E} |\Gamma_x[j, k]|^2$ en fonction de $\log_2 2^j = j$. Bien sûr, en pratique, cela implique, à partir d'une seule réalisation observée, de longueur finie, d'estimer la quantité $\mathbb{E} |\Gamma_x[j, k]|^2$. Etant donné les propriétés des coefficients d'ondelette (stationnarité, faible dépendance statistique), mises en avant au paragraphe précédent, nous proposons simplement d'estimer l'espérance mathématique par la moyenne temporelle :

$$\frac{1}{n_j} \sum_k \Gamma_x[j, k]^2$$

où n_j désigne le nombre de coefficients d'ondelette disponibles à l'octave j .

La regression linéaire étant au centre de la méthode proposée, il convient de rappeler ici quelques résultats classiques, que nous adapterons ensuite aux cas particuliers des coefficients d'ondelette et de leurs statistiques.

Définition 3.2.1 Soient y_j les variables aléatoires telles que $\mathbb{E} y_j = \alpha j + b$ et définissons $\sigma_j^2 = \text{vary}_j$. L'estimateur par régression linéaire pondérée de α s'écrit :

$$\hat{\alpha} = \frac{\sum_{j=j_1}^{j_2} y_j (S_0 j - S_1) / \sigma_j^2}{S_0 S_2 - S_1^2} \equiv \sum_{j=j_1}^{j_2} w_j y_j, \quad \text{avec} \quad S_p = \sum_{j=j_1}^{j_2} j^p / a_j, \quad p = 0, 1, 2 \quad (3.23)$$

où les a_j sont des quantités arbitraires, agissant comme des poids associés aux y_j .

Les coefficients w_j satisfont naturellement les contraintes usuelles, $\sum_{j=j_1}^{j_2} j w_j = 1$ et $\sum_{j=j_1}^{j_2} w_j = 0$. On vérifie aisément que l'estimateur est sans biais, puisque, $\mathbb{E} \hat{\alpha} = \alpha$, et que sa variance s'écrit, dans le cas de variables y_j décorréelées,

$$\text{var} \hat{\alpha} = \sum_{j=j_1}^{j_2} a_j \sigma_j^2.$$

Reste à présent à préciser le choix des poids. On sait que la variance de $\hat{\alpha}$ est minimale si on prend en compte la structure de covariance des y_j dans la définition des a_j . De nouveau, dans le cas de variables y_j décorréliées, cela conduit à choisir $a_j \equiv \sigma_j^2$.

Dans le cas de l'invariance d'échelle, on utilise l'estimateur défini ci-dessus avec les variables

$$y_j = \log_2\left(\frac{1}{n_j} \sum_k \Gamma_x[j, k]^2\right) - g(j), \quad (3.24)$$

où les $g(j)$ sont des termes de correction, destinés à tenir compte du fait que $\mathbb{E}(\log(\cdot))$ n'est pas égale à $\log(\mathbb{E}(\cdot))$ et à assurer que $\mathbb{E} y_j = \alpha j + b$. Cet estimateur consiste donc simplement en une régression linéaire pondérée effectuée dans le diagramme y_j versus j , lequel sera baptisé plus tard *diagramme en échelle* dans [Veitch and Abry, 1999]. Pour que cet estimateur puisse être mis en œuvre facilement, il nous faut déterminer les $g(j)$ et σ_j^2 et de là en déduire les valeurs des a_j . Comme dans [2], nous supposons ici que les coefficients $\Gamma_x[j, k]$ sont des variables aléatoires gaussiennes, ce qui est le cas s'ils résultent de la décomposition en ondelettes d'un processus lui-même conjointement gaussien. Si, de plus, on idéalise la propriété de faible corrélation des coefficients d'ondelettes (cf. Proposition 3.2.4) en « indépendance exacte », alors on peut calculer analytiquement les $g(j)$ et les σ_j^2 pour obtenir les comportements limites suivants :

$$g(j) \sim \frac{-1}{n_j \ln 2}, \quad n_j \rightarrow \infty, \quad ; \quad \sigma_j^2 \sim \frac{2}{n_j \ln^2 2}, \quad n_j \rightarrow \infty. \quad (3.25)$$

Parce que ces quantités ne dépendent que des seuls n_j – entièrement déterminés par la taille de l'échantillon analysé – elles peuvent être facilement évaluées en pratique. Nous obtenons donc, pour les processus gaussiens, un estimateur remarquablement simple à mettre en œuvre, et qui présente d'excellentes performances statistiques. En particulier :

$$\mathbb{E} \hat{\alpha} = \alpha$$

indique que l'estimateur est non-biaisé, et ceci même pour des observations de durée finie, et la variance, calculée avec le choix $a_j = \sigma_j^2$,

$$\text{Var} \hat{\alpha} = \sum_{j=j_1}^{j_2} \sigma_j^2 w_j^2 = \frac{S_0}{S_0 S_2 - S_1^2},$$

atteint la borne de Cramér-Rao, calculée sous les mêmes hypothèses de normalité et d'indépendance [2]. Si, de plus, on précise dans cette expression la forme des $n_j = 2^{-j}n$ inhérente à l'analyse multirésolution d'un échantillon de taille n , alors la variance devient :

$$\text{Var} \hat{\alpha} = \sum_{j=j_1}^{j_2} \sigma_j^2 w_j^2 \simeq \frac{1}{n} \cdot \frac{(1 - 2^{-J})}{2^{1-j_1} F(J)} \quad (3.26)$$

où $F(J) = \ln^2 2 \cdot (1 - (J^2/2 + 2)2^{-J} + 2^{-2J})$, et $J = j_2 - j_1 + 1$ désigne le nombre d'octaves impliquées dans la régression linéaire. Ce résultat analytique montre que la variance de l'estimateur décroît en $1/n$ en dépit de la possible présence de dépendance à longue portée dans le processus analysé. Des simulations numériques effectuées sur un grand nombre de réalisations indépendantes de mouvements browniens fractionnaires, illustrent les excellentes performances de cet estimateurs et justifient *a posteriori* les hypothèses simplificatrices ayant conduit au modèle (3.26) de la variance (cf. figure 3.1).

Dans le cas de processus non gaussiens, la mise en œuvre de l'estimateur est plus délicate puisqu'on ne dispose pas d'expressions analytiques pour les $g(j)$ et les σ_j^2 . Il est néanmoins montré dans [Abry et al., 2000], que dans le cas de processus de variance finie, les variables $(1/n_j) \sum_k \Gamma_x[j, k]^2$ sont asymptotiquement gaussiennes et que par conséquent, les développements asymptotiques du cas gaussien (3.25) restent justifiés.

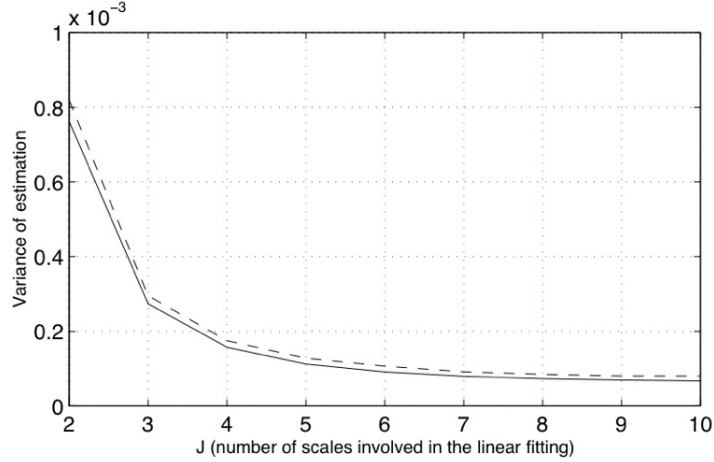


FIGURE 3.1 – Estimation par ondelettes de lois d’échelle : Variance empirique (courbes en pointillés) et théorique (trait plein). Les simulations numériques reposent sur un jeu de 3000 réalisations de bruits gaussiens en $1/f$ d’exposant $\alpha = 5/3$ et de taille $n = 2^{15}$.

Détermination de la gamme d’échelle La section précédente reposait sur l’hypothèse que la quantité $\mathbb{E} \Gamma_x[j, k]^2$ suivait exactement une loi de puissance. Cette situation idéale est rarement celle observée et ceci pour deux raisons. D’une part, des données expérimentales réelles ne seront vraisemblablement qu’approximativement décrites par les modèles d’invariance d’échelle proposés. D’autre part, certains modèles eux-mêmes n’induisent qu’un comportement en loi de puissance approché ou asymptotique — comme c’est le cas pour la dépendance à longue portée ou les processus à trajectoires fractales ; en fait seul le modèle auto-similaire induit une loi de puissance rigoureusement vérifiée.

Pour tirer le meilleur parti de l’estimateur décrit plus haut, il convient de déterminer une gamme « raisonnable » d’octaves $[j_1, j_2]$ sur laquelle effectuer la mesure. Comme il ne s’agit pas nécessairement d’extraire des données une valeur théorique de ces bornes — nous verrons au Chapitre 4 un exemple où celles-ci sont déterminées par le modèle, mais ce n’est en général pas le cas — le terme « raisonnable » résulte plutôt d’une optimisation des performances statistique de l’estimateur. Élargir la gamme d’octaves induit l’usage d’une fraction accrue des coefficients d’ondelettes disponibles résultant en une réduction de la variance d’estimation, comme indiqué par la relation (3.26) ; cela peut, a contrario, signifier une augmentation du biais d’estimation si on effectue la mesure sur une gamme où l’on s’écarte notablement du comportement en loi de puissance. Le choix de la gamme est donc guidé par l’optimisation du compromis biais-variance. Pour la longue dépendance, comme pour la régularité locale, c’est ce principe qui prévaut. Cependant, les intervalles de régression étant dans un cas borné à gauche $[j_1, +\infty[$, et dans l’autre borné à droite $]-\infty, j_2]$, l’erreur quadratique moyenne (EQM=(biais)²+variance) est plus sensible à la minimisation du terme de variance dans le cas de la longue dépendance, et à celui du biais dans le cas de la régularité locale.

Robustesse de l’approche par ondelettes L’une des principales difficultés de l’analyse des phénomènes d’invariance d’échelle est liée au fait que leurs manifestations qualitatives sont proches de celles induites par des non-stationnarités. Pendant longtemps, les modélisations de données par invariance d’échelle ont été boudées, considérées, parfois justement, comme des artefacts dues à des non-stationnarités. Les difficultés sont de deux ordres : d’une part, comme on vient de l’évoquer, identifier de l’invariance d’échelle quand il s’agit en fait de non-stationnarité ; d’autre part, ne pas réussir à déceler de l’invariance d’échelle ou à en estimer correctement les paramètres, quand des effets non stationnaires lui sont superposés. L’analyse en ondelettes a permis de construire des

solutions à ces deux problèmes. Pour le premier, en ayant recours à un test de stationnarité sur la série des exposants α estimés sur des segments temporels adjacents du signal [Veitch and Abry, 2001]. Pour le deuxième, en faisant croître graduellement le nombre de moments nuls de l'ondelette d'analyse (2.13), jusqu'à ce que l'estimation de α , devenant aveugle aux dérives non-stationnaires (régulières) superposées au processus auto-similaire, se stabilise sur l'exposant d'auto-similarité recherché.

3.2.3 Estimation de régularités locales non-stationnaires

Nous avons abordé au paragraphe 3.1.5 les invariances d'échelle non stationnaires sous leurs deux déclinaisons possibles : Celle des processus multifractionnaires où l'exposant H devient une fonction (régulière) du temps $h(t)$, et celle des processus multifractals pour lesquels les variations locales de H sont à tel point erratiques, que seule une description d'ensemble (géométrique ou statistique) permet de rendre compte de la richesse des irrégularités du processus. Dans ces deux cas de figure, nous avons cherché des prolongements, ou des alternatives aux décompositions en ondelettes pour gérer le problème de localité temporelle incompatible avec la moyenne empirique de l'estimateur (3.24).

Processus multifractionnaires

Ainsi, s'agissant de l'estimation ponctuelle de la trajectoire de $h(t)$ des processus multifractionnaires, ou de l'estimation de singularités isolées (de type *cusp*), nous avons proposé deux solutions originales.

Distributions pseudo Wigner-Ville lissées affines. La première, décrite dans une série d'articles [18, 38, 39], exploite la richesse de la classe affine des représentations temps-fréquence (à laquelle appartient d'ailleurs le scalogramme² $|\Gamma_x[j, k]|^2$ utilisée dans (3.24)). Nous avons montré que certains choix judicieux du noyau Π dans (2.36) conduisent les distributions associées $P_x(t, a; \Pi)$ à se comporter comme un lissage affine et local en temps du scalogramme $|\Gamma_x(t, a; \psi)|^2$. Utiliser $P_x(t = 2^{-j}k, a = 2^{-j}; \Pi)$ plutôt que $|\Gamma_x[j, k]|^2$ pour remplacer la somme $\sum_k |\Gamma_x[j, k]|^2$ dans (3.24) s'avère alors particulièrement payant en termes de variance d'estimation de la trajectoire $h(t)$ d'un processus multifractionnaire.

Distributions Wigner affines localisées. Dans cette même série de publications, nous avons aussi défendu un point de vue diamétralement opposé et mis l'accent sur les propriétés de localisation temporelle de certaines distributions de Wigner affines (2.27). En particulier la distribution de Unterberger [Unterberger, 1984], correspondant au paramètre $k = -1$ dans (2.28), calculée sur le *cusp* localisé d'expression $x(t) = |t - t_0|^\beta$, illustre le jeu complexe et parfois surprenant des symétries généralisées, puisqu'alors $P_x^{(-1)}(t, f) = f^{-2(\beta+1)}\delta(t - t_0)$. En concentrant tout l'effet de la singularité β à la verticale du point de rebroussement t_0 , ce comportement s'inscrit précisément en contrepoint du cône d'influence qui caractérise les décompositions en ondelettes³. En revanche, si plusieurs singularités coexistent de façon rapprochée, il faut composer avec les franges interférentielles qui, en se superposant aux lignes de *cusps*, altèrent leur décroissance homogène.

Décompositions en ondelettes multiples. Une autre piste, proposée dans [31], s'appuie sur une généralisation aux décompositions en ondelettes de la méthode d'analyse spectrale dite de « Thomson », et s'adresse aux situations où $h(t)$, tout en restant suffisamment régulière partout ailleurs, est discontinue en $t = t_0$. Pour estimer la régularité de la trajectoire de part et d'autre

2. Ce principe, redécouvert dans [Roueff and Von Sachs, 2009] dans le même contexte d'analyse, introduit la notion de « scalogramme tangent ».

3. La délocalisation liée au cône d'influence s'avère au contraire très payante dans le cas des *cusps* asymétriques où régularité à gauche et à droite du point singulier t_0 diffèrent. Le comportement distinctif des coefficients sous chacun des « demi-cônes » se situant de part et d'autre de la frontière $t = t_0$, permet justement de repérer cette différence entre forces de singularité [Mallat and Hwang, 1992].

de cette discontinuité, on ne peut recourir à une moyenne locale affine de scalogramme autour de t_0 , comme suggéré dans [39], ni a fortiori à la moyenne empirique utilisée par (3.24). Pour maintenir un biais nul, il conviendrait en fait, de limiter l'estimation de $h(t_0^-)$ (respectivement de $h(t_0^+)$) aux seuls coefficients d'ondelette $\Gamma_x(t_0^-, a)$ à gauche (respectivement $\Gamma_x(t_0^+, a)$ à droite) de t_0 . Mais par effet de levier, c'est alors la précision d'estimation qui est pénalisée. Pour compenser cette explosion de la variance sur $\hat{h}(t_0^\pm)$, nous avons proposé dans [31], d'utiliser non pas une ondelette d'analyse, mais une collection $\{\psi^m(t), m = 1, \dots, M\}$ d'ondelettes mères vérifiant entre elles une propriété de quasi-orthogonalité, $\langle \psi^m, \psi^{m'} \rangle \simeq 0, \forall m \neq m'$. Nous avons alors vérifié que cet artifice fournissait une série de coefficients $\{\Gamma_x(t_0^\pm, a; \psi^m), m = 1, \dots, M\}$ faiblement corrélés, chacun agissant moralement comme une réalisation indépendante de la même variable aléatoire. En remplaçant ensuite dans la procédure d'estimation (3.24), la somme temporelle par la « moyenne d'ensemble » $\sum_{m=1}^M |\Gamma_x(t_0^\pm, a; \psi^m)|^2$ – plutôt que par une seule réalisation $|\Gamma_x(t_0^\pm, a; \psi)|^2$ – on réduit la variance d'estimation sur $\hat{h}(t_0^\pm)$ d'un facteur approximativement égal à M .

Décomposition en ondelette et modèles d'arbres de Markov cachés. Toujours dans ce même cas de figure, mais où l'on cherche d'avantage à localiser le point de transition de la régularité locale d'un processus, qu'à estimer les régularités de part et d'autre de cette rupture, c'est une toute autre source d'inspiration qui nous a conduit dans [16, 17] à considérer des modèles markoviens appliqués aux champs aléatoires des coefficients d'ondelettes [Crouse et al., 1998]. Les décompositions en ondelettes orthogonales (2.15) associées à une analyse multirésolution [Mallat, 1989, Daubechies, 1992], s'organisent hiérarchiquement dans le plan temps-échelle selon un arbre dyadique calqué sur le maillage de la figure 2.2(d). Chaque coefficient $\Gamma_x[j, k]$ correspond à un nœud de l'arbre, où l'indice d'échelle j définit la profondeur, et l'indice temporel k la position dans l'arbre. Suivant la terminologie des graphes, à l'exception des feuilles, chaque point de branchement $\Gamma_x[j, k]$ est connecté à ses deux fils $\Gamma_x[j+1, k]$ et $\Gamma_x[j+1, k+1]$ par des arêtes. En toute généralité, on peut donc formaliser la dépendance statistique qui existe entre un nœud et ses descendants par des lois conditionnelles donnant la probabilité de l'amplitude d'un coefficient connaissant la valeur de ses parents, et ainsi propager depuis la racine vers les feuilles de l'arbre la distribution du champ aléatoire d'ondelettes. Cette approche n'est pas nouvelle. Initialement proposée dans [Crouse et al., 1998], elle y est appliquée à la segmentation d'images, à la classification et au débruitage de signaux. Mais elle prend une autre tournure dans la cas des processus auto-similaires d'exposant H homogène, puisqu'on a vu avec la Proposition 3.2.2, que les coefficients d'ondelettes issus de leur décomposition, sont des variables aléatoires identiquement distribuées et de variance $2^{-j(2H+1)} \mathbb{E} \Gamma_{\mathbf{x}}^2[\mathbf{0}, \mathbf{0}]$ à chaque échelle j . Ce comportement très structuré permet alors de paramétrer plus simplement le modèle de dépendance markovienne, en considérant, non plus l'amplitude du coefficient d'ondelette $\Gamma_x[j, k]$ – que l'on sait statistiquement déterminée par la paramètre H et la variance de la racine $\Gamma_{\mathbf{x}}[\mathbf{0}, \mathbf{0}]$ – mais les « états cachés » binaires de cette variable aléatoire :

- $E[j, k] = 0$: le coefficient $\Gamma_{\mathbf{x}}[\mathbf{j}, \mathbf{k}]$ **ne suit pas** la loi $2^{-j(H+1/2)} \Gamma_{\mathbf{x}}[\mathbf{0}, \mathbf{0}]$.
- $E[j, k] = 1$: le coefficient $\Gamma_{\mathbf{x}}[\mathbf{j}, \mathbf{k}]$ **suit** la loi $2^{-j(H+1/2)} \Gamma_{\mathbf{x}}[\mathbf{0}, \mathbf{0}]$;

définissant ainsi l'arbre de Markov caché appliqué à la décomposition en ondelettes d'un processus auto-similaire.

Ce modèle devient encore plus intéressant adapté à notre problème initial de transition brusque de l'exposant $h(t)$ au voisinage du point de rupture $t = t_0$:

$$h(t) = \begin{cases} H_0, & t \leq t_0 \\ H_1, & t > t_0, \end{cases}$$

qui nous permet d'associer un état caché à chacune de ces valeurs de $h(t)$:

- $E[j, k] = 0$: le coefficient $\Gamma_{\mathbf{x}}[\mathbf{j}, \mathbf{k}]$ suit la loi $2^{-j(\mathbf{H}_0+1/2)} \Gamma_{\mathbf{x}}[\mathbf{0}, \mathbf{0}]$;
- $E[j, k] = 1$: le coefficient $\Gamma_{\mathbf{x}}[\mathbf{j}, \mathbf{k}]$ suit la loi $2^{-j(\mathbf{H}_1+1/2)} \Gamma_{\mathbf{x}}[\mathbf{0}, \mathbf{0}]$.

avec pour distribution conditionnelle correspondante :

$$(\Gamma_{\mathbf{x}}[\mathbf{j}, \mathbf{k}] \mid E[j, k] = s) \sim \mathcal{N}(0, 2^{j(2H_s+1)}) = f_s(\Gamma_x[j, k]), \quad s = 0, 1.$$

La décomposition en ondelettes est alors modélisée par l'arbre de Markov caché où :

- la distribution des coefficients est décrite par la loi de mélange

$$f(\Gamma_{\mathbf{x}}[\mathbf{j}, \mathbf{k}] = \Gamma_x[j, k]) = \sum_{s=0}^{s=1} \mathbb{P}(E[j, k] = s) f_s(\Gamma_x[j, k]);$$

- $(E[j, k])_{j,k}$ un arbre dyadique d'états markoviens, de loi a priori $(\pi_j)_j$ et de matrice de transition d'état entre les échelles (profondeurs) j et j' : $(p_{jj'})_{j,j'}$;
- conditionnellement à leurs états cachés, les coefficients d'ondelette sont indépendants.

Notons que ce dernier point est plus souple que l'hypothèse usuellement admise de stricte indépendance entre les coefficients, alors même que la proposition 3.2.4 montre qu'il peut subsister une corrélation résiduelle entre eux.

On estime les paramètres de ce modèle par un algorithme de type EM. L'étape E est réalisée par un algorithme descendant original développé dans [17] et différent de celui proposé dans [Crouse et al., 1998] qui sans adaptation spécifique à la loi d'échelle de la Proposition 3.2.2, est limité à de faibles hauteurs d'arbre (*overflow* au delà de 7 échelles). Les paramètres H_s sont estimés à l'étape M, par une procédure adaptée du principe de maximum de vraisemblance étudié dans [Wornell and Oppenheim, 1992], desquelles on déduit ensuite les estimations $\hat{\mathbb{P}}$ et $\hat{\pi}$. Une procédure de restauration d'états, via l'algorithme revisité de Viterbi, permet enfin d'identifier le point de discontinuité t_0 , comme étant le seul point de branchement où le sous-arbre engendré par le fils gauche est entièrement dans l'état $s = 0$ et celui engendré par le fils droit, dans l'état $s = 1$.

Processus multifractals

S'agissant à présent, des processus multifractals, comme cela a déjà été évoqué, vouloir estimer leur trajectoire hölderienne $h(t)$ n'aurait statistiquement pas de sens car un seul et même exposant dominerait systématiquement. Par conséquent, la contrainte de localisation temporelle des estimateurs ne se pose pas ici avec la même acuité. En revanche, on cherchera une description probabiliste des valeurs possibles de l'exposant local h , en s'intéressant aux moments d'ordres supérieurs du processus accroissement dérivé. Sans remonter aux origines en thermodynamique et en physique des turbulences [Mandelbrot, 1974, Frish and Parisi, 1985] du formalisme multifractal, nous avons vu que la fonction de partition $\zeta(q)$ introduite dans (3.16) décrit les lois d'échelle des estimateurs empiriques $\sum_k |x^{(\tau)}(t+k\tau)|^q \sim |\tau|^{\zeta(q)}$, et que le spectre multifractal $D(h)$ s'en déduisait par simple transformée de Legendre-Fenchel, selon

$$D(h) = \sup_{q \in \mathbb{R}} (hq - \zeta(q)). \quad (3.27)$$

En nous appuyant sur les résultats de [Muzy et al., 1993, Bacry et al., 1993, Arnéodo et al., 1995, Jaffard, 2004b] et sur les travaux antérieurs des auteurs des chapitres 3 et 4 de [4], nous admettrons qu'il existe un formalisme identique construit sur les décompositions en ondelettes et qu'en substance, il suffit pour cela de remplacer les échantillons $x^{(\tau)}(t+k\tau)$ du processus accroissement par les coefficients d'ondelette $\Gamma_x[j = \log_2(\tau), k = t/\tau]$. L'estimateur du spectre multifractal s'écrit alors :

$$\begin{cases} D(h) = \sup_q (hq - \zeta(q)) \quad \text{avec} \quad \zeta(q) = \lim_{j \rightarrow \infty} \frac{1}{j} \log_2 \sum_{k=0}^{2^j-1} |\Gamma_x[j, k]|^q \\ h(q) = \frac{d\zeta(q)}{dq}. \end{cases} \quad (3.28)$$

Dans l'esprit de ce qui a été décrit au paragraphe 3.2.2 sur l'estimation de l'exposant local $h(t)$, nous avons alors proposé dans [46, 49] une première étude analytique⁴ exploratoire des performances statistiques de l'estimateur $(\hat{h}, \hat{D}(h))$. Les biais et variance d'estimation ont été évalués et empiriquement validés dans le cas du mouvement brownien fractionnaire d'exposant H (Définition 3.1.5). C'est un cas d'école simple de processus parfaitement caractérisé permettant un calcul analytique, mais un choix paradoxal puisqu'il s'agit là d'un processus mono-fractal dont le spectre dégénéré $D(h) = \delta(h - H)$, se réduit à un seul point. Ces résultats préliminaires sont donc certainement insuffisants. Cependant, la complexité qu'introduisent les formes non-linéaires, et la quasi absence – malgré des avancées significatives avec les lignes de maxima locaux d'ondelettes [Muzy et al., 1993] et leur contrepartie discrète de coefficients d'ondelette dominants (*wavelet leaders*) [Jaffard et al., 2006] – d'estimateurs multifractals sans biais (i.e non strictement convexes), rendent cet exercice difficile et peut-être même son investissement discutable.

4. Récemment, une approche empirique par *bootstrap* a été proposée pour déterminer numériquement les intervalles de confiance sur les estimations par coefficients d'ondelette dominants (*wavelet leaders*) des quantités h et $D(h)$, simultanément [Wendt et al., 2007, Wendt, 2008].

3.3 EMD et invariances d'échelle

Ce paragraphe montre comment exploiter les décompositions modales empiriques afin de caractériser des lois d'échelle homogènes, et fait écho au paragraphe précédent sur les approches ondelettes. Il s'agit d'une synthèse des articles suivants :

Empirical Mode Decompositions as data-driven wavelet-like expansions

P. FLANDRIN and P. GONÇALVES

Int. J. of Wavelets, Multiresolution and Information Processing Vol.2(4), 2004

Fractal dimension estimation : empirical mode decomposition versus wavelets

P. GONÇALVES, P. Abry, G. Rilling and P. FLANDRIN

Proceedings of IEEE Int. Conf. on Acoust. Speech and Sig. Proc., 2007

3.3.1 EMD analysis of broadband noise

In order to better compare EMD and wavelet analysis, an empirical way is to resort to extensive simulations in situations that are already well-documented from the point of view of wavelets, and which are expected to reveal specific features of EMD in terms of hierarchical extraction of components in a fluctuating signal. One such situation is provided by scaling processes for which wavelets are known to be a naturally fitted analysis tool [4].

EMD equivalent filter banks. We will here report on extensive simulations that have been carried out on fGn processes, with $H = 0.1, 0.2, \dots 0.9$. The data length has been typically set to $N = 512$ and, for each value of H , $J = 5000$ independent sample paths of fGn have been generated via the Wood and Chan algorithm. [Wood and Chan, 1994] It is worth mentioning that the present study, whose first results have been proposed in [27], generalizes therefore the one conducted independently in [Wu et al., 2001] for white noise only ($H = 0.5$) and will support consistently the findings reported in this case.

EMDs have been computed for all sample paths $\{x_H^{(j)}[n]; n = 1, \dots N\}$ (with $j = 1, \dots J$), resulting in a collection of IMFs referred to as $\{d_{k,H}^{(j)}[n]; n = 1, \dots N\}$. The maximum number of those IMFs proved to vary from one realization to the other, but none of them ended up with less than 7 modes. This value has therefore been retained in the study as the maximum considered number of modes ($k = 1, \dots 7$).

Given the so-obtained data set, a spectrum analysis has been carried out mode by mode, on the basis of the power spectrum density (PSD) $\hat{S}_{k,H}(f)$, estimated via the Fourier transform of the empirical estimate of the auto-correlation function $\hat{r}_{k,H}[m]$.

The result of this spectrum analysis is plotted in Figure 3.2, whose graphs reveal a number of striking features :

1. whatever the value of the Hurst exponent H , the behavior of the first IMF (thick line) contrasts with that of the other modes in the sense that, in a first approximation, it presents the characteristics of a high-pass filter, while the modes of higher order are much more of a band-pass nature. The (roughly half-band) high-pass character of the first mode has however to be tempered by the fact that the maximum attenuation in the stop-band is no more than 10 dB (as compared to the maximum which occurs at the Nyquist frequency $1/2$), thus corresponding to a non-negligible contribution in the lower half-band in "ultraviolet" situations ($H < 1/2$).
2. when varying H from 0.1 to 0.9, the spectrum of the last IMF ($k = 7$) is progressively turned from band-pass to more and more low-pass, in accordance with the increasing predominance of low frequencies ("infrared catastrophe").

3. in the same respect, but more generally, the energy balance between the different modes reflects quite well the behavior of the global spectrum (superimposed dashed-dotted curve) described by (3.10) : flat spectrum when $H = 1/2$ (white noise) and increasing (resp. decreasing) power-law spectrum when $H < 1/2$ (resp. $H > 1/2$).
4. for the indices $k = 2$ to 6 corresponding to band-pass IMFs, the spectra all look quite the same, up to some shifts in abscissa and ordinate, in a surprising reminiscence of what is currently observed in wavelet decompositions. [Flandrin, 1998, Mallat, 1998]

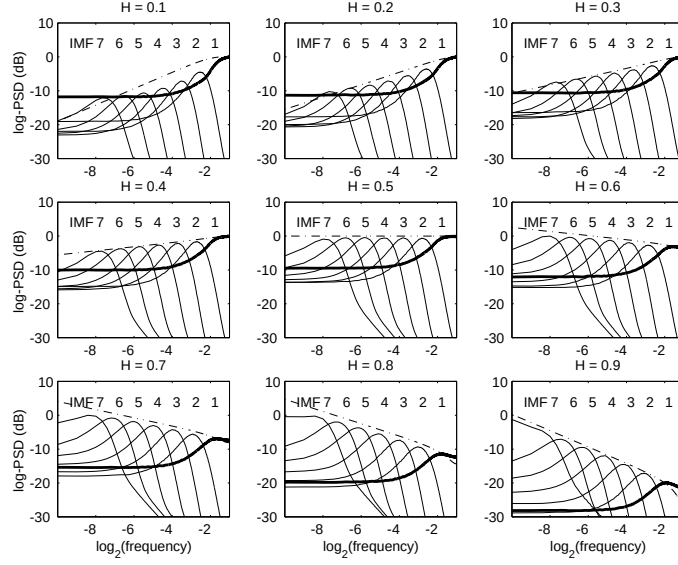


FIGURE 3.2 – *IMF power spectra in the case of fractional Gaussian noise. The logarithm of the estimated power spectrum densities (log-PSD) is plotted as a function of the logarithm of the normalized frequency for the first 7 IMFs. For each of the 9 values of the Hurst exponent H (from 0.1 to 0.9), the spectral estimates have been computed on the basis of 5000 independent sample paths of 512 data points. Theoretical PSDs of the full processes are superimposed as dashed-dotted curves.*

This last observation prompts to consider in greater detail how the different spectra are related to each other, for a given H . To this end, one can make use of the very specific structure of IMFs, according to which all extrema appear as a succession of minima and maxima with one and only one zero-crossing in between them. Measuring the average number of zero-crossings in a mode is therefore a meaningful way to access its mean frequency. A graphical representation of the average number of zero-crossings $z_H[k]$ as a function of the IMF number k is plotted in Figure 3.3, suggesting a functional relation of the form :

$$z_H[k] \propto \rho_H^{-k}, \quad (3.29)$$

with ρ_H very close to 2.

In other words, we can say that the average number of zero-crossings is, in first approximation, divided by 2 when going from one IMF to the next. Based on this result, we can even go further and check for a possible self-similarity in the “filter bank” structures of Figure 3.3. Restricting to the band-pass IMFs ($k = 2$ to 6), self-similarity would mean that

$$\mathcal{S}_{k',H}(f) = \rho_H^{\alpha(k'-k)} \mathcal{S}_{k,H}(\rho_H^{k'-k} f) \quad (3.30)$$

for some α and any $k' > k \geq 2$. As a consequence, the power spectra of all IMFs should overlay onto a single curve, when properly renormalized. Such a collapse via renormalization can indeed be observed on Figure 3.4, obtained with the specific choice $\alpha = 2H - 1$. Even if some low frequency discrepancies can be observed (especially when $H < 1/2$), these diagrams support the claim that, in a first approximation, EMD acts on fGn as a dyadic filter bank of constant- Q band-pass filters.

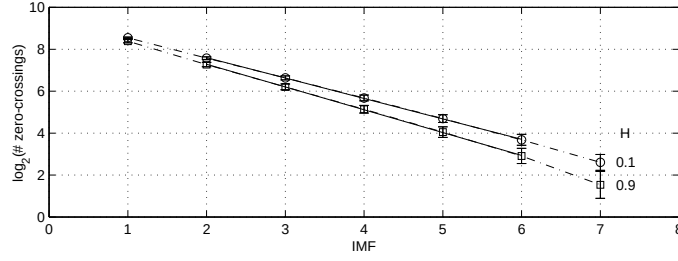


FIGURE 3.3 – IMF average number of zero-crossings in the case of fractional Gaussian noise. For a sake of readability, only the curves corresponding to the extreme indices $H = 0.1$ (circles) and $H = 0.9$ (squares) have been plotted in the diagram, all the other considered cases ($H = 0.2, 0.3, \dots, 0.8$) leading to similar curves. The superimposed full lines correspond to linear fits within the IMF range $k = 2$ to 6.

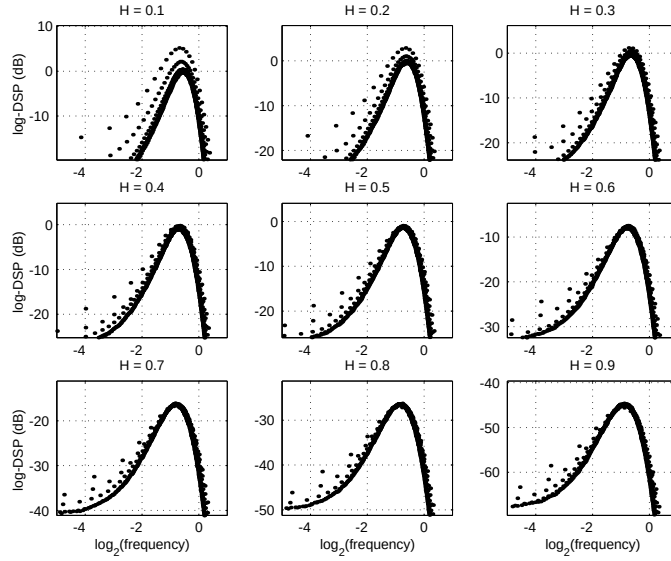


FIGURE 3.4 – Renormalized IMF spectra in the case of fractional Gaussian noise. For each value of H , the band-pass IMFs ($k = 2$ to 6) of Figure 3.2 are plotted after the renormalization given by eq.(3.30) with $\alpha = 2H - 1$, and $\rho_H = 2$

IMF marginal statistics. Although fGn is a Gaussian process, the highly nonlinear structure of EMD cannot a priori guarantee that fGn IMFs are themselves Gaussian. This however turns to be true (except for the first, high-pass, mode), as illustrated in Figure 3.5. Moreover, not only the modes are Gaussian, but they also evidence a general form of self-similarity consistent with the one related to the second-order properties discussed previously : the probability density functions (PDFs) $p(d_{k,H})$ of the different modes $\{d_{k,H}; k = 2, \dots, 6\}$ can all be deduced one from each other through the renormalization equation :

$$p(d_{k',H}) = \beta_H^{k'-k} p(\beta_H^{k'-k} d_{k,H}), \quad (3.31)$$

with $\beta_H := \rho_H^{H-1}$ (the bottom row of Figure 3.5 is displayed on a semi- logarithmic scale, so as to better appreciate the Gaussianity of the superimposed IMFs, characterized by a parabola).

As it is apparent in the top row of Figure 3.5, the PDF of the first IMF is not Gaussian but bimodal. This shape can be simply justified in the white noise case ($H = 1/2$) by the following argument. If we label by $p_+(x)$ and $p_-(x)$, respectively, the PDFs of (Gaussian) white noise maxima and minima, it is easy to show that they read as :

$$p_{\pm}(x) = \left[\frac{1}{2} \pm \left(\frac{1}{2} - \text{erf}(x) \right) \right]^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad (3.32)$$

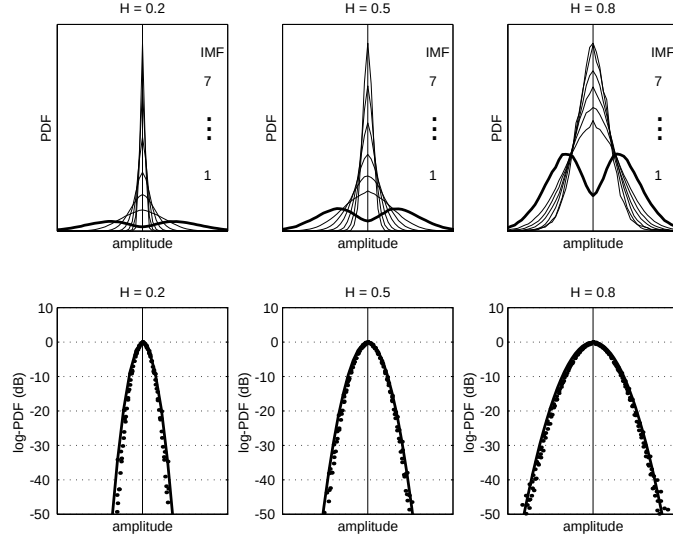


FIGURE 3.5 – IMF probability density functions in the case of fractional Gaussian noise. Top row : PDFs of the first 7 IMFs displayed on a linear scale, for 3 values of H . Bottom row : the corresponding PDFs of the IMFs of indices $k = 2$ to 6 on a semi-logarithmic scale, after the renormalization of eq.(3.31). Actual values are plotted as dots, while the superimposed thick lines correspond to their Gaussian fits.

with $\text{erf}(x)$ the standard error function.

We have seen that, in a first approximation, the first IMF is a high-pass filtered version of the input signal. Since, by construction, EMD only retains positive maxima and negative minima, this amounts to say that, in the white noise case, the PDF of the first IMF is essentially given by $p_*(x) := \max(p_-(x), p_+(x))$. Figure 3.6 shows that this simplified model reproduces the main structure of the actual PDF, with furthermore good agreement even in the cases where $H \neq 1/2$.

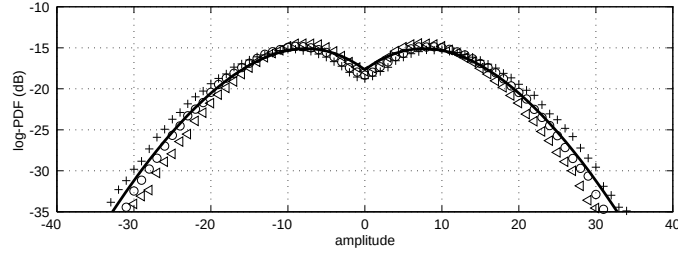


FIGURE 3.6 – Log-probability distribution function of the first IMF in the case of fractional Gaussian noise. The full line corresponds to the simplified model in the white noise case $H = 0.5$, whose actual values are plotted as circles. Crosses and triangles correspond, respectively, to the correlated cases $H = 0.2$ and $H = 0.8$. All PDFs have been normalized to be of unit area.

3.3.2 EMD-based estimation of scaling laws

Regularity exponent

EMD-based Hurst exponent estimation. Given the self-similar relation (3.30) for PSDs of the band-pass IMFs, we can deduce how variance should evolve as a function of the IMF number. Assuming that (3.30) holds for any $k' > k \geq 2$, with $\alpha = 2H - 1$, we easily find that :

$$V_H[k'] := \text{var } d_{k',H}[n] = \rho_H^{(\alpha-1)(k'-k)} V_H[k], \quad (3.33)$$

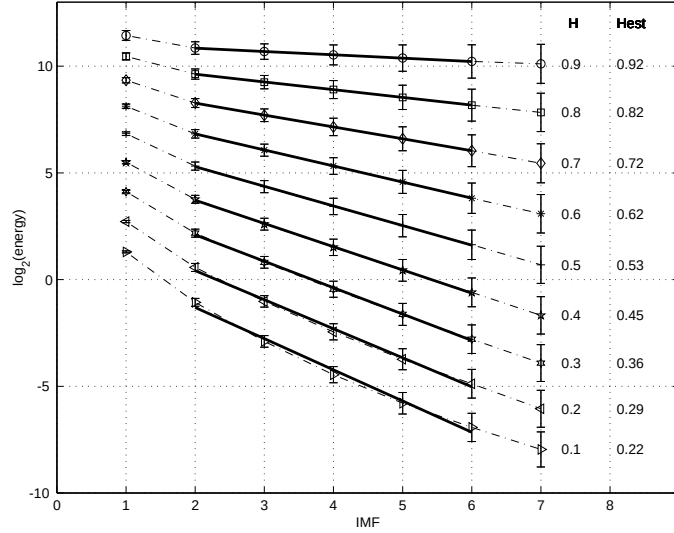


FIGURE 3.7 – Estimated IMF $\log_2$ -variance in the case of fractional Gaussian noise. The actual values of the empirical (energy-based) variance estimates are reported in dotted lines for the different values of the Hurst exponent H , together with error bars corresponding to the standard deviations associated with the 5000 realizations run in the study. The mean value of the estimated Hurst exponents are also given, based on weighted linear fits within the IMF indices range $k = 2$ to 6. For a sake of better readability, all curves have been arbitrarily shifted along the vertical axis to avoid overlapping.

leading to

$$V_H[k] = C \rho_H^{2(H-1)k}. \quad (3.34)$$

The IMF variance is therefore expected to be an exponentially decreasing function of the IMF index, with a decay rate which is a simple linear function of the Hurst exponent H . An experimental evidence for this behaviour is reported in Figure 3.7, where the (energy-based) empirical variance estimate $\hat{V}_H[k]$ has been plotted as a function of the index k in a semi-log diagram (in base 2). In accordance with the logarithmically linearized version of (3.34), it appears that straight lines can be fitted to the different curves, with a slope κ_H such that the estimated Hurst exponent $\hat{H}$ reads

$$\hat{H} = 1 + \frac{\kappa_H}{2}. \quad (3.35)$$

However, it has to be noticed that the predicted relationship (3.34) only holds for IMF indices $k > 1$, with furthermore an increased discrepancy for smaller H 's (typically, the model reasonably fits the data for $H > 1/4$).

EMD vs. wavelets : a quantitative comparison. In order to better appreciate the possibility of estimating H from a slope in a diagram “log-energy vs. IMF index,” and moreover to compare with wavelet-based techniques, it is important to not only consider the variance evolution through modes, but also the possible correlations which may exist within modes and between them. In fact, it is well-known that one of the main features of dyadic wavelet decompositions (WD) is to approximately decorrelate most processes thanks to the number of vanishing moments (2.13) of the mother wavelet, which acts as a tuning parameter. No such tuning parameter exists for EMD, and it is worth investigating the resulting correlation structure when applied to fGn.

To do so, two approaches have been followed. In the first one, we considered a mere generalization of the previous variance study by computing, for each value of the Hurst exponent H , the full variance-covariance matrix defined by

$$C_H[k, k'] := \mathbb{E} d_{k,H}[n] d_{k',H}[n]. \quad (3.36)$$

From a practical point of view, the corresponding estimate $\hat{C}_H[k, k']$ relies on 100 independent realizations, for a data size $N = 2048$ and IMF indices k and k' varying from 1 to 6. Figure 3.8 presents

the obtained normalized result $\hat{C}_H[k, k'] / \hat{C}_H[1, 1]$. As expected, one recovers along the main diagonal (i.e., for $k = k'$) the behaviour reported in Figure 3.7, since $\hat{C}_H[k, k] = \hat{V}_H[k]$. The new feature is that the estimated covariance falls off quickly when moving away from the main diagonal, thus indicating a low level of inter-modes correlation. This observation does not apply however to the first IMF which evidences (in the first row and the first column) a non-negligible amount of correlation with higher-order modes, in clear accordance with the specific low-pass tail reported in Figure 3.2. In good agreement with Figure 3.2 too, we can see that correlations induced by the first IMF are emphasized for low values of H , i.e., those for which the contribution of IMF 1 prevails at low frequencies.

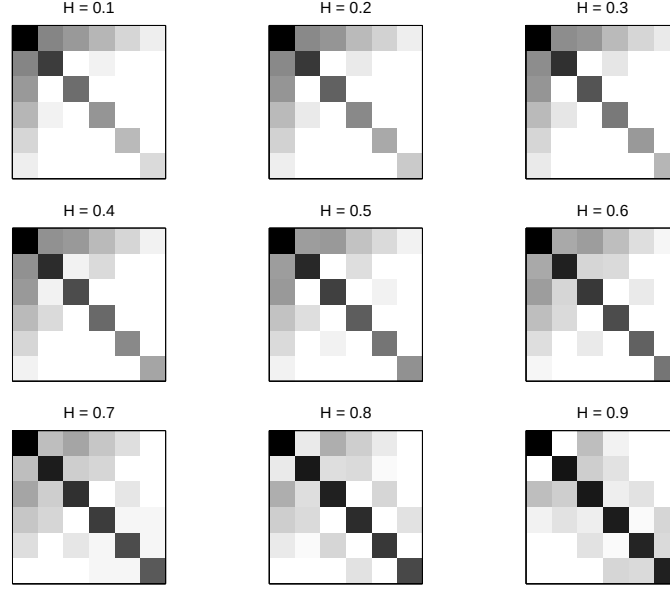


FIGURE 3.8 – Estimated IMF variance-covariance matrix in the case of fractional Gaussian noise. For each value of the Hurst exponent H , the graph displays the quantity $|\hat{C}_H[k, k'] / \hat{C}_H[1, 1]|$ on a logarithmic gray scale with a dynamic range of 10 dB, IMF indices $1 \leq k, k' \leq 6$ running from top to right and from top to bottom.

In a second approach, we focused on band-pass IMFs only ($k > 1$) and evaluated the two-dimensional correlation function

$$D_H[k', n'] := \mathbb{E} d_{k,H}[n] d_{k+k',H}[n + n'] \quad (3.37)$$

by means of the corresponding averaged empirical estimate $\hat{D}_H[k', n']$. The result of this two-dimensional correlation function of the full IMF matrix, considered as a two-dimensional field, is displayed in Figure 3.9.

As for detail sequences at different scales in wavelet decompositions, the graph evidences that modes with different indices are almost uncorrelated. The only significant values of $\hat{D}_H[k', n']$ correspond to $k' = 0$, i.e., to intra-scale correlations, with a correlation decay which becomes slower as H increases.

In terms of estimation of the Hurst exponent H , based on the assumed relation (3.35) with the slope κ_H deduced from the linear regression of $\hat{V}_H[k]$, the consequences of this behavior are twofold. Firstly, because of the non-zero intra-scale correlations, the variance estimate $\hat{V}_H[k]$ is expected to largely fluctuate, especially for large Hurst exponents H and for large IMF indices k . Secondly, the negligible inter-scale correlations yet allows for an estimation of the slope κ_H from a weighted linear regression in the semi-log diagram $\log_2 \hat{V}_H[k]$ vs. k . As far as the variability of the variance estimate is concerned, Figure 3.7 gave a rough, second-order, indication on the basis of the observed standard deviation. A more complete appreciation can be gained from Figure 3.10 where, in 3 typical cases ($H = 0.2, 0.5$ and 0.8), the experimental mean, median and various confidence intervals have been reported, together with the fitted model deduced from (3.34) as :

$$\log_2 V_H[k] = \log_2 \hat{V}_H[2] + 2(H - 1)(k - 2) \log_2 \rho_H \quad (3.38)$$

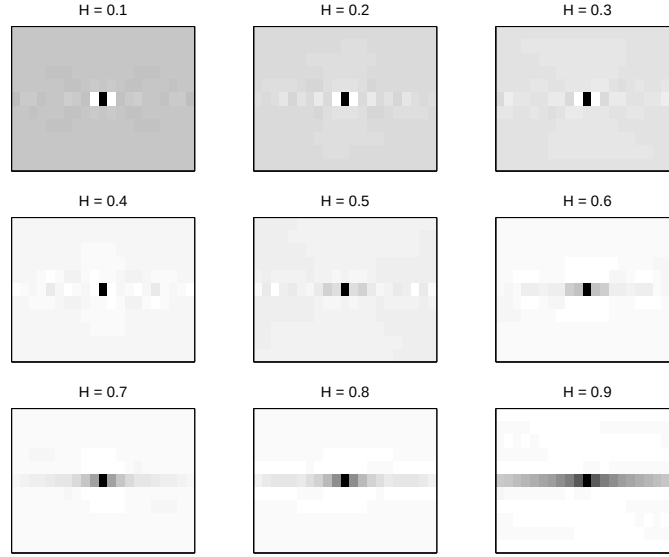


FIGURE 3.9 – Two-dimensional correlation function of the full IMF matrix, considered as a two-dimensional field, in the case of fractional Gaussian noise. For each value of the Hurst exponent H , the graph displays the quantity $|\hat{D}_H[k', n']|$ on a linear gray scale (from white for the minimum to black for the maximum). The horizontal (resp., vertical) axis corresponds to time (resp., IMF index) lags.

for $k \geq 2$. This series of simulations (which has been carried out on 10000 realizations of 2048 data points in each case) highlights larger and larger fluctuations for modes of larger and larger indices, in agreement with (and generalization of) the findings reported in [Wu et al., 2001] for the sole case of white noise. Moreover, the skewed (marginal) distribution of these “modegrams” reveal a better agreement when fitting the linear model (3.38) with the median rather than with the mean of the realizations.

Regularization dimension

Definition. Fractal dimensions are indicators of the (ir-)regularity of function graphs (or of a particular realization of a random process), and mathematically they can be defined in many different ways, associated to as many regularity measurements. For instance, in the case of a 1- d signal $x(t)$ with homogeneous hölder regularity h of Definition 3.1.4, the *fractal regularity dimension* of its graph G_x is simply $\dim_h(G) = 2 - h$.

The *regularization dimension* is another variant that was introduced in [Lévy Véhel and Roueff, 1998]. Let G be the graph of a bounded function u defined on some closed bounded interval $T \subset \mathbb{R}$. Let u_a be the projection of u on some approximation space at lower resolution $a > 0$ (such that $\lim_{a \rightarrow 0} u_a = u$), and define $\mathcal{L}(a) = \int_T \sqrt{1 + u'_a(t)^2} dt$, the finite length of the corresponding graph G_a . Then, the regularization dimension reads :

$$\dim_{\mathcal{R}}(G) = 1 + \lim_{a \rightarrow 0} \frac{\log(\mathcal{L}(a))}{\log(a)}, \quad (3.39)$$

and it derives from the same principle that underlies all fractal dimensions, that is to fit a scale dependent measure (here, the trajectory length) with a polynomial model, to reflect how fast it varies as resolution increases.

Many other definitions of fractal dimension exist, but it is remarkable that for a large class of functions, all these dimensions are equal to each other. This is notably the case for fractional Brownian motions $B_H(t)$ with Hurst exponent $0 < H < 1$, for which we have almost surely $\dim_h(B_H) = \dim_{\mathcal{R}}(B_H) = 2 - H$, and $\forall t, h = H$.

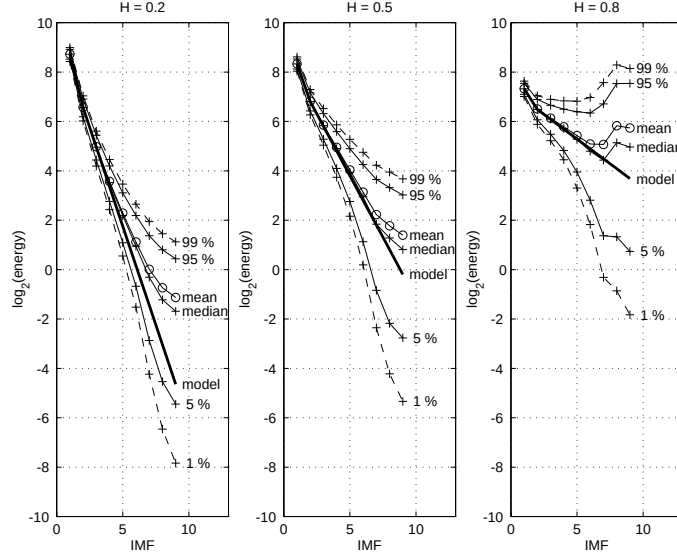


FIGURE 3.10 – Experimental “modegrams” in the case of fractional Gaussian noise. For the 3 considered values of the Hurst exponent, statistical characteristics (mean, median, confidence intervals) of the logarithm of the estimated EMD variance have been plotted as a function of the IMF index, together with the linear model given by (3.38).

EMD-based fractal dimension estimation. Listing all existing fractal dimension estimators is definitely beyond the scope of this document. Yet, we already described two possible approaches for estimating the regularity Hurst exponent (thus the regularity dimension) of a mono-fractal process. Let us recall them, and extend the wavelet-based estimator to its leader companion form :

Estimator 1 is characterized by expression (3.24), where the logarithm of the wavelet empirical variance $\widehat{\text{Var}} \Gamma_x[j, \cdot]$ is (up to a correction factor) linearly regressed against the log-scale j ;

Estimator 2 substitutes in the previous scheme, the wavelet coefficients $\Gamma_x(j, k)$ by their *leader mates* [Jaffard et al., 2006]

$$\Lambda[j, k] := \sup_{(j', k') \in [(k-1)2^{-j}, (k+1)2^{-j}]} |\Gamma_x[j, k]|, \quad (3.40)$$

and yields a more robust estimator [Wendt et al., 2009] ;

Estimator 3 described above (see Eq. (3.34)-(3.38) and (3.35)) is the EMD counterpart of *estimator 1*, where it is the logarithm of the IMF variance that is linearly fitted.

In [Lévy Véhel and Roueff, 1998] where the regularization dimension was originally introduced, the approximation u_a was obtained by smoothing the initial function u with a scale variable low pass kernel $\gamma_a(t) = |a|^{-1}\gamma(a^{-1}t)$. A gaussian window γ is recommended, and for a fBm $B_H(t)$, the Hurst exponent $H = 2 - \dim_{\mathcal{R}}$ is estimated according to (3.39) with :

$$\mathcal{L}(a) = \int_T \sqrt{1 + \left| \frac{d}{dt}(u \star \gamma_a)(t) \right|^2} dt. \quad (3.41)$$

We then suggested in [32] to adapt the regularized dimension (3.39) to the various multiresolution frameworks related to both the wavelet decompositions (2.15) and to the EMD technique. As for the former case, we considered the approximations $A_j(t) \in V_j$ as the regularized trajectories. Doing so, for a discrete fBm sample path, the graph length $\mathcal{L}(a_j)$ at scale $a_j = 2^{-j}$, reads :

$$\mathcal{L}(a_j) = \int_T \sqrt{1 + \left| \frac{d}{dt} \sum_{k=-\infty}^{\infty} \langle B_H, \phi_{j,k} \rangle \phi_{j,k}(t) \right|^2} dt. \quad (3.42)$$

TABLE 3.1 – Mean Square Error of $\hat{H}$ obtained for each estimator (1)–(6) and for different values of the Hurst exponent H .

	(1)	(2)	(3)	(4)	(5)	(6)
$H = 0.1$	.0439	.0044	.0194	.0497	.0057	.0135
$H = 0.2$	.0166	.0010	.0099	.0156	.0058	.0044
$H = 0.3$	.0090	.0051	.0074	.0062	.0079	.0023
$H = 0.4$	.0060	.0059	.0083	.0030	.0102	.0016
$H = 0.5$	.0048	.0049	.0069	.0019	.0123	.0016
$H = 0.6$	.0044	.0041	.0076	.0015	.0139	.0017
$H = 0.7$	.0042	.0034	.0088	.0016	.0163	.0018
$H = 0.8$	.0040	.0031	.0117	.0020	.0181	.0019
$H = 0.9$	.0041	.0030	.0888	.0027	.0175	.0012

Notice that, like expression (3.41), the proposed length (3.42) is parameterized by a sequence of predefined scale values $\{a_j, j = 1, \dots, J\}$, chosen independently of the analyzed process. To alleviate this constraint, we evoke the same rationale that led us to consider EMD in [26, 74], and we use partial reconstruction sums of IMFs to get a series of regularized functions u_{a_k} that live at their “natural scales” $\{a_k, k = 1, \dots, K\}$. More precisely, assuming that $k = K$ corresponds to the IMF at the coarsest scale, we propose to use in (3.39) the following alternative length :

$$\mathcal{L}(a_k) = \int_T \sqrt{1 + \left| \frac{d}{dt} \sum_{i=k}^{i=K} d_i(t) \right|^2} dt, \quad (3.43)$$

with $B_H(t) = m_K(t) + \sum_{i=1}^{i=K} d_i(t)$.

At the final, we end up with three more estimators of the (regularization) fractal dimension :

Estimator 4 stems from linear regression of the lengths of gaussian regularized graphs (3.41), and corresponds to the original estimator proposed in [Lévy Véhel and Roueff, 1998];

Estimator 5 is an adaptation of *Estimator 4* to wavelet multiresolution framework (3.42);

Estimator 6 is the EMD-based replica of *Estimator 4*, enriched with self-adapted analyzing scales (3.43).

Numerical simulations

Experimental setup. To assess and compare the performances of the 6 estimators described above, they are tested on a set of 10.000 independent realizations of fBm's with the same H . Each sequence of size $N = 1024$ points is synthesized by the circulant matrix method of [Wood and Chan, 1994], and the test is repeated for different values of H ranging from 0.1 to 0.9 with a 0.1 stepsize.

To account for a possible variability in the confidence intervals of the quantities to be regressed, all estimators (1)–(6) rely on a weighted least square fit of the *log*-measures against the *log*-scale. In addition, EMD-based estimators (3) and (6), for which the scale a_k is itself a random variable, employ a weighted linear regression with uncertainty along the two axes.

Furthermore and according to previous findings, to cope with finite size effects and spectral aliasings, wavelet based estimators must adapt their regression scale range to the value of H . Then, in our simulations, we systematically disregard the finest and the two coarsest scales from wavelet decompositions. Regarding EMD-based estimators, this limitation does not hold, and they can operate on full IMF's decompositions. Finally, we used a Daubechies wavelet with 2 vanishing moments in estimators (1), (2) and (5), and a gaussian kernel whose width varies from 8 to 256 points over 16 voices for the estimator (4).

Results. Table 3.1 presents the Mean Square Errors $MSE(\hat{H}) = (\mathbf{E}\hat{H} - H)^2 + \text{Var}(\hat{H})$, for each estimator and for each value of H . For most values of the Hurst exponent H , the estimate (6) based

on the partial EMD reconstruction lengths seems to perform better – or slightly worse than companion approaches based on regularization dimension. Only for small H , does the leaders technique (2) seems to significantly outperforms all the others. But a closer view at the individual biases and variances, sheds a quite different light on the results. Indeed, by construction of relation (3.40), a wavelet leader at scale j is necessarily larger or equal to its antecedent at finer scale $j + 1$. Inevitably then, the q -th order moment entering the definition (3.24) of $\hat{H}$ is an increasing function of the scale index j , which in turn mechanically prevents the slope $\hat{H}$ from taking on negative values. Whereas this limitation may be beneficial for large H 's (typically, $H > 0.2$), for small values of H , the distribution of the estimates being artificially truncated below zero, the resulting variance drastically reduces but at the cost of a deleterious bias. This shrinking effect is very clear on figure 3.11 representing the empirical $\hat{H}$ distributions obtained with each method. Ideally, densities corresponding to different values of H should not overlap. Yet, with the leaders based estimator (ii), the densities for $H = 0.1$ and $H = 0.2$ respectively, are totally superimposed, turning impossible to distinguish between the two exponents. This phenomenon does not appear with the other estimators, and although a significant bias systematically occurs for small H 's – even with the partial EMD reconstruction approach – the achieved resolution remains fairly acceptable.

More generally, there are two different angles for looking at the estimators :

1. If we compare the estimates (1)–(3) based on regularity dimension with estimates (4)–(6) relying on regularization dimension, it is quite manifest from Table 3.1, that the second set of estimators performs generally better. One sensible reason for this, lies in the analytical form of the regressed measures themselves. Length estimates $\mathcal{L}(a)$ yield an embedded, hence highly correlated series of measurements, whereas the variances of *details* correspond to almost-independent variables. For relatively small sample sizes then, this correlation stabilizes the scale progression of the measure to be linearly fitted, yielding thus a more robust estimator.
2. Let us now compare the wavelet-based estimators (estimates (1), (2), (4) and (5)) versus the EMD-based estimators (estimates (3) and (6)). According to [Lévy Véhel and Roueff, 1998], we include the gaussian regularized estimator (4) in the group of wavelet-based estimators, for it shares an evident connection with continuous wavelet transforms. From this viewpoint, conclusions seem more ambiguous, and it is not so easy to sketch a sharp frontier between the two groups. However a closer look at the first group of wavelet based estimators and at their corresponding MSE's, reveals that the best general results come from the gaussian kernel regularization method. Surprisingly, this is among all estimators of its group, the sole that relies on a (equivalent) *continuous* wavelet transform. This raises the interesting question of information redundancy and of its usefulness in small data sets scenarios [Percival and Walden, 2000].

Beyond the relative satisfactory performance achieved with EMD for regularization dimension, it is important to stress the *pros and the cons* inherent to this method :

- EMD provides with a totally adaptive multiresolution scheme that, in the case of a fBm (or any self-similar process), does not necessitate to tune the regression scale range, nor it calls for border effects correction ;
- More promisingly, benefits of EMD applied to fractal dimension estimation should become more obvious when addressing the case of discrete rather than continuous scale invariance [Borgnat, 2002] (e.g. Weierstrass processes, Mandelbrot multiplicative cascades, oscillating chirp driven data. . .) In these situations, a dyadic (or any a priori fixed) scale dichotomy may not necessarily be the appropriate thing to do. On the other hand, an adaptive multiresolution approach like EMD, will naturally and automatically lead the analysis along the characteristic scales of the process.
- As far as computational cost is concerned, EMD is incomparably more expensive than a discrete wavelet transform. An obstacle that can severely penalize its use with large sample sizes.

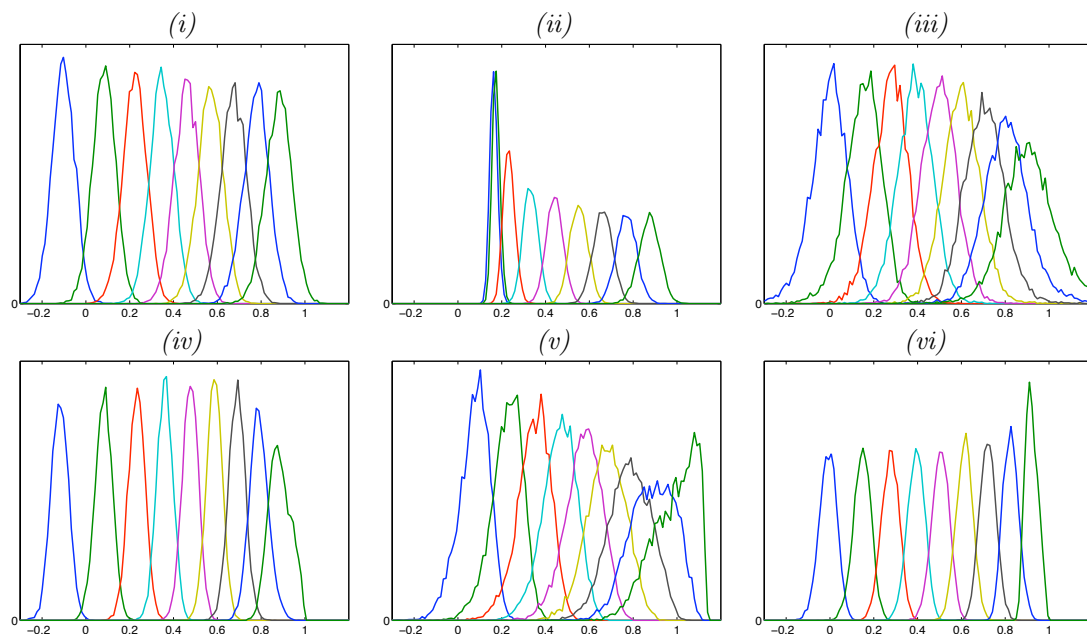


FIGURE 3.11 – For each estimator (1) – (6) of fBm Hurst exponent, empirical distributions of $\hat{H}$ when H varies from 0.1 to 0.9 (from left to right densities on the plots). Y-axes correspond to the repartition of $\hat{H}$ based on a trial of 10.000 independent realizations of discrete fBm samples of size $N = 1024$ points.

3.4 Un test d'existence des moments d'une variable aléatoire

Au paragraphe 3.2, nous avons rappelé le pouvoir naturel des ondelettes à caractériser les lois d'échelle. Plus spécifiquement encore, s'agissant de la régularité locale d'une fonction, le théorème 3.2.1 et la proposition 3.2.5 doublent le trait en valorisant également les propriétés de localisation temporelle de ces décompositions. Cette bonne adéquation est à la base de deux estimateurs que nous avons définis dans un cadre relatif aux distributions « à queues lourdes », et qui tous deux exploitent les spécificités de la fonction caractéristique d'une variable aléatoire.

Le premier, décrit dans l'article :

Wavelet Based Estimation for Univariate Stable Laws

A. ANTONIADIS, A. FEUERVERGER and P. GONÇALVES

Annals of the Institute of Mathematical Statistics, 2006,

propose une estimation empirique des paramètres d'une loi α -stable, par régression des coefficients d'ondelette discrets issus de l'analyse multirésolution de la fonction caractéristique [7].

Dans le deuxième, les propriétés temps-échelle des ondelettes sont au centre d'une procédure proposée pour estimer le seuil d'existence des moments d'une variable aléatoire [29, 48, 72]. Le principe de cet estimateur est résumé ci-dessous, puis suivra une description plus détaillée, issue de l'article original et renvoyant à celui-ci pour les démonstrations :

Diverging moments and parameter estimation

P. GONÇALVES and R. RIEDI

Journal of American Statistical Association, 2005.

3.4.1 Principe

Les distributions dites « à queues lourdes » connaissent un intérêt croissant en partie grâce à l'enrichissement de l'arsenal d'outils numériques et analytiques pour les manipuler. Leur rôle devient de plus en plus incontournable dans la modélisation de données issues de systèmes complexes tels que ceux rencontrés en hydrologie, en finance ou dans les réseaux. Nous reviendrons d'ailleurs en détail sur cette dernière application au chapitre 4. Les lois stables, loi de Pareto, loi de Zipf, ou certaines distributions de valeurs extrêmes sont des paradigmes de distributions à queue lourde, dans lesquelles le paramètre de queue α tient une place centrale. En particulier, celui-ci définit le seuil au delà duquel les moments $\mathbb{E}[|X|^r]$ d'ordre r n'existent pas. Rappelons en effet, qu'une variable aléatoire X est dite « à queue lourde » de paramètre α si

$$P[|X| > x] = x^{-\alpha} L(x) \quad (3.44)$$

où L est une fonction à variation lente telle que $L(tx)/L(x) \rightarrow 1$ lorsque $x \rightarrow \infty$ pour tout $t > 0$. Elle est dite Pareto, si $L(x)$ est constante et si (3.44) est vérifiée pour tout $|x| > \delta$. Dans ce cas précis, il est clair que les moments sont finis jusqu'à l'ordre α exactement, mais c'est un résultat que l'on sait généraliser par des arguments élémentaires, à toute distribution de la forme (3.44). La question de la finitude des moments est particulièrement sensible au regard de la quantité et de l'importance accordée aux estimateurs de moments. Ils interviennent non seulement dans l'estimation de paramètres lorsque la distribution sous-jacente est connue, mais également lorsque celle-ci ne l'est pas, dans l'ajustement de données et la sélection de modèles. C'est par exemple le cas de l'hypothèse statistique du Kurtosis qui permet de trancher sur le caractère gaussien d'une variable aléatoire ; de même que dans le cas d'une variable de Poisson, moyenne et variance doivent être égales. En outre, l'estimation de moments est à la base d'un grand nombre d'applications. Il en est ainsi de l'analyse multifractale, où l'ordre q des moments absolus des accroissements dans l'expression (3.16) ou des coefficients d'ondelettes dans (3.28) portent toute l'information relative à la régularité locale de la trajectoire du processus.

Les choses se compliquent dès que les moments sont infinis ou non définis, comme par exemple pour la distribution de Cauchy qui est de variance infinie et de moyenne indéterminée. Dans la mesure où les moments infinis peuvent dégrader les performances de certains estimateurs (voire introduire un biais), ou ralentir la vitesse de convergence vers les lois limites, il est important de s'assurer de leur existence théorique. L'illustration par l'analyse multifractale est là encore édifiante, puisque les moments infinis peuvent directement être l'expression de transitions de phases caractéristiques de la régularité des processus.

Tout ceci justifie et motive le développement de méthodes statistiques en vue de déterminer la finitude des moments d'une distribution à partir de la seule connaissance d'échantillons de sa variable aléatoire [29]. Plus précisément, il s'agit d'identifier les ordres critiques (positif et négatif) λ^- , λ^+ d'une distribution, tels que :

$$\begin{aligned}\lambda^+ &\triangleq \sup\{r > 0 : \mathbb{E}[|X|^r] < \infty\} \\ \lambda^- &\triangleq \inf\{r < 0 : \mathbb{E}[|X|^r] < \infty\}.\end{aligned}\tag{3.45}$$

Les estimateurs ainsi obtenus caractérisent non seulement la décroissance asymptotique de la distribution à l'infini, mais également son comportement au voisinage de zéro. Pour cela, la procédure que nous avons définie repose sur l'association de deux arguments théoriques.

Tout d'abord, la fonction caractéristique $\phi(u) \triangleq \mathbb{E}[\exp(iuX)]$, transformée de Fourier de la distribution de X , possède autant de dérivées en l'origine que la distribution admet de moments d'ordre positif finis. De plus, sachant que pour tout ordre k pair, on a $\phi^{(k)}(0) = i^k \mathbb{E}[X^k]$ (dès lors que l'un des deux membres de l'égalité existe [Papoulis, 1991]), il nous a fallu démontrer une généralisation de ce résultat aux ordres non entiers. Nous avons pour cela introduit la notion d'exposant caractéristique ρ^+ d'une distribution comme le degré (généralisé) de continuité Lipschitzienne de la partie réelle de la fonction caractéristique en l'origine. Si $\rho^+ \in (0, 2)$, l'exposant caractéristique s'écrit simplement

$$\rho^+ = \sup\{r > 0 : 1 - \operatorname{Re} \phi(u) = O(u^r) \text{ quand } u \rightarrow 0^+, \}\tag{3.46}$$

d'où il découle que

$$\lambda^+ = \rho^+,\tag{3.47}$$

lorsque les deux indices sont compris entre 0 et 2. Estimer l'exposant critique revient donc à mesurer la régularité de ϕ . Le changement de variable $Y = 1/X$ livre suivant la même démarche, l'exposant ρ^- et une estimation de λ^- . Une part importante de notre travail a donc consisté à étendre, du point de vue de l'estimation, cette approche aux ordres supérieurs à 2. Les fondements mathématiques de cette généralisation sont développés dans [47]. Applicable à la sélection de modèles, l'exposant caractéristique fournit une estimation exacte de l'ordre critique $\rho^+ = \lambda^+$ (voir corollaire 3.4.2).

Ensuite, le problème ayant été réduit à l'estimation d'une régularité locale, nous nous sommes naturellement tournés vers le potentiel des décompositions en ondelettes, et en particulier vers l'application du théorème 3.2.1 et de la proposition 3.2.5 à la fonction caractéristique ϕ en 0.

En résumé, le principe de l'estimateur de ρ^+ est extraordinairement simple :

- À partir de l'échantillon $\{X_n, n = 1, \dots, N\}$, on estime la fonction caractéristique empirique $\widehat{\Phi}(u) = N^{-1} \sum_n \exp\{iuX_n\}$, dont on ne calcule ensuite que les seuls coefficients de sa décomposition en ondelettes, localisés en zéro, i.e. $\Gamma_{\widehat{\Phi}}(0, a)$. Plus simplement encore, cela revient en fait, à calculer l'estimateur non paramétrique à noyau et sans biais :

$$\widehat{W}(a) = \frac{1}{N} \sum_{k=1}^N \Psi(a \cdot X_k)\tag{3.48}$$

où le noyau Ψ n'est autre que la transformée de Fourier de l'ondelette (semi-définie positive).

- L'estimation des deux exposants caractéristiques $\widehat{\rho}^+$ et $\widehat{\rho}^-$ résulte d'une simple régression linéaire de $\log \widehat{W}(a)$ en fonction de $\log a$, dans un intervalle d'échelle donné. Ces estimateurs sont invariants par changement d'échelle, peuvent être facilement rendus invariants par translation, et sont asymptotiquement sans biais.

- Les ondelettes ne pouvant pas détecter de régularité plus grande que leur propre régularité n_ψ , il convient d’itérer cette procédure avec des ondelettes de régularité croissante, jusqu’à stabilisation de l’exposant estimé.

Nous avons montré l’efficacité de cet estimateur appliqué aux distributions stables, en le comparant aux estimateurs classiquement utilisés, tels que ceux de Koutrouvelis [Koutrouvelis, 1980], McCulloch [McCulloch, 1986] ou Nolan [Nolan, 2001]. Rappelons que les distributions symétriques stables sont les lois limites d’une somme proprement normalisée de variables aléatoires indépendantes, identiquement distribuées, et le cas échéant de variance infinie. Elles sont définies par leur fonction caractéristique

$$\phi_X(u) = \mathbb{E}[\exp(iuX)] = \exp(-\sigma^\alpha |u|^\alpha + i\mu u) \quad (3.49)$$

et leur paramètre de queue lourde est égal à α [Samorodnitsky and Taqqu, 1994]. En combinant cela, avec le fait que leur densité de probabilité, bien que d’expression analytique inconnue, est uni-modale, on déduit que ses moments absolus d’ordre r ne sont finis que pour $r \in (-1, \alpha)$.

Enfin, pour illustrer en pratique l’apport de cet estimateur, nous l’avons mis en œuvre dans un contexte d’analyse multifractale. Dans ce cadre applicatif précis, nous avons montré qu’il permettait de contourner certains écueils délicats, pour lesquels il n’existait jusqu’alors pas de solution satisfaisante. D’une part, il permet d’améliorer l’estimation du spectre de singularité de certains processus tels que les vols de Lévy, ou les cascades multiplicatives; d’autre part, il constitue un indicateur statistique performant permettant de distinguer ces deux classes de processus.

3.4.2 Detailed rationale

Background

In this section we collect well known facts on the existence of moments as well as the wavelet analysis of irregular signals.

Tail estimators. Most well-known tests for the existence of moments emerge as by-products of tail estimators and appear in parameterized settings. For instance, [Nolan, 2001] proposed a maximum likelihood estimator for general alpha-stable laws (including Gaussian and Cauchy) based on a large sample data set. As no closed form exists for these distribution functions (aside from some particular rational values of the parameters), he proposes an efficient numerical resolution to find the maximum of the likelihood equation.

Previously, [Koutrouvelis, 1980] and [McCulloch, 1986, McCulloch, 1997], among many others, have proposed two different estimators of the parameters of α –stable laws, based either on Pareto approximation for α –stables tails, or on the analytic form of the characteristic function.

More recently, [Bianchi and Meerschaert, 2000] proposed a quadratic estimator of tail index α , based on the asymptotic of the sample variance. This robust estimator has the advantage over Hill estimator [Hill, 1975], to be shift and scale invariant, and also to perform well in situation where the Hill estimator is inefficient, namely for stable distributions with $1.5 < \alpha < 2$.

Starting from a closed form for the characteristic function (recall (3.49)) or in some cases a numerical approximation of the density function, all these methods aim at finding the maximum of the log-likelihood function, given the data. As a result, it is well known that these approaches are optimal in the sense of minimum variance and achieve the Cramér Rao bound [Feuerverger and McDunnough, 1981, Koutrouvelis, 1980, Nolan, 2001]. However, being parametric, these estimators may perform poorly whenever the true underlying distributions do not match the model.

In this work, we propose a *non-parametric* estimation procedure with convincing robustness properties for the characteristic exponents ρ^+ and ρ^- which do not rely on any assumption on the density model. In particular, not even the semi-parametric assumption of heavy tails (3.44) is made and can be tested via this approach. The resulting estimates can be used for estimating the tail parameter and the body parameter.

Notably, both exponents are estimated in the same procedure. Indeed, the problem of existing negative moments could be reformulated with a simple change of variables $x \mapsto x^{-1}$, as a positive moment existence problem. Then, we could apply our estimator to X^{-1} instead of X directly, allowing thus for determining a lower negative bound for the existence of $\int_{-\infty}^{\infty} |x|^{-r} dF(x)$, $r > 0$. However, as we will demonstrate both, the positive and the negative characteristic exponent can be evaluated at once, using the same procedure applied to the same data set of i.i.d. samples $\{X_i\}_{i=1, \dots, N}$.

Characteristic function and moments. Let us recall a well known relation between high order moments of a distribution function $F(x)$ of a random variable X and its so-called *characteristic function* which is defined as :

$$\phi(u) = \mathbb{E}e^{iux} = \int e^{iux} dF(x). \quad (3.50)$$

Using simple duality argument between time and frequency (via the Fourier transform in (3.50)), the behavior of the characteristic function at the origin relates to the tail behavior of the distribution F for large $|x|$. In particular, whenever the integer p -th order moment of F exists, the p -th derivative of ϕ at the origin exists as well and they simply relate as follows

$$\phi^{(p)}(0) = \left. \frac{d^p}{du^p} \phi(u) \right|_{u=0} = i^p \mathbb{E}[X^p] = i^p \int x^p dF(x). \quad (3.51)$$

This justifies ϕ to be also referred to as a *moment generating function*.

Conversely, when p is even, existence of $\phi^{(p)}(0)$ implies existence of $\mathbb{E}[X^p]$. Notably, pathologies can occur when p is odd. As the example of $\phi(u) = C^{-1} \sum_{j=2}^{\infty} \cos ju / (j^2 \log j)$ demonstrates [Kawata, 1972, compare pg. 411], the existence of $\phi^{(1)}(0)$ does not necessarily guarantee the existence of $\mathbb{E}[X]$.

As we strive towards a generalization of a relation between moments and characteristic function to non-integer orders $r > 0$, let us first introduce the absolute moments of order $r \in \mathbb{R}$:

$$\mathcal{M}_r \triangleq \mathbb{E}[|X|^r] = \int |x|^r dF(x) \quad (3.52)$$

where we allow the value ∞ . Let us emphasize that $\mathbb{E}[X^p]$ exists if and only if $\mathcal{M}_p$ is finite, in other words, if and only if both $\mathbb{E}[\max(X, 0)^p]$ and $\mathbb{E}[\max(-X, 0)^p]$ are finite.

We first recall the definition of λ^+ in (3.45) and note a simple fact :

Lemma 3.4.1 *For any distribution F we have*

$$\begin{aligned} \lambda^+ &= \sup\{r > 0 : \mathcal{M}_r < \infty\} \\ &= \sup\{r > 0 : 1 - F(x) + F(-x) = O(x^{-r}) \text{ as } x \rightarrow \infty\} \end{aligned} \quad (3.53)$$

Note that a priori there is no information on the behavior in (3.53) for r exactly equal to λ^+ .

Next, we apply a theorem⁵ due to [Binmore and Stratton, 1969] (see e.g. [Kawata, 1972]) which relates the Lipschitz regularity of ϕ at the origin to the tail decay of F for orders less than 2. Recalling the definition of the Lipschitz exponent of ϕ given in (3.46) we find :

Corollary 3.4.1 *If either λ^+ or ρ^+ is known to be strictly less than 2 then :*

$$\lambda^+ = \rho^+. \quad (3.54)$$

With this in mind, we present wavelet theory in the next section with particular emphasis on their natural abilities to detect and estimate the local regularity of a function.

Wavelet analysis of characteristic functions

We start by demonstrating how the wavelet analysis of characteristic functions can be simplified tremendously.

5. Let $0 < r < 2$. Then, $1 - \text{Re} \phi(u) = O(u^r)$ if and only if $P[|X| > x] = O(x^{-r})$.

Semi-definite wavelets. As it turns out it is particularly simple to estimate the wavelet coefficients of a characteristic function provided the wavelet ψ is *semi-definite* by which we mean that its Fourier transform $\Psi(\nu) = \int \psi(t) \exp(-it\nu) dt$ is real and does not change sign. In other words, ψ is either positive semi-definite, i.e., $\Psi(\cdot) \geq 0$, or it is negative semi-definite, i.e., $\Psi(\cdot) \leq 0$. Examples of such wavelets are the derivatives of even order of the Gaussian kernel : set

$$\psi_p(t) \triangleq c_p \frac{d^{2p}}{dt^{2p}} \exp(-\sigma^2 t^2) \quad (3.55)$$

where c_p is a normalization constant and p is a positive integer. One finds the semi-definite Fourier transform

$$\Psi_p(\nu) = C_p (-1)^p \nu^{2p} \exp\left(\frac{-\nu^2}{4\sigma^2}\right). \quad (3.56)$$

Lemma 3.4.2 *If the Fourier transform Ψ of the wavelet ψ is real, square integrable and semi-definite then*

$$|W[\text{Re } \phi](t, a)| \leq |W[\text{Re } \phi](0, a)| \quad (3.57)$$

In other words, for fixed scale a the modulus of the wavelet transform of the real part of a characteristic function is maximal at $t = 0$ for semi-definite Ψ and equals

$$W[\text{Re } \phi](0, a) = \int \Psi(ax) dF(x) = \mathbb{E}[\Psi(aX)] \quad (3.58)$$

Critical orders smaller than 2. We are now in a position to combine the above results into the anticipated tight connection between a wavelet analysis and the critical order λ^+ . We will then extend this connection to orders larger than 2 (see Section 3.4.2).

Theorem 3.4.1 *Consider an admissible, semi-definite wavelet ψ of regularity $n_\psi \geq 2$. Then,*

$$\lambda^+ = \rho^+ = \sup\{r > 0 : |W[\text{Re } \phi](0, a)| \leq C a^r \text{ for } a \rightarrow 0^+\}, \quad (3.59)$$

provided that either term is known a priori to be strictly less than 2.

From a wavelet point of view we can not stress enough that the above result owes its simplicity to the fact that the wavelet coefficients of $\text{Re } \phi$ are maximal at 0. Also, $\lambda^+ = \rho^+$ was noted earlier.

Critical orders higher than 2. Attempting to extend the appealing three-fold connection of theorem 3.4.1 to orders higher than 2 we face two hurdles, one surmountable due to special properties of the characteristic function, the other more profound.

For a better understanding, we need to extend the concept of Lipschitz continuity to higher orders. To this end, we define the Taylor rest-term of order $2p$ at zero as :

$$Q_{2p}(t) = \text{Re } \phi(t) - 1 - \sum_{k=1}^p \frac{t^{2k}}{(2k)!} \phi^{(2k)}(0) \quad (3.60)$$

whenever it exists. Thus, the general definition of ρ^+ reads then as

$$\rho^+ = \sup\{r > 0 : Q_{2p}(u) = O(u^r) \text{ as } u \rightarrow 0^+, \text{ for } 2p \leq r < 2(p+1)\} \quad (3.61)$$

Also, we require a more general version of corollary 3.4.1. The higher order extension of [Binmore and Stratton, 1969] is found in [Kawata, 1972] and relates the finiteness of moments, i.e., the value of λ^+ to a smoothness condition of Q_{2p} .

The first hurdle concerns the fact that the wavelet analysis is a powerful tool for assessing the local degree of regularity, but does in general not allow to make conclusions on differentiability of the analyzed function. To make the point, functions which behave at zero as $|\cdot|^{2.5}$ (modulo a polynomial) but have only one derivative are easily constructed. In other words, the corrective polynomial does not have to be

the Taylor polynomial as in (3.60). This difficulty is overcome by proving existence of moments directly via monotone convergence from the decay of appropriate wavelet coefficients. Finite moments imply then that ϕ was indeed differentiable and that wavelet analysis indeed reflects the regularity ρ^+ of Q_{2p} .

The second hurdle stems from the fact that Kawata's smoothness condition⁶ (which allows to compute λ^+) is in terms of an integral and weaker than the Lipschitz condition (3.61) (which is the one resulting from wavelet analysis). However, using results on the expansion of characteristic functions by [Lukacs, 1983] we are able to obtain exact bounds. We state only the final result and leave mathematical details to a companion paper [47].

Corollary 3.4.2 *In general, the Lipschitz regularity of a characteristic function (3.61) is related to the critical order of moments (3.45) via*

$$\rho^+ = \lambda^+ = \sup\{r > 0 : |W[\text{Re } \phi](0, a)| \leq C a^r \text{ for } a \rightarrow 0^+\}. \quad (3.63)$$

Negative critical orders. We are now interested in estimating the negative critical order λ^- defined in (3.45), of a random variable X with density $dF_X(x)$. Let us define a new random variable Y using the one to one mapping from $\mathbb{R}$ to $\mathbb{R} : Y = g(X) = X^{-1}$. Fixing $Y = y$, equation $y = g(x)$ has only one root $x = y^{-1}$, and $|g'(x)| = y^2$, from which we deduce the distribution of Y , $dF_Y(y) = y^{-2}dF_X(y^{-1})$. The negative q th order of X simply corresponds to the positive $-q$ th order of random variable Y :

$$\mathbb{E}[|X|^q] = \mathbb{E}[|1/Y|^q] = \mathbb{E}[|Y|^{-q}]. \quad (3.64)$$

Therefore, to estimate $\lambda^-(X)$ of X , we can directly apply general results obtained in Section 3.4.2 for positive higher orders, to get

$$\lambda^-(X) = -\lambda^+(1/X) \quad (3.65)$$

Estimation procedure

In this section, we elaborate on the implementation of our estimator for λ^+ , in particular the choice of wavelet and scales to consider, its bias, robustness and optimality properties.

Implementation. Given X_i , ($i = 1, \dots, N$) a set of N observed i.i.d. samples of the distribution $dF(x)$, we use the empirical estimator for the characteristic function

$$\hat{\phi}(u) \triangleq \phi_N(u) \triangleq \frac{1}{N} \sum_{k=1}^N \exp\{iuX_k\} \quad (3.66)$$

For our purpose, we need to evaluate this function on a properly sampled interval $u_j = j \cdot \delta u$, $j = 0, \dots, K-1$, that we will make more precise later.

We now recall some convergence properties of this empirical characteristic function (see [Feuerverger and Mureika, 1977, Feuerverger and McDunnough, 1981] for details), justifying its use in the rest of our method. First, $\phi_N(u)$ converges almost surely when N goes to infinity towards $\phi(u)$ in the L^∞ sense, over some finite interval T

$$\sup_{|u| \leq T} |\phi_N(u) - \phi(u)| \rightarrow 0. \quad (3.67)$$

Second, consider the random process $Y_N(u) = \sqrt{N}(\phi_N(u) - \phi(u))$ and let $Y(u) = \bar{Y}(-u)$ be a zero mean complex Gaussian process with covariance structure $\mathbb{E}Y(u)Y(v) = \phi(u+v) - \phi(u)\phi(v)$. Then, the sequence $Y_N(u_1), Y_N(u_2), \dots, Y_N(u_m)$ converges in distribution to $Y(u_1), Y(u_2), \dots, Y(u_m)$.

It is also shown in [Feuerverger and Mureika, 1977] that $Y_N(u)$ converges weakly towards $Y(u)$ in any finite interval, provided that $\mathbb{E}|X|^{1+\delta} < \infty$.

6. Assume that $\phi^{(2p)}(0)$ exists. Then $\mathcal{M}_r$ exists if and only if [Kawata, 1972]

$$\int_0^\infty \frac{1}{t^{r+1}} |Q_{2p}(t)| dt < \infty. \quad (3.62)$$

Next to consider is the wavelet decomposition of $\phi_N(u)$ which readily simplifies to

$$\begin{aligned}\Gamma_{\phi_N}(t, a) &= \int \psi_{t,a}(v) \phi_N(v) dv \\ &= \frac{1}{N} \sum_k \Psi(a \cdot X_k) \exp\{iX_k t\}.\end{aligned}$$

Two-step estimation procedure.

(1) Assuming that Ψ is real, semi-definite we finally arrive at the surprisingly simple estimator for the maximal wavelet coefficient of $\text{Re } \phi$ of scale a , which is the main ingredient in our method :

$$\widehat{W}(a) \triangleq \Gamma_{\text{Re } \phi_N}(0, a) = (1/N) \sum_{k=1}^N \Psi(a \cdot X_k). \quad (3.68)$$

(2) Finally, according to theorem 3.4.1 the characteristic exponent ρ^+ is estimated from the power-law exponent which steers the decay of $\widehat{W}(a)$. An estimator of the critical moment order results from either corollary 3.4.1 or corollary 3.4.2. Taking the logarithmic of this powerlaw model yields the linear trend

$$\log \widehat{W}(a) \approx \widehat{\rho^+} \log a + \log C, \quad (3.69)$$

where $\widehat{\rho^+}$ is simply obtained via a standard (weighted) linear regression procedure of $\log \widehat{W}(a)$ against a restricted to some scaling interval $(a_{\min}, a_{\max})$ to be specified.

Robustness. Since we assume nothing on the distribution we obtain thus a non-parametric estimator. We also note immediately, that the estimation can be *made shift invariant* by subtracting the sample average from the data X and that it is *scale invariant*.

Indeed, consider the data $X'_i = bX_i$. Then $\widehat{W}[X'](a) = \widehat{W}[X](ba)$. Rewriting $\log(a)$ as $\log(ba) - \log(b)$ we find that the regression data of X' and X differ merely by a shift, leading to the same estimated least square slope.

Statistics of the estimator

Let us study the bias of both, the simple estimator of the wavelet coefficient (3.68) itself, as well as the derived estimation of the scaling exponent (3.69).

Since all observations are drawn from the same distribution, we may write

$$\mathbb{E}[\widehat{W}(a)] = \frac{1}{N} \sum_{k=1}^N \mathbb{E}[\Psi(aX_k)] = \mathbb{E}[\Psi(a \cdot X)] = \int \Psi(ax) dF(x). \quad (3.70)$$

This shows that as an estimator of the wavelet coefficient $W(0, a)$ itself, $\widehat{W}(a)$ is unbiased. However, as we will show, a bias is introduced as we estimate the powerlaw decay of $W(0, a)$ through the powerlaw decay of $\widehat{W}(a)$. This result is similar to the one obtained in [2] where it is shown that using log-periodograms (Welch estimator) to analyze processes with spectra of the type $\mathcal{S}_X(f) \sim \sigma^2 |f|^{-\alpha}$ leads to a systematic bias on the estimate of α . On the other hand, using a wavelet-based spectral analysis (similar to the one depicted in Section 3.2.2) yields an asymptotically unbiased estimator for exponent α .

Example : Estimating the critical order. To explore the properties of an estimator of the characteristic exponent ρ^+ through the wavelet coefficients we first treat a simple case where we assume that

- the distribution is Pareto, i.e., $F'(x) = p_X(x) = c_0 x^{-\alpha-1}$ for $|x| > \delta$ and vanishes elsewhere, with $c_0 = \alpha \delta^\alpha$;

- the wavelet is bandlimited, actually require that $\Psi(\nu) = 0$ for $|\nu| \leq \underline{\nu}$, where $\underline{\nu} > 0$ is some constant.

Such wavelets are known to exist. For instance, by construction, the auto-correlation function of any admissible and band-limited wavelet is itself a symmetric in time, band limited and positive definite admissible wavelet.

Inserting the particular form of $p_X(x/a)$ into the bias formula (3.70) we can extract the scale a through a substitution. Provided a is *small enough*, i.e., $a < \underline{\nu}/\delta$, the remaining integral is independent of the scale due to the band limitation of Ψ . This reads as :

$$\begin{aligned}\mathbb{E}[\widehat{W}(a)] &= \int_{\delta}^{\infty} \Psi(xa) c_0 x^{-\alpha-1} dx = a^{\alpha} \cdot \int_{\delta a}^{\infty} \Psi(y) c_0 y^{-\alpha-1} dy \\ &= a^{\alpha} \cdot \int_{\underline{\nu}}^{\infty} \Psi(y) c_0 y^{-\alpha-1} dy = C_{\Psi}(\alpha) \cdot a^{\alpha}.\end{aligned}\quad (3.71)$$

Thus, the exact powerlaw of the density translates into one of $\widehat{W}(a)$ thanks to the band limitation of the wavelet. Apart from this showcase, approximatively the same decay of $\widehat{W}(a)$ can be observed under much less restrictive assumptions, as we are about to show.

Estimating the critical order of fat tail distributions. We do not stress the demonstration here (refer to the original article [48] for details), and directly jump to the following result :

Proposition 3.4.1 *Assume that Ψ is positive semi-definite. Assume, the distribution has a density p_X which can be bounded as follows :*

$$p_X(x) \begin{cases} \leq a|x|^{-\alpha-1} & \text{for } |x| \geq \delta, \\ \geq b|x|^{-\alpha-1} & \text{for } |x| \geq \delta, \\ \leq c & \text{for } |x| \leq \delta. \end{cases} \quad (3.72)$$

Assume that the regularity of the wavelet ψ is larger than the critical order, i.e., $n_{\psi} \geq \alpha$. Then,

$$\tilde{a} \cdot a^{\alpha} + O(a^{n_{\psi}}) \geq \mathbb{E}[\widehat{W}(a)] \geq \tilde{b} \cdot a^{\alpha} + O(a^{n_{\psi}}), \quad (3.73)$$

with $\tilde{a}/\tilde{b} = a/b$.

Numerical robustness. Provided that the observations X_k ($k = 1, \dots, N$) are un-correlated one finds easily

$$\text{var } \widehat{W}(a) = (1/N) \text{var } (\Psi(aX)). \quad (3.74)$$

Moreover, under the assumptions of proposition 3.4.1 we conclude that $\mathbb{E}\Psi(aX) \sim a^{\alpha}$ and, considering Ψ^2 as a wavelet, $\mathbb{E}\Psi^2(aX) \sim a^{\alpha}$. Thus,

$$\begin{aligned}\text{var } (\Psi(aX)) &= \mathbb{E}\Psi^2(aX) - (\mathbb{E}\Psi(aX))^2 \\ &\sim a^{\alpha} \mathbb{E}\Psi^2(X) - a^{2\alpha} (\mathbb{E}\Psi(X))^2.\end{aligned}$$

Now, regarding the error estimate on $\widehat{W}(a)$, its variance readily takes on the form :

$$\text{var } \widehat{W}(a) \sim \frac{c}{N \nu_{\psi}^{\alpha+1}} a^{\alpha} \left(1 - \frac{a^{\alpha} c}{\nu_{\psi}^{\alpha+1} \sqrt{n_{\psi}}} \right) \quad (3.75)$$

where c , ν_{ψ} and n_{ψ} are fixed and depend only on the chosen wavelet. For small scales ($a \rightarrow 0$), the variance behaves like $\text{var } \widehat{W}(a) \sim O(a^{\alpha})$. Figures 3.12(a)–(c) show empirical variance $\text{var } \widehat{W}(a)$ varying with parameters N , a and n_{ψ} , attesting the good agreement between experimental and theoretical results.

Let us now consider the new variable $\log \widehat{W}(a)$. With a central limit theorem argument, we can say that $\widehat{W}(a)$ is asymptotically normal with mean $\delta_s \approx Aa^{\alpha}$ and variance $\sigma_s^2 \approx Ca^{\alpha}$. Then, in first

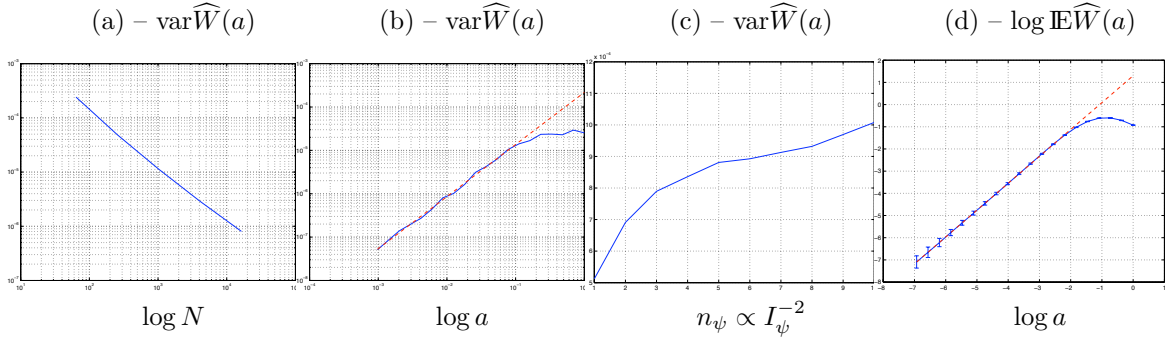


FIGURE 3.12 – Experimental verification of expressions (3.75) and (3.76). (a)–(c) : Empirical estimates of $\text{var } \widehat{W}(a)$ estimated over a set of 100 independent realizations of α -stable processes of length N . Evolution of $\text{var } \widehat{W}(a)$ is plotted as a function of : (a) $\log N$ ($\alpha = 1.2$, $n_\psi = 4$, $a = 0.0087$); (b) $\log a$ ($\alpha = 1.2$, $n_\psi = 4$, $N = 2^{14}$); (c) n_ψ ($\alpha = 1.2$, $a = 0.0087$, $N = 2^{14}$). (d) Empirical estimation of $\log \mathbb{E} \widehat{W}(a)$ versus $\log a$. The error bars correspond to the standard deviation of $\log \widehat{W}(a)$. The dashed line materializes the theoretical law $\log \mathbb{E} \widehat{W}(a) = \alpha \cdot \log(a) + C'$ ($\alpha = 1.2$, $n_\psi = 4$, $N = 2^{14}$).

approximation, using a result on functions of asymptotically Gaussian variables [Sirfling, 1980, Papoulis, 1991], we conclude that $\log \widehat{W}(a)$ is asymptotically Gaussian and

$$\begin{cases} \mathbb{E} \log \widehat{W}(a) \approx \log \mathbb{E} \widehat{W}(a) \approx \log A + \alpha \cdot \log(a) \\ \text{var } \log \widehat{W}(a) \approx |\mathbb{E} \widehat{W}(a)|^{-2} \text{var } \widehat{W}(a) \approx B/A \cdot a^{-\alpha} \end{cases} \quad (3.76)$$

See figure 3.12.

To summarize the above, we propose to estimate the characteristic exponent ρ^+ via the estimator of the scaling exponent α of the wavelet coefficients in (3.76). For (asymptotically) Gaussian random variables such as (3.76), the maximum likelihood estimator of α is simply obtained from a linear regression of $\log \widehat{W}(a)$ against $\log a$, as already suggested in (3.69). Asymptotically, the resulting estimate converges to ρ^+ . In practice though, the finite size data set limits the regression range to some interval $a \in (a_{\min} > 0; a_{\max} < \infty)$. The important issue of properly choosing this scaling region is treated in the next section.

Choice of the scale range. We have defined an estimator for ρ^+ via a log-log linear fit. While in theory the wavelet coefficients should decay as a powerlaw of the scale, we are in practice faced with the fact that the scaling deviates significantly from the theoretical ideal case for both large and small scales. Here we discuss the reasons for this deviation and explain how to choose the scaling region.

Lower bound of scaling region. We now determine a lower bound for the scale range of the linear regression $\log \widehat{W}(a)$ versus $\log a$ in (3.69).

The first one is based on a Shannon-like theorem. Our estimator estimates the singularity of the characteristic function at the origin. In practice, we use the empirical estimator for the characteristic function, i.e., $\hat{\phi}(u_k) = N^{-1} \sum_j \exp(iu_k X_j)$. The maximum variation of $\hat{\phi}$ is controlled by the maximum value of X_j . Therefore sampling $\hat{\phi}$ at a higher rate than approximately the Nyquist rate $(\bar{X})^{-1}$ with $\bar{X} = \max\{X_j, j = 1, \dots, N\}$, does not bring any finer information on the regularity of $\phi(u)$ at $u = 0$. On the contrary, when the analyzing scale goes below the minimum bound $a_{\min} = (\bar{X})^{-1}$, the measured regularity is overestimated, as the function under analysis reduces to the sole C^∞ component $\exp(iu_k \bar{X})$, oversampled at the vicinity of the origin. Thus, concordantly with theorem 3.4.1, when $\hat{\alpha}$ is estimated from data below this minimum scale bound it reflects the regularity n_ψ of the analyzing wavelet rather than the targeted regularity of the characteristic function.

We propose in [48] a second approach which, based on a probabilistic argument, leads to the same

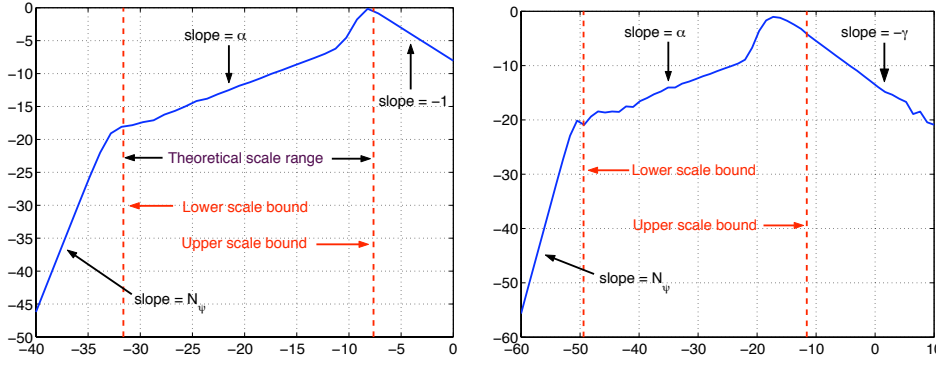


FIGURE 3.13 – Scaling region and cutoff scales : Choosing the scale too small, the resolution is fine enough for the wavelet to analyze the individual exponentials that form the estimator $\hat{\phi}$. According to section 3.4.2 the wavelet coefficients decay (at least) with exponent n_ψ . Choosing the scale too large, the estimator samples the body instead of the tail of the distribution ; thus, the wavelet coefficients adhere to a powerlaw with exponent λ^- .

conclusion that the lower cut-off scale should be chosen equal to

$$a_{\min} = 1/\bar{X}. \quad (3.77)$$

Using a stable law with index of stability (or shape parameter) α , we present in figure 3.13 the theoretical lower scale bound $a_{\min} = (\bar{X})^{-1}$. A linear regression of $\log \widehat{W}(a)$ versus $\log a$ for $a > a_{\min}$ yields an accurate estimate of the characteristic exponent $\rho^+ = \alpha$. Moreover, on this same plot, we verify that for $a < a_{\min}$, the wavelet estimator $\widehat{W}(a)$ behaves like a^{n_ψ} , in accordance with the aforementioned argument that the function under analysis is now the C^∞ exponential $\exp(iu\bar{X})$.

Upper bound and negative moments. As we saw, existence of moments is dictated by the tail decay of the distribution $F(x)$ for $x \rightarrow \infty$. For instance, it is shown in [Samorodnitsky and Taqqu, 1994], that the asymptotic tail behavior of a stable law is Pareto when $0 < \alpha < 2$. Defining when exactly this asymptotic behavior starts seems to be a tough problem (see [Nolan, 2001]), as it depends heavily on the parameterization that is used to model the distribution (in the parametric context). We just pretend here, that the upper cutoff scale $a_{\max}$ below which $\widehat{W}(a)$ behaves like a^α is also determined by this cutoff value of X separating the tail behavior (as a Pareto law for instance) from the *body* of the distribution. We illustrate this with a compound distribution, made out of a uniform distribution for $|X| \leq \delta$ and α -stable distribution for $|X| > \delta$. We show with this simple example (see Figure 3.13), that the upper cutoff scale is of order :

$$a_{\max} = \delta^{-1} \quad (3.78)$$

where δ marks the transition from body to tail behavior in the distribution. In practical applications one might choose δ from prior knowledge (rendering the estimator semi-parametric) or estimate δ itself from the scaling plots (see Figure 3.13).

Beyond this upper limit, the wavelet estimator $\log \widehat{W}(a)$ decays with slope -1 . This particular value of the slope depends only upon the distribution we have chosen for the body of our compound distribution. In our example, the uniform distribution has negative moments only for $p > \lambda^-(X) = -1$. That is precisely this bound that is estimated in figure 3.13, when $a > a_{\max}$.

The same value for $\lambda^-(X)$ would have been estimated, if instead of X directly we had analyzed the new random variable $Y = X^{-1}$ as discussed in Section 3.4.2, and estimated $\lambda^+(Y) = -\lambda^-(X)$ from the tail decay of the transformed distribution.

This observation bears a convenient consequence as far as negative moments are concerned : We can fully exploit the behavior of $\widehat{W}(a)$ for $a > a_{\max}$, leading us to a simple estimator of λ^- in (3.45).

To summarize, given K i.i.d. random variables $\{X_j, j = 1, \dots, K\}$, the wavelet estimator (3.68) behaves like :

- $\widehat{W}(a) \sim a^{n_\psi}$, for $a < a_{\min} = (\max_j \{X_j\})^{-1}$,
- $\widehat{W}(a) \sim a^{\rho^+}$, for $a_{\min} < a < a_{\max}$, where $a_{\max}$ corresponds to the inverse of the cut-off value separating the tail from the body of the underlying distribution,
- $\widehat{W}(a) \sim a^{-\rho^-}$, for $a > a_{\max}$.

This is impressively demonstrated in Figure 3.13. In fine, both ρ^+ and ρ^- can be deduced from a linear regression of $\log \widehat{W}(a)$ versus $\log a$, over the corresponding scale intervals.

As elaborated in section 3.4.2 choosing an appropriate wavelet and according to corollary 3.4.2, we have $\lambda^+ = \rho^+$ and $\lambda^- = \rho^-$.

Choice of the wavelet. The theoretical results of section 3.4.2 form the basis of our estimator. For them to hold the analyzing wavelet ψ is required to have a semi-definite Fourier transform as well as a number of vanishing moments n_ψ larger than $\mathcal{H}_{\text{Re } \phi}(0)$.

In practice, we suggest to start with a low regularity wavelet such as the second derivative of the gaussian window $\psi_2(t)$, corresponding to $n_\psi = 2$. If the slope $\widehat{\rho}^+$ obtained from the linear regression of $\log \widehat{W}(a)$ versus $\log a$ is smaller than $n_\psi = 2$, then corollary 3.4.1 immediately posits that the positive critical order λ^+ is equal to ρ^+ . Now, if the measured slope $\widehat{\rho}^+$ equals $n_\psi = 2$, we need to verify whether the regularity ρ^+ is actually larger than two or not.

To this end, we increase the number of vanishing moments $n_\psi = p$ of $\psi_p(t)$, and repeat the estimation of ρ^+ for increasing integer p as long as the slope $\widehat{\rho}^+$ hits the bound n_ψ . Once we get a $\widehat{\rho}^+ < n_\psi$, we should recall corollary 3.4.2 which again guarantees that $\lambda^+ = \rho^+$.

We can find in the original article [48] an application of particular interest which concerns the parameter estimation of stable laws. Yet, we shall not reproduce it here, for we deem more pertinent to illustrate the strength of the proposed estimator in the context of multifractal analysis.

3.4.3 Application to the multifractal partition function estimation

As we already mentioned it in 3.2.2, the multifractality of a process lies in the non-linearity of its partition function $\zeta(q)$ as defined in (3.28). A great difficulty and source of inaccuracy in multifractal analysis stems from infinite moments : to an innocent empirical estimator of moments, divergence might not be apparent. Consequently, a numerical procedure might return finite estimates of $\zeta(q)$ from (3.28) even if the moment of order q does not exist. As we are about to demonstrate with the simple examples of self-similar processes, this estimation inaccuracy may lead to a concave *estimate* of $\zeta(q)$, although the actual process is mono-fractal with a *linear* partition function.

In order to distinguish multifractal from mono-fractal processes, we must, therefore, restrict the support of $\zeta(q)$ to the actual range of orders q – with finite moments $\mathbb{E}|\Gamma_X(t, a)|^q$.

We now present some examples of well-studied processes commonly used as fractal and multifractal models.

Fractional Brownian motion. As defined in 3.1.5, the fractional Brownian motion $B_H(t)$ of index $0 < H < 1$, is a mono-fractal process whose partition function reads :

$$\zeta(q) = \begin{cases} qH & \text{for } q > -1, \\ -\infty & \text{else.} \end{cases} \quad (3.79)$$

Self-Similar stable processes. Self-similar stable processes are obtained when replacing normal by stable distributions in definition 3.1.5 of H -fBm. Such processes have discontinuous paths with interesting fractal properties and they are well adapted for infinite variance stochastic modelling. A rigorous definition reads as follows : [Samorodnitsky and Taqqu, 1994]

$$X(t) \triangleq \int f(t, u) M(du). \quad (3.80)$$

In this integral representation, M is a symmetric stable measure with characteristic exponent α and $f(t, u)$ is an integral kernel that controls the fractal properties of X . In particular, choosing $f(t, u) = (t - u)_+^{H-1/\gamma} - (-u)_+^{H-1/\gamma}$, with $(u)_+ = \max(u, 0)$, $\gamma \in (0, 2)$ and $H \neq 1/\gamma \in (0, 1)$, yields the *linear fractional stable motion* [Samorodnitsky and Taqqu, 1994]. A detailed study of this process [Delbeke and Abry, 2000, Pesquet-Popescu, 1999] showed that at each scale, the wavelet coefficients of its decomposition are identically distributed with a stable law of same index α as X . Then, taking into account the range of existence ($\lambda^- = -1 < q < \alpha = \lambda^+$) for the moments of a stable random variable and following the lines of H -fBm analysis, we get [Riedi, 2002] :

$$\zeta(q) = \begin{cases} qH & \text{for } -1 < q < \alpha, \\ -\infty & \text{else.} \end{cases} \quad (3.81)$$

Here again, $\zeta(q)$ is linear over the range $q \in (\lambda^-, \lambda^+)$ of finite moments $\mathbb{E}|\Gamma_X(t, a)|^q$.

Multiplicative Cascades. The distinguishing and defining common feature of cascades consists in an underlying multiplicative construction which is iterated across scales. The well known *canonical binomial cascade* as introduced by Mandelbrot [Mandelbrot, 1974, Mandelbrot, 1974] may be viewed as the archetype of multifractal random measures. Among the several ways of defining the binomial cascade the most useful in our context is via the iterative products

$$\beta_n(t) = \prod_{\{(j,k): 1 \leq j \leq n, t \in I_{j,k}\}} W_{j,k} = \beta_{n-1}(t) \cdot \prod_{\{k: t \in I_{n,k}\}} W_{n,k}. \quad (3.82)$$

Here, $I_{j,k}$ stands for the nested dyadic intervals $[k2^{-j}, (k+1)2^{-j})$ and $W_{j,k}$ denotes i.i.d. positive random variables of mean one ($\mathbb{E}W_{j,k} = 1$). By construction, β_n is constant over each interval $I_{n,k}$.

An equivalent construction of the binomial cascade emphasizes the measure-theoretic aspect by considering the β_n as densities and studying their distribution functions $A_n(t) = \int_0^t \beta_n(s) ds$. As positive martingales, these converge weakly to a limiting distribution X [Kahane and Peyrière, 1976, Kahane, 1987]. Also we find immediately that

$$\mathbb{E}[|A_n((k+1)2^{-n}) - A_n(k2^{-n})|^q] = \mathbb{E}[(2^{-n}\beta_n(k2^{-n}))^q] = 2^{-nq} \mathbb{E}[W^q]^n \quad (3.83)$$

which translates in a weaker form to the limiting process $A = \lim_n A_n$ and eventually to (see [Riedi, 2002] and references therein) :

$$\zeta(q) = q - 1 - \log_2 \mathbb{E}[W^q] \quad (3.84)$$

for all q where $\mathbb{E}[W^q]$ as well as $\mathbb{E}[A(1)^q]$ are finite. Clearly, this scaling behavior provides a more flexible and in some applications a more appropriate framework for modeling.

Attractive from a signal processing point of view is the iterative aspect of (3.82) which allows for fast, tree-based synthesis algorithms as the one used for the so-called MWM model [Riedi et al., 1999]. This underlying tree structure is inherited from the nested arrangement of the $I_{j,k}$ which may be represented by the points $((k+1/2)2^{-j}, 2^{-j})$ in the (time, scale)-plane.

However, such cascades have two major drawbacks. They are *not strictly stationary* since the construction is not time-shift invariant; this may result in “blocking effects”. Further, by construction the scaling of moments is log-periodic and favors, in particular, the scale ratio equal to 2. Following a more recent idea of Mandelbrot [Barral and Mandelbrot, 2002] one may overcome both drawbacks by replacing the rigid, nested arrangement of multipliers $W_{j,k}$ of the binomial cascade by a stationary random arrangement based on a Poisson process (see [Bacry and Muzy, 2003, Barral and Mandelbrot, 2002, Chainais et al., 2003, Chainais et al., 2004]). Alternative ways to construct stationary processes based on iterated multiplication have been put forward [Mannersalo et al., 2002]. Discussing such approach, however, is beyond the scope of present manuscript.

Multifractal Processes. *Mono-fractals* like H -fBm or self-similar stable processes, and multiplicative cascades like the Binomial, albeit providing us with elegant fractal paradigms, often lack flexibility to serve as models in real-world situations, due to their simplicity in scaling or due to non-stationarities. To

meet the need for closed form models that are stationary with zero mean increments, and yet present simple non-linear partition functions, the following multifractal processes were proposed in [Riedi, 2002], based on an idea of B. Mandelbrot :

$$\mathcal{B}(t) \triangleq B_H(\mathcal{M}(t)). \quad (3.85)$$

In this definition $\mathcal{M}(t)$ stands for a monotonic multiplicative cascade with flexible multifractal properties (i.e. non-linear multifractal partition function $\zeta_{\mathcal{M}}(q)$), and B_H represents an H -fBm (or more generally any self-similar process with index H). Based on a multifractal time warping, $\mathcal{B}(t)$ combines in one the rich multifractal structure of a multiplicative cascade along with the self-similarity and the non-monotonicity of a H -fBm paths. For this compound process we have the fundamental result [Riedi, 2002] :

$$\zeta_{\mathcal{B}}(q) = \begin{cases} \zeta_{\mathcal{M}}(qH) & \text{for } q > -1 \\ -\infty & \text{else.} \end{cases} \quad (3.86)$$

This partition function is clearly non-linear as it inherits from the multifractal structure of the time measure $\mathcal{M}(t)$ (in [46] we elaborated on the wavelet analysis of fBm in multifractal time).

Empirical evidence for multifractals. Using the self-similar stable process (3.80) and the time warped fBm (3.85) as paradigms for fractal and multifractal processes respectively, we now show how ambiguous it can be to decide between true multi-scaling and mono-scaling behaviors from linearity of the empirical partition function.

In [Chainais et al., 2000], a thorough study on self-similar stable processes stresses that empirical estimators of $\mathbb{E}|\Gamma_X(t, a)|^q$, based on a finite length data set, are not sensitive to theoretical divergence of moments. More precisely, whereas q -th order moments of a self-similar stable process should not exist for $q \notin (-1, \alpha)$, there is no clear evidence of this divergence in practice. Moreover, the non-linearity of the logarithmic transform used in (3.28), creates an artificial scaling $\mathbb{E} \log \widehat{\mathbb{E}}|\Gamma_X(t, a)|^q \sim \widehat{\zeta}(q) \log a$ for $q > \alpha$, very akin to the ones occurring for $q \in (-1, \alpha)$. As a result, it seems completely legitimate in practice, to stretch out $\zeta(q)$ outside its theoretical support and get the non-linear *concave empirical partition function* for self-similar stable processes :

$$\widehat{\zeta}(q) = \begin{cases} qH & \text{for } q \in (-1, \alpha) \\ 1 + q(H - \frac{1}{\alpha}) & \text{for } q > \alpha. \end{cases} \quad (3.87)$$

In Figure 3.14, the power law scalings observed for the empirical estimates $\widehat{\mathbb{E}}|\Gamma_X(t, a)|^q$, are plotted for both $q \in (-1, \alpha)$ and $q \notin (-1, \alpha)$. In the same figure, the corresponding empirical multifractal partition function is clearly non-linear over the real line, and prompts to conclude that process X is multifractal.

The same analysis applied to the multifractal process $\mathcal{B}(t)$ of definition (3.85) yields an empirical partition function that clearly deviates from linearity as well (Fig. 3.15).

- In summary, these two trivial examples emphasize how misleading it is to simply rely on linearity of the partition function, to decide whether a process is fractal or multifractal.

To be fully reliable this linearity criterion must focus on the true support of $\zeta(q)$, which is uniquely determined by the range $q \in (\lambda^-, \lambda^+)$ of converging moments : that is where our wavelet-based characteristic exponent estimator (3.68)-(3.69) comes in as an intermediate ingredient of fractal analysis. We refer to [47] for details on the implementation of this estimator.

Back to the self-similar stable process X of (3.80), we know from [Chainais et al., 2000], that for any fixed scale a_0 , its wavelet decomposition $\Gamma_X(t, a_0)$ has stable distribution with index α imposed by the integrant symmetric stable measure M used in (3.80). We then simply run the procedure described in Section 3.4.2 on $\Gamma_X(t, a_0)$, and get the characteristic interval (λ^-, λ^+) on which $\zeta(q)$ is theoretically defined. Doing so, we find for our self-similar stable example, that the empirical estimate $\widehat{\zeta}(q)$ restricted now to the predetermined interval $(\widehat{\lambda}^- = -1, \widehat{\lambda}^+ = \alpha)$ displays a linear behavior coherent with the mono-fractality of the underlying process (Fig. 3.14).

Regarding our multifractal example $\mathcal{B}(t)$, the normality of the process transposes to its wavelet decomposition $\Gamma_{\mathcal{B}}(t, a_0)$, meaning that all moments $\mathbb{E}|\Gamma_{\mathcal{B}}(t, a_0)|^q$ exist for $q \in (-1, +\infty)$. In this particular case, restricting the empirical partition function to the characteristic interval $(\widehat{\lambda}^- = -1, \widehat{\lambda}^+ = +\infty)$ does not linearize $\widehat{\zeta}(q)$, which remains the footprint of $\mathcal{B}(t)$ multifractality (Fig. 3.15).

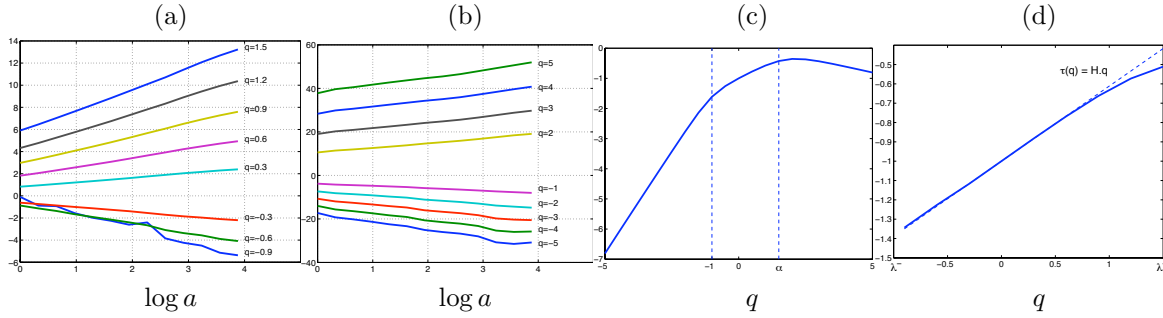


FIGURE 3.14 – Multifractal analysis of a self-similar stable process (3.80) with $\alpha = 1.5$, $H = 0.47$. We use a $N = 2^{13}$ sample points realization. (a) $\log \widehat{\mathbb{E}}|\Gamma_X(t, a)|^q$ versus $\log a$ for $q = -0.9$ (bottom curve) to $q = \alpha$ (top curve). Within this range of q 's, log-moments scale linearly with slope $\zeta(q) = qH$. (b) $\log \widehat{\mathbb{E}}|\Gamma_X(t, a)|^q$ versus $\log a$ for $q \notin (-1, \alpha)$. While in theory they should diverge, due to finite sample size effect and to non-linearity of the log function, log-moments keep scaling linearly with a different slope $\zeta(q) = 1 + q(H - 1/\alpha)$, $q > \alpha$. (c) Empirical partition function $\widehat{\zeta}(q)$ computed over a non supervised interval $(-5, 5)$. (d) Empirical partition function $\widehat{\zeta}(q)$ restricted to the predetermined characteristic interval $(\widehat{\lambda}^- = -1, \widehat{\lambda}^+ = \alpha)$.

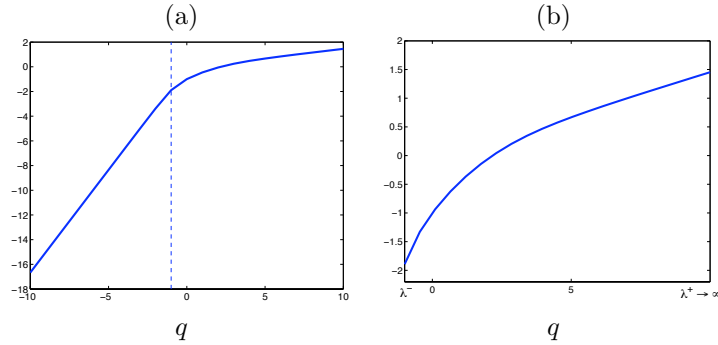


FIGURE 3.15 – Multifractal analysis of a H -fBm in multifractal time (3.85) with $H = 0.47$, and $\mathcal{M}$ a binomial multiplicative cascade. (a) Empirical partition function $\widehat{\zeta}(q)$ computed over a non supervised interval $(-5, 5)$. (b) Empirical partition function $\widehat{\zeta}(q)$ restricted to the predetermined characteristic interval $(\widehat{\lambda}^- = -1, \widehat{\lambda}^+ \rightarrow \infty)$.

3.5 Bilan et perspectives : un point de vue personnel

Un grand nombre d'autres travaux, directement ou indirectement liés à l'estimation par ondelettes de lois d'échelle, ont été publiés. Certains sont contemporains, voire antérieurs aux résultats présentés dans ce chapitre [Jaffard, 1991, Mallat and Hwang, 1992, Flandrin, 1992, Wornell and Oppenheim, 1992, Tewfik and Kim, 1992, Muzy et al., 1993, Bacry et al., 1993], mais la fraction la plus importante de ceux-ci a germé sur les concepts et sur les principes de base établis dans ces premières publications.

Parmi les rayonnements les plus marquants de l'estimateur défini dans [2], le travail de vulgarisation et d'adaptation aux décompositions discrètes d'ondelette mené dans [Veitch and Abry, 1999, Abry et al., 2000] a fourni une procédure opérationnelle et numériquement très robuste pour estimer la longue dépendance sur des données réelles. Cet estimateur est aujourd'hui un standard de référence en la matière.

Dans [Coeurjolly, 2000], l'estimation de lois d'échelle est abordée sous le même angle mais en faisant usage des différences d'ordres finis calculées sur les séries temporelles. Cette approche originale est néanmoins réductible à un formalisme de type ondelettes.

Un peu plus tard, une partie de la communauté théoricienne des statisticiens s'est à son tour intéressée aux outils ondelettes pour l'estimation de la longue dépendance. Dans le cas des processus gaussiens à mémoire longue, les performances statistiques de l'estimateur, telles que les propriétés de normalité asymptotiques des coefficients d'ondelette, l'efficacité des estimateurs obtenus par régression linéaire des log-variances, leur vitesse de convergence, le lien avec l'estimation au sens du maximum de vraisemblance, ou les bornes sur les variances d'estimation de H – autant d'aspects qui avaient été démontrés sans le même rigorisme dans [1, 2] – ont été reformulées en des termes précis qu'il n'est à présent plus question de réfuter [Moulines et al., 2007a, Moulines et al., 2007b, Moulines et al., 2008]. De même, avec l'étude menée dans [Roueff and Von Sachs, 2009], l'estimateur semi-paramétrique de l'exposant d'auto-similarité locale $h(t)$ des processus multifractionnaires – dont le principe avait été posé dans [18, 28, 38, 39] – gagne en épaisseur théorique.

S'agissant encore des lois d'échelle non homogènes, des travaux récents sur les coefficients d'ondelette dominants (*wavelet leaders*) ont réactivé l'intérêt porté à l'estimation des spectres multifractals. En généralisant aux schémas discrets la notion de lignes de maxima locaux des champs continus d'ondelettes, et en simplifiant le principe, les *leaders* sont bien plus qu'une procédure d'analyse numérique permettant de stabiliser l'estimation de spectres multifractals ou de faciliter la détection de singularités oscillantes [Arnéodo et al., 1997]. Ils apportent aussi des éléments mathématiques relatifs aux problèmes théoriques liés à la validité du formalisme multifractal lorsque celui-ci est énoncé sur les coefficients d'ondelette [Jaffard et al., 2006, Wendt et al., 2009, Wendt et al., 2007, Wendt, 2008].

Sur un autre plan, mais toujours en lien avec l'estimation des spectres de singularités, nos activités avec J. Barral et S. Seuret dans le cadre de l'action DMASC, se concentrent autour des non-convexités spectrales. Ces « pathologies » sont les signatures de ruptures de phase dans les signaux, auxquelles les spectres de Legendre strictement convexes, sont totalement aveugles. La théorie des grandes déviations contourne ce handicap et permet de définir des spectres non nécessairement convexes, qui dans le principe sont donc sensibles à ce type d'information. La difficulté qui se pose alors est d'ordre numérique et est liée au couplage fonctionnel qui doit exister entre les variables d'une double limite. Le choix de ce couplage est d'autant plus sensible qu'aucun lien analytique raisonnable n'a encore été établi. Mal déterminé il peut converger vers un spectre biaisé, auquel échappe toute information relative aux ruptures de phases. L'ajustement systématique auquel nous travaillons s'appuie sur deux raisonnements indépendants qui se recoupent : l'un issu de la théorie de la mesure et l'autre inspiré d'arguments classiques en théorie de l'estimation. Si la validité de cette proposition se confirme, c'est la nature même de l'information mise à jour par l'analyse multifractale qui s'en trouvera changée. Et, dans cette évolution aussi, les ondelettes continuent de jouer un rôle majeur.

Enfin, notons que la place des ondelettes en analyse multifractale ne se limite pas aux seuls problèmes d'estimation. Pour la définition de modèles et la synthèse de processus multifractals [Sellan, 1995, Arnéodo et al., 1998, Riedi et al., 1999, Jaffard, 2000, Seuret and Barral, 2005b, Seuret

and Barral, 2005a], comme pour la caractérisation d'espaces fonctionnels de régularité [Jaffard, 2004a, Jaffard, 2005], l'outil « ondelettes » a beaucoup contribué au développement théorique et à la vulgarisation de l'analyse multifractale.

L'objectif de ce chapitre n'était pas de faire un tour d'horizon des nombreuses approches et points de vue existants, encore moins de dresser une bibliographie précise et exhaustive sur le sujet. Néanmoins, l'ampleur de la littérature dans ce domaine, traduit un intérêt pour les objets fractals qui ne se dément pas, et qui est autant attisé par la curiosité mathématique qu'ils inspirent que par leur pertinence à décrire un grand nombre de phénomènes physiques, économiques ou sociétaux.

4

Applications

4.1 Réseaux

4.1.1 Motivations

Les réseaux informatiques sont des systèmes complexes d'échanges où interagissent un grand nombre et une toute aussi grande diversité d'agents, de médias, de codes et de mécanismes de communication. La cohésion de ce tout est maintenue grâce à un modèle stratifié (*Open System Interconnection model*) où chaque éléments d'architecture du réseau est (virtuellement) cloisonné dans une couche, et où chacune de ces couches ne peut transmettre, ou recevoir d'ordres que de la couche immédiatement supérieure ou inférieure. Cette intelligibilité hiérarchique, bien que de plus en plus contestée par les approches de *cross-layering*, structure les interactions entre processus réagissant à des échelles de temps très différentes, mais qui *in fine*, se retrouvent par le jeu de segmentations et de multiplexages, enchevêtrés dans le trafic agrégé (souvent le seul aisément mesurable). Comprendre la nature du trafic observé, l'associer à l'organisation du système et la corréler aux performances du réseau, est une démarche éprouvée dans beaucoup de disciplines, et qui met au premier plan la théorie des probabilités pour l'analyse et la modélisation des données. Cette conscience n'est pas non plus nouvelle dans l'étude des réseaux de communication, puisque dès le début du XX^{ème} siècle, les travaux de Erlang [Erlang, 1909, Erlang, 1917] démontrent le caractère poissonien des communications téléphoniques, et qu'en 1994 ce même modèle est dénoncé dans le cas des réseaux de type Internet au profit de modèles auto-similaires [Paxson and Floyd, 1994, Leland et al., 1994].

Depuis, la place des méthodes probabilistes et statistiques dans les réseaux n'a cessé d'évoluer : d'abord descriptives, elles sont maintenant incontournables dans la conception des systèmes, et feront dans un futur proche partie intégrante des mécanismes de contrôle à travers les concepts de *traffic awareness* et de *semantic networking*. Face aux défis que pose l'Internet du futur :

- limites du surdimensionnement des infrastructures ?
- stratégies de planification des transmissions et de réservation de la bande passante pour lisser les non-stationarités ;
- factorisation des ressources ;
- quel degré de fiabilité et de sûreté dans un environnement de plus en plus rapide et hétérogène ?
- quelle égalité de traitement face à la multiplication des applications et des services ?
- devant le poids économique et stratégique, quel compromis entre maximisation des gains financier pour les opérateurs, et de qualité de service pour les usagers ?

il semble difficile d'imaginer que le Traitement du Signal n'apportera pas sa part de réponses à toutes ces questions.

Pour satisfaire à ces nombreuses attentes, l'analyse du trafic dans les réseaux recourt à une large gamme de méthodologies et d'outils de mesure. Actuellement, l'environnement expérimental privilégier d'une grande majorité des acteurs dans ce domaine, repose sur des technologies de

simulation et d'émulation, au risque parfois que les limitations propres à ces approches ne particularisent, voire ne biaisent les résultats obtenus. Les simulateurs se focalisent en général sur un comportement ou un mécanisme spécifique du réseau et abstraient le reste du système, ce qui rend particulièrement difficile leur validation et leur passage à l'échelle. Dans l'exemple des simulateurs à événements discrets, le réseau est une abstraction, et le nombre de sources de trafic ou les débits simulables ne sont fonction que de la puissance de calcul de la machine sur laquelle s'exécute le simulateur. Lorsqu'il devient difficile d'identifier et d'isoler les facteurs influençant les protocoles, les émulateurs permettent quant à eux, d'exécuter sur une plate-forme entièrement contrôlée (mais figée), les logiciels dans toute leur complexité. Mais là encore, on s'éloigne de la réalité car toute la dynamique, toute la variété et toute la complexité des conditions et des équipements physiques est occultée. C'est donc pourquoi les liens physiques de véritables réseaux sont de plus en plus utilisés pour observer et pour analyser le trafic réel. L'Internet (à travers les projets tels que PlanetLab ou RON...) offre également un bon support pour déployer des expériences de métrologie du trafic dans des conditions réalistes de fonctionnement. Comme à un certain nombre d'autres équipes, il nous a alors semblé que pour mieux comprendre la vraie nature et mieux surveiller l'évolution du trafic dans les réseaux, il fallait développer en parallèle des outils classiques de mesure à faible granularité, des sondes capables de capturer le trafic à l'échelle temporelle plus fine des paquets. Cette nécessité devient même impérative si l'on souhaite identifier des lois d'échelle dans les séries temporelles. Pour estimer des invariances statistiques sur une large gamme d'échelle, le débit instantané doit être mesuré à différentes résolutions d'agrégation : de l'échelle fine des temps d'inter-arrivée entre paquets, aux échelles caractéristiques plus grandes des applications, chaque gamme d'échelle étant représentative d'une des couches du système. Réciproquement, ces lois d'échelle n'ont d'impact sur les performances du réseau qu'à travers des mécanismes sensibles aux mêmes échelles. C'est par exemple le cas de la dépendance à longue portée (Définition 3.1.3) qui n'implique que les échelles supérieures à la durée moyenne des flux, et donc généralement plus grandes que les tailles usuelles des files d'attente. De sorte que l'occupation et la saturation des *buffers* sont vraisemblablement plus sensibles aux rafales du trafic à l'échelle du RTT (*Round Trip Time*) qu'aux propriétés de longue mémoire de celui-ci.

Cependant cette métrologie au niveau du paquet est coûteuses en ressources de calcul et nécessite des dispositifs opérationnels aux bandes passantes des réseaux hauts débits ($\geq 10\text{Gbps}$) qui caractériseront l'Internet de demain (l'« Internet du futur » est au centre d'un grand nombre de programmes de recherche internationaux). Garantir une analyse fiable du trafic généré par ces nouvelles applications, impose donc de relever ces défis technologiques.

Enfin, même si l'analyse du trafic sur des liens de production réels demeure une finalité, il est important de disposer d'une plate-forme expérimentale réaliste (de par sa complexité, son hétérogénéité et ses débits) capable de générer dans des conditions contrôlées et reproductibles des signaux « stationnaires » et « paramétrables », qui permettent d'isoler et d'expliquer l'effet des différentes variables du système (niveau d'agrégation, congestion, statistique des connexions,...).

4.1.2 Une plate-forme de simulation et de métrologie

Grid5000 est une grille expérimentale nationale de recherche offrant un ensemble de plus de 5000 CPU à la communauté de chercheurs travaillant sur les thématiques de réseaux. Grid5000 reproduit un environnement similaire à celui des dispositifs de grande taille utilisés en physique, en astronomie et en biologie. C'est un outil de recherche original par ses capacités de reconfigurabilité, de contrôle et de *monitoring*, conçu pour étudier des systèmes distribués à grande échelle, et en complément des modèles théoriques et des simulateurs. Près de 17 laboratoires français participent à ce projet, hébergeant chacun une ou plusieurs grappes d'environ 500 coeurs chacune. Une infrastructure de réseau optique privé et dédié, mise à disposition par RENATER (réseau national français pour la recherche et l'enseignement), interconnecte les différents sites. Grid5000 offre également plusieurs interconnexions internationales, dont une à 10 Gbps avec DAS3, son homologue hollandais, et une à 1 Gbps avec NAREGI, le centre japonais de recherche sur les grilles (lien en phase de migration vers 10 Gbps avec l'université d'Osaka). Sur la plate-forme Grid5000, le *backbone* du réseau est composé d'un lien Ethernet privé à 10 Gbps, connecté à un cœur DWDM offrant des lambdas

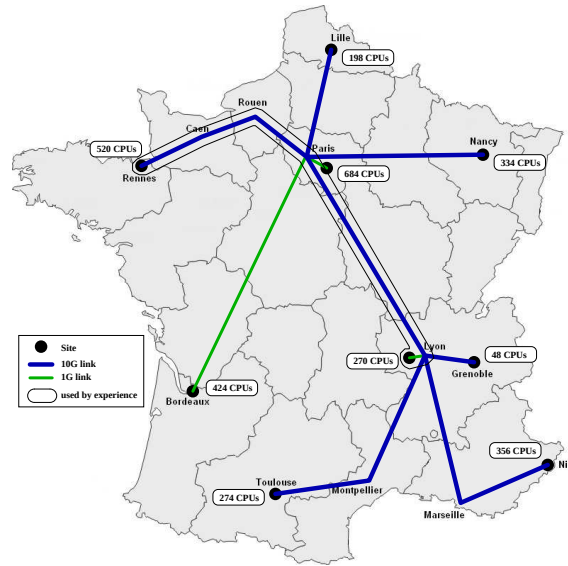


FIGURE 4.1 – Backbone de Grid5000

dédiées à 10 Gbps. Enfin, Grid5000 présente deux points de congestion à 1 Gbps à l'entrée des sites de Lyon et de Bordeaux (voir la topologie représentée à la figure 4.1).

Grid5000 laisse aux utilisateurs l'entière maîtrise des ressources nécessaires à leurs expériences. Il leur est ainsi possible de réserver les mêmes ensembles de ressources pour chaque répétition d'expérience, d'exécuter leur échanges sur des liens entièrement dédiés, d'imposer leur propres conditions expérimentales et d'installer leurs propres logiciels de mesure. Pour les y aider, ils ont à leur disposition deux outils réseau : OAR, un outil de réservation de ressources, et Kadeploy un système de déploiement d'environnement. Grâce à ces deux fonctionnalités, logiciels, modules de noyaux pour limiter les débits, mécanismes de qualité de service, ou variantes de contrôle de congestion peuvent être déployées automatiquement au cœur des systèmes d'exploitation d'origine d'un grand nombre de nœuds communicants. OAR permet la planification des expériences et des réservations d'équipements sur plusieurs heures.

Grid5000 permet donc aux chercheurs de répéter leurs expériences en garantissant à chaque fois les mêmes conditions expérimentales, ce qu'il serait pratiquement impossible de faire sur un réseau partagé non contrôlé. Il devient également possible de maintenir un environnement stationnaire sur de grandes périodes d'observation, ce que ne permet évidemment pas Internet. Enfin, Grid5000 étant un banc d'essai privé, l'installation d'équipements expérimentaux tels que des sondes de capture en certains points d'agrégation du trafic, devient relativement aisée.

MetroFlux est une sonde de métrologie que nous avons développée en collaboration avec le GTRC (Grid Technology Research Center) de l'AIST (Advanced Industrial Science and Technology - Japan). *MetroFlux* associe capture des paquets, extraction des en-têtes et analyse spécifique des données.

Il existe deux versions de *MetroFlux*, l'une permettant la capture de trafic sur un lien (cuivre) à 1 Gbps, et l'autre sur un lien (optique) à 10 Gbps. La version à 1 Gbps est capable de monitorer simultanément et de façon synchrone deux liens de capacité 1 Gbps chacun, alors que la version 10 Gbps ne permet le monitoring que d'un seul lien 10 Gbps. Pour le reste, peu de caractéristiques les distinguent.

Le cœur de *MetroFlux* est une carte FPGA programmable (**GtrcNET-1** ou **GtrcNET-10**), capable d'extraire en temps réel et sans perte, les en-têtes de paquets capturés, et de les rediriger pour traitement vers un serveur dédié. Les **GtrcNET-1** et **-10** sont configurées pour isoler les 52 octets d'en-tête (composés des 14, 20 et 18 octets des adresses respectives Etherneer, IP et TCP) des

paquets reçus sur le port (1 Gbps ou 10 Gbps) instrumenté. Chaque en-tête est estampillé d'un horodatage (*timestamp*) avec une précision de 60 ns (2^{-24} s), avant d'être encapsulé par groupe de 25 dans un paquet UDP envoyé ensuite vers un port gigabit de sortie.

Les en-têtes concaténés sont sauvegardés sur un serveur équipé d'un *quad core processor* à 2.66 GHz, de 4 GB de mémoire, de 2 ports ethernet gigabits, d'un disque système type SAS de 300 MB, d'un contrôleur RAID sur une baie RAID 0 de 5×300 GB de disques SAS offrant ainsi une capacité totale de sauvegarde de 1500 GB. Enfin, nous avons développé un pilote permettant de lire les paquets GtrcNET-1 et -10, de les désencapsuler et de les écrire dans un fichier au format pcap.

4.1.3 Validation expérimentale et extension du théorème de Taqqu

Théorie. Le théorème de Taqqu [Leland et al., 1994, Taqqu et al., 1997, Willinger et al., 1997] établit un lien fonctionnel entre deux propriétés statistiques largement observées dans les réseaux de communication : d'un côté l'auto-similarité (telle que définie au paragraphe 3.1.1) présentes dans les séries temporelles de débit agrégé, et de l'autre les décroissances à queue lourde (3.44) des distributions de tailles de flux.

Ce résultat a initialement été établi dans le cas d'une file d'attente M/G/ ∞ servie par N sources indépendantes, dont les activités $Z_i(t)$, $i \in \{1, \dots, N\}$ sont décrites par des processus binaires de type ON/OFF. La durée des périodes ON d'activité (correspondant à l'émission d'un train de paquets par la source) sont des variables aléatoires τ_{ON} distribuées selon une loi à queue lourde P_{ON} , d'exposant $\alpha = \alpha_{ON}$. Alternant avec ces périodes ON, les périodes OFF de silence (la source n'émet pas de trafic) sont de durées aléatoires τ_{OFF} distribuées selon une loi P_{OFF} qui, le cas échéant, peut aussi être à queue lourde d'indice $\alpha = \alpha_{OFF}$. Chaque source $Z_i(t)$ est donc un processus 0/1 de renouvellement avec récompense (*renewal reward process*) dont les périodes d'activité sont indépendantes et identiquement distribuées.

Soit $Y_N(t) = \sum_{i=1}^N Z_i(t)$ la série temporelle correspondante au trafic agrégé, et définissons le processus cumulé $X_N(Tt)$:

$$X_N(tT) = \int_0^{Tt} Y_N(u) du = \int_0^{Tt} \left(\sum_{i=1}^N Z_i(u) \right) du. \quad (4.1)$$

Le théorème de Taqqu [Taqqu et al., 1997] dit alors que dans la limite d'un nombre infini de sources ($N \rightarrow \infty$), et dans la limite d'un temps d'observation infini ($T \rightarrow \infty$) – l'ordre des limites a ici son importance – alors $X_N(tT)$ se comporte comme :

$$X_N(tT) \sim \frac{\mathbb{E} \tau_{ON}}{\mathbb{E} \tau_{ON} + \mathbb{E} \tau_{OFF}} NTt + C\sqrt{NT}^H B_H(t). \quad (4.2)$$

Dans cette relation, C est une constante et B_H désigne un mouvement Brownien fractionnaire (Définition 3.1.5) de paramètre de Hurst :

$$H = \frac{3 - \alpha^*}{2}, \text{ où } \alpha^* = \min(\alpha_{ON}, \alpha_{OFF}, 2). \quad (4.3)$$

Ce qu'il ressort de ce théorème, c'est que dans des conditions limites de nombre de sources et de durée d'observation infinis, un mBf superposé à une tendance linéaire est un modèle asymptotique satisfaisant pour décrire l'évolution temporelle du nombre de paquets transmis sur un lien agrégé. L'expression (4.3) de H montre quant à elle, que seules des distributions P_{ON} ou P_{OFF} de variance infinie (i.e., $1 < \min(\alpha_{ON}, \alpha_{OFF}) < 2$) peuvent générer de la longue dépendance (i.e. $H > 1/2$ dans la Définition 3.1.3). À l'inverse, des durées d'activité ON et d'inactivité OFF de variances finies ($\alpha^* = 2$) induisent systématiquement $H = 1/2$ correspondant à un processus de Wiener $X_N(t)$ décorrélé, et par conséquent sans effet de longue mémoire.

Validation expérimentale. Ce sont les différentes frontières et conditions d'applicabilité de cette relation que nous avons dans un premier temps, cherché à identifier expérimentalement dans

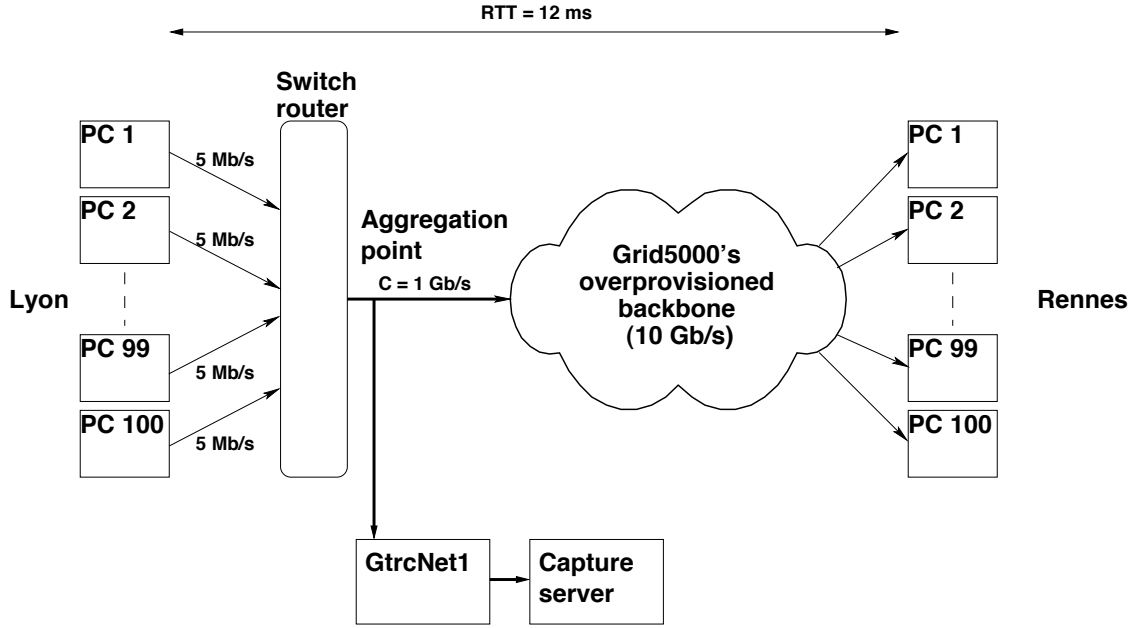


FIGURE 4.2 – Topologie du réseau de métrologie.

le travail de thèse de Patrick Loiseau et partiellement publiées dans [59]. Cette étude cristallise bon nombre des développements théoriques et méthodologiques présentés aux chapitres précédents, et les précisions atteintes sont dans une large mesure dues aux choix judicieux et à la qualité de ces outils d'analyse :

- l'estimateur de loi d'échelle présenté au paragraphe 3.2.2 (et son extension dans [Veitch and Abry, 1999]) ;
- l'estimateur d'indice de queue lourde du paragraphe 3.4, complétée dans le cas d'un trafic sous-échantillonné, par une procédure d'estimation de l'indice α_{ON} au sens du maximum de vraisemblance, développée dans [60] ;
- la plate-forme de simulation et de métrologie décrite au paragraphe 4.1.2.

Nous avons choisi la topologie de réseau en *butterfly* schématisée figure 4.2, mettant en jeu 100 clients et 100 serveurs indépendants, et un seul point de congestion à 1 Gbps.

L'entrée du système est ici la distribution statistique des tailles de flux (périodes ON) et des durées de silence (périodes OFF) que l'on sait efficacement prescrire. La sortie correspond quant à elle, aux lois d'échelle observées dans le trafic agrégé mesuré. Chaque expérience fournit une trace de 8 heures de trafic généré et enregistré sous différentes conditions de fonctionnement contrôlé et stationnaire. Le tableau 4.1 résume les différents scénarios considérés, où μ_{ON} (resp. μ_{OFF}) correspond à la durée moyenne des périodes ON d'émission de flux (resp. à celle des périodes OFF de silence) ; α_{ON} (resp. α_{OFF}) à l'indice de queue lourde de la distribution des durée ON (resp. des durées OFF) ; $\hat{H}$ le paramètre de Hurst de longue dépendance (3.20) mesuré sur les traces de trafic agrégé et $\hat{h}_{loc}$ l'exposant de régularité locale (3.21) mesuré sur ce même signal ; Δ_j représente l'échelle temporelle seuil au delà de laquelle on observe effectivement un comportement en loi d'échelle lié à de la dépendance à longue portée dans le trafic agrégé.

Les résultats issus de chacune de ces expériences sont interprétés en détail dans la thèse de P. Loiseau [Loiseau, 2009] et résumés ci-dessous.

Exp. A. En l'absence de pertes, quel que soit le protocole de transport utilisé, et quel que soit le mécanisme de limitation de débit des sources implémenté, le théorème de Taqqu, qu'il s'agisse de la gaussianité des processus de débit agrégé ou de la relation (4.3), est parfaitement vérifié. Pour

TABLE 4.1 – Vérification du théorème de Taquq : conditions expérimentales.

	Protocole	Méc. de lim. de débit	α_{ON}	α_{OFF}	μ_{ON}	μ_{OFF}	Type de perte	Sortie
A	TCP	PSP HTB TCP	1.1 – 4	exp.	0.24 s.	0.24 s.	0	$\hat{H}$ $\hat{\alpha}$
	UDP	iperf						
B	TCP	TCP	1.1 – 4	exp.	0.24 s.	0.24 s.	0	Δ_j
					2.4 s.	2.4 s.		
C	TCP	TCP	exp.	1.1 – 4	2.4 s.	2.4 s.	0	$\hat{H}$
D	TCP	TCP	1.1 – 4	exp.	0.24 s.	0.24 s.	0	$\hat{h}_{\text{loc}}$
	UDP	iperf						
E	TCP	TCP	1.5	exp.	0.55 s.	0.77 s.	0.7% (lossy)	$\hat{H}$
					6.14 s.	0.70 s.	5% (lossy)	
F	TCP	TCP	1.1 – 3	exp.	1000 pkts	0.48 s.	$\sim 0.45\%$ (congestion)	$\hat{H}$

toutes les configurations testées, les courbes empiriques donnant $\hat{H}$ en fonction de $\hat{\alpha}^*$ suivent sans différences notables les prédictions théoriques, comme l'illustre les deux tracés typiques correspondant aux cas UDP et TCP de la figure 4.3.

Exp. B. Il est avancé dans [Hohn et al., 2003], que la gamme d'échelle sur laquelle doit se manifester la longue dépendance induite par la décroissance lente des distributions de taille de flux, est bornée inférieurement par la durée moyenne de ces flux. Les résultats de cette série d'expériences confirment cette intuition, les Δ_j mesurés (0.64 s et 10.28 s) étant dans un rapport sensiblement de même ordre que celui des paramètres μ_{ON} (0.24 s et 2.4 s).

Exp. C. Cette expérience, duale de l'expérience A, montre l'effet similaire de la décroissance à queue lourde des distributions de temps OFF. Dès lors que $\alpha_{\text{OFF}} < \alpha_{\text{ON}}$, ce sont effectivement les longues périodes de silence qui imposent la longue dépendance (les tracés de $\hat{H}$ en fonction de $\hat{\alpha}^*$ sont comparables à ceux de la figure 4.3) et par conséquent non reproduits ici.

Exp. D. Le but recherché ici est de montrer que d'autres lois d'échelle, distinctes de la longue dépendance et donc étrangères au théorème de Taquq, coexistent dans les traces de débit agrégé. Ces lois d'échelle, mesurées aux échelles de temps inférieures au RTT, rendent compte de la dynamique d'émission des paquets par les protocoles (ou par les mécanismes de limitation de débit), et caractérisent la régularité locale (3.12) du débit instantané (à l'échelle d'agrégation). Les valeurs estimées de $\hat{h}_{\text{loc}}$ soulignent alors le caractère « en rafales » (*bursty*) du protocole TCP ($\hat{h}_{\text{loc}} \simeq 0.63$) par rapport à celui plus dilué (mais moins régulier) du protocole UDP ($\hat{h}_{\text{loc}} \simeq 0.4$).

Exp. E. Nous considérons à présent l'effet de la boucle de contre-réaction TCP sur la relation (4.3). Pour activer le mécanisme d'auto-régulation de TCP, nous avons provoqué des pertes de type Bernoulli en aval du point de congestion (situation correspondante à un *lossy link*). Pour un taux de perte raisonnable (i.e. 0.7%), la probabilité de perte de paquets successifs est faible et le protocole TCP reste dans sa phase de *congestion avoidance* n'interrompant pas la transmission. Dans ces conditions, la connexité des flux reste intacte et la longue dépendance observée conforme à la prédiction de la relation (4.3). En revanche, face à un taux de perte élevé (i.e. 5%), le protocole TCP réagit en passant à l'état d'*exponential backoff*, et suspend l'émission de paquets sur des durées pouvant excéder celle du RTT de plusieurs ordres de grandeur. Les flux longs ainsi fragmentés, sont perçus à la macro-échelle du système comme une série de flux plus courts, masquant la distribution à queue lourde sous les apparences d'une distribution exponentielle. Conformément, le paramètre de Hurst mesuré $\hat{H} \simeq 0.5$ reste cohérent avec la relation de Taquq (4.3), laquelle, dans ces conditions, n'est pas remise en cause par le contrôle actif de TCP.

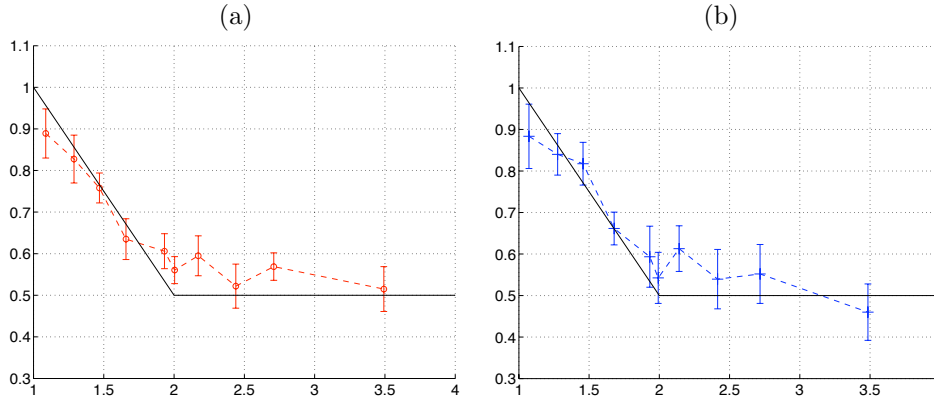


FIGURE 4.3 – Vérification expérimentale de la relation de Taquq (4.3) dans le cas des protocoles : (a) UDP – (b) TCP. La courbe en trait plein correspond à la prédiction théorique : l’axe des abscisses désigne la valeur prescrite α_{ON} , l’axe des ordonnées, la valeur estimée $\hat{H}$. Les marqueurs $\circ$ représentent la moyenne des points expérimentaux, les barres d’erreur superposées symbolisent l’écart-type de la mesure.

Exp. F. Il en va différemment dans le cas de la présence d’un point de congestion (*bottleneck*). Le mécanisme d’AIMD (*Additive Increase Multiplicative Decrease*) de TCP a pour but l’occupation maximale de la bande passante disponible, tout en imposant un principe d’équité entre flux. Cette optimisation sous contrainte d’un critère global induit une forte corrélation entre les sources et produit un débit agrégé qui sature à la capacité C du lien. Le principe d’indépendance des sources n’étant plus garanti, pas plus que le caractère gaussien du processus $X_N(t)$ (et ce, même à une très grande échelle d’agrégation), la longue dépendance dont serait responsable la distribution des tailles de flux est masquée par l’effet « stabilisateur » de la saturation, et par conséquent difficile à mesurer sur l’ensemble du trafic agrégé. Toutefois, si l’on considère un sous-ensemble aléatoire de sources (en nombre suffisamment grand pour approcher le cas limite $N \rightarrow \infty$), le trafic partiel agrégé correspondant, soumis à une saturation elle-même variable dans le temps, gagne en « degré de liberté », et retrouve un caractère normal. Sous ces conditions, le paramètre de longue dépendance $\hat{H}$ estimé à partir de ce trafic partiel, évolue avec $\hat{\alpha}_{ON}$ selon une loi comparable à la relation (4.3) du théorème de Taquq (l’accord entre courbes expérimentale et théorique étant cependant moins tangible que dans les tracés de la figure 4.3).

Extension du théorème de Taquq. Le modèle de sources ON/OFF qui conduit à la relation (4.3) éclaire le rôle de l’indice α_{ON} des durées de flux sur la génération de longue dépendance. Si les N sources émettent toutes avec le même débit R , les variables aléatoires de durée et de taille Q ont (au facteur R près) la même distribution à queue lourde. La relation (4.3) s’applique alors indifféremment à l’un ou l’autre de ces indices puisque $\alpha_Q = \alpha_{ON}$. En fait, cette hypothèse est rarement vérifiée car on observe souvent sur des trafics réels, des distributions de tailles et de durées qui, si elles sont toutes deux à queue lourde, n’ont pas les mêmes exposants α_{ON} et α_Q . Les mécanismes de TCP (notamment la phase de *slow start*), ou des latences point-à-point différentes dans les réseaux, introduisent une corrélation entre le débit moyen R_i d’un flux i et sa taille Q_i , ou parce qu’elles sont linéairement liées, sa durée D_i . Nous avons donc cherché à étendre le théorème de Taquq en précisant les rôles spécifiques de chacun des indices α_{ON} et α_Q sur la génération de longue dépendance dans le trafic agrégé.

Pour cela, nous avons commencé par reformuler le modèle de sources ON/OFF à partir d’un processus (planaire) de Poisson à deux dimensions (T_i, D_i) de densité Λ , où les T_i représentent les instants d’arrivée des flux i , supposés indépendants et uniformément distribués (Figure 4.4).

La distribution à queue lourde des durées D_i est ici portée par l’expression particulière de la

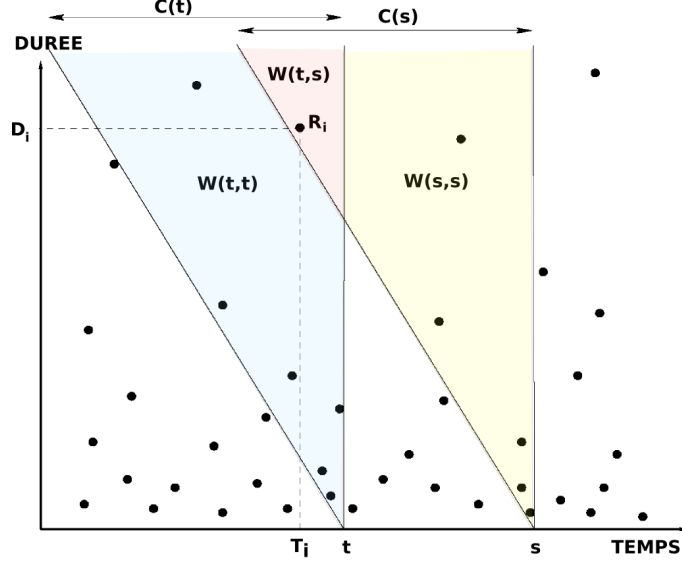


FIGURE 4.4 – Modélisation de sources ON/OFF par un processus de Poisson planaire.

densité qui s'écrit

$$\Lambda(dt, dd) = \frac{C dt dd}{d^{\alpha_{ON}+1}}.$$

À chaque point (T_i, D_i) est ensuite affecté un poids (une récompense) R_i correspondant au débit moyen du flux i , et corrélé à la durée D_i via l'espérance conditionnelle suivante :

$$\mathbb{E}\{R_i \mid D_i\} = K D_i^{\beta-1} \quad \text{et} \quad \text{Var}\{R_i \mid D_i\} = V \quad (= \text{Cte}).$$

Cette forme est le seul lien compatible avec des distributions de durée D_i et de taille Q_i à queues lourdes, d'indices respectifs α_{ON} et α_Q , et vérifiant la contrainte linéaire $D_i = R_i \cdot Q_i$ (on retrouve le cas usuel $\alpha_{ON} = \alpha_Q$ correspondant à un débit R_i constant lorsque $\beta = 1$). Posons alors $\alpha' = \alpha_{ON} - 2(\beta - 1)$, et notons

$$W_t = \sum_{(T_i, D_i) \in C(t)} R_i,$$

le débit instantané à l'instant t , résultant de la contribution de tous les points (T_i, D_i) contenus dans le cône $C(t)$ de la Figure 4.4. On peut alors calculer (voir [Loiseau, 2009] pour les démonstrations) la fonction d'auto-corrélation du débit instantané

$$\rho_W(\tau) = \frac{CK^2}{\alpha'(\alpha' - 1)} \tau^{1-\alpha'} + \frac{CV}{\alpha_{ON}(\alpha_{ON} - 1)} \tau^{1-\alpha_{ON}},$$

dont la décroissance aux grandes échelles (i.e. $\tau \rightarrow \infty$) est dominée par la plus petite des valeurs entre α_{ON} et α' . Le théorème de Taqqu s'étend alors à la proposition suivante :

Proposition 4.1.1 *Le débit agrégé résultant de la superposition d'un nombre infini de sources ON/OFF indépendantes de durée et de taille à queues lourdes d'indices respectifs α_{ON} et α_Q est un processus asymptotiquement à longue dépendance d'exposant de Hurst*

$$H = \frac{3 - \alpha_H}{2}, \quad \text{avec} \quad \alpha_H = \alpha_{ON} - 2 \max\left(\frac{\alpha_{ON}}{\alpha_Q} - 1, 0\right).$$

Suivant les circonstances, il est donc possible que la corrélation entre la taille et le débit des flux, accentue la dépendance à long terme du débit agrégé, au delà du seul effet lié aux distributions à queue lourde des durées de flux.

4.1.4 Nouvelles propriétés d'invariance d'échelle des flux TCP

La longue dépendance discutée au paragraphe précédent, n'est pas la seule loi d'échelle qui ait été observée dans les signaux de trafic. Aux échelles de temps plus fines, comme celles du RTT par exemple, la régularité locale (au sens de la définition 3.1.4) des séries temporelles est un indicateur possible de caractéristiques importantes du trafic telle que la *burstiness*, ou encore (mais cela reste une hypothèse) l'empreinte de certains protocoles. Les auteurs de [Riedi and Lévy Véhel, 1997] ont été parmi les premiers à identifier des propriétés multifractales, telles que celles évoquées aux paragraphes 3.1.5 et 3.2.3, sur les liens d'accès à certains réseaux. Depuis lors, l'origine de cette multifractalité, et notamment le rôle joué par le protocole TCP, n'a cessé d'intriguer et inspiré (avec divers niveaux de justification) un grand nombre de modèles mathématiques pour rendre compte de ces invariances d'échelle [Riedi et al., 1999, Feldmann et al., 1999, Feldmann et al., 1998, Maulik and Resnick, 2003, Sarvotham et al., 2001, Sarvotham et al., 2005, Barral and Lévy Véhel, 2004, Baccelli and Hong, 2002].

Alors que tous ces travaux traitent du support multifractal de la régularité hölderienne des trajectoires, c'est une toute autre mesure qui a été considérée dans le Chapitre 8 de [Loiseau, 2009]. En nous restreignant au cas des flux TCP à long terme (*long lived TCP flows*), nous nous sommes intéressés à l'évolution du débit moyen d'un flux, calculé sur des fenêtres temporelles de largeur variables. Sans rentrer dans les démonstrations, détaillées dans [Loiseau, 2009], fixons simplement quelques notations :

w_i : la taille de la fenêtre de congestion d'un flux TCP au i -ème RTT ;

n : la taille (en nombre de RTT) de l'intervalle glissant sur laquelle est calculé le débit moyen du flux TCP ;

k_n : le nombre d'intervalles $j = 1, \dots, k_n$ adjacents de taille n disponibles sur la durée N finie de l'observation ;

$\eta_j^{(n)}$: la moyenne des fenêtres de congestions w_i calculé sur le j -ème intervalle de taille n , i.e.

$$\eta_j^{(n)} = \frac{\sum_{i=jn}^{i=(j+1)n} w_i}{n}, \quad j = 1, \dots, k_n.$$

On peut alors montrer que la mesure $\eta_j^{(n)}$ joue le rôle d'un exposant de grain qui a pour support un ensemble multifractal. La distribution de cet exposant est invariante par changement de l'échelle n , et est associée à un principe de grandes déviations selon lequel :

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{\#\{j \in \{1, \dots, k_n\} : \eta_j^{(n)} \in (\alpha - \varepsilon, \alpha + \varepsilon)\}}{k_n} = f(\alpha). \quad (4.4)$$

Le spectre des grandes déviations $f(\alpha) < 0$ fixe la proportion d'exposants $\eta_j^{(n)}$ proches (à ε près) de α , et indique que celle-ci décroît comme $\exp\{nf(\alpha)\}$ lorsque l'échelle n tend vers l'infini.

Ce résultat, démontré dans des conditions très générales de mesures $(w_i)_{i \in \mathbb{N}}$ mélangeantes, ouvre des perspectives sur trois fronts :

- 1) Il démontre que le protocole TCP, via le mécanisme de AIMD, est naturellement à l'origine d'une invariance d'échelle multifractale.
- 2) Cette propriété multifractale est intrinsèque aux chaînes de Markov irréductibles et apériodiques, couramment utilisées pour modéliser l'évolution de la fenêtre w_i d'un flux TCP. Le spectre des grandes déviations $f(\alpha)$ est alors lié à la pression (ou fonction de partition) $P(q)$ de la chaîne de Markov, par la transformée de Legendre (identique au cas hölderien de la relation (3.27)) :

$$f(\alpha) = -\sup_{q \in \mathbb{R}} \{\alpha q - P(q)\}. \quad (4.5)$$

- 3) Le principe des grandes déviations (4.4) nous dit enfin que le spectre $f(\alpha)$ peut être empiriquement estimé à partir d'une seule réalisation de la chaîne de Markov $\mathbf{w} = (w_1, w_2, \dots)$ (moyennant une condition sur la séquence $(k_n)_{n \in \mathbb{N}}$ du nombre d'intervalles). Ce résultat important d'ergodicité nous permet en particulier de vérifier expérimentalement les deux assertions précédentes.

Dans [Loiseau, 2009], la généralisation aux chaînes de Markov vectorielles, formellement à un nombre infini de sources, en pratique à deux sources partageant le même *bottleneck*, permet de tenir compte dans le modèle, et de quantifier par l'estimation des spectres, les effets de certaines interactions entre les flux. On y verra illustrés les cas de la synchronisation des sources, et l'équité du partage de la bande-passante.

4.2 Analyse du rythme cardiaque

Le système cardiovasculaire est au même plan que les réseaux de communication, un système dynamique complexe, et à eux comparable par bien des aspects :

- Le tissu de vascularisation constitue un réseau de transport densément ramifié où coexistent une grande diversité de débits et de longueurs de chemins ;
- Le cœur est une pompe qui fonctionne sur un mode impulsif (polarisation/dépolarisation) et qu'il est commun de modéliser par un processus ponctuel ;
- Pour adapter le flux sanguin aux besoins instantanés et aux ressources disponibles, un mécanisme d'auto-régulation du rythme cardiaque est assuré par le système nerveux autonome, lui-même renseigné par un ensemble de capteurs physiques et chimiques répartis à tous les niveaux de l'organisme ;
- Ce sont des systèmes multi-échelles, où les temps de réaction des nombreux échanges couvrent une très large gamme. Dans le circuit de transport, les temps d'acheminement du sang du cœur aux extrémités varient sur plusieurs ordres de grandeur ; dans le circuit de contrôle, les fréquences caractéristiques des branches sympathiques et parasympathiques du système nerveux autonome, occupent des bandes spectrales différentes.

Sans pousser plus loin l'analogie entre système cardiovasculaire et réseau de communication – il serait facile de relever au moins autant de dissemblances structurelles et fonctionnelles entre les deux systèmes – aucun des deux n'est réductible à un jeu d'équations dont la résolution décrirait le comportement ou l'évolution. Dans les deux cas, seule l'expérience et l'observation permettent d'inférer des règles de fonctionnement (dont la seule connaissance ne suffit cependant pas à prédire le résultat final) pour modéliser et mieux comprendre le système.

4.2.1 Détermination de la nature du tonus cardiaque chez la souris

L'essentiel des tests préliminaires médicamenteux est actuellement réalisé sur le rat. Comme l'humain, le rythme cardiaque de base de ces petits mammifères est dominé par la branche parasympathique (ou vagale) du système nerveux autonome (SNA), laquelle tend en permanence à ralentir la fréquence des pulsations cardiaques. Cette régulation de fond est appelée « tonus vagal ». À l'effort, la branche sympathique du SNA est stimulée et, sans prendre totalement le relais de la branche parasympathique qui n'est jamais totalement inhibée, permet au rythme cardiaque d'accélérer pour répondre aux besoins énergétiques. C'est l'action combinée de ces deux boucles de contrôle antagonistes qui permet de réguler en permanence le rythme cardiaque.

Contrairement au rat, la souris supportant assez bien la vie en captivité, son taux de survie jusqu'à l'âge adulte est beaucoup plus élevé. Pour des raisons économiques évidentes, l'industrie pharmaceutique s'intéresse donc de plus en plus à ces autres petits mammifères pour conduire leurs études expérimentales de protocoles médicamenteux. L'un des projets de recherche menés à l'Unité INSERM U572 de l'Hôpital Lariboisière (Paris) avait pour objectif de déterminer le type de tonus, vagal ou sympathique, dominant le rythme cardiaque de base chez la souris. Notre collaboration¹ avec Bernard Swynghedauw et Pascale Mansier entre 1998 et 2003, s'est articulée autour de trois axes.

Boîte à outil d'analyse du rythme cardiaque

La grande diversité des appareils utilisés pour l'acquisition et le traitement des signaux ECG, chacun associé à un format spécifique de données, rendait très difficile l'inter-opérabilité des systèmes. De même, le développement et l'intégration de fonctionnalités nouvelles aux logiciels commercialisés étant soumise à l'intervention (coûteuse) de leurs concepteurs, l'adaptation des traitements aux besoins manquait de réactivité. Pour toutes ces raisons, nous avons décidé de

1. Cette étude est malheureusement restée inachevée pour des raisons d'ordre institutionnel. J'ai néanmoins tenu à présenter les résultats préliminaires que nous avons obtenus, car bien qu'incomplets, je pense qu'ils soulèvent une question de fond et suggèrent un mode original de régulation du rythme cardiaque.

TABLE 4.2 – Valeur moyenne du RR (écart-type) [nombre d’individus] pour chaque classe. Les temps sont exprimés en ms.

	T_0 [n=50]	T_{20} [n=40]	T_{40} [n=41]	T_{60} [n=40]
C	114.4 (17.7) [n=50]	95.9 (7.8) [n=12]	98.6 (7.4) [n=13]	102.7 (8.7) [n=12]
A		92.8 (6.0) [n=10]	96.9 (9.4) [n=10]	100.0 (9.5) [n=10]
P		131.3 (22.0) [n=9]	139.3 (33.3) [n=9]	146.6 (41.3) [n=9]
AP		123.0 (22.0) [n=9]	128.0 (15.4) [n=9]	130.2 (12.8) [n=9]

développer en interne, un ensemble d’interfaces graphiques et de programmes d’analyse du signal ECG répondant mieux aux impératifs de nos activités de recherche. Cette boîte à outils « U127 », développée en langages C et Matlab, nous a offert une grande souplesse de traitement en nous laissant l’entière maîtrise des opérations de détection, de filtrage, d’estimation, de modélisation et de représentations du rythme cardiaque.

Application de modèles linéaires mixtes à l’étude statistique du rythme cardiaque

Pour essayer de déterminer le type de tonus particulier chez la souris, nous avons procédé à une série d’expériences sur une base de 18 individus mâles adultes de race C57bl/6. Un capteur bio-compatible sous cutané mesurant l’activité électrique du cœur et transmettant l’ECG par télémetrie, est implanté sur chaque souris, la laissant ainsi libre de tous mouvements. Trois jours après l’intervention, on enregistre les ECG correspondants aux réponses à quatre type d’injection :

- C** : une dose placebo de solution saline qui servira de contrôle ;
- A** : une dose d’atropine pour inhiber l’influence du tonus parasympathique ;
- P** : une dose de propranolol, un bêta-bloquant des fonctions du système nerveux sympathique ;
- AP** : une dose d’atropine et de propranolol combinés, reproduisant les conditions d’un cœur dénervé.

Le rythme cardiaque, déterminé à partir du signal ECG, est donné par la série temporelle des intervalles RR entre deux battements successifs. La valeur moyenne du RR est calculée sur une fenêtre de 5 minutes au temps T_0 (avant l’injection), puis aux temps T_{20} , T_{40} et T_{60} correspondant respectivement à 20, 40 et 60 minutes après l’injection. Les valeurs obtenues sont regroupées dans le tableau 4.2.1.

Les réponses du rythme cardiaque aux différents blocages du système autonome montrent :

- un important effet placebo dû au stress de l’injection ;
- l’absence de tachycardie notable après blocage du système parasympathique (vagal) ;
- un ralentissement très marqué après désactivation de la branche sympathique ;
- un rythme sinusal après dénervement, comparable à celui produit par l’inhibition du système sympathique.

En première lecture, ces observations semblent réfuter l’hypothèse d’un tonus vagal chez la souris, puisque l’inactivation de la branche parasympathique ne produit pas l’accélération cardiaque attendue. En revanche, bloquer le rôle accélérateur du système sympathique se traduit bien par un allongement de l’intervalle RR moyen. L’atropine ne semble donc pas avoir d’influence plus significative que celle d’un placebo et laisse penser que le rythme cardiaque de la souris fonctionne en permanence en mode accéléré. . .

Si l'on regarde de plus près le plan d'expérience, on voit qu'il est *(i)* déséquilibré, car la population étudiée n'est pas la même pour chaque classe; *(ii)* avec répétitions, puisque la même souris peut servir à différentes expériences. Pour tenir compte de ces conditions expérimentales, une simple analyse de la variance ne suffit pas, et c'est pourquoi nous avons recouru dans [47] aux modèles linéaires mixtes (LMM) pour caractériser le rôle précis de chacune des variables du système. Sans entrer dans le détail, les LMM permettent d'isoler dans l'explication de la variance, les effets aléatoires dus :

- aux variabilités inter-individus (indépendantes de l'état et conditionnellement liées à ceux-ci) ;
- aux corrélations intra-individus générées par les répétitions (temporelles et inter-classes) ;
- à l'influence du temps sur les mesures ;

des effets fixes liés aux seules variables explicatives que sont le type d'injection $\{C, A, P, AP\}$ et la date $\{T_0, T_{20}, T_{40}, T_{60}\}$. Cette étude a donc confirmé les conclusions tirées du tableau 4.2.1, en démontrant notamment que l'injection d'atropine n'entraîne pas d'effet fixe significatif sur le rythme cardiaque moyen, et qu'elle est donc à exclure des variables explicatives. De plus, les modèles linéaires mixtes ont permis de quantifier l'influence des facteurs aléatoires liés aux conditions expérimentales, et de formaliser l'effet déterministe des facteurs fixes retenus : injection de propranolol (seul ou associé à l'atropine) et facteurs temps T_{20} et T_{40} .

Une conjecture simple pour expliquer le mécanisme de contrôle cardiaque chez la souris

Si l'injection d'atropine ou de placebo ne produisent pas d'accélération notable du rythme cardiaque moyen, l'observation des séries temporelles de RR révèle toutefois des différences de variabilité très claires. La Figure 4.5 montre les rythmes cardiaques (séries d'intervalles de RR) de souris de contrôle, après injection d'atropine et après injection d'un placebo. Autour d'une ligne de base sensiblement constante (environ 120 ms), on voit nettement que l'amplitude des variations est fortement atténuée par l'atropine. En particulier, il est intéressant de noter que les allongements de période très soudains et limités dans le temps (correspondant à des bradycardies supérieures à 100 battements par minutes sur une dizaine de secondes) disparaissent totalement sous l'effet de l'atropine, mais persistent dans le cas du placebo.

D'un point de vue spectral, il est communément admis que chez la plupart des mammifères, les deux branches du SNA coopèrent à travers un mécanisme de *push-pull*, mais en réagissant chacune dans des gammes de fréquences caractéristiques différentes. Le rapport entre les puissances activées dans une bande spectrale haute – par la branche parasympathique – et basse – par la branche sympathique, mais aussi parasympathique – est un indice couramment utilisé pour mesurer l'équilibre sympatho-vagal [Akselrod et al., 1981]. Nous avons alors vérifié dans le cas de la souris, que la présence ou l'absence de ces « bradycardies impulsionnelles » modifiaient le bilan énergétique du rythme cardiaque, au point de pouvoir expliquer la valeur prise par cet indice sous l'effet des différentes injections. Ces observations nous ont donc amenés à la conjecture suivante :

Proposition 4.2.1 *Chez la souris, le rôle régulateur de la branche parasympathique n'agit pas en continu, mais sur un mode impulsionnel provoquant des bradycardies transitoires quasi-périodiques.*

Pour consolider notre intuition, nous avons procédé à une autre campagne de mesures chez la souris au repos (enregistrements de jour) et en activité (enregistrement nocturnes). La figure 4.6 montre quelques exemples de séries temporelles de RR filtrées (les dérives basses-fréquences ont été supprimées par un filtrage adaptatif basé sur la décomposition modale empirique présentée au paragraphe 2.4, afin d'isoler les régimes transitoires).

Pour vérifier qu'il existe bien une différence significative entre les bradycardies diurnes et nocturnes, nous nous intéressons aux statistiques des valeurs extrêmes des intervalles RR. Nous notons y^+ la série des maxima locaux (strictement) positifs, et y^- la série des minima locaux (strictement) négatifs. Chacune des séries ainsi obtenue est normalisée par son écart inter-quartiles. Les distributions des valeurs extrêmes $\mathbf{Y} = [y^+, y^-]$ pour chacun des deux états jour et nuit sont représentées figure 4.7-(a).

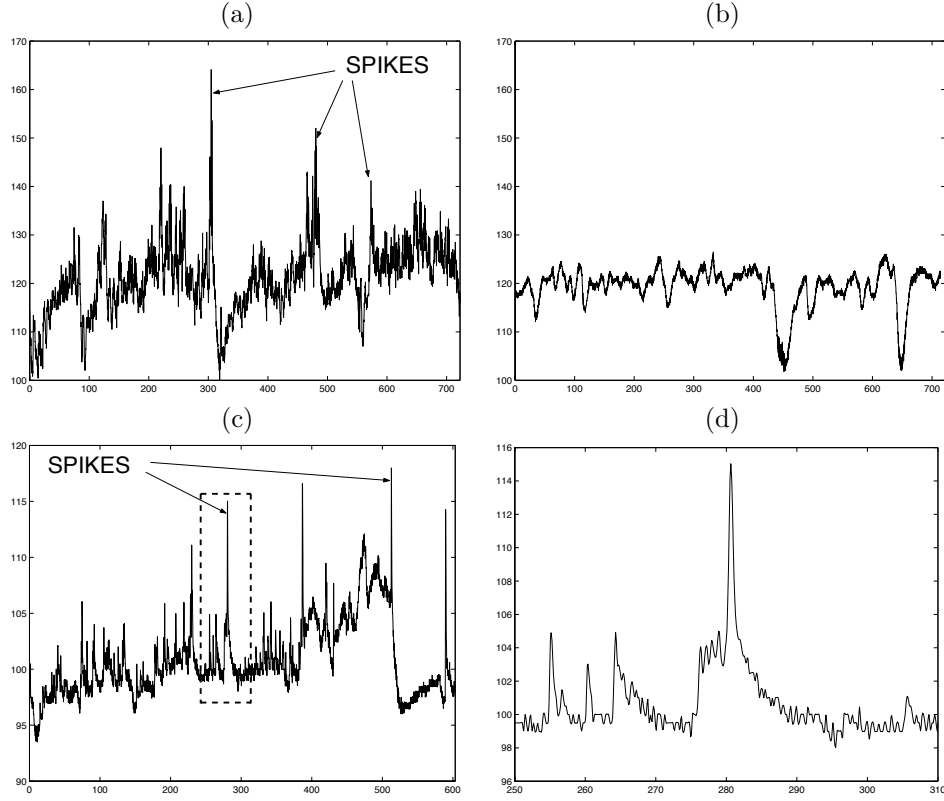


FIGURE 4.5 – Rythme cardiaque de souris enregistré par télémétrie sous-cutané. (a) Rythme de contrôle (aucune injection). (b) Rythme 20 minutes après injection d’atropine. (c) Rythme 20 minutes après injection d’une solution saline placebo. (d) Agrandissement d’un *spike* repéré dans l’encadré en pointillés de la figure (c).

Les deux densités sont pratiquement confondues pour les valeurs négatives de $\mathbf{Y}$, indiquant en cela que les tachycardies ont des dynamiques d’amplitudes relatives comparables aux états de repos et d’activité.

S’agissant des bradycardies, de jour, les intervalles entre deux battements consécutifs peuvent subir des allongements (par rapport à la ligne de base) jusqu’à trois fois supérieures à l’amplitude des contractions. La distribution de type « queue lourde » (3.44) qui en résulte montre qu’au repos le rythme cardiaque de la souris peut, avec une probabilité non négligeable, subir un ralentissement considérable (jusqu’à 50%) par rapport au rythme moyen de base.

Ces mêmes densités (jour et nuit) représentées dans un schéma correctement normalisé, permettent de visualiser l’écart des données à la normalité (figure 4.7-(b)). Dans les deux cas, superposées aux régressions linéaires caractéristiques des lois gaussiennes, les densités modifiées mettent en évidence le caractère non gaussien des valeurs extrêmes des intervalles RR. Lié à l’existence de queue lourde, l’écart à la normalité est nettement plus prononcé pour les valeurs extrêmes y^+ diurnes que nocturnes.

Pour conduire un test d’hypothèse entre les deux états (jour *versus* nuit), nous avons modélisé les données y^+ et $-y^-$ par des lois Gamma d’expression :

$$P_Y(y) = \frac{c^b}{\Gamma(b)} y^{b-1} e^{-cy} U(y). \quad (4.6)$$

Les lois Gamma ne sont pas à queue lourde mais permettent néanmoins de modéliser des variables aléatoires strictement positives pouvant prendre avec une probabilité non négligeable, des valeurs « accidentellement » élevées par rapport à la variance. Les valeurs estimées des paramètres

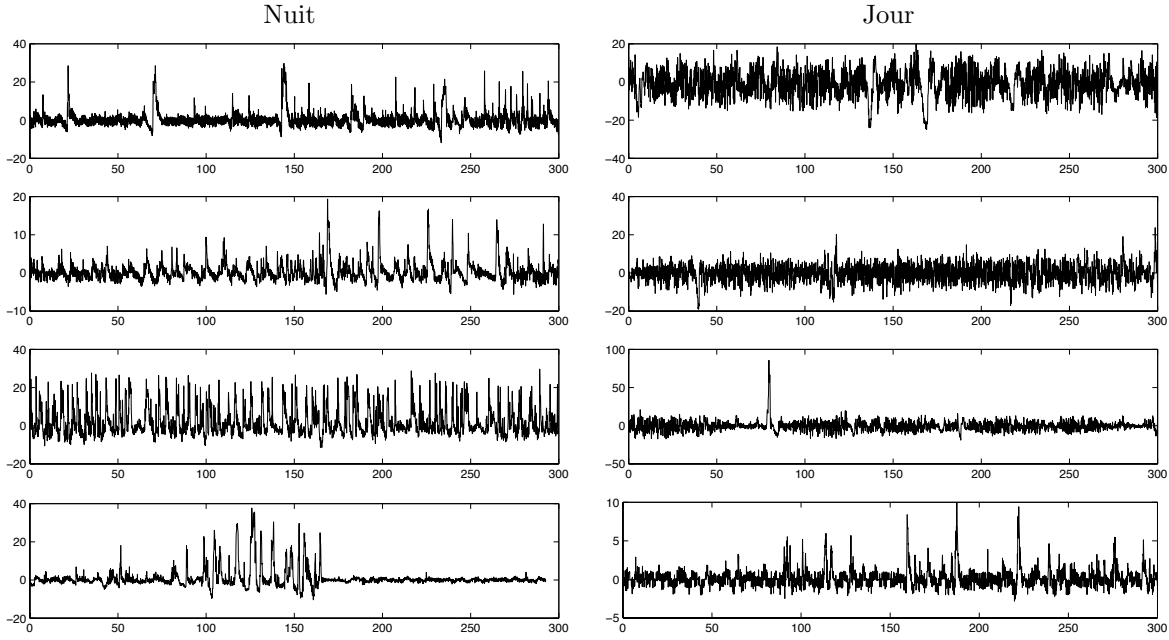


FIGURE 4.6 – Exemples caractéristiques de rythmes cardiaques chez la souris : séries d’intervalles inter-battements (RR) filtrées. Enregistrements de nuit, lorsque les souris sont en pleine phase d’activité et de jour, dans la phase de repos. L’axe des abscisses est gradué en seconde (5 minutes d’enregistrement). L’axe des ordonnées est gradué en ms.

	y^+		y^-	
	Jour	Nuit	Jour	Nuit
c	0.93 (0.3)	1.51 (0.49)	1.64 (0.57)	1.51 (0.46)
d	1.35 (0.74)	0.57 (0.24)	0.46 (0.17)	0.58 (0.16)

TABLE 4.3 – Valeur moyenne (écart-type) des paramètres c et d d’une loi Gamma d’expression (4.6), estimés à partir des valeurs extrêmes (positives et négatives) des intervalles inter-battements enregistrés chez la souris au repos (de jour) et en activité (de nuit).

b et c correspondant à chacune des deux séries (y^+ et y^-) dans chacun des deux états (jour *vs* nuit) sont consignées dans la table 4.3.

Pour vérifier l’existence de différences significatives entre ces quatre distributions, nous procédons à une analyse de la variance sur les deux paramètres estimés c et d . Les p -values correspondant à chacun des test d’hypothèse sont présentées dans les tableau croisés 4.4.

Les vraisemblances associées à chacun de ces tests d’hypothèses montrent que seules les bradycardies au repos se distribuent selon une loi significativement (voire très significativement) différente des autres statistiques de variation de l’intervalle RR. Le rythme cardiaque de jour subit des ralentissements d’amplitudes et d’occurrences très significativement supérieures aux variations du rythme cardiaque nocturne. Par ailleurs, les tachycardies au repos et en activité, comme les décélérations en phase active, sont identiques et présentent peu de variations « anormalement » élevées par rapport à la variabilité chiffrée par l’écart-type des séries RR.

Enfin, nous avons voulu caractériser la dynamique des séries temporelles de RR et en particulier la morphologie des tachycardies transitoires. Pour cela nous avons modélisé les tracés de la figure

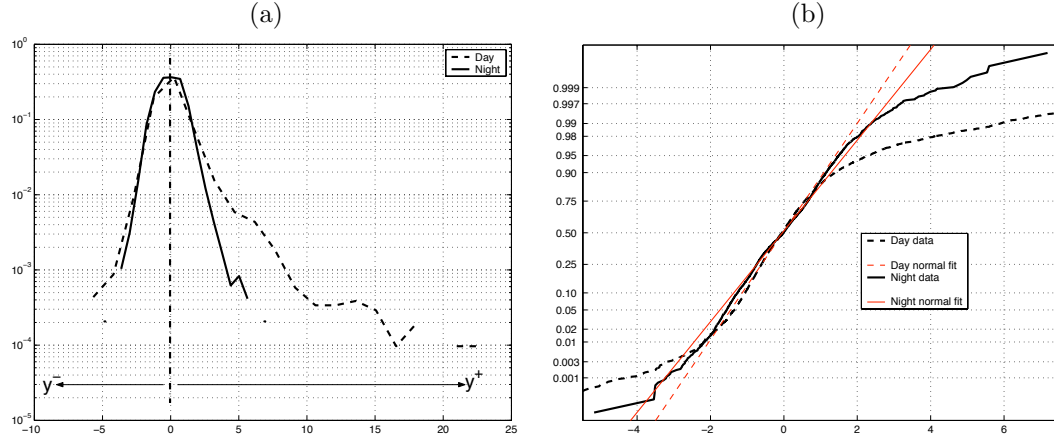


FIGURE 4.7 – (a) Profils moyen des densités des valeurs extrêmes $\mathbf{Y}$ enregistrées de jour (tirets) et de nuit (trait plein). L’axe des abscisses correspond aux valeurs des données normalisées par l’inter quartile, l’axe des ordonnées représente la log-probabilité estimée (par histogramme). (b) Affichage normal de ces mêmes distributions. Les droites superposées correspondent aux régressions des données par un modèle gaussien. L’axe des abscisses correspond aux valeurs des données renormalisées sous hypothèse gaussienne, l’axe des ordonnées correspond à la distribution (probabilité cumulée) des données.

Param. c		y^+		y^-	
		Jour	Nuit	Jour	Nuit
y^+	Jour	-	-	-	-
	Nuit	0.01*	-	-	-
y^-	Jour	0.0018**	0.68	-	-
	Nuit	0.0089**	0.99	0.68	-

Param. b		y^+		y^-	
		Jour	Nuit	Jour	Nuit
y^+	Jour	-	-	-	-
	Nuit	0.04*	-	-	-
y^-	Jour	0.001**	0.32	-	-
	Nuit	0.0422*	0.93	0.2	-

TABLE 4.4 – Test d’hypothèse (analyse de la variance) entre les quatre lois Gamma estimées au tableau 4.3. Les valeurs indiquées correspondent aux p -values. Une valeur de $p < 0.05^*$ indique une différence significative entre les deux distributions correspondantes, et $p < 0.05^{**}$ indique une différence très significative.

TABLE 4.5 – Ajustement du rythme cardiaque (RR) de la souris au repos, par un processus ponctuel filtré. Valeurs empiriques des paramètres du modèle estimés sur chaque individu.

Souris	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9
$\langle \tau_i \rangle$ [s]	0.77	0.65	0.78	0.66	0.86	1.02	0.95	0.53	0.85
$\langle A_i \rangle$ [ms]	17.5	32.3	31	19.5	23.15	30.3	29.1	16	24.5
$\langle (t_{i+1} - t_i) \rangle$ [s]	20.4	12.4	22.7	19.8	15.4	11.3	4.9	134.7	12.6

4.6 par un processus ponctuel filtré (*shot noise*) d'expression :

$$s(t) = \sum_{i \in \mathcal{I}} A_i h(t - t_i; \tau_i).$$

Dans ce modèle, la série $\{t_i, i \in \mathcal{I}\}$ correspond aux instants aléatoires d'arrivée des impulsions, A_i l'amplitude de ces impulsions et τ_i un paramètre de forme du filtre linéaire h que l'on identifie à un système du premier ordre de réponse impulsionnelle :

$$h(t; \tau_i) = \exp(-t/\tau_i) U(t).$$

Les constantes de temps τ_i sont estimées par ajustement du modèle aux données (illustration figure 4.8).

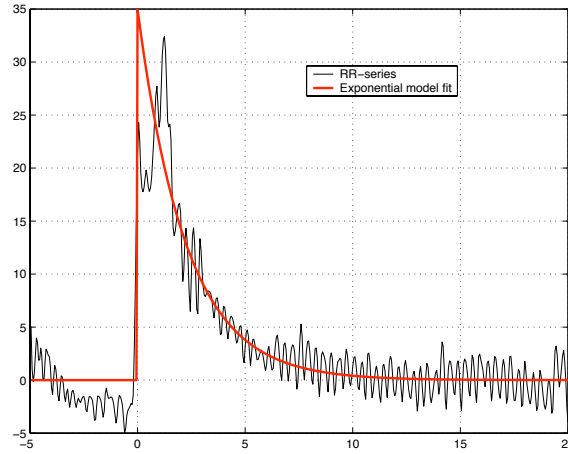


FIGURE 4.8 – Ajustement d'une bradycardie (observée chez une souris au repos) à un modèle exponentiel du premier ordre.

Le tableau 4.5 présente pour chacune des souris à l'état de repos, les valeurs moyennes estimées sur 5 minutes d'enregistrement, des paramètres τ_i , A_i et du temps moyen d'inter-arrivée entre impulsions $(t_{i+1} - t_i)$. Ces valeurs moyennes sont données à titre indicatif, car à quelques rares exceptions, les séries de RR étant fortement non-stationnaires, la densité des événements, ainsi que l'amplitude des impulsions sont très variables au cours du temps (nous n'avons d'ailleurs pas pu vérifier si les instants d'arrivée des bradycardies suivaient une loi de Poisson). Seules les constantes de temps τ_i semblent relativement stables au cours du temps. De même, d'un individu à l'autre, les temps d'inter-arrivée des impulsions peuvent varier dans de très grandes proportions (comparer

par exemple les souris S_7 et S_8).

4.2.2 Analyse multifractale du rythme cardiaque

Les circonstances et les pistes qui nous ont conduits à nous intéresser à l'analyse multifractale du rythme cardiaque ont déjà été présentées au paragraphe 1.4.3. Ces deux projets sont encore trop récents pour avoir produit la matière nécessaire à un développement plus détaillé. Soulignons simplement que dans l'étude du rythme cardiaque chez le fœtus, comme chez le sujet diabétique, les objectifs visés d'aide au diagnostic s'appuient sur des altérations subtiles de la variabilité des intervalles RR [6]. Et qu'en particulier, nous nous interrogeons sur les dissymétries entre les statistiques (que l'on mesure avec les spectres multifractals, par exemple) des accélérations et des décélérations du rythme cardiaque. Une démarche que l'on peut, par un raccourci du raisonnement, mettre en perspective de l'analyse descriptive du paragraphe précédent sur la dynamique des bradycardies.

Conclusions, Bilan et Perspectives de Recherche

Dans ce document, j'ai souhaité éclairer avec la même intensité les trois composantes de mon activité professionnelle :

- développement d'outils d'analyse harmonique adaptés aux cas non stationnaires ;
- caractérisation de propriétés statistiques d'invariance d'échelle ;
- applications à l'étude de systèmes complexes.

Le traitement du signal est au centre de chacune de ces parties et maintient la cohésion du tout. Parce qu'il est fondamentalement pluridisciplinaire et qu'il émarge sur plusieurs champs scientifiques, le traitement du signal m'a amené à interagir avec diverses communautés. Soit pour son propre enrichissement, comme ce fut le cas pour les représentations temps-fréquence, soit pour augmenter l'arsenal des outils et des méthodes d'autres collègues, tels que celui des statistiques, de la bio-informatique, des sciences de la terre ou des réseaux ; mais dans tous les cas, il s'est agi à chaque fois d'un véritable partage scientifique.

En introduisant les équations aux dérivées partielles dans les représentations temps-fréquence, nous avons ouvert la classe de Cohen et la classe affine à un potentiel de lissages adaptatifs et anisotropes jusqu'alors inexploré. Nous n'avons montré les avantages de cette approche que pour optimiser le compromis entre réduction des interférences et localisation des composantes, mais le principe peut facilement s'étendre à toutes autres formes de contraintes. Comme cela a été le cas en traitement d'images, la diffusion permettrait alors de renouveler un certain nombre de points de vue sur les représentations temps-fréquence et sur leur utilité dans les applications réelles. Après que leur côte de popularité se soit passablement essoufflée au cours des dernières années, on pourrait voir dans cette nouvelle source d'adaptativité, un possible regain de dynamisme et d'intérêt pour les classes de distributions quadratiques.

Dès 1990, avec le début de mon travail de thèse, j'ai eu la chance d'accompagner les premiers résultats sur l'utilisation des ondelettes pour la caractérisation des lois d'échelle dans les signaux. Depuis lors, cette activité n'a pas cessé de s'intensifier. Notre travail sur les estimateurs de longue dépendance, avec en particulier l'étude comparée entre analyse spectrale de Fourier et celle construite sur les opérateurs d'ondelette, a servi de marche-pied et reste l'une des contributions importantes dans ce domaine. Repris à différents comptes, adapté au cadre des décompositions discrètes, étendu et perfectionné au cas des processus non gaussiens, statistiquement reconnu, opérationnellement démontré, cet estimateur est devenu un standard ayant même réussi à forcer la porte de certaines communautés longtemps dubitatives et qui en redécouvrent aujourd'hui toutes les vertus.

L'activité de recherche sur les thèmes d'invariance d'échelle est toujours très vivace, et les perspectives sont de plusieurs ordres. Sur le plan des processus multifractionnaires, nous avons listé un certain nombre de variations autour des représentations temps-échelle pour estimer la

régularité locale des trajectoires. Cette préoccupation semble revenir dans le rond de lumière de la communauté statistique, relançant ainsi le défi de l'estimation de lois d'échelles localement stationnaires. Mais l'enjeu n'est pas que théorique, en pratique aussi, certaines situations non triviales (par exemple rencontrées en finance), semblent correspondre à des cas de figure où l'indice de longue dépendance varie avec le temps, justifiant ainsi de l'effort qu'il faut continuer d'investir dans cette direction.

Avec l'analyse multifractale et deux microlocale, les ondelettes ont retrouvé une seconde jeunesse dans les *leaders*. Mais d'autres approches plus anciennes sur les champs continus et les maxima locaux d'ondelettes pourraient aussi être ramenées sur le devant de la scène. C'est le cas des « oscillations » mais ce pourrait être aussi celui de la transformée de Legendre sur des lignes de coefficients d'ondelettes pour caractériser les singularités oscillantes, ou du formalisme grand-canonique.

S'agissant des spectres multifractals, nous n'avons pas abordé les spectres des grandes déviations. Mais ceux-ci sont au cœur du projet DMASC avec deux prolongements possibles : l'analyse multifractale vectorielle et les spectres multifractals signés. Sur ce dernier point, les résultats obtenus, bien qu'encore préliminaires, sont très encourageants puisque sur les signaux de rythme cardiaque, accélérations et décélérations semblent obéir à des dynamiques d'échelle différentes.

Une autre déclinaison possible pour l'analyse multifractale est bien sûr motivée par les nouvelles propriétés d'invariance d'échelle identifiées sur les flux TCP et naturellement reproduites par les chaînes de Markov. Plus généralement, les réseaux sont, aux côtés de la turbulence ou du système cardiovasculaire, un autre paradigme de systèmes complexes, riche en interactions multi-échelles et dans lesquels la recherche en analyse multifractale trouve matière à se développer.

Comme on l'a vu, toujours avec l'application aux réseaux, les distributions de variables aléatoires à queues lourdes sont intimement couplées à ces problématiques de lois d'échelle. Confronté à des problèmes d'estimation statistique, avec l'aide de spécialistes du domaine, j'ai apporté un certain nombre de contributions méthodologiques en inférence statistique. C'est le cas du test non paramétrique d'existence des moments finis d'une variable aléatoire, qui répondant à un besoin ressenti dans un grand nombre d'applications distantes (nous en avons étudié deux), comble indiscutablement un manque dans la panoplie des outils statistiques. De façon plus marginale, car répondant à un besoin moins pressant dans les domaines applicatifs, la régression des paramètres d'une loi stable, l'estimation de l'indice de queue lourde à partir de données incomplètes, ou la modélisation des coefficients d'ondelette par arbres de Markov cachés, sont d'autres exemples de collaborations qui ont enrichi le répertoire méthodologique des estimateurs statistiques.

Au niveau des applications, et plus particulièrement de celle visant à une meilleure compréhension du trafic dans les réseaux, nous sommes nombreux à penser que métrologie, traitement des données et évaluation de performances, forment un triptyque qui fera partie intégrante des systèmes de communications de demain [Brabaux, 2009]. Pour prévenir la dégradation des performances d'Internet, l'allocation dynamique des ressources en fonction de l'activité mesurée sur la bande passante, semble une solution de plus en plus plausible. Elle nécessitera alors l'intégration de sondes pour mesurer le trafic, l'extraction d'attributs pertinents et la définition de lemmes pour permettre au système de s'auto-configurer et de s'adapter aux conditions d'utilisation. Sur l'ensemble de ces points, mon activité au sein du projet RESO mise sur cette évolution des infrastructures, et je pense que les orientations scientifiques choisies nous placent en bonne position pour relever ce défi. C'est vrai pour l'analyse sémantique menée dans le cadre du « Laboratoire Commun INRIA - Alcatel Lucent Bell Labs », dont l'objectif final est d'optimiser le partage des ressources via un traitement différencié des connexions selon une typologie judicieuse des flux. Et c'est également dans ce but que la plate-forme expérimentale GRID5000 et le banc de métrologie METROFLUX ont été pensés. Inférer à partir de l'observation des modèles théoriques puis les confronter à la validation empirique est une démarche éprouvée en physique, sur laquelle nous comptons pour mieux comprendre les liens entre mécanismes de transport et profils statistiques des données. Les résultats déjà obtenus sur la caractérisation du trafic nous confortent dans ce choix, et je suis convaincu qu'il faut poursuivre dans cette voie. À double titre. Tout d'abord, parce que les

modèles mathématiques se complexifient, que les infrastructures de réseaux évoluent vers toujours plus d'hétérogénéité, que les couches systèmes sont de plus en plus poreuses, que la virtualisation est amenée à créer une dissociation entre ressources physiques et services mis à disposition des utilisateurs. . . l'expérimentation sur des plate-formes réalistes et fiables est un réflexe qui pourrait s'intensifier et se banaliser aux cotés des simulateurs et des approches numériques. L'autre raison tient aux nouveaux enjeux économiques et technologiques qui transforment le regard porté sur les réseaux et sur leur rôle dans la société. Encore essentiellement utilisés pour communiquer, ils sont dans certaines optiques de l'Internet du futur des supports modulables et reconfigurables de calculs intensifs et distribués. Cette nouvelle dimension s'accompagne de contraintes supplémentaires liées à la fiabilité, au partage des ressources ou en amont au contrôle d'admission. Elle rend encore plus critiques des questions qui se posent déjà, comme par exemple : Comment anticiper sur les besoins des utilisateurs et adapter dynamiquement l'offre à la demande ? Selon quels critères hiérarchiser les besoins et comment utiliser l'information disponible pour contrôler efficacement la gestion des requêtes ? Quels sont les voyants qui peuvent alerter sur les risques de « crises » et sur quels leviers agir pour les éviter ? Les politiques de réservation et de planification offrent quelques éléments de réponse à ces questions, mais déportent au second plan les contraintes de dynamicité posées par les interactions homme-machine. La prédiction (dynamique ou par le raisonnement) des processus aléatoires, avec toutes les difficultés qu'elle soulève, est alors d'autant plus concevable que les informations collectées sur l'état du système sont fiables et de bonne qualité. Pour satisfaire à cette exigence, instrumentation, métrologie et théorie du signal ont donc toutes les raisons de progresser ensemble pour dépasser le simple rôle d'aide au dimensionnement des liens et des équipements. C'est sur la base de ces éléments de réflexion que j'inscris mon projet de recherche : renforcer les liens entre expérimentation et analyse statistique des données appliquées aux réseaux, tout en veillant à maintenir l'équilibre entre sciences de l'ingénierie et recherche en mathématiques appliquées. Si ce tandem nous a déjà donné quelques satisfactions, c'est en grande partie grâce au travail d'une équipe qui couvre un large spectre de compétences complémentaires. Pour être acteur des transformations qui s'annoncent dans le domaine des réseaux, cette pluridisciplinarité est indispensable, autant que peut l'être la confrontation d'idées issue de collaborations avec des chercheurs d'autres cultures, et en particulier avec des communautés scientifiques centrées sur d'autres systèmes complexes. En l'occurrence, mon travail sur le système cardiovasculaire, jusque là concomitant de l'étude des réseaux, pourrait bien nous guider dans la conception de modèles bio-inspirés. Mais ce transfert de connaissances peut être réciproque, notamment s'agissant de l'identification de certains facteurs ou mécanismes responsables des lois d'échelle observées dans le rythme cardiaque, toutes choses étant égales par ailleurs.

Bibliographie Personnelle

- [1] P. Abry, P. Gonçalves, and P. Flandrin. Wavelet-based spectral analysis of $1/f$ processes. In *IEEE, Int. Conf. on Acoust., Speech and Signal Proc.*, pages III.237–III.240, Minneapolis (US), April 1993.
- [2] P. Abry, P. Gonçalves, and P. Flandrin. Wavelets, spectrum analysis and $1/f$ processes. *Wavelets and statistics*, 1995. Lecture Notes in Statistics.
- [3] P. Abry, P. Gonçalves, and J. Lévy Véhel. *Lois d'Échelle, Fractales et Ondelettes*, volume 1 and 2 of *Information - Commande - Communication*. Hermès Sciences Publications, 2002.
- [4] P. Abry, P. Gonçalves, and J. Lévy Véhel. *Scaling, Fractals and Wavelets*. Digital signal and image processing series. ISTE – John Wiley & Sons, Inc., London (UK), 2009.
- [5] P. Abry, H. H. Wendt, S. Jaffard, H. Helgason, P. Gonçalves, E. Pereira, Cl. Gharib, P. Gaucherand, and M. Doret. Methodology for multifractal analysis of heart rate variability : From $1/f$ ratio to wavelet leaders. In *32nd Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, Buenos Aires (Argentina), Sept. 2010.
- [6] P. Abry, H. Helgason, P. Gonçalves, E. Pereira de Souza Neto, M. Doret, and C. Gharib. Multifractal analysis for in partum fetal-ECG diagnosis. In *IEEE Int. Conf. on Acoust. Speech and Sig. Proc.*, Dallas, TX (US), March 2010.
- [7] A. Antoniadis, A. Feuerverger, and P. Gonçalves. Wavelet based estimation for univariate stable laws. *Annals of the Institute of Mathematical Statistics*, 58(4) :779–807, 2006.
- [8] F. Auger, P. Flandrin, and P. Gonçalves. *Time-frequency in action with Matlab*. 2000. Freeware tutorial.
- [9] S. Baykut, P. Gonçalves, P.-H. Luppi, P. Abry, E. Pereira de Souza Neto, and D. Gervasoni. Emd-based analysis of rat eeg data for sleep state classification. In *Biosignals*. Springer, January 2009.
- [10] H. Carrão, A. Araújo, P. Gonçalves, and M. Caetano. Multitemporal meris images for land cover mapping at national scale : the case study of portugal. *International Journal of Remote Sensing*, 2008. To appear.
- [11] H. Carrão, M. Caetano, and P. Gonçalves. Meris based land cover characterization : a comparative study. In *Proceedings of the ASPRS international conference*, Reno (NV, US), May 2006.
- [12] H. Carrão, P. Gonçalves, and M. Caetano. Use of intra-annual satellite imagery time-series for land cover characterization purpose. In *Proceedings of the 26th EARSeL Symposium*, Warsaw (Poland), 2006.
- [13] H. Carrão, P. Gonçalves, and M. Caetano. Land cover characterization through parametric modeling of intra-annual reflectance time series : a comparative study with meris data. In *SPIE Europe Symposium on Remote Sensing*, Firenze (Italy), Sept. 2007.
- [14] H. Carrão, P. Gonçalves, and M. Caetano. Contribution of multispectral and multitemporal information from modis images to land cover classification. *Elsevier, Remote Sensing of Environment*, 112(3) :986–997, 2008.
- [15] H. Carrão, P. Gonçalves, and M. Caetano. A nonlinear model for satellite images time series : analysis and prediction of land cover dynamics. *IEEE Trans. in Geosciences and Remote Sensing*, 2009. To appear.

- [16] J.-B. Durand, P. Gonçalves, and Y. Guédon. Arbres de markov cachés : inférence et applications. In *Proceedings of XXXVIèmes Journées de Statistique*, Montpellier (France), 2004.
- [17] J.-B. Durand, P. Gonçalves, and Y. Guédon. Statistical inference for hidden markov tree models and application to wavelet trees. *IEEE Trans. on Signal Processing*, 52(9) :2551–2560, September 2004. Also Inria Technical Report RR-4248, 2002.
- [18] P. Flandrin and P. Gonçalves. From wavelets to time-scale energy distributions. In G. W. LL. Schumaker, editor, *Recent Advances in the Theory of wavelets*, pages 309–334. Academic Press, 1992.
- [19] P. Flandrin and P. Gonçalves. Geometry of affine distributions. In *IEEE, Int. Symp. Time-Frequency and Time-Scale Analysis*, pages 80–83, Philadelphia (US), May 1994.
- [20] P. Flandrin and P. Gonçalves. Geometry of affine time-frequency distributions. *Applied Computational Harmonic Analysis*, 3 :10–39, January 1996.
- [21] P. Flandrin and P. Gonçalves. Sur la décomposition modale empirique,. In *Proceedings of the 19th Colloquium GRETSI*, Paris (France), 2003.
- [22] P. Flandrin and P. Gonçalves. Empirical mode decompositions as data-driven wavelet-like expansions. *Int. J. of Wavelets, Multiresolution and Information Processing*, 2(4) :477–496, 2004.
- [23] P. Flandrin, P. Gonçalves, and P. Abry. Invariance d’échelle et ondelettes. In P. Abry, P. Gonçalves, and J. Lévy Véhel, editors, *Lois d’échelle, Fractales et Ondelettes*, pages 19–51. Hermès Sciences Publications, 2002.
- [24] P. Flandrin, P. Gonçalves, and P. Abry. Scale Invariance and Wavelets. In P. Abry, P. Gonçalves, and J. Lévy Véhel, editors, *Scaling, Fractals and Wavelets*, pages 71–102. ISTE – John Wiley & Sons, Inc., 2009.
- [25] P. Flandrin, P. Gonçalves, and G. Rilling. Detrending and denoising with empirical mode decompositions. In *Proceedings of Eusipco*, pages 1581–1584, Wien (Austria), 2004.
- [26] P. Flandrin, P. Gonçalves, and G. Rilling. Emd equivalent filter banks, from interpretation to applications. In N.E. Huang and S.S.P. Shen, editors, *The Hilbert-Huang Transform and Its Applications*, pages 55–74. Interdisciplinary Mathematical Sciences, 2005.
- [27] P. Flandrin, G. Rilling, and P. Gonçalves. Empirical mode decomposition as a filter bank. *IEEE, Signal Processing Letters*, 11(2) :112–114, February 2004.
- [28] P. Gonçalves. *Représentations Temps-Fréquence et Temps-Echelle Bilinéaires : Synthèse et Contributions*. Thèse de doctorat 3ème cycle, Inst. Nat. Polytechnique de Grenoble, 1993.
- [29] P. Gonçalves. Existence test of moments : Application to multifractal analysis. In *Proceedings Int. Conf. on Telecommunications*, Acapulco (Mexico), May 2000.
- [30] P. Gonçalves. *Analyse temps-fréquence et modèles d’invariance d’échelle : Contribution à l’étude de systèmes complexes*. Habilitation à diriger des recherches, École doctorale de mathématiques et d’informatique fondamentale de Lyon, 2010.
- [31] P. Gonçalves and P. Abry. Multiple window wavelet transform and local scaling estimation. In *IEEE Int. Conf. on Acoust. Speech and Sig. Proc.*, Munich (Germany), April 1997.
- [32] P. Gonçalves, P. Abry, G. Rilling, and P. Flandrin. Fractal dimension estimation : empirical mode decomposition versus wavelets. In *IEEE Int. Conf. on Acoust. Speech and Sig. Proc.*, Honolulu, Hawaii (US), April 2007.
- [33] P. Gonçalves and R. Baraniuk. Pseudo affine wigner distributions. In *IEEE, Int. Conf. on Acoust., Speech and Signal Proc.*, Atlanta (US), April 1996.
- [34] P. Gonçalves and R. Baraniuk. A pseudo-bertrand distribution for time-scale analysis. *IEEE Signal Processing Letters*, 3(3) :82–84, March 1996.
- [35] P. Gonçalves and R. Baraniuk. Pseudo affine wigner distributions : Definition and kernel formulation. *IEEE Trans. on Signal Processing*, 46(6) :2305–2314, June 1998.
- [36] P. Gonçalves, H. Carrão, and M. Caetano. Parametric model for intra-annual reflectance time series. In *Proceedings of the 26th EARSel Symposium*, Warsaw (Poland), 2006.

- [37] P. Gonçalves, H. Carrão, A. Pinheiro, and M. Caetano. Land cover classification with support vector machine applied to modis imagery. In *Proceedings of the 25th EARSeL Symposium*, Porto (Portugal), June 2005.
- [38] P. Gonçalves and P. Flandrin. Bilinear time-scale analysis applied to local scaling exponents estimation. In *Progress in Wavelet Analysis and Applications (Y. Meyer and S. Roques, eds.)*, pages 271–276, Toulouse (France), June 1992.
- [39] P. Gonçalves and P. Flandrin. Scaling exponents estimation from time-scale energy distributions. In *IEEE, Int. Conf. on Acoust., Speech and Signal Proc.*, pages V.157–V.160, San Francisco (US), April 1992.
- [40] P. Gonçalves and P. Flandrin. Sur la localisation et la géométrie des distributions affines. In *Proceeding of the 14th colloquium GRETSI*, pages 355–358, Juan-les-pins (France), September 1993.
- [41] P. Gonçalves and P. Flandrin. Distributions affines : de la géométrie des interférences à la localisation des distributions. In *Proceedings of Colloque T.O.M.*, INSA Lyon (France), March 1994.
- [42] P. Gonçalves, P. Flandrin, and E. Chassande-Mottin. Time-frequency methods in time series data analysis. In M Davier and P. Hello, editors, *Proceedings of Gravitational Wave Data Analysis Workshop 2*, pages 35–46. Éditions Frontières, 1997.
- [43] P. Gonçalves, C. Lenoir, C. Heymes, B. Swynghedauw, and C. Lavergne. Statistical modelling of cardiovascular data. an introduction to linear mixed models. Research Report 5787, INRIA Rhone-Alpes, Grenoble, France, Dec. 2005.
- [44] P. Gonçalves, J.-Ph. Ovarlez, and R. Baraniuk. Quadratic Time-Frequency Analysis III : The Affine Class and Other Covariant Classes. In F. Hlawatsch and F. Auger, editors, *Time-Frequency Analysis : Concepts and Methods*, pages 193–226. ISTE – John Wiley & Sons, Inc., 2008.
- [45] P. Gonçalves and E. Payot. Adaptive diffusion equation for time frequency representations. In *IEEE Digital Signal Processing Workshop*, Utah (US), August 1998.
- [46] P. Gonçalves and R. Riedi. Wavelet analysis of fractional brownian motion in multifractal time. In *Proceedings of the 17th Colloquium GRETSI*, Vannes (France), September 1999.
- [47] P. Gonçalves and R. Riedi. Diverging moments and parameter estimation. Research Report 4647, INRIA Rhone-Alpes, Grenoble, France, Nov. 2002.
- [48] P. Gonçalves and R. Riedi. Diverging moments and parameter estimation. *Journal of American Statistical Association*, 100(472) :1382–1393, December 2005. Also Inria Technical Report RR-4647, 2002.
- [49] P. Gonçalves, R. Riedi, and R. Baraniuk. A simple statistical analysis of wavelet-based multifractal spectrum estimation. In *Proceedings of the 32nd Conference on Signals, Systems and Computers*, Asilomar (US), November 1998.
- [50] J. Gosme, P. Gonçalves, C. Richard, and R. Lengellé. Adaptive diffusion and discriminant analysis for complexity control of time-frequency detectors. In *Proc. XI European Signal Processing Conference*, Toulouse (France), 2002.
- [51] J. Gosme, C. Richard, and P. Gonçalves. Anisotropic diffusion equations for adaptive quadratic representations. In *Proceedings of Eusipco*, Wien (Austria), 2004.
- [52] J. Gosme, C. Richard, and P. Gonçalves. Diffusion equations for adaptive affine distributions. In *Proceedings of Int. Conf. Acoustics Speech and Signal Processing*, Montréal (Canada), 2004.
- [53] J. Gosme, C. Richard, and P. Gonçalves. Adaptive diffusion as a versatile tool for time-frequency and time-scale representations processing : a review. *IEEE Trans. on Signal Processing*, 53(11) :4136–4146, November 2005.
- [54] O. Grémillet, P. Gonçalves, P. Vicat-Blanc Primet, and A. Dupas. Traffic classification techniques supporting semantic networks. In *1st Int. Wireless Comm. and Mobile Comp. Conf., TRaffic Analysis and Classification W.*, Caen (France), Jun-jul 2010.

- [55] D. Johnson, P. Gonçalves, and R. Baraniuk. Improved type-based detection of analog signals. In *IEEE Int. Conf. on Acoust. Speech and Sig. Proc.*, Munich (Germany), April 1997.
- [56] P. Loiseau, P. Goncalves, J. Barral, and P. Vicat-Blanc Primet. Modeling tcp throughput : an elaborated large-deviations-based model and its empirical validation. In *IFIP, Performance*, Namur (Belgium), Nov. 2010. Submitted.
- [57] P. Loiseau, P. Gonçalves, G. Dewaele, P. Borgnat, P. Abry, and P. Vicat-Blanc Primet. Investigating self-similarity and heavy-tailed distributions on a large scale experimental facility. Technical Report 6472, INRIA, March 2008.
- [58] P. Loiseau, P. Gonçalves, G. Dewaele, P. Borgnat, P. Abry, and P. Vicat-Blanc Primet. Vérification du lien entre auto-similarité et distributions à queues lourdes sur un dispositif grande échelle, June 2008. 9 ième Atelier en Evaluation de Performances, Aussois, France.
- [59] P. Loiseau, P. Gonçalves, G. Dewaele, P. Borgnat, P. Abry, and P. Vicat-Blanc Primet. Investigating self-similarity and heavy-tailed distributions on a large scale experimental facility. *IEEE/ACM Transactions on Networking*, May 2009. To appear.
- [60] P. Loiseau, P. Gonçalves, S. Girard, F. Forbes, and P. Vicat-Blanc Primet. Maximum likelihood estimation of the flow size distribution tail index from sampled packet data. In *ACM Sigmetrics*, June 2009.
- [61] P. Loiseau, P. Gonçalves, R. Guillier, M. Imbert, Y. Kodama, and P. Vicat-Blanc Primet. Metroflux : A high performance system for analyzing flow at very fine-grain. In *TridentCom*, April 2009.
- [62] P. Loiseau, P. Gonçalves, Y. Kodama, and P. Vicat-Blanc Primet. Metroflux : A fully operational high speed metrology platform, September 2008. Euro-NF workshop : New trends in modeling, quantitative methods and measurements, in cooperation with Net-coop, THOMSON Paris Research Labs, France.
- [63] P. Loiseau, P. Gonçalves, and P. Vicat-Blanc Primet. A comparative study of different heavy tail index estimators of the flow size from sampled data. In *MetroGrid Workshop, GridNets*, Lyon, France, October 2007. ACM Press.
- [64] P. Loiseau, P. Gonçalves, and P. Vicat-Blanc Primet. How TCP can kill self-similarity, September 2008. Euro-NF workshop : Traffic Engineering and Dependability in the Network of the Future, VTT, Finland.
- [65] P. Loiseau, P. Goncalves, and P. Vicat-Blanc Primet. Impact of the correlation between flow rates and durations on the large-scale properties of aggregate network traffic. *Annals of Telecommunications*, Special issue on Traffic and Performance Issues in Networks of the Future, 2010. Submitted.
- [66] P. Loiseau, R. Guillier, O. Goga, M. Imbert, P. Goncalves, and P. Vicat-Blanc Primet. Automated traffic measurements and analysis in grid'5000. Best Demonstration award at ACM SIGMETRICS/PERFORMANCE (Seattle, US), June 2009.
- [67] P. Oliveira, P. Gonçalves, and M. Caetano. Land cover time profiles from linear mixture models applied to MODIS images. In *Proceedings of the 31st International Symposium on Remote Sensing of Environment*, St. Petersburg (Russian Federation), June 2005.
- [68] J.-Ph. Ovarlez, P. Gonçalves, and R. Baraniuk. Analyse temps-fréquence quadratique III : La classe affine et autres classes covariantes. In F. Hlawatsch and F. Auger, editors, *Temps-Fréquence : concepts et outils*, pages 201–235. Hermès Sciences Publications, 2005.
- [69] M. Pasquier, P. Gonçalves, and R. Baraniuk. Hybrid linear/bilinear time-scale analysis. In *3rd IEEE Sig. process. Int. Symposium on Time-Frequency and Time-Scale Analysis*, Paris (France), June 1996.
- [70] M. Pasquier, P. Gonçalves, and R. Baraniuk. Hybrid linear/bilinear time-scale analysis. *IEEE Trans. on Signal Processing*, 47(1) :254–259, January 1999.
- [71] A. Pinheiro, P. Gonçalves, H. Carrão, and M. Caetano. A toolbox for multi-temporal analysis of satellite imagery. In *Proceedings of the 26th EARSel Symposium*, Warsaw (Poland), 2006.

- [72] R. Riedi and P. Gonçalves. Diverging moments, wavelets and non-parametric tail estimation. In *Proceedings of Joint Statistical Meeting*, Minneapolis (USA), 2005.
- [73] G. Rilling, P. Flandrin, and P. Gonçalves. On empirical mode decomposition and its algorithms. In *Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado (Italy), 2003.
- [74] G. Rilling, P. Flandrin, and P. Gonçalves. Empirical mode decomposition, fractional gaussian noise and hurst exponent estimation. In *IEEE Int. Conf. on Acoust. Speech and Sig. Proc.*, Philadelphia (US), March 19–23 2005.
- [75] G. Rilling, P. Flandrin, and P. Gonçalves. Une extension bivariée pour la décomposition modale empirique : Application à des bruits blancs complexes. In *Proceedings of the 21th Colloquium GRETSI*, Troyes (France), September 2007.
- [76] G. Rilling, P. Flandrin, P. Gonçalves, and J. M. Lilly. Bivariate empirical mode decomposition. *IEEE, Signal Processing Letters*, 14(12) :936–939, 2007.

Bibliographie Générale

- [Abry et al., 2000] Abry, P., Flandrin, P., Taqqu, M., and Veitch, D. (2000). Wavelets for the analysis, estimation, and synthesis of scaling data. In Park, K. and Willinger, W., editors, *Self-Similar Network Traffic and Performance Evaluation*, pages 39–88. Wiley.
- [Akselrod et al., 1981] Akselrod, S., Gordon, D., Ubel, F. A., Shannon, D. C., Barger, A. C., and Cohen, R. J. (1981). Power spectrum analysis of heart rate fluctuation : a quantitative probe of beat-to-beat cardiovascular control. *Science*, 2 :213–220.
- [Alvarez et al., 1993] Alvarez, L., Guichard, F., Lions, P. L., and Morel, J. M. (1993). Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, 123(3) :199–257.
- [Arnéodo et al., 1995] Arnéodo, A., Bacry, E., and Muzy, J.-F. (1995). The thermodynamics of fractals revisited with wavelets. *Physica A*, 213 :232–275.
- [Arnéodo et al., 1998] Arnéodo, A., Bacry, E., , and Muzy, J.-F. (1998). Random cascades on wavelet dyadic trees. *J. of Mathematical Physics*, 39(8) :4142–4164.
- [Arnéodo et al., 1997] Arnéodo, A., Bacry, E., Jaffard, S., and Muzy, J.-F. (1997). Oscillating singularities on Cantor sets : A grand canonical multifractal formalism. *J. of Statistical Physics*, 87 :179–209.
- [Averkamp and Houdré, 1998] Averkamp, R. and Houdré, C. (1998). Some distributional properties of the continuous wavelet transform of random processes. *IEEE Trans. on Info. Theory*, 44(3) :1111–1124.
- [Baccelli and Hong, 2002] Baccelli, F. and Hong, D. (2002). Aimd, fairness and fractal scaling of tcp traffic. In *IEEE INFOCOM*.
- [Bacry and Muzy, 2003] Bacry, E. and Muzy, J.-F. (2003). Log-infinitely divisible multifractal processes. *Comm. in Math. Phys.*, 236 :449–475.
- [Bacry et al., 1993] Bacry, E., Muzy, J.-F., and Arnéodo, A. (1993). Singularity spectrum of fractal signals : exact results. *Journal of Statistical Physics*, 70(3/4) :635–674.
- [Baraniuk et al., 2001] Baraniuk, R. G., Flandrin, P., Jansen, M., and Michel, O. (2001). Measuring time-frequency information content using the Rényi entropies. *IEEE Transactions on Information Theory*, 47(4) :1391–1409.
- [Barral and Lévy Véhel, 2004] Barral, J. and Lévy Véhel, J. (2004). Multifractal analysis of a class of additive processes with correlated non-stationary increments. *Elec. J. Probab.*, 9 :508–543.
- [Barral and Mandelbrot, 2002] Barral, J. and Mandelbrot, B. B. (2002). Multiplicative products of cylindrical pulses. *Probability Theory and Related Fields*, 124 :409–430.
- [Benassi et al., 2000] Benassi, A., Bertrand, P., Cohen, S., and Istas, J. (2000). Identification of the hurst exponent of a step multifractional brownian motion. *Statistical inference for Stochastic Processes*, 3 :101–110.
- [Benassi et al., 1998a] Benassi, A., Cohen, S., and Istas, J. (1998a). Identifying the multifractional function of a gaussian process. *Statistic and Probability Letters*, 39 :337–345.
- [Benassi et al., 1998b] Benassi, A., Cohen, S., Istas, J., and Jaffard, S. (1998b). Identification of filtered white noises. *Stoch. Proc. Appl.*, 75 :31–49.

- [Beran, 1994] Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.
- [Bertrand and Bertrand, 1991] Bertrand, J. and Bertrand, P. (1991). Some practical aspects of the affine time-frequency distributions. In *Proceedings of 13ème colloque GRETSI*, Juan-Les-Pins.
- [Bertrand and Bertrand, 1992] Bertrand, J. and Bertrand, P. (1992). A class of affine Wigner functions with extended covariance properties. *J. Math. Phys.*, 33(7).
- [Bianchi and Meerschaert, 2000] Bianchi, K. and Meerschaert, M. (2000). Scale and shift invariant estimators for the heavy tail index α . *AMS 2000 Subject Classification : 62F12, 60F05*.
- [Binmore and Stratton, 1969] Binmore, K. G. and Stratton, H. H. (1969). A note on characteristic functions. *Ann. Math. Statist.*, 40 :303–307.
- [Borgnat, 2002] Borgnat, P. (2002). *Modèles et outils pour les invariances d'échelle brisées : variations sur le transformation de Lamperti et contributions aux modèles statistiques de vortex en turbulence*. Thèse de doctorat 3ème cycle, École Normale Supérieure de Lyon.
- [Borgnat et al., 2009] Borgnat, P., Flandrin, P., Honein, P., Richard, C., and Xiao, J. (2009). Testing stationarity with surrogates : A time-frequency approach. *IEEE transactions on signal processing*. Submitted.
- [Brabaux, 2009] Brabaux, A. (2009). Construire l'internet de demain. *L'Usine Nouvelle*, (3167) :18–21.
- [Candès et al., 2006] Candès, E., Romberg, J., and Tao, T. (2006). Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. on Info. Theory*, 52(2) :489–509.
- [Castaing, 1996] Castaing, B. (1996). The temperature of turbulent flows. *J. Phys. II France*, 6 :105–114.
- [Chainais et al., 2000] Chainais, P., Abry, P., and Veitch, D. (2000). Multifractal analysis and α -stable processes : a methodological contribution. In *ICASSP*, Istanbul (Turkey).
- [Chainais et al., 2003] Chainais, P., Riedi, R., and Abry, P. (2003). Scale invariant infinitely divisible cascades. In *Int. Symp. on Physics in Signal and Image Processing, Grenoble, France*.
- [Chainais et al., 2004] Chainais, P., Riedi, R., and Abry, P. (in print 2004). On non scale invariant infinitely divisible cascades. In *Traitement du Signal*.
- [Chassande-Mottin and Flandrin, 1999] Chassande-Mottin, E. and Flandrin, P. (1999). On the time-frequency detection of chirps. *Appl. Comp. Harm. Anal.*, 6(9) :252–281.
- [Claasen and Mecklenbräuker, 1980] Claasen, T. A. C. M. and Mecklenbräuker, W. F. G. (1980). The wigner distribution – a tool for time-frequency signal analysis. part I : Continuous time signals. *Philips J. Res.*, 35(3) :217–250.
- [Coeurjolly, 2000] Coeurjolly, J.-F. (2000). *Inférence statistique pour les mouvements browniens fractionnaires et multifractionnaires*. Thèse de doctorat 3ème cycle, Université de Grenoble 1.
- [Cohen, 1995] Cohen, L. (1995). *Time-Frequency Analysis*. Prentice-Hall.
- [Crank, 1957] Crank, J. (1957). *The mathematics of diffusion*. Oxford University Press, London.
- [Crouse et al., 1998] Crouse, M. S., Nowak, R. D., and Baraniuk, R. G. (1998). Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing*, 46 :886–902.
- [Daoudi et al., 1998] Daoudi, K., Lévy Véhel, J., and Meyer, Y. (1998). Construction of continuous functions with prescribed local regularity. *Constructive Approximation*, 14(3) :349–386.
- [Daubechies, 1992] Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, New York.
- [Delbeke and Abry, 2000] Delbeke, L. and Abry, P. (2000). Stochastic integral representation and properties of the wavelet coefficients of linear fractional stable motion. *Stochastic processes and their applications*, 86(2) :177–182.
- [Erlang, 1909] Erlang, A. K. (1909). The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik*, pages 33–39.

- [Erlang, 1917] Erlang, A. K. (1917). Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Electroteknikeren*, pages 5–13.
- [Feldmann et al., 1999] Feldmann, A., Gilbert, A. C., and Willinger, W. (April 1999). Scaling analysis of conservative cascades, with applications to network traffic. *IEEE Trans. Info. Theory, Special issue on multiscale statistical signal analysis and its applications*, 45(3) :971–991.
- [Feldmann et al., 1998] Feldmann, A., Gilbert, A. C., Willinger, W., and Kurtz, T. (1998). The changing nature of network traffic : Scaling phenomena. *Computer Communication Review*, 28 :5–29.
- [Ferraro et al., 1999] Ferraro, M., Boccignone, G., and Caelli, T. (1999). On the representation of image structures via scale space entropy conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11) :1199–1203.
- [Feuerverger and McDunnough, 1981] Feuerverger, A. and McDunnough, P. (1981). On the efficiency of empirical characteristic function procedures. *J. R. Statist. Soc. B*, 43(1) :20–27.
- [Feuerverger and Mureika, 1977] Feuerverger, A. and Mureika, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.*, 5 :88–97.
- [Flandrin, 1992] Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Info. Theory*, IT-38(2) :910–917.
- [Flandrin, 1998] Flandrin, P. (1998). *Time-Frequency/Time-Scale Analysis*. Academic Press. Traduction anglaise de *Temps-Fréquence*, 2ème éd.
- [Flandrin and Borgnat, 2008] Flandrin, P. and Borgnat, P. (2008). Sparse time-frequency distributions of chirps from a compressed sensing perspective. In *8th Int. Conf. on Math. in Signal Processing*, pages 16–19, Cirencester (UK).
- [Förstner and Gülch, 1987] Förstner, W. and Gülch, E. (1987). A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. of the ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken Switzerland.
- [Frisch, 1995] Frisch, U. (1995). *Turbulence. The legacy of A. Kolmogorov*. Cambridge University Press, Cambridge, UK.
- [Frish and Parisi, 1985] Frish, U. and Parisi, G. (1985). Fully developed turbulence and intermittency. *Proc. Int. Summer on Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, pages 84–88.
- [Gilboa et al., 2002] Gilboa, G., Sochen, N., and Zeevi, Y. (2002). Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing*, 11(7) :689–703.
- [Grossmann and Morlet, 1984] Grossmann, A. and Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15(4) :723–736.
- [Hill, 1975] Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5) :1163–1173.
- [Hlawatsch and Auger, 2008] Hlawatsch, F. and Auger, F. (2008). *Time-Frequency Analysis : Concepts and Methods*. Digital signal and image processing series. ISTE – John Wiley & Sons, Inc., London (UK).
- [Hlawatsch et al., 1993] Hlawatsch, F., Papandreou, A., and Boudreaux-Bartels, G. F. (1993). Regularity and unitarity of affine and hyperbolic time-frequency representations. *IEEE Int. Conf. on Acoust., Speech and Signal Proc.*, pages 245–248. Minneapolis (MN), USA.
- [Hohn et al., 2003] Hohn, N., Veitch, D., and Abry, P. (2003). Cluster processes, a natural language for network traffic. *IEEE Trans. on Sig. Proc. – Special Issue on Sig. Proc. in Networking*, 8(51) :2229–2244.
- [Holschneider and Tchamitchian, 1990] Holschneider, M. and Tchamitchian, P. (1990). Régularité locale de la fonction “non-différentiable” de Riemann. In Lemarié, P., editor, *Les Ondelettes en 1989*. Springer-Verlag.

- [Honeiné and Richard, 2009] Honeiné, P. and Richard, C. (2009). Distribution temps-fréquence à paramétrisation radialement gaussienne optimisée pour la classification. *Traitement du Signal*, 26(9) :1–12.
- [Honeiné et al., 2005] Honeiné, P., Richard, C., and Flandrin, P. (2005). Reconnaissance des formes par méthodes à noyau dans le domaine temps-fréquence. In *Gretsi*, Louvain (Belgium).
- [Honeiné et al., 2007] Honeiné, P., Richard, C., and Flandrin, P. (2007). Time-frequency learning machines. *IEEE transactions on signal processing*, 55(7) :3930–3936.
- [Huang et al., 1998] Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., Yen, N., Tung, C., and Liu, H. (1998). The empirical mode decomposition and hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. London A*, 454 :903–995.
- [Iijima, 1971] Iijima, T. (1971). Basic equation of figure and observational transformation. *Systems, Computers, Controls*, 2(4) :70–77.
- [Jaffard, 1989] Jaffard, S. (1989). Exposants de Hölder en des points donnés et coefficients d’ondelettes. *C. R. Acad. Sc. Paris*, 308.
- [Jaffard, 1991] Jaffard, S. (1991). Pointwise smoothness, two-microlocalization and wavelet coefficients. *Publicacions Matemàtiques*, 35.
- [Jaffard, 2000] Jaffard, S. (2000). On lacunary wavelet series. *Annals of Applied Probability*, 10(1) :313–329.
- [Jaffard, 2004a] Jaffard, S. (2004a). Beyond Besov spaces part 1 : Distribution of wavelet coefficients. *J. Four. Anal. Appl.*, 10(3) :221–246.
- [Jaffard, 2004b] Jaffard, S. (2004b). Wavelet techniques in multifractal analysis. *Fractal Geometry and Applications : A Jubilee of Benoît Mandelbrot, Proc. of Symposia in Pure Mathematics*, 72(2) :91–152.
- [Jaffard, 2005] Jaffard, S. (2005). Beyond Besov spaces part 2 : Oscillation spaces. *J. Constr. Approx.*, 21(1) :29–61.
- [Jaffard et al., 2006] Jaffard, S., Lashermes, B., and Abry, P. (2006). Wavelet leaders in multifractal analysis. In T Qian, M. I. and Vai, X. Y., editors, *Wavelet Analysis and Applications*, pages 219–264. Birkhäuser Verlag.
- [Jones and Baraniuk, 1991] Jones, D. L. and Baraniuk, R. G. (1991). Efficient approximation of continuous wavelet transforms. *IEEE Electronic Letters*, 27(9) :984–750.
- [Jones and Baraniuk, 1995] Jones, D. L. and Baraniuk, R. G. (1995). An adaptive optimal-kernel time-frequency representation. *IEEE Transactions on Signal Processing*, 43(10) :2361–2371.
- [Kahane, 1987] Kahane, J.-P. (1987). Positives martingales and random measures. *Chinese Ann. Math.*, 8B :1–12.
- [Kahane and Peyrière, 1976] Kahane, J.-P. and Peyrière, J. (1976). Sur certaines martingales de Benoît Mandelbrot. *Adv. Math.*, 22 :131–145.
- [Kawata, 1972] Kawata, T. (1972). *Fourier Analysis in Probability Theory*. Probability and Mathematical Statistics. Academic Press.
- [Koenderink, 1984] Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, 50 :363–370.
- [Koutrouvelis, 1980] Koutrouvelis, I. A. (1980). Regression-type estimation of the parameter of stable law. *J. American Statistical Association*, 75 :918–928.
- [Leland et al., 1994] Leland, W. E., Taqqu, M. S., Willinger, W., and Wilson, D. V. (1994). On the self-similar nature of Ethernet traffic (Extended version). *IEEE/ACM Trans. on Networking*, 2 :1–15.
- [Lévy Véhel and Roueff, 1998] Lévy Véhel, J. L. and Roueff, F. (1998). A regularization approach to fractional dimension estimation. In *proceedings of Fractals 98*, Malta.
- [Loiseau, 2009] Loiseau, P. (2009). *Contributions to the Analysis of Scaling Laws and Quality of Service in Networks : Experimental and Theoretical Aspects*. Thèse de doctorat 3ème cycle, Ecole Normale Supérieure de Lyon.

- [Lukacs, 1983] Lukacs, L. (1983). *Developments in Characteristic Function Theory*. MacMillan Co., New York.
- [Mallat, 1989] Mallat, S. (1989). A theory for multiresolution signal decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11 :674–693.
- [Mallat, 1998] Mallat, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press. Second edition.
- [Mallat and Hwang, 1992] Mallat, S. G. and Hwang, W. L. (1992). Singularity detection and processing with wavelets. *IEEE Trans. on Info. theory*, 38(2) :617–643.
- [Mandelbrot, 1974] Mandelbrot, B. B. (1974). Intermittent turbulence in self similar cascades : Divergence of high moments and dimension of the carrier. *J. Fluid Mech.*, 62(331).
- [Mandelbrot, 1997] Mandelbrot, B. B. (1997). *Fractals and Scaling in Finance*. Springer, New-York.
- [Mandelbrot and Van Ness, 1968] Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motion, fractional noises and applications. *SIAM Reviews*, 10 :422–437.
- [Mannensalo et al., 2002] Mannensalo, P., Norros, I., and Riedi, R. (2002). Multifractal products of stochastic processes : construction and some basic properties. *Applied Probability Trust*, 34(4) :1–16.
- [Maulik and Resnick, 2003] Maulik, K. and Resnick, S. (2003). The self-similar and multifractal nature of a network traffic model. *Stochastic Models*, 19(4) :549–577.
- [McCulloch, 1986] McCulloch, J. H. (1986). Simple consistent estimator of stable distribution parameter. *Communication on Statistical Simulation*, 15 :1109–1136.
- [McCulloch, 1997] McCulloch, J. H. (1997). Measuring tail thickness to estimate the stable index alpha : A critique. *J. American Statistical Association*, 15 :74–81.
- [Molla and Torr  sani, 2005] Molla, S. and Torr  sani, B. (2005). An hybrid audio scheme using hidden markov models of waveforms. *Applied and Computational Harmonic Analysis*, 18(2) :137–166.
- [Moulines et al., 2007a] Moulines, E., Roueff, F., and Taqqu, M. S. (2007a). Central limit theorem for the log-regression wavelet estimation of the memory parameter in the gaussian semi-parametric context. *Fractals*, 15(4) :301–313. 0218-348X.
- [Moulines et al., 2007b] Moulines, E., Roueff, F., and Taqqu, M. S. (2007b). 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. of Time Series Analysis*, 28(2) :155–187. 0143-9782.
- [Moulines et al., 2008] Moulines, E., Roueff, F., and Taqqu, M. S. (2008). A wavelet whittle estimator of the memory parameter of a nonstationary gaussian time series. *Ann. Statist.*, 36(4) :1925–1956.
- [Muzy et al., 1993] Muzy, J.-F., Bacry, E., and Arn  odo, A. (1993). Multifractal formalism for fractal signals : The structure-function approach versus the wavelet-transform modulus-maxima method. *Phys. Rev. E*, 47(2) :875–884.
- [Nolan, 2001] Nolan, J. P. (2001). Maximum likelihood estimation and diagnostics for stable distributions. In Barndorff-Nielsen, O., Mikosch, T., and Resnick, S., editors, *L  vy Processes*. Birkh  user, Boston.
- [Olver et al., 1999] Olver, P. J., Sapiro, G., and Tannenbaum, A. (1999). Affine invariant detection : edge maps, anisotropic diffusion, and active contours. *Acta Applicandae Mathematicae*, 59 :45–77.
- [Ovarlez et al., 1992] Ovarlez, J.-P., Bertrand, J., and Bertrand, P. (1992). Computation of affine time frequency representations using the fast Mellin transform. *IEEE Int. Conf. on Acoust. Speech and Signal Proc.* San-Francisco (CA), USA.
- [Papoulis, 1991] Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, 3 edition.

- [Park and Willinger, 2000] Park, K. and Willinger, W., e. (2000). *Self-Similar Network Traffic and Performance Evaluation*. Wiley (Interscience Division).
- [Paxson and Floyd, 1994] Paxson, V. and Floyd, S. (1994). Wide area traffic : The failure of poisson modeling. *SIGCOMM*, pages 257–268.
- [Peltier and Levy-Vehel, 1995] Peltier, R. and Levy-Vehel, J. (1995). Multifractional brownian motion : Definition and preliminary results. Technical Report RR-2645, INRIA.
- [Percival and Walden, 2000] Percival, D. and Walden, A. (2000). *Wavelet methods for time series analysis*. Cambridge series in statistical and probabilistic mathematics. Cambridge Univ. Press.
- [Perona and Malik, 1990] Perona, P. and Malik, J. (1990). Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7) :629–639.
- [Pesquet-Popescu, 1999] Pesquet-Popescu, B. (1999). Statistical properties of the wavelet decomposition of certain non-gaussian self-similar processes. *IEEE Trans. on Sig. Proc.*, 75 :303–322.
- [Riedi, 2002] Riedi, R. (2002). Multifractal processes. In Doukhan, O. and Taqqu, editors, *Long range dependence : theory and applications*. Birkhauser.
- [Riedi et al., 1999] Riedi, R., Crouse, M. S., Ribeiro, V., and Baraniuk, R. G. (April 1999). A multifractal wavelet model with application to TCP network traffic. *IEEE Trans. Info. Theory, Special issue on multiscale statistical signal analysis and its applications*, 45 :992–1018.
- [Riedi and Lévy Véhel, 1997] Riedi, R. and Lévy Véhel, J. (1997). Multifractal properties of TCP traffic : A numerical study. *Technical Report No 3129, INRIA Rocquencourt, France*.
- [Rilling, 2007] Rilling, G. (2007). *Décompositions Modales Empiriques*. Thèse de doctorat 3ème cycle, École Normale Supérieure de Lyon.
- [Rioul and Flandrin, 1992] Rioul, O. and Flandrin, P. (1992). Time-scale energy distributions : A general class extending wavelet transform. *IEEE Trans. on Signal Proc.*, SP-40(7) :1746–1757.
- [Roueff and Von Sachs, 2009] Roueff, F. and Von Sachs, R. (2009). Locally stationary long memory estimation. Preprint.
- [Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60 :259–268.
- [Samorodnitsky and Taqqu, 1994] Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes, Stochastic Models with Infinite Variance*. Chapman & Hall, New-York, London.
- [Sarvotham et al., 2001] Sarvotham, S., Riedi, R., and Baraniuk, R. (2001). Connection-level analysis and modeling of network traffic. In *IMW 01 : Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement*, pages 99–103, New York, NY, USA. ACM Press.
- [Sarvotham et al., 2005] Sarvotham, S., Riedi, R., and Baraniuk, R. (2005). Network and user driven alpha-beta on-off source model for network traffic. *Computer Networks*, 48(3) :335–350.
- [Sellan, 1995] Sellan, F. (1995). Synthèse de mouvements browniens fractionnaires à l’aide de la transformation par ondelettes. *Compte Rendu de l’Académie des Sciences, Série I*, 321 :351–358.
- [Seuret and Barral, 2005a] Seuret, S. and Barral, J. (2005a). From multifractal measures to multifractal wavelet series. *J. of Fourier Anal. and App.*, 11(5) :589–614.
- [Seuret and Barral, 2005b] Seuret, S. and Barral, J. (2005b). Multifractal wavelet series built using multifractal measures. *Compte Rendu de l’Académie des Sciences, Série I*, 341 :353–356.
- [Sirfling, 1980] Sirfling, R. (1980). *Approximation theorems of Mathematical Statistics*. Wiley.
- [Smolka, 2003] Smolka, B. (2003). Combined forward and backward anisotropic diffusion filtering of color images. In *Proc. of the 24th DAGM Symposium on Pattern Recognition*, pages 314–321.
- [Sporring and Weickert, 1999] Sparring, J. and Weickert, J. (1999). Information measures in scale-spaces. *IEEE Transactions on Information Theory*, 45(3) :1051–1058.
- [Taqqu et al., 1997] Taqqu, M. S., Willinger, W., and Sherman, R. (1997). Proof of a fundamental result in self-similar traffic modeling. *SIGCOMM CCR*, 27(2) :5–23.

- [Teich et al., 2001] Teich, M., Lowen, S., Jost, B., Vibe-Rheymer, K., and Heneghan, C. (2001). Heart rate variability : Measures end models. *Nonlinear Biomedical Signal Processing*, II, Dynamic Analysis and Modeling :159–213.
- [Tewfik and Kim, 1992] Tewfik, A. H. and Kim, M. (1992). Correlation structure of the discrete wavelet coefficients of fractional Brownian motions. *IEEE Trans. on Info. Theory*, IT-38(2) :904–909.
- [Thorne, 1987] Thorne, K. S. (1987). Gravitational radiation. In Hawking, A. W. and Israel, W., editors, *300 Years of Gravitation*, pages 330–458. Cambridge Univ. Press.
- [Unterberger, 1984] Unterberger, A. (1984). The calculus of pseudo-differential operators of Fuchs type. *Comm. in Part. Diff. Eq.*, 9 :1179–1236.
- [Veitch and Abry, 1999] Veitch, D. and Abry, P. (1999). A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory special issue on "Multiscale Statistical Signal Analysis and its Applications"*, 45(3) :878–897.
- [Veitch and Abry, 2001] Veitch, D. and Abry, P. (2001). A statistical test for the time constancy of scaling exponents. *IEEE Trans. on Signal Processing*, 49(10) :2325–2334.
- [Voci et al., 2004] Voci, F., Eiho, S., Sugimoto, N., and Sekibuchi, H. (2004). Estimating the gradient in the Perona-Malik equation. *IEEE Signal Processing Magazine*, 21(3) :39–65.
- [Weickert, 1997] Weickert, J. (1997). A review of nonlinear diffusion filtering. In *Proc. of the 1st International Conference on Scale-Space Theories in Computer Vision*, volume 1252, pages 3–28.
- [Weickert, 1998] Weickert, J. (1998). *Anisotropic diffusion in image processing*. B. G. Teubner, Stuttgart.
- [Weickert et al., 1999] Weickert, J., Ishikawa, S., and Imiya, A. (1999). Linear scale-space has first been proposed in japan. *Journal of Mathematical Imaging and Vision*, 10 :237–252.
- [Weickert and Scharr, 2002] Weickert, J. and Scharr, H. (2002). A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. *Journal of Visual Communication and Image Representation*, 13 :103–118.
- [Wendt, 2008] Wendt, H. (2008). *Contributions of wavelet leaders and bootstrap to multifractal analysis : images, estimation performance, dependence structure and vanishing moments*. Thèse de doctorat 3ème cycle, Ecole Normale Supérieure de Lyon.
- [Wendt et al., 2007] Wendt, H., Abry, P., and Jaffard, S. (2007). Bootstrap for empirical multifractal analysis. *IEEE Signal Proc. Magazine*, 24(4) :38–48.
- [Wendt et al., 2009] Wendt, H., Abry, P., Roux, S., Jaffard, S., and Vedel, B. (2009). Analyse multifractale d'image : l'apport des coefficients dominants. *Traitement du Signal*, 1.
- [Willinger et al., 1997] Willinger, W., Taqqu, M. S., Sherman, R., and Wilson, D. V. (1997). Self-similarity through high-variability : statistical analysis of ethernet lan traffic at the source level. *IEEE/ACM Trans. on Networking*, 5(1) :71–86.
- [Wood and Chan, 1994] Wood, A. T. and Chan, G. (1994). Simulation of stationary processes in $[0, 1]^d$. *J. Comp. Graph. Stat.*, 3 :409–432.
- [Wornell and Oppenheim, 1992] Wornell, G. and Oppenheim, A. V. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Processing*, 40 :611–623.
- [Wornell, 1996] Wornell, G. W. (1996). *Signal Processing with Fractals — A Wavelet-Based Approach*. Prentice-Hall.
- [Wu et al., 2001] Wu, Z., Schneider, E., Hu, Z., and Cao, L. (2001). The impact of global warming on enso variability in climate records. Technical Report CTR 110, COLA.