



**SAMOS-MATISSE**

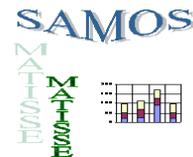
(Statistique Appliquée et MODélisation Stochastique)

**Centre d'Economie de la Sorbonne**

**Université Paris 1 - Panthéon-Sorbonne, CNRS**

90, rue de Tolbiac  
75634-PARIS CEDEX 13

Tél. : +33 (0) 1 44 07 89 22 ou 89 35  
Fax : +33 (0) 1 44 07 89 22



**Etablissement : Université Paris 1**  
**Responsable : Marie Cottrell**

**CES**  
**CENTRE D'ECONOMIE DE LA SORBONNE**

**Unité Mixte de Recherche CNRS 8174**  
**Université Paris 1 Panthéon-Sorbonne,**

*<http://samos.univ-paris1.fr>*  
90, rue de Tolbiac  
75013 PARIS

## **RAPPORT SCIENTIFIQUE 2005–2008**

### **SOMMAIRE**

1. BILAN EN 2008 DES OPÉRATIONS PRÉVUES DANS LE CONTRAT EN 2005.....	3
<i>AXE A. Réseaux neuronaux et méthodes connexionnistes, analyse de données complexes.....</i>	<i>3</i>
<i>AXE B. Statistiques, processus et méthodes aléatoires.....</i>	<i>17</i>
2 PRINCIPAUX RÉSEAUX FRANÇAIS, EUROPÉENS ET INTERNATIONAUX DEPUIS 2005.....	42
3 LES ACTIVITÉS D’ANIMATION SCIENTIFIQUE 2005-2008.....	43
3.1 <i>Les colloques et conférences internationales organisées par des membres du SAMOS.....</i>	<i>43</i>
3.2 <i>Appartenance à des comités scientifiques de conférences.....</i>	<i>45</i>
3.3 <i>Le séminaire SAMOS 2005-2008.....</i>	<i>46</i>
3.4 <i>Le séminaire Mathématiques des systèmes complexes (novembre 2006 – mai 2008).....</i>	<i>49</i>
4 LES CONVENTIONS DE RECHERCHE ET D’ÉTUDE 2005-2008.....	51
5 LA COMMUNICATION : LE SITE WEB DU SAMOS .....	52

## 1. BILAN EN 2008 DES OPÉRATIONS PRÉVUES DANS LE CONTRAT EN 2005.

Les projets 2005-2008 du précédent rapport d'activités 2001-2004 étaient structurés en deux axes:

1. **A (Méthodes connexionnistes et analyse de données complexes)**
2. **B (Statistiques, processus et méthodes aléatoires).**

Avec l'arrivée de nouvelles compétences au sein du SAMOS, l'Axe B a été découpé en deux axes en 2 axes,

**B : Statistique des processus et Statistique Spatiale**

**C : Probabilités et Processus Stochastiques**

### AXE A :

RÉSEAUX NEURONAUX ET MÉTHODES CONNEXIONNISTES,  
ANALYSE DE DONNÉES COMPLEXES

#### Coordination : Marie Cottrell

Nous avons poursuivi l'étude mathématique des algorithmes connexionnistes, tout en développant les applications de ces méthodes en analyse de données, en économétrie, en traitement de données qualitatives, en statistique. Nous avons commencé à travailler sur des systèmes complexes en essayant de proposer des solutions au problème de la reconstruction de trajectoires.

Nous avons en même temps continué à publier des synthèses d'une partie de nos travaux, dans un article paru dans une revue internationale :

2005 COTTRELL M., LETRÉMY P., How to use the Kohonen algorithm to simultaneously analyze individuals and modalities in a survey, *Neurocomputing*, 63, p. 193-207

et dans un chapitre d'un ouvrage complet sur les méthodes connexionnistes :

2006 COTTRELL M., IBBOU S., LETRÉMY P., ROUSSET P., Cartes auto-organisatrices de Kohonen, in *Apprentissage connexionniste*, Younès Bennani Ed., Hermès, Paris, p. 141-184

et édité deux numéros spéciaux de revues :

2005 COTTRELL M., HAMMER B., VILMANN T., Edition de "New aspects in Neurocomputing: 11th European Symposium on Artificial Neural Networks 2003, Special Issue", *Neurocomputing* Vol. 63

2006 COTTRELL M., VERLEYSSEN M., Edition de "Advances in Self-Organising Maps, WSOM 05 Special Issue", *Neural Networks* Vol. 19, N° 6-7

Nous avons également travaillé dans le cadre de plusieurs contrats industriels avec GDF, EDF, la STSI/CTSI, et commencé deux partenariats de longue durée avec l'entreprise

NOVACYT (partenariat commencé en décembre 2007) et la SNECMA (deux contrats signés en juin 2008).

Dans le dossier correspondant à la contractualisation 2006-2009, les projets de recherche de l'axe A avaient été répartis en 12 opérations regroupées en 4 sous-ensembles. Nous pouvons les passer en revue pour faire le point sur le travail réalisé et sur ce qui reste à faire (ce qui sera détaillé dans la partie « projet »). Les paragraphes en italique reprennent en résumé la rédaction du projet écrit en 2005.

## **I) Projets plutôt théoriques**

### **I-1) - Etude théorique de l'algorithme de Kohonen, vu comme un algorithme de quantification, comparaison avec les autres méthodes, accélération, algorithme batch, étude de l'algorithme à 0 voisin, étude de la distorsion étendue (Marie Cottrell, SAMOS, Jean-Claude Fort, SAMOS, Patrick Letrémy, SAMOS, Joseph Rynkiewicz, SAMOS, Michel Verleysen, ML, Louvain-la-Neuve)**

*[Rappel : Il s'agit de continuer l'étude mathématique de l'algorithme de Kohonen considéré comme un algorithme stochastique particulier qui a pour caractéristique de ne pas pouvoir s'identifier à un algorithme du gradient dans le cas général]*

**Bilan :** Sur l'algorithme de Kohonen proprement dit, nous n'avons pas obtenu de résultats vraiment nouveaux, mais nous avons continué à mettre en évidence les propriétés mathématiques « négatives » qui rendent l'étude de l'algorithme originel de Kohonen ardu. Par exemple, cet algorithme organise « en pratique », mais il est possible de trouver un chemin de probabilité strictement positive qui « désorganise » lorsque la dimension est supérieure à 1. On ne connaît pas la forme de la décroissance du paramètre d'adaptation qui assure la convergence vers une situation organisée, même en dimension 1, etc.

Le point sur ces aspects théoriques a été réalisé dans un article

2006 FORT J.C., SOM's mathematics, *Neural Networks*, 19, n°. 6-7, p. 812-816

Ce travail a également donné lieu à plusieurs communications dans des conférences internationales (WSOM 05, à Paris en 2005, OR7 à La Havane en 2006).

Un aspect a été approfondi par Joseph Rynkiewicz qui a étudié la distorsion étendue. Comme on le sait, l'algorithme de Kohonen a souvent été vu comme une minimisation approximative du critère de variance intra étendue aux voisins (distorsion étendue), au moins pour le cas discret (nombre fini d'observations). Cependant, on sait que ce n'est plus vrai pour le cas asymptotique (nombre infini d'observations). Il donc étudié les propriétés asymptotiques de la distorsion étendue et montré notamment que l'estimateur du minimum de distorsion empirique converge vers le minimum de distorsion étendue théorique quand le nombre d'observations tend vers l'infini.

Il a aussi donné aussi un contre exemple qui montre que l'algorithme de Kohonen n'approxime pas le minimum de distorsion étendue, puisque ce minimum est atteint sur les points de discontinuité de cette fonction. Cela explique l'apparente contradiction entre le comportement empirique et théorique de l'algorithme de Kohonen vis à vis de cette distorsion.

2005 RYNKIEWICZ J., Consistance d'un estimateur de minimum de variance étendue, *Comptes Rendues de l'Académie des Sciences – Series 1 - Mathematics*, 341, p. 133-136

2006 RYNKIEWICZ J., Self Organizing Map algorithm and distortion measure, *Neural Networks*, 19, n°. 6-7, p. 671-678

La rédaction par Marie Cottrell et Jean-Claude Fort de l'ouvrage prévu sur l'algorithme de Kohonen n'a pu être achevée, elle fera l'objet d'une priorité dans le contrat quadriennal à venir.

Un travail alternatif a consisté à étudier l'algorithme « Neural Gas » qui peut être vu comme un algorithme du gradient, et à en proposer une version batch. Ce travail a donné lieu à une collaboration avec Barbara Hammer (Clausthal University, Allemagne), en particulier à l'occasion de son séjour dans notre laboratoire.

2006 COTTRELL M., HAMMER B., HASENFUB A., VILLMANN T., Batch and median neural gas, *Neural Networks*, 19, n°. 6-7, p. 762-771.(ACL13)

C'est dans le cadre de ce point de notre projet scientifique du contrat précédent que nous avons accepté d'organiser la conférence WSOM 05, qui s'est tenue à Paris 1 (Sorbonne et Carré des Sciences) en septembre 2005. Cette conférence a été un grand succès en accueillant sensiblement plus de participants (140 inscrits) que les éditions précédentes. Les participants se sont félicités de la qualité scientifique et humaine de cette manifestation, bien organisée grâce à l'aide de Paris 1 (BQR) et du ministère, et aussi au dévouement de tous les membres de l'équipe SAMOS.

Les meilleurs papiers ont été (après un second processus de review peer-to-peer) publiés dans un numéro spécial de la revue *Neural Networks* (ISI Impact Factor en 2008 = 2).

**I-2) - Perceptrons multi-couches combinés avec des chaînes de Markov cachées, identification du nombre de régimes, sélection de modèles et d'architectures, application à des données financières (Catherine Aaron, SAMOS, Madalina Olteanu, SAMOS, Joseph Rynkiewicz, SAMOS)**

*[Rappel : il s'agit de déterminer le nombre de régimes à l'aide d'une classification préalable. Un grand nombre de séries économiques ou financières peuvent être modélisées à l'aide des modèles auto-régressifs à changement de régimes markovien. Afin de spécifier ces modèles, on doit déterminer le nombre de régimes, mais les tests du nombre de régimes ont une distribution non standard et ne sont donc pas directement utilisables (problème de paramètres de nuisance sous l'hypothèse nulle d'égalité des coefficients de deux régimes). Un préalable aux tests pourrait être de déterminer a priori, sur la base d'une approche empirique, le nombre de régimes existant. Il s'agirait d'adapter les algorithmes de classification existants (algorithme de Kohonen, classification de type K-means) aux spécificités de notre problème.]*

L'essentiel ici est que l'ensemble de ce thème a donné lieu à la thèse de Madalina Olteanu, qui a ensuite été recrutée comme MCF au SAMOS. Ce travail a fait l'objet de nombreuses publications et on peut distinguer plusieurs aspects.

- A) *Caractérisation des crises financières à l'aide de modèles à changements de régime*

Dans un premier temps, nous nous sommes intéressés à la caractérisation des crises financières à l'aide de modèles à changements de régime intégrant des perceptrons multicouches. Nous avons utilisé un indice de crise introduit par Maillet et Michel (2002). Calculé à partir de la volatilité historique, l'indice IMS est construit par analogie avec l'échelle de Richter en géophysique comme une transformation logarithmique de l'agrégation de plusieurs volatilités observées pour des fréquences différentes.

La modélisation par des modèles à changements de régime n'a pas permis d'améliorer les erreurs de prévision (par rapport à un modèle linéaire, par exemple) ce qui est conforme à l'hypothèse d'efficacité faible des marchés. En revanche, on a obtenu une séparation intéressante entre deux états relatifs à deux comportements différents de l'indice et du marché. Un premier régime correspond aux périodes de crise et de fortes turbulences, tandis que le deuxième s'adapte aux périodes moins agitées. De plus, si on représente graphiquement l'évolution des probabilités conditionnelles des régimes, on est capable d'extraire précisément les dates correspondant aux décisions ou aux événements générant de larges fluctuations et plus généralement d'évaluer la durée totale d'une crise.

2004 MAILLET B., OLTEANU M., RYNKIEWICZ J., Caractérisation des crises financières à l'aide de modèles hybrides (HMC-MLP), *Revue d'économie politique*, 114, n° 4, p. 489-506

2004 MAILLET B., OLTEANU M., RYNKIEWICZ J., Nonlinear Analysis of Shocks when Financial Markets are Subject to Changes in Regime, *Proceedings of ESANN 2004 (European Symposium on Artificial Neural Networks)*, p.87-92

2006 OLTEANU M., Modèles à changements de régime, applications aux données financières, chapitre 3, *Thèse de doctorat*, <http://tel.archives-ouvertes.fr/tel-00133132/fr/>

Le travail appliqué sur les données financières nous a mené vers des questions plus théoriques, notamment celle de la sélection du nombre de régimes. La question de la sélection du nombre de régimes soulève des problèmes de non-identifiabilité du modèle. Ceci rend la matrice de Fisher non inversible et ne permet donc pas d'appliquer la théorie usuelle sur la convergence du rapport de vraisemblance. Notons que de plus ce rapport devient divergent dans certains cas comme celui des chaînes de Markov cachées (Keribin et Gassiat, 2000).

Plus particulièrement, nous nous intéressons aux modèles autorégressifs à changements de régime (indépendants ou markoviens). Ces modèles sont largement utilisées dans les séries temporelles et permettent de modéliser des observations non linéaires.

- B) *Sélection du nombre de régimes avec un algorithme de classification non supervisée*

Afin de contourner le problème de non identifiabilité, nous avons proposé une première approche empirique. Il s'agit d'une approche issue des méthodes de classification non supervisée et qui peut être utilisée comme traitement préliminaire des données. L'idée est de transformer le problème d'estimation des paramètres dans un modèle autorégressif en problème de classification. Dans ce cadre, on ne fait plus d'hypothèse sur les changements de régime qui peuvent être indépendants ou markoviens, on impose seulement que les fonctions de régression dans chaque régime soient linéaires. Avec cette condition et en appliquant une fenêtre mobile de taille  $p+1$  à la série des observations, trouver le nombre de régimes est équivalent à trouver le nombre d'hyperplans de régression qui s'ajustent le mieux aux données.

Nous avons proposé un algorithme de classification mixte qui combine les cartes de Kohonen pour faire une classification initiale des données, suivie d'une classification hiérarchique pour regrouper ensuite les clusters obtenus. Pour la classification hiérarchique, nous avons introduit une nouvelle mesure de dispersion : on regroupe ainsi les classes qui minimisent le gain en dispersion intra classes, en définissant cette dispersion comme le total des sommes des carrés résiduels issus des régressions linéaires ajustées dans chaque classe.

2006 OLTEANU M., A descriptive method to evaluate the number of regimes in a switching autoregressive model, *Neural Networks*, 19, n°. 6-7, p. 963-972

2006 OLTEANU M., Modèles à changements de régime, applications aux données financières, chapitre 6, *Thèse de doctorat*, <http://tel.archives-ouvertes.fr/tel-00133132/fr/>

- C) *Sélection du nombre de régimes avec un critère de log-vraisemblance pénalisée*

Une deuxième approche proposée pour répondre à la question sur la sélection du nombre de régimes consiste à utiliser un critère de vraisemblance pénalisée. Nous avons démontré la consistance faible d'un tel critère pour un modèle autorégressif comportant des perceptrons multicouches et des changements de régime indépendants. Des simulations effectuées à l'aide d'un algorithme de type EM nous ont permis ensuite d'illustrer le résultat, sa vitesse de convergence et la stabilité de l'algorithme proposé.

2008 OLTEANU M., RYNKIEWICZ J., Estimating the number of components in a mixture of multilayer perceptrons, *Neurocomputing/ EEG Neurocomputing*, 71, n°. 7-9, p. 1321-1329

- D) *Etude théorique des perceptrons multicouches à sortie vectorielle*

Lorsque la variable à expliquer par un modèle de régression non linéaire est multidimensionnelle, les erreurs sont elles aussi multidimensionnelles. Dans ces conditions, minimiser l'erreur quadratique moyenne revient à minimiser la trace de la matrice de covariance empirique de l'erreur. Cependant si la vraie matrice de covariance de l'erreur n'est pas l'identité, cette méthode n'est pas optimale (sauf si le modèle de régression est linéaire sans contrainte). On a montré ici que minimiser le déterminant de la matrice de covariance empirique de l'erreur est asymptotiquement optimal même pour des erreurs non-gaussiennes. On généralise aussi ce résultat aux régressions linéaires avec contraintes.

2006 RYNKIEWICZ J., Efficient estimation of multidimensional regression model using multilayer perceptrons, *Neurocomputing/ EEG Neurocomputing*, 69, n°. 7-9, p. 671-678

2007 RYNKIEWICZ J., Estimation and test for multidimensional regression models, *Communications in Statistics – Theory and Methods*, 36, n°. 14, p. 2655-2671

- E) *Identification de l'architecture des perceptrons multicouches*

On considère des modèles de régression impliquant des perceptrons multicouches (MLP) avec une couche cachée et un bruit gaussien. L'estimation des paramètres du MLP peut être faite en maximisant la vraisemblance du modèle. Dans ce cadre, il est difficile de déterminer le vrai nombre d'unités cachées parce que la matrice d'information de Fisher n'est pas inversible si ce nombre est surestimé. Cependant, si les paramètres du MLP sont dans un ensemble compact, nous prouvons que la minimisation d'un critère d'information convenable permet l'estimation consistante du vrai nombre d'unités cachées.

2006 RYNKIEWICZ J., Estimation consistante de l'architecture des perceptrons multicouches, *Comptes Rendus de l'Académie des Sciences – Series 1 – Mathematics*, 342, n°. 9, p. 697-700

2007 DUTOT A.-L., RYNKIEWICZ J., STEINER F. E., RUDE J., A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions, *Environmental Modelling and Software*, 22, n°. 9, p. 1261-1269

**I-3) - Prévision, méthodes de double quantification, étude théorique et applications (Marie Cottrell, SAMOS, Jean-Claude Fort, SAMOS et Toulouse 3, Geoffroy Simon, Machine Learning, Louvain-la-Neuve, Michel Verleysen, Machine Learning, Louvain-la-Neuve)**

*[Rappel : Nous souhaitons consolider les résultats obtenus (théoriques et expérimentaux) pour l'algorithme de double quantification, en pratiquant une comparaison plus approfondie de cette méthode avec les méthodes plus classiques de type perceptron ou ARMA vectorielle, ou même avec les méthodes de simulation plus courantes. En effet, même si les résultats que nous avons obtenus sont bons, et même si la stabilité de la méthode a pu être prouvée dans le cas de dimension 2, il reste à en faire un paradigme incontestable, et à le replacer parmi l'ensemble des méthodes adaptées à la prévision à long terme en dimension quelconque.]*

Nous avons poursuivi l'étude de la double quantification et obtenus des résultats théoriques de stabilité en dimension quelconque et proposé plusieurs applications. En particulier la technique a été appliquée aux prévisions de séries temporelles. Ces travaux ont donné lieu à de nombreuses publications et abouti à la soutenance de thèse de Geoffroy Simon, dirigée par Michel Verleysen, en collaboration avec Marie Cottrell (SAMOS)

2004 SIMON G., LENDASSE A., COTTRELL M., FORT J.-C., VERLEYSSEN M., Double Quantization of the Regressor Space for Long-Term Time Series Prediction: Method and Proof of Stability, *Neural Networks*, 17, p. 1169-1181

2004 SIMON G., LEE J.A., COTTRELL M., VERLEYSSEN M., COTTRELL M., .Double Quantization Forecasting Method for Filling Missing Data in the CATS Time Series, *Proceedings of the conference IJCNN'2004*, International Joint Conference on Neural Networks, Budapest, Hungary, p. 1635-1640

2005 SIMON G., LENDASSE A., COTTRELL M., FORT J.-C., VERLEYSSEN M., Time series forecasting: Obtaining long term trends with self-organizing maps, *Pattern Recognition Letter, Elsevier*, Vol. 26, n°12, p. 1795-1808

2007 SIMON G., LEE J.A., COTTRELL M., VERLEYSSEN M., Forecasting the CATS benchmark with the Double Vector Quantization method, *Neurocomputing*. **70**, 13-15 (2007) 2400-2409

**I-4) - Connexité et analyse des données (Catherine Aaron, SAMOS, Jean-Claude Fort, SAMOS et Toulouse 3)**

*[Rappel : En partant du principe que la connexité représente, topologiquement, le pendant non linéaire de la convexité, il est tenté d'établir des méthodes d'analyses des données reposant sur cette notion. Les problèmes liés à l'hétérogénéité en terme de dispersion intra classes d'un tel algorithme nous poussent, dans les travaux à venir à considérer les algorithmes de classification reposant sur l'estimation de la densité  $f$  du nuage de point :*

*En conséquence l'estimation de densité (noyaux, ondelettes, algorithme EM..) fait aussi partie de nos centres d'intérêt. En travaillant sur le plus petit arbre connexe nous avons construit un algorithme de normalisation et de recherche d'indicateur central dont les propriétés sont actuellement étudiées. Cette partie se fait en relation avec la théorie des graphes. Enfin un travail sur l'analyse de la dimension et la réduction de dimension est lui aussi commencé.]*

Ce programme de travail a été réalisé et a conduit notamment à la soutenance de thèse de Catherine Aaron en 2005 (actuellement maître de conférences à l'Université Clermont II).

2005 AARON C., Connexité et analyse des données non linéaires, *Thèse de doctorat*, <http://tel.archives-ouvertes.fr/tel-00308495/fr/>

Le travail a porté sur trois points :

**- comment obtenir des classes connexes par une classification hiérarchique**

Pour cela, C. Aaron a défini et utilisé une nouvelle distance intra classes qui reflète la connexité des classes choisies. Ce travail a fait l'objet de l'article suivant :

2004 AARON C., Clustering under connectivity hypothesis, *Student*, Vol 5, n°1, p. 43-58.

**- classification par attraction des modes de la densité**

En partant du cadre de la classification en composantes connexes, C. Aaron a construit un algorithme adaptatif où on calcule simultanément un estimateur de la densité et une classification fondée sur les domaines d'attraction des différents modes trouvés. Les résultats en classification sont plutôt modestes, alors que l'algorithme adaptatif pour l'estimation de densité fonctionne très bien. Son étude a été poursuivie par une étude théorique démontrant sa convergence en toute dimension et son optimalité en dimension 1. La classification obtenue n'est vraiment bonne que lorsque les classes recherchées sont aussi convexes. Ces résultats ont été présentés dans une conférence avec actes et un article est en préparation.

2005 AARON C., Couplage d'un problème de classification et d'estimation de densité, *Actes du 37<sup>ème</sup> Congrès de la SFC Montréal*, p. 27-30

**- dimension intrinsèque et réduction de dimension**

On se place sur une composante connexe des données étudiées. On cherche à déterminer sa dimension, puis à la "déplier" dans un espace euclidien de la dimension "réelle" de la classe qui est généralement beaucoup plus petite que celle de l'espace où se trouvent initialement les données. La méthode utilise trois outils : une normalisation originale des données, la distance géodésique et un dépliage sur une carte de Kohonen (SOM).

Ces travaux ont été présentés au congrès WSOM05 et publiés dans l'article qui suit :

2005 AARON C., Graph-based normalization for non-linear data analysis (I), *Proceedings of the 5<sup>th</sup> Workshop on Self Organizing Maps*, Paris, France, p.645-652

2005 AARON C., Graph-based normalization for non-linear data analysis (II), *Proceedings of the 5<sup>th</sup> Workshop on Self Organizing Maps*, Paris, France, p.203-210

2006 AARON C., Graph-based normalization and whitening for non linear data analysis, *Neural Networks*, 19, n°. 6-7, p. 864-876

## **II) Application des méthodes définies à la prévision**

### **II-1) - Prévision, classification de courbes, cartes de Kohonen et analyse dynamique (Catherine Aaron, SAMOS, Corinne Perraudin, SAMOS, Joseph Rynkiewicz, SAMOS.**

[Rappel : Analyse des dynamiques individuelles. D'un point de vue descriptif nous avons modifié légèrement l'algorithme de base de construction de cartes de Kohonen pour l'adapter à l'observation et à la description des trajectoires individuelles, -soit de manière « brute » pour cause de faible effectif de données, dans le cadre d'un

*travail sur la convergence aux critères de Maastrich avec Corinne Perraudin et Joseph Rynkiewicz, -soit en préliminaire à la classification de courbes dans le cas de l'analyse des stratégies de gérants de fond.]*

## **Analyse des dynamiques individuelles**

Nous avons poursuivi les travaux visant à adapter l'algorithme de Kohonen dans le cas de données temporelles et spatiales, dans la lignée des travaux effectuant une classification de courbes permettant d'étudier la convergence des pays européens selon les critères de Maastricht :

2003 AARON C., PERRAUDIN C., RYNKIEWICZ J, Curves Based Kohonen Map and Adaptive Classification: an Application to the Convergence of the European Countries, *WSOM 2003 Proceedings*, Kitakyushu, Japan, 2003.

Afin d'étudier plus particulièrement les dynamiques et trajectoires individuelles, on peut également modifier sensiblement l'algorithme de Kohonen en définissant une double notion de voisinage, spatial et temporel. Pour chaque année, une carte SOM est construite et les différentes cartes sont juxtaposées les unes sous les autres. Afin d'obtenir une continuité non seulement dans la dimension spatiale, mais également dans la dimension temporelle, on contraint l'algorithme à respecter les deux dimensions, spatiale (horizontalement) et temporelle (verticalement). La mise en œuvre d'une classification ascendante hiérarchique sur l'ensemble de la carte contrainte permet ensuite d'étudier si les super-classes se forment dans le temps ou alors, si pour une date donnée, le nombre de super-classes par année se réduit, indiquant alors un processus de convergence. Cette méthodologie a été appliquée à l'évolution des performances européennes en matière d'emploi entre 1992 et 2003 aux 25 pays de l'Union Européenne, ainsi qu'à plusieurs pays candidats. Les résultats permettent d'illustrer l'évolution des performances européennes en matière d'emploi (processus de convergence et d'amélioration des performances globales) ainsi que les trajectoires suivies par les pays. Ces travaux ont été utilisés dans d'autres applications économiques, comme l'analyse dynamique de la qualité de l'emploi en Europe (Davoine et Erhel, 2007, Document de travail CEE, No. 86).

Après les premiers résultats présentés à la conférence WSOM 03 au Japon, la méthode a été développée dans la communication suivante :

2005 AARON C., PERRAUDIN C. ET RYNKIEWICZ J., Adaptation de l'algorithme SOM à l'analyse de données temporelles et spatiales : application à l'étude de l'évolution des performances en matière d'emploi, *ASMDA 2005 Proceedings*, Brest, France

## **II-2) - Modèles non linéaires de type perceptrons multi-couches, appliqués à des données financières (Yacine Jerbi, SAMOS et Université de Sfax.)**

*[Rappel : Le travail de thèse de Yacine Jerbi porte sur l'évaluation des options et gestion des risques financiers par les réseaux de neurones et des modèles à volatilité stochastique]*

Le programme prévu pour le travail de thèse a été réalisé, et la thèse soutenue en février 2006. La durée de la thèse a été exceptionnellement longue (de 1998 à 2006), mais cela s'explique par la très lourde charge d'enseignement de Yacine Jerbi en tant qu'assistant à Sfax. Il a été promu Maître-Assistant à Sfax à la suite de la soutenance.

La thèse consiste à comparer des modèles d'évaluation d'options européennes, aussi bien au niveau de l'évaluation (Black & Scholes, réseaux de neurones, modèles à volatilité stochastique), qu'au niveau de la gestion des risques (Black & Scholes et réseaux de

neurones), en se basant sur deux bases d'options européennes, sur l'indice CAC 40, cotées sur le MONEP: la première base est une base intraday s'étalant du mois de janvier 1998 au mois de juin 1998 et la seconde est journalière s'étalant du mois de janvier 1997 au mois de décembre 1999). Après traitement, ces bases sont découpées par contrats et par classes, selon la parité et la durée de vie résiduelle.

On prend en compte le modèle de Black & Scholes et on expose les différentes méthodes de calcul des volatilités implicite et historique, aussi bien dans le cas de données intraday, que pour les données journalières. Ensuite on étudie les modèles à volatilité stochastique, en considérant aussi bien les données intraday que les données journalières, aussi bien la volatilité implicite que la volatilité historique selon les trois processus classiques de la volatilité. Les résultats générés sont comparés à ceux générés par des simulations de Monte Carlo, appliquées aux mêmes données et au mêmes processus de la volatilité.

Ensuite à l'aide de réseaux de neurones, on évalue des options européennes, en se basant sur l'algorithme « cascade corrélation » et sur les mêmes données utilisées pour les modèles à volatilités stochastiques.

Enfin, la méthodologie de calcul de l'erreur de couverture moyenne absolue relative est exposée, dans le cas d'un portefeuille autofinancé, et en considérant quatre stratégies de couverture dynamiques. Ces calculs sont appliqués pour déterminer la matrice des risques et comparer les modèles de Black & Scholes et le modèle neuronal, en terme de couverture.

La comparaison des performances des différents modèles utilisés, aussi bien au niveau de l'évaluation qu'au niveau de la gestion des risques, fait l'objet de la conclusion générale.

2006 JERBI J., Evaluation des options et gestion des risques financiers par les réseaux de neurones et par les modèles à volatilité stochastique, *Thèse de doctorat*, <http://tel.archives-ouvertes.fr/tel-00308623/fr/>

### III) Analyse de données complexes

#### III-1) - Méthodes neuronales pour la visualisation des liens entre variables catégorielles, KACM, KDISJ, comparaison avec les méthodes développées par Samuel Kaski (Marie Cottrell, SAMOS, Patrick Letrémy, SAMOS, Elisabeth Roy, SAMOS)

*[Rappel : Nous allons continuer à approfondir le travail de classification, segmentation, visualisation d'observations principalement décrites par des variables qualitatives. En utilisant l'algorithme de Kohonen, on obtient en général un nombre de classes trop élevé, et se pose alors la question de leur regroupement en un nombre réduit de classes qu'on peut décrire. La méthode de regroupement proposés pour l'instant est une méthode de classification hiérarchique des vecteurs-codes des classes, avec pondération et utilisation de la distance de Mahalanobis. Cette méthode a fait ses preuves lorsque les variables sont numériques, mais n'est pas complètement adaptée au cas qualitatif. Nous souhaitons travailler à l'élaboration d'une méthode plus performante pour résoudre ce problème.]*

En fait, nos essais successifs pour trouver une méthode plus performante pour regrouper les classes en super-classes n'ont pas permis de mettre en évidence un paradigme incontestable. Cependant si on a pour but (principal ou accessoire) de définir une graduation ou une échelle entre les classes, nous avons montré que la classification des vecteurs codes par une seconde classification de Kohonen sur une topologie de dimension 1 donne d'excellents résultats.

Nous avons rédigé deux publications qui font la synthèse de ces méthodes basées sur les cartes de Kohonen (algorithme SOM) et qui permettent de classer et de visualiser simultanément les individus et leurs caractéristiques qu'elles soient quantitatives ou qualitatives. La première est un article dans la revue *Neurocomputing* (ISI Impact Factor I 2007: 0.865)

2005 COTTRELL M., LETRÉMY P., How to use the Kohonen algorithm to simultaneously analyze individuals and modalities in a survey, *Neurocomputing*,. 63, p. 193-207. (ACL14)

et la seconde consiste en un chapitre d'un ouvrage complet sur les méthodes connexionnistes :

2007 COTTRELL M., IBBOU S., LETRÉMY P., ROUSSET P., Cartes auto-organisatrices de Kohonen, in *Apprentissage connexionniste*, Younès Bennani Ed., Hermès, Paris, p. 141-184.

#### III-2) - Construction de nouveaux algorithmes et développement de ces algorithmes en SAS et en Visual Basic pour les associer à Excel (Joseph Rynkiewicz, SAMOS, Patrick Letrémy, SAMOS, Catherine Aaron, SAMOS)

*[Rappel : Dans l'environnement SAS (avec des Macros SAS combinées au langage de programmation IML), nous allons continuer à développer l'ensemble des programmes de traitements de données quantitatives et qualitatives fondés sur l'algorithme de Kohonen. Nous nous proposons d'implanter des traitements pour données manquantes aussi bien quantitatives que qualitatives. Parallèlement Catherine Aaron continuera le développement des traitements de données issus de l'algorithme de Kohonen et de ses variantes en EXCEL, pour en faire un macro facilement utilisable par un non spécialiste. Enfin, pour permettre au plus grand nombre d'utiliser les techniques statistiques qui seront étudiées dans les projets précédents, nous fournirons des programmes informatiques utilisés pour la modélisation par des perceptrons multicouches combinés avec des chaînes de Markov cachées en les mettant en œuvre, essentiellement sous forme de bibliothèques pour le logiciel "R".]*

Les trois collègues cités ont continué le développement et les tests de leurs programmes.

<b>Auteur</b>	<b>Objet</b>	<b>Langage et plateforme</b>
Aaron Catherine	Cartes de Kohonen	VBA (Excel), Windows
Letrémy Patrick	Analyse de données Techniques neuronales	SAS-IML, Windows
Rynkiewicz Joseph	Perceptrons multicouches et chaînes de Markov cachées	C, Linux

Ces logiciels sont largement utilisés autant pour l'enseignement en M2 que pour les travaux de recherche des membres de l'équipe, comme par exemple dans

2008 SISTACHS V., COTTRELL M., LETRÉMY P., Aplicacion de las redes de Kohonen a un estudio sobre condiciones de bienestar y salud de los trabajadores, XVI SIMMAC, San José, Costa Rica, p. 166 -172.

### **III-3) - Analyses de données réelles, études économiques, Emploi du temps, Semainier (Patrick Letrémy, SAMOS, Corinne Perraudin, SAMOS)**

*[Rappel : Poursuite du travail sur la gestion de l'emploi et du travail à partir de l'enquête Emploi du Temps (INSEE 1998-1999). Poursuite de l'étude sur l'économie de la Turquie]*

Le travail entrepris sur les modes différenciés de gestion de l'emploi et du travail des établissements français à partir d'une classification issue d'une analyse de données par la méthode de Kohonen a été poursuivi., A partir de l'enquête REPONSE 1998-99 de la DARES, Perraudin et Petit (2006) proposent d'étudier les conséquences pour les salariés des modes de gestion adoptés par les établissements dans lesquels ils travaillent. La typologie permet d'apporter un éclairage sur les pratiques des établissements en matière d'utilisation d'emplois atypiques, de mise en place de nouvelles formes d'organisation du travail ou de création et destruction d'emplois.

Ce travail a également été poursuivi sur la base de la nouvelle vague de l'enquête REPONSE de la DARES (2004-2005). Une typologie des formes d'organisation du travail dans les établissements français a permis de remettre en cause la thèse d'un nouveau modèle organisationnel dominant, celui de la *lean production*, dont l'efficacité ferait un nouveau *one best way*, qui supplanterait le modèle taylorien. En effet, les résultats de Bunel, Dayan, Desage, Perraudin et Valeyre (2006) et Perraudin et Valeyre (2007) indiquent qu'à coté de la forme taylorienne et de la forme dite simple d'organisation du travail, coexistent deux nouvelles formes d'organisation, qui se distinguent par l'adoption de dispositifs organisationnels innovants et la délégation d'autonomie dans le travail. A ces formes d'organisation du travail sont associées des pratiques spécifiques de gestion de l'emploi et des ressources humaines.

2006 LEMIERE S., PERRAUDIN C., PETIT H., Les pratiques de gestion du travail et de l'emploi en France et leurs conséquences sur les salariés », *Document de travail CEE N° 75*, Décembre 2006.

2006 BUNEL M., DAYAN J.-L., DESAGE G., PERRAUDIN C., VALEYRE A., Formes d'organisation et relation de travail : configurations, complémentarités et dynamiques *Contrat de recherche pour l'exploitation secondaires de l'enquête REPONSE 2004-2005 pour la DARES* (2006)

2007 PERRAUDIN C., VALEYRE A., New pay systems and management of employment: differences between new forms of work organisation. *International Industrial Relations Association*, Royaume-Uni

2007 PERRAUDIN C., PETIT H., REBÉRIOUX A., Marché boursier et gestion de l'emploi : analyse sur données d'entreprises françaises, *document de travail*.

Afin d'étudier la diversité des situations en matière de conciliation entre vie professionnelle et vie familiale des mères de jeunes enfants en France, Perraudin et Pucci (2007 et 2008) proposent une typologie des mères de jeunes enfants sur la base de l'enquête Modes d'accueil et de garde des jeunes enfants, menée par la Drees en 2002. Cette démarche permet d'enrichir la définition de l'activité au-delà de la distinction entre personnes en emploi et sans emploi, et de distinguer les différentes combinaisons entre modes de garde selon leur durée. Elle offre la possibilité d'identifier un nombre important de modes d'organisation, qui illustrent les situations complexes des familles avec de jeunes enfants. Cette démarche permet enfin de montrer les limites des représentations économiques traditionnelles de la conciliation, dans lesquelles les choix reposent sur un arbitrage financier, les différents modes de garde payants ou les situations de non emploi ne sont pas distingués, et les modes de conciliation observés correspondent à des choix optimaux des familles.

2008 PERRAUDIN C., PUCCI M., Activité des mères de jeunes enfants et organisation de la garde : des choix complexes et souvent contraints, *Revue française des affaires sociales*, 1, p. 57-84

2007 PERRAUDIN C., PUCCI M., Diversité des modes de conciliation entre vie professionnelle et vie familiale des mères de jeunes enfants, Document de travail CEE, N° 94

#### **III-4) - Profils des consommateurs de GDF, aide à la détermination des tarifs (Marie Cottrell, SAMOS, Patrick Letrémy, SAMOS)**

[Rappel : Poursuite du le travail commencé en partenariat avec le laboratoire de recherche de Gaz-de-France à Saint-Denis. Il s'agit de réaliser une classification des profils de consommation des clients et de proposer une méthode d'allocation d'un nouveau consommateur parmi les classes définies.]

Tout d'abord, à l'occasion de ce travail, il s'est avéré qu'il était très important de pouvoir utiliser l'algorithme de Kohonen pour des bases de données avec des valeurs manquantes. En effet, le nombre de relevés de consommation des clients présents dans la base de données varie beaucoup d'un client à un autre.

Nous avons donc proposé une adaptation de l'algorithme d'auto-organisation de Kohonen originel pour

- Travailler avec des matrices de données incomplètes
- Estimer les valeurs manquantes

Ceci a conduit à une communication :

2005 COTTRELL M., LETRÉMY P., Missing values : processing with the Kohonen algorithm, *Proc. ASMDA*, Brest, mai 2005, [asmda2005.enst-bretagne.fr](http://asmda2005.enst-bretagne.fr), p.489-496.

Par ailleurs, le contrat avec GDF (d'une durée de 2 ans) s'est achevé par la remise d'un rapport et d'une maquette de logiciel permettant de définir les profils des consommateurs et d'affecter un nouveau venu à une de ces classes pour déterminer sa tarification.

Ce travail a donné lieu à plusieurs communications et publications :

2005 LETRÉMY P., ESPOSITO E., LAFFITE V., SHOWK S., COTTRELL M., The “Profilographe” : a Toolbox for the Analysis and the Segmentation of Gas Load Curves, WSOM 05 Conference, Paris, Sept. 2005, *Proceedings*, p. 447-454.

2007 LETRÉMY P., ESPOSITO E., LAFFITE V., SHOWK S., COTTRELL M., Consumer Profile Identification and Allocation, *Computational and Ambient Intelligence 9th International Work-Conference on Artificial Neural Networks, IWANN 2007*, San Sebastian (2007), p.530-538

Dans la même période, le SAMOS a été sollicité par GDF pour réaliser une étude de faisabilité sur la prévision de consommation de gaz au moyen de perceptrons multi-couches. Cela a abouti au rapport de recherche suivant :

2006 RYNKIEWICZ J., AARON C., COTTRELL M., Prévision court terme de la consommation de GAZ par les perceptrons multicouches : Etude de faisabilité.

où nous avons montré que l’utilisation des perceptrons n’améliore pas toujours la prévision par rapport aux modèles linéaires classiques, étant donné que l’essentiel de l’explication est contenu dans des termes d’analyse de variance, c’est-à-dire dans un modèle linéaire.

### **III-5) - Construction de pseudo-panels au moyen d’une carte de Kohonen, (Marie Cottrell, SAMOS, Patrice Gaubert, SAMOS et LEMMA)**

*[Rappel : Nous voulons explorer une nouvelle méthode de formation de pseudo-panels (pseudo-cohortes) cohortes, qui repose sur les cartes auto-organisées de Kohonen, dont les propriétés nous semblent particulièrement bien adaptées. Nous appliquerons cette méthodologie à un ensemble d’enquêtes canadiennes, portant sur les profils de consommation, pour estimer les élasticités des fonctions de consommations de chaque bien considéré.]*

Sur ce point, nous n’avons pas avancé, nous avons eu du mal à obtenir les données, et le changement de laboratoire de Patrice Gaubert (départ à l’université du Littoral) n’a pas permis que ce travail se poursuive. Nous le retrouvons dans la partie PROJET du présent rapport.

### **IV) ACI Systèmes complexes. Equilibre Dynamique, étude de trajectoires dans un système économique. Segmentation du marché du travail. (Marie Cottrell, SAMOS, Patrice Gaubert, SAMOS et LEMMA, Patrick Letrémy, SAMOS, Joseph Rynkiewicz, SAMOS)**

*[Rappel : L’objet de cette recherche est d’utiliser l’approche en termes de complexité pour mettre en évidence certaines dynamiques propres au marché du travail. L’arrière-plan théorique est la théorie du marché du travail segmenté. Pour cela on part d’une analyse préalable d’un marché réel, le marché américain, pour lequel on dispose d’un grand nombre d’informations longitudinales sur la situation de salariés grâce au panel PSID (Panel Study of Income Dynamics), créé et mis à jour depuis 1968. Sur ces données réelles, à l’aide de cartes de Kohonen, on met en évidence une typologie des situations occupées par les salariés, ainsi que leurs trajectoires. En confrontant ces trajectoires réelles aux mouvements observés dans un marché simulé construit à partir des comportements et caractéristiques identifiés sur le marché réel, nous déduisons des enseignements sur la dynamique du marché du travail. Les outils utilisés seront des réseaux de neurones combinés avec des chaînes de Markov cachées.]*

Ce travail a été financé par une ACI, de 2005 à 2006, mais il n’est pas terminé. Dans un premier temps, nous avons travaillé sur la base de données PSID qui est un panel où les personnes interrogées sont présentes tout au long de la durée de l’enquête. Mais nous n’avons

pas pu nous procurer la base actualisée (les délais de mise à disposition étaient trop grands) et nous avons décidé de travailler sur l'enquête EMPLOI de l'INSEE en France.

Cette enquête est de très grand volume et nous avons d'abord choisi les variables utiles en nombre le plus réduit possible, en distinguant les variables liées à la situation sur le marché de l'emploi, et d'autre part les variables qui décrivent les personnes.

Ensuite, toutes années confondues, nous avons classé les couples (individu, année) à l'aide d'une carte de Kohonen. En regroupant convenablement les classes de la carte de Kohonen, on a défini des super-classes qui peuvent être décrites en terme de segments du marché de l'emploi.

Ces résultats ont été présentés dans un colloque « Vers une science des systèmes complexes » organisé dans le cadre de l'ACI en 2007, puis dans une conférence internationale :

2007 COTTRELL M., GAUBERT P., RYNKIEWICZ J., LETRÉMY P, Equilibre dynamique, étude de trajectoires dans un système économique. Le cas du marché du travail, *MASHS 07*, Brest, *Conference Proceedings*, 2007

Mais dans cette enquête, chaque personne n'est présente que trois années consécutives, ce qui fait que nous avons dû estimer pour chaque couple d'années les probabilités de transition d'un segment à l'autre. Ensuite il est facile de simuler des trajectoires, en utilisant ces probabilités. Ces trajectoires ainsi reconstruites peuvent être analysées, regroupées en classes, pour décrire les trajectoires typiques. Ces résultats ont fait l'objet de deux communications :

2007 LETRÉMY P., COTTRELL M., GAUBERT P., RYNKIEWICZ J., Dynamical Equilibrium, trajectories study in an economical system. The case of the labor market, - *WSOM 2007, 6th International Workshop on Self-Organizing Maps Conference*, Bielefeld, Allemagne, CD-ROM Proceedings TUP-2.

2008 LETRÉMY P., COTTRELL M., Reconstruction and Classification of the trajectories in the market labor, *Communication à la Conférence OR8*, La Havane.

Reste un aspect qui n'a pas été traité pour l'instant et qui fera partie des projets de recherche : obtenir une définition « d'individus types » construits à partir des variables personnelles et qui pourraient être associés à tel ou tel segment.

## AXE B

### STATISTIQUES, PROCESSUS ET MÉTHODES ALEATOIRES

**Coordination : Jean-Marc Bardet et Annie Millet**

Les projets 2005-2008 du précédent rapport d'activités 2001-2004 étaient structurés en deux axes: A (méthodes Réseaux neuronaux et méthodes connexionnistes, analyse de données complexes) et B (Statistiques, processus et méthodes aléatoires). Avec l'arrivée de nouvelles compétences au sein du SAMOS et le découpage de l'Axe B en 2 axes (B et C), certains des projets 2005-2008 à consonance plutôt probabilistes sont du ressort du nouvel Axe C, les autres pouvant être regroupés dans le nouvel Axe B. Nous avons donc repris les projets de cet axe en les regroupant en deux sous-groupes qui correspondent aux nouveaux axes B et C. Les numéros de l'action alloués dans le précédent rapport sont repris.

### Actions dans des thématiques du nouvel axe B: Statistiques des processus et Statistiques spatiales

1. Estimation de phénomènes dynamiques et pseudo-panels (Action B.1),
2. Méthodes d'appariements et leur application à différents problèmes de modélisation (Action.B.2),
3. Estimation paramétrique pour des modèles à temps continu incomplètement observés (Action.B.3),
4. Dynamique spatiale, choix de standard et processus spatiaux (Action 1.B.4), Auto-modèles à états mixtes (Action B.5),
5. Théorème Central Limite (TCL) pour les fonctionnelles de Champ de Markov (Action B.6),
6. Modèles autorégressifs à changements de régimes (Action B.8)
7. Estimation d'un modèle de système de particules (Action.B.9),
8. Etude des variations quadratiques, outils d'identification de champs gaussiens sur  $R^2$  (Action B.12),
9. Analyse par ondelettes de processus longue mémoire (Action B.16),
10. Sélection de modèle, détection de ruptures et analyse multispectre (Action B.17),
11. La statistique paramétrique et non-paramétrique pour des processus à faible dépendance (Action B.18),
12. Etude du supremum de processus (Action.B.19),
13. Dépendance et processus alpha-stables (Action.B.20),

Voyons maintenant en détails les aboutissements de ces différentes actions:

#### **ACTION B.1. Estimation de phénomènes dynamiques et pseudo-panels (G. Duncan, Northwestern University, F. Gardes, Paris 1, P. Gaubert, SAMOS, M. Gurgand, Delta-CREST, C. Starzec, Team CNRS Paris 1).**

*[Rappel du projet : Nous avons mis en évidence des biais d'endogénéité présents dans l'estimation de phénomènes dynamiques qui se retrouvent dans l'estimation des paramètres que l'on obtient aussi bien avec des séries temporelles ou des coupes instantanées; ces biais sont évités par le recours aux estimateurs de panels appliqués aussi bien à des vrais panels qu'à des pseudo-panels.*

*D'autre part, nous avons effectué une vérification réalisée simultanément avec un panel américain (PSID) et un panel polonais, ainsi qu'en créant des pseudo-panels. à partir de ces vrais panels; l'utilisation de deux pays à des niveaux de développement très différents permettra de répondre à l'objection de conditions particulières d'un pays donné.]*

**Bilan:** Ce travail est terminé et a déjà fait l'objet de nombreuses présentations dans des universités françaises et américaines ainsi qu'à de nombreux colloques internationaux.

2005 GARDES, F., P. GAUBERT, P., DUNCAN, G., GURGAND, G. et STARZECH, C. Panel and Pseudo-Panel Estimation of Cross-Sectional and Time Series Elasticities of

Food Consumption: *The Case of U. S. and Polish Data*, Journal of Business and Economic Statistics, 23, 2.

**ACTION B.2. Méthodes d'appariements et leur application à différents problèmes de modélisation. (F. Gardes, CES et P. Gaubert, SAMOS).**

*[Rappel du projet: Dans le cadre d'un petit contrat avec la Direction de la Prévision (Ministère des Finances) sur des méthodes d'appariements et leur application à différents problèmes de modélisation.]*

**Bilan:** Nous construisons des pseudo-panels avec une application aux enquêtes françaises Budgets de Famille. L'issue prévue est un numéro spécial de la revue "Economie et Prévision" consacré à l'étude théorique et empirique de méthodes d'appariement.

**ACTION B.3. Estimation paramétrique pour des modèles à temps continu incomplètement observés. (X. Guyon, SAMOS, S. Souchet, SAMOS)**

*[Rappel du projet: Notre projet consiste à estimer les paramètres de processus de sauts markoviens et de modèles de diffusion à régimes cachés en se basant sur une observation discrétisée de ces processus. Nous projetons notamment d'aborder les trois questions suivantes :*

- *Obtention de conditions garantissant l'identifiabilité de ces modèles.*
- *Caractéristiques asymptotiques des estimateurs obtenus (loi et efficacité).*
- *Filtrage des régimes cachés.]*

**Bilan:** A ce jour, le projet n'a pas évolué et reste à l'étude. Il consiste à estimer deux types de processus: (1) les processus de sauts markoviens et (2) les modèles de diffusion à régimes cachés. On se base sur une observation discrétisée et incomplète de ces processus. Il s'agira: (a) d'obtenir des conditions garantissant l'identifiabilité de ces modèles, (b) d'estimer leurs paramètres, en précisant leur comportement asymptotiques et enfin (c) de filtrer le régime caché.

**ACTION B.4. Dynamique spatiale, choix de standard, processus spatiaux (C. Hardouin et X. Guyon, SAMOS)**

*[Rappel du projet: Nous travaillons sur la modélisation markovienne spatio-temporelle par potentiels de Gibbs pour des problèmes de choix de standards (i.e. choix entre deux technologies). Distribution spatiale des agents. Des agents placés sur un réseau choisissent entre deux standards suivant une règle contextuelle. Deux études sont envisagées : (i) La première considère que les choix sont itérés soit de façon séquentielle, soit de façon synchrone. On étudie la distribution stationnaire de ces modèles. Celle des modèles synchrones nécessite le recours aux simulations (en préparation).(ii) Reprenant les paradigmes d'adoption et de diffusion de standards en économie (cf. Arthur, Orléan, Paulré), nous souhaitons formaliser l'interprétation « chaîne de Markov » d'un certain nombre de dynamiques. Une dynamique est caractérisée par sa transition. La discrétisation en temps et en espace permet une approximation diffusion à partir de laquelle il est plus facile de décrire les caractéristiques analytiques de la loi stationnaire et qui devrait permet d'interpréter certains comportements de la dynamique observés empiriquement.]*

**Bilan:** Une première étude a été réalisée dans le contexte où l'on considère un seul balayage de l'ensemble des sites. Ce qui revient à considérer que le choix des agents est unique ; cela correspond à un marché de premier équipement, ou bien au contexte de deux technologies incompatibles en compétition. Les résultats obtenus ont été repris et complétés. Les deux aspects cités ci-dessus en rappel sont toujours en projet, les sujets de l'action B.5. ayant fait l'objet d'une attention privilégiée.

2007 HARDOUIN C. Dumping influence on a non iterative dynamics, in Recent Advances in Stochastic Modelling and Data Analysis, Editor C. H. Skiadas, World Scientific Publishing.

2007 HARDOUIN C. (2007) Influence d'une contamination initiale sur une dynamique spatiale non itérative, Soumis.

### **ACTION B.5. Auto-modèles à états mixtes (X.Guyon et C. Hardouin, SAMOS, J.F. Yao, IRISA et Univ. de Rennes 1)**

*[Rappel du projet: J. Besag a proposé en 1974 une famille importante de modèles à états réels caractérisés par leur famille de lois conditionnelles, les auto-modèles de Besag. Nous les généralisons au cas multivarié. Ceci permet de définir, entre autres, des modèles pour des variables à espace d'état mixte, par exemple  $E=\{0\}UR^+$ . Un équivalent en économétrie est le modèle TOBIT. Les états mixtes se rencontrent dans divers domaines d'applications, par exemple en économie (dépenses de loisir d'un ménage), en analyse de données pluviométriques (0 signifie qu'il n'a pas plu, une donnée est  $>0$  donnant la hauteur de pluie), en analyse du mouvement (0 repère un objet immobile et sinon la valeur  $>0$  est le module de la vitesse enregistrée). Ces modèles s'étendent en champ de Markov sur réseau spatial  $S$  et en dynamique markovienne de champs de Markov. Ils se manipulent bien sur le plan statistique au travers de la pseudo-vraisemblance conditionnelle facile à expliciter et à manipuler là où la vraisemblance exacte est incalculable. Reste à identifier les propriétés des estimateurs, convergences, normalités et tests asymptotiques. Les auto-modèles markoviens exponentiels mixtes, bien adaptés à des données pluviométrique, seront étudiés et comparés à d'autres modèles sur la base de données réelles. Une première limitation de ces modèles est de ne pas permettre la coopération spatiale : des variantes sont étudiées, permettant la coopération. Une deuxième limitation de ces modèles est d'être rapidement de très grande dimension paramétrique si le champ est inhomogène spatialement ; nous proposons diverses alternatives permettant une forte réduction de la dimension, en particulier par le biais de la prise en compte de variables exogènes.]*

**Bilan:** L'objectif était l'étude de modèles de champs spatiaux dont les observations appartiennent à un espace d'état  $E$  de nature mixte, par exemple des valeurs nulles accompagnées de valeurs réelles positives. Ces données se rencontrent par exemple en météorologie (absence ou hauteur de précipitation), en analyse du mouvement d'une séquence vidéo, en épidémiologie (absence d'une maladie et gravité si la maladie est présente).

*Généralisation des auto-modèles de Besag:* Une première étape était de mettre en place une structure adéquate à la modélisation de telles variables mixtes. Les "auto-modèles" de Besag (1974) sont basés sur des propriétés des lois conditionnelles. Ils supposent que premièrement les interactions ne concernent que les singletons et les paires, deuxièmement que les lois conditionnelles sont dans une famille exponentielle. Une limitation majeure de cette famille de modèles est que les lois conditionnelles en question sont en dimension 1 dans la mesure où la statistique exhaustive et le paramètre naturel associé sont scalaires. Un second inconvénient réside dans les conditions d'utilisation de ces modèles. Il faut s'assurer que les lois conditionnelles se recollent en une loi jointe bien définie, intégrable. Cette condition d'intégrabilité va entraîner des contraintes sur les paramètres qui induisent des comportements de compétition entre sites voisins dans beaucoup de modèles, ce qui les rend inintéressants dans les applications. Nous avons travaillé sur l'idée de considérer des familles exponentielles pour lesquelles la statistique exhaustive ainsi que le paramétrage naturel associé sont en dimension supérieure ou égale à 2. Une première étape a été de décrire cette extension, et de donner un résultat de consistance pour l'estimateur de pseudo-vraisemblance conditionnelle. Un exemple important dans les applications de cette extension est l'utilisation de lois Beta conditionnelles pour des modèles coopératifs.

*Auto-modèles mixtes:* La notion de données mixtes a tout d'abord été étendue au cas où les observations ont des composantes de nature différente, à valeurs dans un espace discret ou continu. Dans un second temps, le cadre des auto-modèles multivariés a permis une modélisation directe, non hiérarchique de ces données de nature mixte. Comme dans le cadre univarié, l'intégrabilité de la loi jointe peut entraîner des contraintes sur les paramètres du modèle imposant un principe de compétition spatiale entre sites. Nous détaillons sur l'exemple exponentiel mixte (les lois conditionnelles présentent une masse en zéro et une densité

exponentielle sur  $\mathbb{R}^+$ ) des variantes, soit tronquées à droite, soit censurées à droite, qui elles, permettent la coopération. Enfin, cette nouvelle modélisation a été expérimentée sur des données réelles d'analyse de mouvement dans des séquences vidéo.

2006 BOUTHEMY, P., HARDOUIN C., PIRIOU, G. et YAO J.-F. Mixed state auto-models and motion texture modeling, *J. M. I. V.*, 25, 3, 387--402.

2007 HARDOUIN, C. et YAO J.-F. Multi-parameter auto-models with applications to cooperative systems, *C. R. Acad. Sci. Paris, Ser. I* 345.

2008 HARDOUIN, C. et YAO J.-F. Multi-parameter auto-models and their application, A paraître dans *Biometrika*.

2008 HARDOUIN, C. et YAO J.-F. Spatial modelling for mixed state observations. *Electronic Journal of Statistics* 2, 213-233.

### **ACTION B.6. Théorème Central Limite (TCL) pour les fonctionnelles de Champ de Markov (C. Gaetan, Univ. de Venise, et X. Guyon, SAMOS)**

[Rappel du projet: Les Champs de Markov (CM) interviennent dans la modélisation de nombreux problèmes, données longitudinales, imagerie, sciences de l'environnement, épidémiologie, etc. Les résultats mathématiques sont donnés généralement pour un champ défini sur un réseau régulier, typiquement  $\mathbb{Z}^d$ , comme en mécanique statistique. En fait, dans la plupart des applications citées précédemment, les réseaux spatiaux sont irréguliers, la structure de voisinage n'est pas invariante par translation tout comme les potentiels qui définissent les champs de Gibbs sous-jacent. Notre objectif est d'obtenir des résultats asymptotiques dans ce cadre non-régulier, TCL en particulier. Nous généralisons les résultats de Comets-Janzura à ce cadre pour une fonctionnelle conditionnellement centrée d'un champ de Markov. Une difficulté dans l'obtention d'un TCL est de s'assurer que la matrice de variance-covariance limite est définie positive. Nous donnons des outils pour vérifier cette condition en utilisant une structure d'indépendance conditionnelle pour la fonctionnelle aux sites définissant une partition conditionnellement séparante. Ces résultats permettent d'établir la normalité asymptotique d'estimateur de pseudo-vraisemblance conditionnelle d'un CM. Rappelons que sans hypothèse particulière (du type absence de transition de phase) nous ne savons pas si l'estimateur du MV est asymptotiquement gaussien. Ces résultats devraient permettre la construction de tests du Chi2 de pseudo-vraisemblance pour des modèles emboîtés.]

**Bilan:** Très souvent (épidémiologie, environnement, sciences de la terre, etc.), les réseaux spatiaux sur lesquels sont définis des champs de Markov sont irréguliers, de structure de voisinage non invariante par translation tout comme les potentiels des champs de Gibbs sous-jacent. Notre objectif est toujours d'obtenir des résultats asymptotiques dans ce cadre non-régulier, TCL en particulier. Il s'agit donc de généraliser les résultats de Comets-Janzura à ce cadre «potentiels non-invariants par translation» pour une fonctionnelle conditionnellement centrée d'un champ de Markov. Une première difficulté dans l'obtention d'un TCL est d'abord de s'assurer que la matrice de variance-covariance limite est définie positive. Nous avons obtenu des outils qui permettent de vérifier cette condition en utilisant une structure d'indépendance conditionnelle pour la fonctionnelle aux sites définissant une partition conditionnellement séparante (Gaetan-Guyon). L'existence d'un TCL dans ce contexte est un problème difficile; nous ne disposons pas en effet des outils classiques de la mécanique statistique qui considère toujours un réseau régulier et des modèles de Gibbs à potentiels invariants par translations.

2004 GAETAN, C. et GUYON, X. A Central Limit Theorem for conditionally centred functional of a Markov random field. 14 pages.

## **ACTION B.8 Modèles autorégressifs à changements de régimes. (B. Maillet, TEAM, Paris I, M. Oltéanu, SAMOS et J. Rynkiewicz, SAMOS).**

*[Rappel du projet: a) Détermination du nombre de régimes : Les modèles à changements de régimes sont de plus en plus utilisés pour la description des séries temporelles. Une question récurrente est la détermination statistique du nombre de régimes. Des débuts de réponse ont été apportés par Elisabeth Gassiat et Stéphane Boucheron, nous nous proposons d'étudier l'extension de leurs résultats aux modèles auto-régressifs à changement de régimes.*

*b) Modélisation « réaliste » de la durée des régimes. Outre l'identification du modèle, la deuxième critique que l'on puisse faire aux modèles utilisant des chaînes de Markov cachée est la pauvreté de la modélisation du temps passé dans chacun des régimes. Nous nous proposons d'introduire des modèles de durée pour obtenir des modèles semi-markoviens qui s'adaptent mieux aux données réelles. Les principales difficultés seront de trouver des méthodes de calculs qui ne consomment pas trop de ressources et d'étudier les propriétés statistiques des estimateurs obtenus.]*

**Bilan:** Ce travail constitue une partie de la thèse de M. Olteanu, soutenue en décembre 2006. Le problème de la sélection du nombre de régimes représente un cas typique de non-identifiabilité du modèle qui apparaît aussi, par exemple, pour les mélanges de densités. Ceci rend la matrice de Fisher non inversible et ne permet pas d'appliquer la théorie usuelle sur la convergence du rapport de vraisemblance. Notons aussi que ce rapport devient divergent dans certains cas comme celui des chaînes de Markov cachées (Keribin et Gassiat, 2000). Afin de contourner ce problème, nous avons proposé un critère basé sur la log-vraisemblance pénalisée et nous avons montré sa consistance. Le résultat a été démontré dans le cas où les changements de régime sont indépendants (mélanges de modèles autorégressifs). Les hypothèses du résultat de convergence ont été ensuite vérifiées dans le cas particulier des fonctions de régression linéaires et du bruit gaussien. La partie « difficile » a été de montrer que la classe des fonctions scores généralisés vérifie une condition sur sa complexité, exprimée en termes d'entropie à crochets. Afin de contourner les difficultés, nous avons introduit une reparamétrisation du modèle, inspirée d'un article de Liu et Shao (2003) et permettant un développement de Taylor à l'ordre deux autour de la vraie valeur des paramètres identifiables.

Des simulations effectuées à l'aide d'un algorithme de type EM nous ont permis ensuite d'illustrer le résultat, sa vitesse de convergence et la stabilité de l'algorithme proposé.

2008 OLTEANU, M. et RYNKIEWICZ, J. Estimating the number of regimes in a switching autoregressive model. HAL archives hal-00137438, en révision pour *ESAIM PS*.

## **ACTION B.9. Estimation d'un modèle de système de particules (X. Guyon, SAMOS, et B. Pumo, INH, Angers)**

*[Rappel du projet: Les systèmes de particules constituent un domaine de la théorie des probabilités faisant appel à des techniques spécifiques originales et difficiles (i.e. cf le cours de St. Flour de Durrett). Ils modélisent des évolutions temporelles de configurations spatiales dont les transitions sont locales et invariantes par translation dans le temps et l'espace. Un PC modélise par exemple l'évolution d'une espèce végétale sur un réseau spatial, des phénomènes de coopération / compétition en économie spatiale. En dehors d'un travail de Fiocco et Van Zwett, il y a peu d'étude statistique sur ces modèles. Considérant un système de particules particulier, le processus de contact au 4-plus proches voisins et à temps discret (Durrett et Levin) et disposant d'une suite d'observations temporelles sur un domaine fini et fixé  $S$ , nous construisons une pseudo-vraisemblance marginale qui s'explique aisément là où la vraisemblance exacte devient incalculable. Cette fonctionnelle permet d'identifier les paramètres du modèle et conduit à une estimation convergente et asymptotiquement normale pour l'asymptotique temporelle. Une propriété cruciale est la sous ergodicité du processus. Il faut maintenant généraliser ces résultats à d'autres systèmes de particules, mettre en place les outils statistiques et numériques permettant d'estimer et de valider un modèle de système de particules.]*

**Bilan:** Nous avons obtenu des résultats asymptotiques pour l'estimation d'un processus de contact (PC) au sens de Durrett-Levin lorsque nous disposons d'une séquence de  $T$  observations du PC sur une partie finie du réseau sur lequel est définie la dynamique (2

publications). La vraisemblance étant numériquement incalculable, on utilise une pseudo-vraisemblance basée sur une «transition marginale» égale au produit des transitions en chaque sites.

2007 GUYON, X. et PUMO, B. (2007), Space-time estimation of a particule system model, *Statistics*, Vol. 41, 395-407

2005 GUYON, X. et PUMO, B. (2005), Estimation spatio-temporelle d'un système de particule, *CRAS, Sér. I-340*, 619-622.

**ACTION B.12. Variations quadratiques, outils d'identification de champs gaussiens sur  $R^2$  (S. Cohen et M. Pontier, LSP à l'UPS, Toulouse, O. Perrin, GREMAQ à Toulouse 1, X. Guyon, SAMOS)**

*[Rappel du projet: On sait (cf. thèse de Guyon-Prum) qu'il existe une grande variété de variations quadratiques pour un champ sur  $R^2$  : variations sur un segment, variations curvilignes, variations produit, variations superficielles, etc. Cette situation contraste avec celle des processus à un indice où il n'existe qu'une seule variation quadratique non triviale. Mieux, pour les champs, ces variations peuvent être paramétrées par un facteur qui est sous notre contrôle, par exemple l'angle que fait le segment avec l'axe des  $x$  pour une variation linéaire, la géométrie du rectangle (ratio des deux côtés) pour une variation superficielle, etc. Ceci explique pourquoi ces variations sont un outil d'identification de modèle efficace (cf Guyon, PTRF 1987, Leon et Ortega 1989). Notre objectif est d'enrichir et de préciser ces résultats pour deux modèles gaussiens particulier : le drap brownien fractionnaire standard et le drap brownien fractionnaire. Nous montrons pour le premier que les variations sur segments permettent d'identifier une transformation isométrique du drap. Pour le deuxième modèle, et en nous limitant à une rotation, nous identifions l'angle de cette rotation ainsi que les deux paramètres fractionnaires. Une question importante qui subsiste est la mise en œuvre statistique et numériquement ces résultats théoriques : quels schémas d'estimation associés à quel bon jeu de fonctionnelles d'identification lorsque le champ n'est observé que sur un maillage fin, mise en place des intervalles de confiance d'un paramètre, procédures de tests. Ces résultats s'appliquent potentiellement à l'identification d'images observées finement, comme par exemple les images médicales.]*

**Bilan:** Deux articles sont parus. Le premier montre comment l'utilisation d'une bonne fonction de singularité permet, entre autre, d'identifier les deux paramètres d'un drap Brownien fractionnaires. Un deuxième explicite le fait que les variations sur segments permettent d'identifier une transformation isométrique du drap brownien. Une question pratique importante subsiste, à savoir celle de la mise en œuvre statistique et numérique de ces résultats. Les domaines d'applications potentiels concernent l'identification d'images observées à haute résolution, comme par exemple les images médicales.

2006 COHEN, S., GUYON, X., PERRIN, O. et PONTIER, M., Identification of an isometric transformation of the standard Brownian sheet, *Jour. Stat. Planning and Inf.*, 136, 1317-1330

2006 COHEN, S., GUYON, X., PERRIN, O. et PONTIER, M., Singularity functions for a fractional process: application to the fractional Brownian sheet, *Ann. I.H.P.*, 42, 187-205

**ACTION B.16. Analyse par ondelettes de processus longue mémoire. (J.-M. Bardet, H. Bibi, SAMOS et G. Teysnière, ENSAI et SAMOS)**

*[Rappel du projet: L'analyse par ondelettes des processus autosimilaires et des processus longue mémoire a été initiée dans une optique de traitement du signal par Flandrin (IEEE, 1992) et Abry et Veitch (JTSA, 1998) et a été reprise dans un cadre plus mathématique par Bardet, Lang, Moulines et Soulier (SISP, 2002). Il s'agit d'utiliser la propriété d'invariance d'échelle (asymptotique dans le cadre des processus longue mémoire) des moments d'ordre 2 du processus, en la retrouvant sur les moments d'ordre 2 empirique des coefficients d'ondelette du processus pour certaines échelles choisies. Plusieurs questions demeurent, notamment les*

*propriétés statistiques des estimateurs du paramètre (d'autosimilarité ou de Hurst) pour des processus non-gaussiens, le choix adaptatif de la gamme d'échelles permettant de l'estimation et la généralisation de cette technique à des processus alpha-stables. Ceci est le projet de thèse de doctorat de Hatem Bibi.]*

**Bilan:** Une première étape de la thèse de Hatem Bibi a été menée à bien: celle de la construction d'un estimateur adaptative du paramètre de mémoire pour les séries chronologiques gaussiennes, et ceci dans un cadre semi-paramétrique très général. Le comportement asymptotique de l'estimateur a été étudié, et un théorème de la limite centrale a été démontré. Des simulations ont été réalisées, montrant la rapidité de convergence de cet estimateur en comparaison de celles des estimateurs semi-paramétriques les plus performants (méthodes du log-périodogramme, contraste de Whittle,...). Une publication est issue de ce travail: Bardet, Bibi et Jouini (2008).

Ceci a permis également de mettre en place et à disposition de tous (voir le site internet du SAMOS) toute une gamme de logiciels (écrits en Matlab) permettant la simulation et l'estimation (paramétrique ou semi-paramétrique) de processus à longue mémoire.

Un deuxième travail portant sur les processus linéaires à paramètre de mémoire est en cours et devrait être achevé avant la fin 2008.

Sur la même thématique, Teyssière et Abry (2007) ont étudié la robustesse de l'estimateur du paramètre de longue mémoire basé sur les ondelettes de Abry et Veitch (1998) dans le cas des modèles de volatilité les plus utilisés en statistique/économétrie de la finance. Il apparaît que cet estimateur est peu sensible aux non-linearités, ruptures et changement de régime observés sur des séries financières habituelles. Ceci permet d'établir un diagnostic sur la présence supposée de longue portée dans les processus de volatilité en finance ou deux points de vue radicalement opposés prévalent: processus fortement dépendant, ou processus stationnaire. L'estimateur du paramètre de longue portée basé sur les ondelettes nous permet de conclure qu'il est plus vraisemblable que les deux phénomènes coexistent. Ceci a des conséquences sur les méthodes de valorisation des options. Ce travail a été publié dans Teyssière et al. (2007) *Long Memory in Economics*, Springer Verlag.

2008 BARDET, J.-M., BIBI, H. et JOUINI, A. Adaptive wavelet based estimator of the memory parameter for stationary Gaussian processes. A paraître dans *Bernouilli*.

2005 TEYSSIERE, G. et ABRY, P. Wavelet Analysis of Nonlinear Long-Range Dependent Processes. Applications to Financial Time Series, dans *Long-Memory in Economics*, G. Teyssière and A. Kirman editors, 173-238, Springer Verlag.

### **ACTION B.17. : Sélection de modèle, détection de ruptures et analyse multispectre (J.-M. Bardet, SAMOS, Véronique Billat, Laboratoire d'Etude de la PHysiologie de l'Exercice, et I. Kammoun, SAMOS)**

*[Rappel du projet: Nous avons entrepris un travail en commun avec le Laboratoire d'Etude de la PHysiologie de l'Exercice de l'Université Evry Val d'Essonne. Disposant d'un grand nombre de relevés de signaux (fréquence cardiaque et pulmonaire, vitesse en fonction du temps) de coureurs au cours d'un marathon, l'objectif de ce travail, et en particulier de la thèse de I. Kammoun, est d'analyser ces signaux pour en obtenir de possibles modélisations et de détecter de potentielles ruptures lors de l'épreuve. Ceci qui pourrait aboutir à plus long terme à un outil d'aide au diagnostic et de prévision de « dysfonctionnement » physiologique (arrêt cardiaque par exemple) pour les athlètes. La recherche de potentielles ruptures et leur détection pourront être d'abord étudiée sur les indicateurs classiques que sont la moyenne ou la variance du signal, à partir de technique de contrastes pénalisés, mais également sur l'irrégularité même de ces signaux, ceci à l'aide de l'analyse par ondelettes (voir également le travail précurseur de V. Billat et Y. Meyer, 2002). L'étude du comportement « multifractal » de ces signaux est ainsi un objectif à plus long terme de ce projet.]*

**Bilan:** Ce travail en collaboration constitue la thèse de doctorat de Imen Kammoun, soutenue en décembre 2007. Trois directions de recherche ont finalement été suivies:

1/ un premier travail a consisté à montrer que l'estimateur DFA (Detrended Fluctuations Analysis) utilisé fréquemment par les physiciens, physiologistes,..., pour estimer le paramètre de longue mémoire pour des séries avec tendance, n'est finalement pas un choix raisonnable. Un travail théorique montre d'abord que dans un cadre semi-paramétrique, la vitesse de convergence de l'estimateur DFA pour un processus gaussien stationnaire (sans tendance) n'atteint pas toujours la vitesse de convergence mini-max, atteinte pourtant par les estimateurs log-périodogramme, à ondelettes, à contraste de Whittle local,... Ensuite, on montre que sur des tendances les plus élémentaires la méthode DFA conduit à des estimations du paramètre de longue mémoire totalement erronées. Des simulations illustrent les résultats théoriques obtenus. Cette étude a amené la publication Bardet et Kammoun (2008a).

2/ un second travail, centré sur l'objectif de la modélisation des données de fréquences cardiaques instantanées pendant une épreuve d'endurance a conduit à définir un nouveau type de processus: les bruits gaussiens localement fractionnaires, pour lesquels, dans une gamme de fréquence finie, le comportement de la densité spectrale est celui d'une loi de puissance. La puissance de cette loi étant une fonction affine du paramètre de fractalité locale, qui est une généralisation du paramètre de Hurst, ayant la vertu de pouvoir être tout nombre réel et non seulement un nombre dans  $[0,1]$ . Ensuite, une méthode d'estimation de ce paramètre est proposée, utilisant l'analyse par ondelettes, et un test d'adéquation est enfin proposé (les comportements asymptotiques sont également étudiés). Cette méthode est appliquée aux données de fréquences cardiaques instantanées, après que celles-ci aient été découpées en 3 zones: début de course, milieu de course et fin de course. On montre alors que la modélisation par un bruit gaussien localement fractionnaire est validée dans chacune des zones (alors qu'elle ne l'est pas sur l'ensemble de la course) et on montre également que le paramètre de fractalité locale augmente significativement au cours de la course. Ce travail a conduit à la prépublication : Bardet *et al.* (2008).

3/ enfin, une troisième étude traite de la détection des paramètres de longue-mémoire, de local fractalité ou d'autosimilarité pour des séries chronologiques gaussiennes. On montre ainsi la convergence vers les instants de ruptures (supposés en nombre fini connu), et un théorème de la limite centrale pour l'estimation des différents paramètres dans chaque zone estimée. Des simulations montrent la qualité des estimateurs et l'application aux données physiologiques de fréquences cardiaques instantanées permet de détecter l'apparition de la fatigue chez un coureur, ce qui se manifeste par une rupture sur le paramètre de fractalité locale. Une note succincte découle de ce travail (voir Bardet et Kammoun, 2008b) ainsi qu'une prépublication Bardet et Kammoun (2007).

2007 BARDET, J.-M., BILLAT, V. et KAMMOUN, I. A new stochastic process to model Heart Rate series during exhaustive run and an estimator of its fractality parameter. HAL archives hal-00176298.

2007 BARDET, J.-M. et KAMMOUN, I. Detecting changes in the fluctuations of a Gaussian process and an application to heartbeat time series. HAL.archives hal-00194909.

2008 BARDET, J.-M. et KAMMOUN, I. Asymptotic Properties of the Detrended Fluctuation Analysis of Long Range Dependence Processes. *IEEE Trans. on Info. Theory* 54, 2041-2052.

2008 BARDET, J.-M. et KAMMOUN, I. Detecting abrupt changes of the long-range dependence or the self-similarity of a Gaussian process. A paraître dans *Notes aux C.R.A.S.*

## **ACTION B.18. Statistique paramétrique et non-paramétrique pour des processus à faible dépendance (J.-M. Bardet, SAMOS, P. Doukhan SAMOS et Université de Cergy-Pontoise, J. Léon, Université de Caracas)**

*[Rappel du projet: Les processus faiblement dépendants ont été introduits par Doukhan et Louichi (SPA, 1999) pour définir une nouvelle forme de dépendance plus générale et effective que celle des différentes classes de processus mélangeant. Hors les processus longue mémoire, la plupart des exemples classiques de séries chronologiques stationnaires (ARMA, GARCH,...) sont des processus faiblement dépendants. De nombreuses études probabilistes et statistiques sont en cours sur ces processus (Rio, 2002, Dedecker et Doukhan, SPA 2003, Dedecker et Prieur, 2004). Nous nous sommes tout d'abord intéressés au comportement asymptotique du périodogramme intégré de tels processus. Nous avons ainsi obtenu des théorèmes de limite centrale uniformes pour une classe de fonctions dans un espace de Sobolev. L'utilisation naturelle de ceci est l'obtention de théorèmes de la limite centrale pour les estimateurs paramétriques (technique du maximum de vraisemblance de Whittle) de ces processus. Par ailleurs, le traitement des processus faiblement dépendants non-causaux a nécessité l'étude poussée de la vitesse de convergence vers la loi normale de sommes de fonctions du processus. Un tel résultat est la source de nombreuses conséquences possibles en statistique non-paramétriques, telles l'estimation semi-paramétrique et des tests d'adéquation. L'exploitation et l'extension de ces résultats est l'objet de la thèse de O. Winterberger (codirection, J.-M. Bardet et P. Doukhan) qui débute en septembre 2004.]*

**Bilan:** De ce projet, et avec l'arrivée de P. Doukhan au sein du SAMOS, de très nombreuses études ont été développées.

En premier lieu, le livre Dedecker *et al.* (2007) (Lecture Notes 190 in Statistics, 350 pages), est maintenant la référence de la théorie en plein développement de la dépendance faible initiée par Doukhan et Louhichi dès 1996. Dans Doukhan et Neumann (2008) nous présentons aussi de riches applications des notions de dépendance faibles au ré-échantillonnage. Dans Doukhan, Neumann (2007), des inégalités exponentielles du type de Bernstein et des inégalités de Rosenthal précises sont prouvées par des techniques de cumulants sous des hypothèses de dépendance faible. Leur application aboutit à de nombreux énoncés liés à la loi du logarithme en probabilité. Ainsi des vitesses minimax sont obtenues dans ce contexte dépendant.

Dans Bardet, Doukhan, Lang et Ragache (2008), nous étendons la méthode de Lindeberg à un cas dépendant, son applications à des questions de sous échantillonnage et d'estimation fonctionnelle montre l'attrait de notre énoncé, pourtant très simple. Dans Doukhan et Winterberger (2007), la méthode de Lindeberg précédemment citée permet d'aboutir à des résultats étendant le théorème de Donsker sous des conditions de dépendance faible. On trouve également dans Bardet, Doukhan et Leon (2008a) un théorème de la limite centrale pour des fonctions de séries chronologiques faiblement dépendantes, avec des applications pour l'estimation des moments. Par ailleurs, des théorèmes limite uniformes sont démontrés pour les formes quadratiques de processus faiblement dépendants. Une application à l'estimation paramétrique de Whittle est ensuite proposée, et un théorème de la limite centrale de l'estimateur est montré; de très nombreux exemples de processus faiblement dépendants sont ensuite proposés (voir Bardet, Doukhan et. Leon, 2008b).

Dans Doukhan, Fermanian, Lang (2008), un TLC empirique multidimensionnel permet d'obtenir les propriétés asymptotiques des copules empiriques, très utilisés par les financiers. Le comportement de la densité de copule y est aussi abordé ainsi que l'étude asymptotique de copules paramétriques. Dans P. Doukhan, N. Mayo, L. Truquet. (2008), loin de se limiter à un article de revue ce travail exhibe les propriétés asymptotiques d'estimateurs de moindres carrés pour des modèles ARCH infinis. Une dernière partie montre aussi des inégalités de moments et un théorème de Donsker pour des champs aléatoires.

Dans Doukhan, Latour et Oraichi (2006), un modèle de type ARCH à valeurs entières est construit en utilisant les idées de A. Latour, Ses propriétés de dépendance faible permettent de déduire la consistance et la normalité asymptotique d'estimateurs empiriques de ses paramètres.

Dans Doukhan, Madré et Rosenbaum (2007), nous prouvons les propriétés de dépendance faible d'un modèle à mémoire infini introduit par Giraitis et Surgailis en 2002; de plus l'étude de la régularité de ses répartitions finies permet d'en envisager les propriétés asymptotiques

d'estimateurs non paramétriques. Dans Doukhan, Teyssière et Winant (2006), une large classe de modèles de séries temporelles est mise en lumière ici, elle étend de nombreux modèles économétrique; nous en déterminons aussi les propriétés de dépendance faible. Ces modèles généralisent le précédent. Dans Doukhan et Wintenberger (2008), des modèles non Markoviens et non linéaires à mémoire infinie étendent ici la plupart des modèles utilisés en statistique des séries temporelles et en économétrie: nous prouvons qu'ils satisfont aussi à des conditions de dépendance faible. Ces derniers résultats sont utilisés avec succès dans Bardet et Wintenberger (2007), où l'on définit un modèle causal multidimensionnel et on étudie la convergence du quasi-maximum de vraisemblance (loi des grands nombres fortes et TLC). Les résultats obtenus améliorent tous ceux existant, et permettent de traiter de nouveaux processus par cette méthode d'estimation (processus TARCH par exemple).

2008 BARDET, J.-M., DOUKHAN, P. et LEON, J. A functional limit theorem for  $\eta$ -weakly dependent processes and its applications. A paraître dans *Statistical Inference for Stochastic Processes*.

2008 BARDET, J.-M., DOUKHAN, P. et LEON, J. Uniform limit theorems for the integrated periodogram of weakly dependent time series and their applications to Whittle's estimate. A paraître dans *Journal of Time Series Analysis*.

2008 BARDET, J.-M., DOUKHAN, P., LANG, G. et RAGACHE, N. Dependent Lindeberg central limit theorem and some applications. *ESAIM Probability and Statistic*, 12, 154-172.

2008 BARDET, J.-M. et WINTENBERGER, O. Asymptotic normality of the Quasi Maximum Likelihood Estimator for multidimensional causal processes. En révision pour *Annals of Stats*.

2007 DEDECKER, J., DOUKHAN, P., LANG, G., LEON, J., LOUHICHI, S. et PRIEUR C. Weak dependence: with examples and applications. *Lecture Notes in Statistics*, 190, Springer.

2008 DOUKHAN, P., FERMANIAN, J.-D. et LANG, G., An empirical central limit theorem with applications to copulas under weak dependence. A paraître dans *Statistical Inference for Stochastic Processes*.

2008 DOUKHAN, P. et LANG, G. Evaluation for moments of a ratio with application to regression estimation. HAL

2006 DOUKHAN, P., LATOUR, A. et ORAICHI, D. A simple integer-valued bilinear time series model. *Adv. in Appl. Probab.*, 38, 559-578.

2007 DOUKHAN, P., MADRE, H. et ROSENBAUM, M., ARCH type bilinear weakly dependent models. *Statistics*, 41, 31-45.

2008 DOUKHAN, P., MAYO, N. et TRUQUET, L., Weak dependence, models and some applications. A paraître dans *Metrika (2008)*.

2007 DOUKHAN, P. et NEUMANN, M., Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Process. Appl.* 117, 878-903.

- 2008 DOUKHAN, P. et NEUMANN, M., The notion of weak dependence and its applications to bootstrapping time series. *Probability Surveys* 5.
- 2006 DOUKHAN, P., TEYSSIERE, G. et WINANT, P., A LARCH\$(\infty\$-vector valued process. *Dependence in probability and statistics*, 245-258, Lecture Notes in Statistics, 187, Springer.
- 2007 DOUKHAN, P. et WINTENBERGER, O., An invariance principle for weakly dependent stationary general models. *Probab. Math. Statist.*, 27, 45--73.
- 2008 DOUKHAN, P. et WINTENBERGER, O. (2008)., Weakly dependent chains with infinite memory. A paraître dans *Stochastic Process. Appl.*

**ACTION B.19. Etude du supremum de processus (J.-M. Bardet, SAMOS, J.-M. Azaïs Université de Toulouse III et M. Wschebor Université de Montevideo)**

[Rappel du projet: La méthode dite « de Rice » permet de calculer l'espérance du nombre de dépassement de niveau (« crossings ») d'un processus stochastique suffisamment régulier. L'utilisation de cette méthode a amenée J.-M. Azaïs et M. Wschebor à obtenir de nouveaux résultats sur la densité de probabilité et la probabilité de dépassement de niveau du supremum d'un processus gaussien de classe C2. Dans un précédent travail, nous nous sommes intéressé au comportement asymptotique de la loi du supremum d'un processus gaussien stationnaire régulier (Azaïs, Bardet et Wschebor, ESAIM, 2002). Cependant, si la méthodologie développée est riche en potentialité, elle se heurte au passage aux processus non réguliers. Par des techniques de régularisation de trajectoires, nous travaillons sur le problème ouvert du comportement asymptotique en 0 de la fonction de répartition du supremum du mouvement brownien fractionnaire de paramètre H pris sur un intervalle fermé [0,T] (des premiers éléments de réponse ont été apportés par d'autres techniques et par Sinai, 2000 et Molchoi, 2002)]

**Bilan:** Malheureusement ce projet n'a pas été poursuivi.

**ACTION B.20. Dépendance et processus alpha-stables (J.-M. Bardet, SAMOS et P. Bertrand, Université Blaise Pascal)**

[Rappel du projet: Les mouvements browniens fractionnaires multi-échelles sont une généralisation des mouvements browniens fractionnaires et sont tels que le paramètre de Hurst définissant en quelque sorte l'irrégularité de la trajectoire soit dépendant de la fréquence (« l'échelle ») à laquelle on se place. L'étude de tels processus introduits par Benassi et Deguy (1999) a déjà fait l'objet de deux travaux communs (J.-M. Bardet et P. Bertrand) qui sont en voie de publication. La définition de tels processus est obtenue à partir de l'expression dite « harmonisable » du mouvement brownien fractionnaire. Une telle écriture garantit la stationnarité des accroissements du processus, mais limite au cas gaussien. Notre projet de recherche est une généralisation de ce procédé à des processus alpha-stables (on trouve des éléments en vue de cette démarche dans le livre de Samorodnisky et Taqqu, 1994). Un exemple qui se prête naturellement à ce projet est le processus alpha-stable linéaire fractionnaire, pour lequel la mesure brownienne est remplacée par une mesure alphastable dans l'écriture sous forme de moyenne mobile du mouvement brownien fractionnaire. L'étude des propriétés d'autosimilarité et d'irrégularité locale devra être effectuée, ainsi que les propriétés de dépendance des accroissements d'un tel processus. Enfin, la méthode d'analyse par ondelettes pourrait amener à des estimations semi-paramétriques des paramètres du modèle ainsi qu'à un possible test d'adéquation.]

**Bilan:** Si une poursuite des travaux réalisés avec P. Bertrand (qui se sont conclus par 2 publications, Bardet et Bertrand, 2007a et 2007b) a bien eu lieu, l'extension considérée n'a pas été celle des processus alphastables. Nous nous sommes penchés sur l'étude de trajectoires de processus gaussiens à accroissement stationnaires (ou stationnaires) observés suivant des instants aléatoires. La méthode d'analyse par ondelettes permet de s'affranchir de la difficulté d'un échantillon à pas irréguliers, et on peut montrer que la variance empirique des coefficients d'ondelettes vérifie un théorème de la limite centrale. Ce résultat permet la construction d'un estimateur non-paramétrique de la densité spectrale en considérant des

ondelettes dont le support de la transformée de Fourier est compact et symétrique. Une première application a conduit à une première publication (Bardet *et al.* 2008) avec un travail numérique réalisé sur des données de fréquences cardiaques instantanées. Une prépublication (Bardet et Bertrand, 2008) se place dans un cadre plus général et offre de nombreuses simulations et applications numériques sur des données réelles.

- 2007 BARDET, J.-M. et BERTRAND, P., Definition, properties and wavelet analysis of multiscale fractional Brownian motion. *Fractals*, 15, 73-87.
- 2007 BARDET, J.-M. et BERTRAND, P., Identification of the multiscale fractional Brownian motion with biomechanical applications. *Journal of Time Series Analysis*, 28, 1-52.
- 2008 BARDET, J.-M. et BERTRAND, P., A nonparametric estimation of the spectral density of a continuous-time Gaussian Process observed at random times. HAL archives hal-00276735.
- 2008 BARDET, J.-M., BERTRAND, P. et BILLAT, V. (2008). Estimation non-paramétrique de la densité spectrale d'un processus gaussien échantillonné aléatoirement. A paraître dans *Ann. I.S.U.P.*

Mis à part les actions projetées, les travaux suivants ont été réalisés pendant cette période.

**ACTION B.24. Modèles spatiaux et dépendance faible (P. Doukhan, SAMOS et L. Truquet, SAMOS)**

**Bilan:** Dans le domaine des processus spatiaux, de nouvelles recherches ont été également entreprises avec l'arrivée de Paul Doukhan au SAMOS. Ainsi Coupier *et al.* (2006) montrent une application de la dépendance faible aux champs aléatoires, en frontière de la logique, par le biais de lois du 0-1; prouvant ainsi que toute forme de motif se retrouve avec la probabilité 1 dans de grandes images sous des hypothèses de stationnarité et de dépendance adaptées. Doukhan *et al.* (2008) met en oeuvre la machinerie de Liggett pour définir des champs aléatoires; utilisant et prouvant leurs caractéristiques de dépendance faible, nous en déduisons un théorème de limite centrale fonctionnelle. Le résultat est adapté à l'étude de la fiabilité de systèmes multicomposants dépendants. Doukhan et Truquet (2007) construisent de nombreux modèles de champs aléatoires naturellement issus d'équations implicites et nous en déterminons les propriétés de dépendance faible. De plus une notion de causalité adaptée permet de préciser les hypothèses permettant de les construire.

2006            COUPIER, D., DOUKHAN, P. et YCART, B., Zero-one laws for binary random fields. *A.L.E.A.*, 157-175.

2008            DOUKHAN, P., LANG, G., LOUHICHI, S. et YCART, B., A functional central limit theorem for interacting particle systems on transitive graphs. *Markov Processes Relat. Fields.* 14, 79-114.

2007            DOUKHAN, P. et TRUQUET, L., A fixed point approach to model random fields. *A.L.E.A.*, 111-132.

**ACTION B.25. Détection de ruptures (G. Teyssièrè, SAMOS et M. Lavielle, Paris 11)**

**Bilan:** L'arrivée de G. Teyssièrè a été un clair apport sur la thématique de la détection de ruptures. Ainsi, Lavielle et Teyssièrè (2006a et b) ont considéré des méthodes adaptatives pour détecter des ruptures en moyenne et/ou variance de séries chronologiques univariées et multivariées. Ces méthodes permettent de sélectionner de façon assez précise et fiable la dimension d'un modèle et semblent plus fiables que les méthodes paramétriques (telles que le test du rapport de vraisemblance généralisé) et les critères standard de parcimonie (critères d'Akaike et de Bayes).

2006            LAVIELLE, M. et TEYSSIERE, G., Adaptive Detection of Multiple Change-Points in Asset Price Volatility, dans Long-Memory in Economics, G. Teyssièrè and A. Kirman editors, 129-156, Springer Verlag.

2006            LAVIELLE, M. et TEYSSIERE, G., Detection of Multiple Change-Points in Multivariate Time Series, *Lithuanian Mathematical Journal*, 46, 287-306.

**ACTION B.26. Longue mémoire et applications (J.-M. Bardet, SAMOS, P. Doukhan, SAMOS, D. Surgailis, Académie Sciences de Lituanie, et G. Teyssièrè, SAMOS)**

**Bilan:** P. Doukhan a obtenu de nombreux résultats concernant la longue mémoire. En particulier, l'ouvrage «Dependence in Probability and Statistics», avait pour objectif de recueillir des travaux actuels portant sur la dépendance en probabilité et en statistiques, il a suivi l'organisation en 2005 d'un colloque à l'ENSAE. L'article Doukhan *et al.* (2007) a pour objectif d'exhiber des comportements fortement dépendants pour des processus essentiellement non linéaires qui étendent les processus fractionnaires usuels.

2006 BERTAIL, P., DOUKHAN, P. et SOULIER, P., Dependence in Probability and Statistics, Springer.

2007 DOUKHAN, P., LANG, G. et SURGAILIS, D., Limit theorems for sums of non linear function of ARFIMA processes with random Hurst exponents and Gaussian innovations, *Lith. J. of Math.* 47, 1-25.

**ACTION B.27. Comportement asymptotique d'une moyenne de rapports d'accroissements pour des processus stochastiques et applications (J.-M. Bardet, SAMOS, D. Surgailis, Académie Sciences de Lituanie, et G. Teyssière, SAMOS)**

Donatas Surgailis a depuis 3 ans introduit et étudié une statistique basée sur des rapports d'accroissements, permettant d'étudier la régularité, mais aussi la longue mémoire de processus stochastiques. A la différence des statistiques existantes ayant le même objet, cette statistique ne dépend pas de statistiques globales, et est donc non affectées par les non-linéarités et non stationnarités les plus couramment rencontrées dans des séries chronologiques en finance. Ainsi, dans Bardet et Surgailis (2008), on étudie la régularité de processus stochastiques à temps continu à l'aide de cette statistique. Des théorèmes limite sont montrés, en particulier pour des diffusions et des processus gaussiens à accroissements non stationnaires. Pour arriver à de telles fins, un théorème de la limite centrale pour les tableaux triangulaires multidimensionnels gaussiens est montré, ce qui généralise des résultats obtenus par Taqqu (1977) et Arcones (1994).

Surgailis *et al.* (2008) ont également étudié la statistique du rapport des incréments, qui est une statistique «locale» qui permet de tester l'hypothèse nulle que la série étudiée a un paramètre de longue portée fixé  $d$ , contre des hypothèses alternatives fractionnaires. Ce test peut être utilisé pour tester la présence de racine unitaire, de longue portée, de stationnarité, etc.

2008 BARDET, J.-M. et SURGAILIS, D.. Measuring the roughness of random paths by increment ratios. HAL archives hal-00238556.

2008 SURGAILIS, D., TEYSSIERE, G. et VAICIULIS, M. The Increment Ratio Statistic. *Journal of Multivariate Analysis*, 99, 510-541.

**ACTION B.28. Inférence statistique pour les équations fractionnaires (Equations stochastiques dirigées par le mouvement Brownien fractionnaire (T. Sottinen, Tommi Reykjavik University Iceland, C.A. Tudor SAMOS et F. Viens, Purdue University)**

L'étude du mouvement brownien fractionnaire a naturellement mené à une étude statistique et de problèmes d'estimation. Nous avons étudié la question de l'estimation du paramètre de drift dans des modèles fractionnaires en abordant principalement des approches en temps continu basées sur une transformation de Girsanov fractionnaire.

- 2007 TUDOR, C.A. et VIENS, F., Statistical aspects of the Fractional Stochastic Calculus, *The Annals of Statistics*, 25(5), p. 1183-1212.
- 2007 SOTTINEN, T. et TUDOR, C.A. Parameter estimation for stochastic equations with fractional Brownian sheet, *Statistical Inference for Stochastic Processes* (hal-00176226).

**ACTION B.29. Modélisation et Statistique Spatiales (C. Gaetan, Venise et X. Guyon, SAMOS)**

La statistique spatiale connaît un développement important de fait de son utilisation dans de nombreux domaines : environnement, géographie, économie, épidémiologie, sciences de la terre, imagerie, ... Nous avons écrit un livre qui présente les principaux modèles spatiaux et leurs différences avec les séries temporelles. On distingue trois types de données spatiales : les données géostatistiques, les données sur réseau discret et les données ponctuelles. L'objectif visé est de décrire de façon précise, succincte mais complète les modèles classiques (variogramme, auto-régression spatiale, champ de Markov, processus ponctuel), leur statistique et leur simulation par méthode de Monte Carlo. De nombreux exemples utilisant le logiciel R illustrent les sujets abordés. On trouvera un descriptif précis de ce livre dans les sites

<http://www.dst.unive.it/~gaetan/ModStatSpat/>  
et [http://matisse.univ-paris1.fr/fr/article.php3?id\\_article=240](http://matisse.univ-paris1.fr/fr/article.php3?id_article=240).

Ce livre sera publié en anglais chez Springer à l'été 2009.

- 2008 GAETAN, C. et GUYON, X. *Modélisation et Statistique Spatiales*. Springer Heidelberg, Collection: Mathématiques et Applications (SMAI), Vol 63, 308 pages.

**Actions dans des thématiques du nouvel axe C :  
(Faisait partie de l'axe B dans le rapport précédent)  
Probabilités et Processus stochastiques**

Les actions projetées dans cette thématique lors du précédent contrat quadriennal étaient les suivantes (les numéros se réfèrent au contrat 2004-2008) :

- Franchissements par des processus gaussiens (Action B.7),
- Modèles dits de champs moyens (Action B.10),
- Diffusions en milieu aléatoire (Action B.11),
- Formule d'Itô et temps local du drap Brownien fractionnaire de petit paramètre de Hurst et généralisation (Action B.13),
- Sur l'équivalence en loi des draps gaussiens (Action B.14),
- Trajectoires irrégulières géométriques du drap Brownien et application aux EDPS (Action B.15),
- Etude trajectorielle des quasi-helix (Action B.21),
- Caractérisation du support des trajectoires irrégulières géométriques du Brownien fractionnaire (Action B.22),
- Schémas de discrétisation pour des équations d'évolution non linéaires (Action B.23).

Détaillons ci-dessous la réalisation de ces projets.

**ACTION B.7 Franchissements par des processus gaussiens (M. Kratz, SAMOS, J. Léon et W. Urbina, UCV Caracas).**

*[Rappel du projet: Il s'agit de poursuivre avec J. Léon l'étude sur les franchissements par des processus et champs gaussiens, à savoir retravailler les méthodes développées dans nos articles (1997, 2000 et 2001) sur le sujet pour trouver des conditions plus souples sur la fonction de corrélation du processus pour obtenir des théorèmes limites. Cela permettrait peut-être également de comprendre un peu mieux un problème très difficile, toujours ouvert, qui est la question de la condition nécessaire et suffisante pour avoir un second moment fini pour le nombre de franchissements d'un niveau non nul ou d'une courbe différentiable. Une autre étude en cours, menée avec W. Urbina, est celle concernant les franchissements et le temps local pour des solutions d'EDS sans drift par un mouvement brownien fractionnaire. Finalement, il s'agirait de reprendre et clore une étude menée en 2000 (preprint samos127) sur la régularité du nombre de franchissements en un niveau donné par un processus gaussien stationnaire, dans laquelle reste une conjecture.]*

**Bilan:** Le travail réalisé dans le cadre de l'étude de fonctionnelles non linéaires de processus Gaussiens a permis de s'attaquer à différents problèmes (certains figurant dans le rappel du projet, d'autres nouveaux) en vue d'applications diverses.

Le problème de la condition nécessaire et suffisante pour avoir une variance finie du nombre de franchissements d'un niveau non nul ou d'une courbe différentiable par un processus stationnaire Gaussien a été finalement résolu (K. & L., Ann. Probab., 2006).

Par ailleurs, un important travail de synthèse des études faites sur le nombre de dépassements ou franchissements de niveau par un processus Gaussien (moments (factoriels ou non), conditions de finitude de ces moments, méthode de Rice, approximation du temps local par le nombre de franchissements, comportements asymptotiques du nombre de franchissements, et bibliographie exhaustive) a constitué la 1<sup>ère</sup> partie du document de HDR de M. Kratz (2005) et a fait l'objet d'une publication dans *Probability Survey* (2006).

Avec J. Leon, nous nous sommes également intéressés aux problèmes de réflexion et réfraction sur des surfaces aléatoires traités dans les années 60 par l'ingénieur physicien Longuet-Higggin ; nous avons démontré et complété les résultats qu'il avait obtenus de façon heuristique, en utilisant, entre autres, les études faites sur les franchissements de courbes différentiables par des champs gaussiens stationnaires. Un premier travail a fait l'objet d'une prépublication (preprint Samos-226, 2005). Nous avons complété l'étude afin d'obtenir les

comportements limites et avons développé quelques applications, en particulier pour la modélisation de la mer (preprint 2008, hal-00239290).

Une autre étude utilisant l'approche par les franchissements et par la méthode de Rice concerne la modélisation des milieux poreux par un champ seuillé (à un niveau donné) et a été menée avec A. Estrade (Univ. Paris Descartes) et I. Iribarren (UCV, Caracas) dans le cadre du projet « mipomodim » ANR-05-BLAN-0017 (Janvier 2006-Juin 2009). Nous avons considéré un milieu poreux comme bi-phasique, avec une phase solide et une poreuse correspondant aux points de l'espace en lesquels un champ stationnaire fixé est au-dessus ou au-dessous d'un seuil donné, et avons observé les longueurs de cordes traversant ce milieu, technique couramment mise en oeuvre pour modéliser des milieux poreux et ayant une bonne adaptabilité aux données réelles. Les travaux existants sur la question étant jusqu'à présent essentiellement empiriques, il s'agissait de faire une étude analytique pour établir les lois exactes des cordes dans les deux phases, la loi jointe de deux cordes successives, ainsi que la corrélation à deux points. Les résultats obtenus (preprint 2007) ne nécessitent pas l'hypothèse d'indépendance entre les cordes, habituellement rencontrée dans la littérature sur le sujet.

- 2005            KRATZ, M. et LEON, J., Curve crossings and specular points, d'après Longuet-Higgins, *preprint Samos-226 - Map5-2005-18*.
- 2005            KRATZ, M., Etude des comportements extrêmes de processus stochastiques (**HDR**, Univ. Paris I, 2005).
- 2006            KRATZ, M. et LEON, J., On the second moment of the number of crossings by a stationary Gaussian process, *Ann. Probab.* 34.
- 2006            KRATZ, M., Level crossings and other level functionals of stationary Gaussian processes, *Probab. Survey* 3, 230-288..
- 2008            KRATZ, M. et LEON, J., Level curves, crossings and specular points for Gaussian models, preprint hal-00239290 (submitted).
- 2007            ESTRADE, A., KRATZ, M. et IRIBAREN, I., Chord-distribution functions and Rice formulae. Application to random media, preprint MAP5-2007-10 ou hal-00161806.

#### **ACTION B.10. Modèles dits de champs moyens (M. Kratz, SAMOS et P. Picco, CNRS, CPT Luminy).**

*[Rappel du projet : On classe ces modèles de champs moyens selon 2 types: les processus Gaussiens et les modèles d'Hopfield. Dans un article récent avec P. Picco, nous avons étudié la mesure de Gibbs d'un modèle à énergie aléatoire (REM), modèle le plus désordonné car les interactions sont indépendantes. Nous souhaitons traiter ce problème en levant cette condition d'indépendance.*

*Par ailleurs, j'ai été amenée à prendre connaissance (comme membre du jury) des travaux de thèse de M. Atencia, sous la direction de G. Joya (Univ. de Malaga), sur le problème de stabilité des systèmes neuronaux de Hopfield. Nous allons réfléchir sur d'éventuels problèmes permettant de combiner nos 2 approches pour les résoudre.]*

**Bilan:** En collaboration avec M. Atencia et G. Joya (Univ. Malaga), nous avons traité des modèles ou réseaux de Hopfield avec bruit aléatoire que nous utilisons à des fins d'optimisation combinatoire.

Dans sa thèse, M. Atencia a montré que, dans un cadre déterministe, les réseaux de neurones de Hopfield constituent un outil intéressant pour obtenir la solution de problèmes d'optimisation et d'ingénierie de contrôle ; il en a étudié les différentes formulations dont celle d'Abe qu'il a démontré être la plus adaptée à la résolution des problèmes d'optimisation.

Suite à ce travail, nous étudions l'influence du bruit aléatoire dans l'application des réseaux de Hopfield (formulation de Abe) à l'optimisation combinatoire. Dans un premier temps, nous avons défini les conditions pour avoir la convergence de la formulation de Abe, qui s'exprime par une équation différentielle stochastique; le modèle se révèle stable. Nous avons déterminé l'ensemble de convergence des états, en supposant l'intensité du bruit bornée (ESANN 2005). Puis nous avons étudié la relation entre l'ensemble de convergence du modèle déterministe et celui du modèle stochastique (ICANN, 2007).

2007 ATENCIA, M., JOYA, G. et KRATZ, M., Fixed points of the Abe formulation of Stochastic Hopfield Networks ICANN.

2005 ATENCIA, M., JOYA, G. et KRATZ, M., Stochastic analysis of the Abe formulation of Hopfield networks, Proceedings ESANN.

### **ACTION B.11. Diffusions en milieu aléatoire (A. Budhiraja, UNC Chapel Hill, Usa et M. Kratz, SAMOS).**

*[Rappel du projet : L'étude des processus aléatoires en milieux aléatoire est fondamentale pour l'étude des systèmes désordonnés. Récemment, de nombreux auteurs se sont intéressés aux questions en temps discret concernant les marches aléatoires en milieux aléatoire (cf. par exemple Ofer Zeitouni, St. Flour lecture notes on random walks in random environment), pour tenter de bien comprendre leurs comportements asymptotiques, théorèmes ergodiques, Lois des Grands Nombres, principes d'invariance, grandes déviations, etc.... En particulier le problème des grandes déviations a été étudié par Comets, Gantert and Zeitouni (Quenched, annealed and functional large deviations for a one-dimensional random walk in random environment, 2000). Se posent alors ces questions en temps continu, en particulier pour des processus de diffusion en milieu aléatoire. Nous nous proposons d'y réfléchir en rebondissant en particulier sur le travail de M. Taleb (Large deviations for a Brownian motion in a drifted Brownian potential, 2001).]*

**Bilan:** Ce projet n'a pas pu être mené à bien pendant la période du contrat.

### **ACTION B.13 Formule d'Itô et temps local pour le drap Brownien fractionnaire de petit indice de Hurst (C. Tudor, SAMOS et F. Viens, Department of Statistics, Purdue University)**

*[Rappel du projet : Il s'agit de continuer l'étude commencée dans l'article Electronic Journal of Probability 2003, où il a été introduit un calcul stochastique par rapport au drap fractionnaire avec les indices de Hurst supérieurs à  $\frac{1}{2}$ . Nous envisageons de regarder la situation quand les paramètres sont petits et de la généraliser: c.a. d, si le mouvement Brownien fractionnaire a un module de continuité d'ordre " $rH$ ", nous voulons intégrer par rapport à des processus gaussiens ayant un module de continuité d'ordre " $\log(r)$ ". Ceci est une continuation de l'article de O. Mocioalca, F. Viens (preprint 2004).]*

La formule d'Itô pour les draps browniens fractionnaires est maintenant complètement démontrée, quelque soit la valeur des paramètres de Hurst fractionnaires.

2006 TUDOR, C.A, VIENS, F. Ito formula for the two-parameter fractional Brownian motion using the extended divergence integral. *Stochastics*, 78(6), p. 443-462.

### **ACTION B.14. Sur l'équivalence de la loi de draps Gaussiens (T. Sottinen, University of Helsinki et C. Tudor, SAMOS)**

*[Rappel du projet : Nous voulons caractériser les processus gaussiens qui sont équivalents en loi (dans le sens de l'absolue continuité) avec un drap brownien ou avec un drap fractionnaire. Il s'agit de donner des formules et type "Hida-Hitsuda" et de type Shepp [6]. Le cas unidimensionnel a été étudié par Baudouin et Nualart (2003) et par Sottinen (2004).]*

**Bilan:** Notre recherche a abouti à un travail publié qui donne une représentation comme intégrale stochastique pour les processus stochastiques qui sont équivalents en loi avec un drap brownien ou avec un drap brownien fractionnaire. Notre critère d'équivalence est obtenu à l'aide des propriétés des espaces gaussiens associés aux draps brownien et brownien fractionnaire.

2006 SOTTINEN, T. et TUDOR, C. :On the equivalence of multiparameter Gaussian processes, *Journal of Theoretical Probability*, 19(2), p. 461-485.

**ACTION B.15. Trajectoires irrégulières géométriques du drap Brownien (A. Millet SAMOS et M. Sanz-Solé, Universitat de Barcelona)**

*[Rappel du projet : Nous souhaitons proposer une définition possible de trajectoires géométriques pour le drap Brownien comme un processus indexé par deux paramètres jouant des rôles symétriques, et pas par une approche fonctionnelle du type de celle de M. Ledoux, T.J. Lyons, Z. Qian. En effet, une telle définition devrait permettre de mettre en oeuvre toute une étude pour les EDP stochastiques similaire à celle des EDS par rapport au Brownien. Cependant, on sait (Millet et Sanz-Solé, Bally-Millet et Sanz-Solé, ...) que la caractérisation du support des équations paraboliques ou hyperboliques est assez différente de celle des diffusions. En effet, dans le cas hyperbolique (le plus proche de celui des diffusions), il faut jouer sur toute une famille d'interpolées linéaires adaptées non homogènes du drap Brownien. Si les ensembles d'équations contrôlées sont les mêmes, les deux inclusions du théorème du support sont prouvées en utilisant es approximations différentes du drap Brownien. De même, dans Nualart-Sanz la condition d'ellipticité est formulée non pas à l'aide es crochets de Lie, mais des dérivées covariantes des coefficients e diffusion. Cette façon d'écrire des éléments à l'aide de produits d'exponentielles est à la base de la définition de norme proposée dans Friz-Victoir et devrait également permettre une définition des trajectoires géométriques. Le cas des équations paraboliques est pire car c'est une suite divergente de coefficients qui apparaît dans la suite des équations contrôlées approximante. Il est clair que c'est une formule de Taylor stochastique, du genre formule d'Itô montrée par Hajek, qui devrait fournir la bonne notion de trajectoire multiplicative, mélangeant des accroissements rectangulaires et des intégrales doubles sur des rectangles emboîtés et sur des rectangles "non-comparables".]*

**Bilan:** Ce projet n'a pas été réalisé dans la forme prévue. Nous avons en effet rencontré des difficultés techniques pour définir une « bonne » notion de « rough path » au-dessus d'un processus à deux paramètres tel qu'un drap brownien. Cependant, l'action B.22 centrée sur les trajectoires rugueuses au-dessus du brownien fractionnaire d'indice de Hurst  $H$  tel que  $1/4 < H < 1/2$  a été faite.

**ACTION B.21. Etude trajectorielle des quasi-helix (F. Russo, LAGA Paris 13 et C. Tudor, SAMOS)**

*[Rappel du projet : Nous voulons étudier un processus Gaussien de covariance  $R(s,t) = 2^{-K} [(t^{2H} + s^{2H})^K - |t-s|^{2HK}]$  où  $H \in (0,1)$ , et  $K \in (0;1]$ . Ce processus et un quasi-helix dans le sens de J.P. Kahane *Adv. Math*, 1981 et il a été introduit dans C. Houdré and J. Villa (2003) *Contemporary Mathematics*. Si  $K = 1$ , on retrouve le brownien fractionnaire standard. Nous envisageons d'intégrer au sens trajectorielle par rapport à ce processus.]*

**Bilan:** Nous avons abouti à une série des résultats intéressants pour les mouvements browniens bi-fractionnaires : étude des propriétés de base, introduction du calcul stochastique par rapport à ces processus (calcul stochastique de type « Malliavin » ou bien de type « trajectorielle »). L'étude de ces processus a été ensuite reprise par plusieurs chercheurs de différents pays. Les participants à cette action ont finalement été : C. Tudor et K. Es-Sebaiy (SAMOS), I. Kruk et Francesco Russo (Paris 13) et Y. Xiao (University of Michigan). Enfin une partie importante de la thèse de K. Es-Sebaiy est consacrée aux temps locaux et aux formules de Tanaka pour les processus fractionnaires (par exemple le mouvement brownien bifractionnaire et les processus gaussiens avec drift.

- 2006 RUSSO, F. et TUDOR, C., On the bifractional Brownian motion. *Stochastic Processes and their Applications* 116(5), p. 830-856.
- 2007 KRUK, I., RUSSO, F. et TUDOR, C., Wiener integrals, Malliavin calculus and covariance measure structure. *Journal of Functional Analysis*, 249, p. 92-142.
- 2007 TUDOR, C. et XIAO, Y. , Sample path properties of the bifractional Brownian motion. *Bernoulli* 13(4), p. 1023-1052.
- 2007 ES-SEBAI, K. et TUDOR, C., Multidimensional Ito and Tanaka formula for the bifractional Brownian motion, *Stochastics and Dynamics* 3, p. 365-388.

### **ACTION B.22. Support des trajectoires irrégulières géométriques du Brownien fractionnaire (A. Millet, SAMOS et M. Sanz-Solé, Universitat de Barcelona)**

[Rappel du projet : Dans le cadre du mouvement Brownien  $d$ -dimensionnel, une topologie légèrement différente de celle de la  $p$ -variation, a permis à Friz et Victoir de retrouver les résultats usuels sur la caractérisation du support et les grandes déviations des diffusions dans la topologie usuelle de la norme uniforme ou des fonctions Höldériennes d'ordre  $\alpha \in ]0, 1/2[$ .

Nous avons commencé à étudier le mouvement Brownien géométrique  $W^H$  au-dessus du mouvement Brownien fractionnaire  $W^H$  d'indice de Hurst  $H$ . Dans le travail de L. Coutin et Z. Qian, l'approximation choisie consiste à utiliser les interpolées linéaires du Brownien fractionnaire  $W^H$ . A l'ordre 1 ces approximations n'appartiennent pas à l'espace auto-reproduisant de processus Gaussien  $W^H$ , ce qui pose problème pour utiliser un théorème de Girsanov même au premier ordre. Dans le but de transposer la technique de caractérisation du support développée par A. Millet et M. Sanz-Solé et d'utiliser une topologie du type de celle de Friz et Victoir, nous avons introduit des approximations  $W^H(n)$  de  $W^H$ , différentes et appartenant à l'espace auto-reproduisant de  $W^H$  et avons montré la convergence de  $W^H(n)$  vers  $W^H$  dans la norme de la  $p$ -variation pour  $pH > 1$ . Dans le cas  $1/4 < H < 1/2$ , ces approximations permettent de définir des intégrales stochastiques itérées par rapport à  $W^H(n)$ . Nous avons établi la convergence de cette suite vers le Brownien fractionnaire au premier niveau dans la topologie de la  $p$ -variation pour  $pH > 1$ . Suivant les valeurs de  $H$  par rapport à  $1/3$ , qui déterminent la régularité des trajectoires des processus, il nous reste à établir la convergence des intégrales stochastiques doubles (et triples) vers les composantes correspondante du Brownien fractionnaire géométrique  $W^H$  construit au-dessus de  $W^H$ . Ceci fournira directement la caractérisation du support trajectoires de  $W^H$  dans une topologie de la  $p$ -variation (et dans une topologie plus naturelle de norme Höldérienne). Nous souhaitons ensuite définir une application d'Itô étendant l'intégrale stochastique par rapport au mouvement Brownien fractionnaire et, de nouveau, le cas  $H < 1/2$  risque d'être beaucoup plus délicat, ne serait-ce que par le type d'intégrale stochastique qu'il faudra considérer.]

**Bilan:** Nous avons introduit des approximations qui permettent de caractériser le support de la loi du brownien fractionnaire. Signalons que les approximations de type Stooch-Varadhan que nous avons proposées, à l'aide des rough paths au-dessus des éléments de l'espace auto-reproduisant, ne sont pas des interpolées linéaires, contrairement à l'approche classique. Elles sont plus proches des travaux de D. Feyel et A. de la Pradelle et permettent de trouver la même caractérisation du support de la loi des trajectoires rugueuses au-dessus du brownien fractionnaire que celle obtenue indépendamment par L. Coutin, P. Friz et N. Victoir. Les démonstrations reposent sur du calcul stochastique anticipatif et des estimations précises du noyau permettant une représentation intégrale de  $W^H$  à l'aide du brownien. Nous avons également prouvé un principe de grandes déviations pour les rough paths au-dessus du mouvement Brownien fractionnaire d'indice de Hurst  $H > 1/4$ , qui s'étend ensuite aux diffusions dirigées par ce processus grâce au théorème de continuité universelle de T. Lyons.

Ce travail complète l'article de M. Ledoux, Z. Qian et T. Zhang pour les trajectoires rugueuses au-dessus du mouvement Brownien.

2006 MILLET, A. et SANZ-SOLE, M., Large deviations for rough paths of the fractional Brownian motion, *Annales de l'Institut Henri Poincaré (B), Probabilités et Statistiques* 42-2, p. 245-271.

2008 MILLET, A. et SANZ-SOLE, M. Approximations and support for the rough paths of the fractional Brownian motion, *Seminar on Stochastic Analysis, Random Fields and Applications V, Series Progresses in Probability* 59, p. 275-304.

### **ACTION B.23. Schémas de discrétisation pour des équations d'évolution non linéaires (I. Gyöngy, University of Edimburgh et A. Millet, SAMOS)**

[Rappel du projet : Dans un article écrit en collaboration avec I. Gyöngy lors d'un séjour qu'il a fait à Paris en mai 2003, nous avons montré la convergence de schémas de discrétisation explicite et implicite de l'équation d'évolution

$$u_t = u_0 + \int_0^t A_s(u_s) ds + \sum_{j \leq r} \int_0^t B_s^j(u_s) dW_s^j,$$

où  $V$  est un espace de Banach,  $H$  est un espace de Hilbert tels que  $V(H \subset V^*)$ . Les coefficients  $A: [0, T] \times V \times \Omega \rightarrow V^*$  et  $B: [0, T] \times V \times \Omega \rightarrow H^r$  satisfont des conditions de mesurabilité, de monotonie et d'hémicontinuité, ainsi que de coercivité et de restriction sur la croissance qui peuvent dépendre de  $t$ .

Nous avons montré qu'un schéma implicite de discrétisation espace-temps  $u^{n,m}$  basé sur une méthode d'éléments finis linéaires existe et que, lorsque  $m, n$  tend vers  $+\infty$ ,  $u^{n,m}$  converge faiblement dans  $L^p_V(\lambda)$  vers la solution  $u$  de l'équation d'évolution, où  $\lambda(t)$  est un poids qui apparaît dans la formulation de la coercivité et de la restriction sur la croissance. De plus, à un instant  $T > 0$  fixé, les variables aléatoires  $u^{n,m}(T)$  convergent fortement vers  $u_T$  dans  $L^2_H$ . La convergence d'un schéma implicite en temps (ainsi que celle d'un schéma explicite espace temps) est également prouvée ; la convergence du schéma explicite demande une condition supplémentaire sur les rapports entre discrétisation spatiale et temporelle.

Contrairement au cas de schémas d'approximations pour des équations paraboliques semi-linéaires étudiées dans par Gyöngy, Millet-Morien, Printemps, ... ou à dérive polynomiale (telle que l'équation de Cahn-Hilliard étudiée par Cardon-Weber), aucun résultat de vitesse de convergence n'est connu pour les schémas d'approximation de ces équations non-linéaires. Nous souhaiterions, au moins dans un cas particulier d'opérateur  $A$ , obtenir une vitesse de convergence forte pour les divers schémas étudiés. Nous pensons nous inspirer d'un article de Krylov qui établit une vitesse de convergence dans le contexte parabolique purement non-linéaire de l'équation de Bellman.]

**Bilan:** Après l'étude théorique sur la convergence de schémas dans un cadre fortement non-linéaire, nous avons montré que dans le cas  $p=2$ , lorsque la solution a la propriété de régularité Höldérienne  $E\|u_t - u_s\|_V^2 \leq (t-s)^{2\nu}$ , sous des hypothèses de monotonie forte du couple  $(A, B)$ , de Lipschitz sur les opérateurs  $B(t, \cdot)$  et de Lipschitz sur les opérateurs  $B(\cdot, x)$ , le schéma implicite en temps converge vers la solution dans  $L^2(\Omega, X)$ , où  $X = C([0, T]; H) \cap L^2([0, T]; V)$ , avec la vitesse de convergence  $\nu$  ou  $\nu/2$  suivant que  $t \rightarrow A(t, \cdot)$  a ou non la propriété de Lipschitz. Nous avons ensuite complété l'étude précédente en temps et en espace-temps pour des schémas explicites ou implicites avec des conditions générales de compatibilité entre les opérateurs définissant les processus approximant et limite. Les hypothèses faites englobent les approximations par ondelettes ou éléments finis. Des conditions suffisantes pour des équations quasi-linéaires donnent des exemples d'opérateurs satisfaisant les conditions abstraites lorsque  $H = L^2(\mathbb{R}^d)$  et  $V = H^1(\mathbb{R}^d)$  (resp.  $H = H^r(\mathbb{R}^d)$  et  $V = H^{r+1}(\mathbb{R}^d)$  dans un cas linéaire).

2005 GYONGY, I. et MILLET, A., On discretization schemes for stochastic evolution equations, *Potential Analysis* 23-2, p.99-134.

2007 GYONGY, I. et MILLET, A., Rate of convergence of Implicit Approximations for stochastic evolution equations, *Stochastic Differential Equations: Theory and Applications* (A volume in the honor of B. Rosovskii), *World Scientific Interdisciplinary Sciences* Vol. 2, p. 281-310.

2008, GYONGY, I. et MILLET, A., Rate of Convergence of Space-Time Approximations for stochastic evolution equations, *à paraître dans Potential Analysis*, arXiv:0706.1404v1 et serveur hal-00153372.

<b>Autres travaux pendant la période 2005-2008</b>
--

Mis à part les actions projetées, les travaux suivants ont été réalisés pendant cette période. Ils ont comme thème l'analyse stochastique par rapport au fbm, le calcul de Malliavin, l'étude d'équations d'évolution en dimension infinie. Un autre thème a été l'analyse stochastique d'EDP non linéaires de l'hydro-dynamique en dimension 2 perturbées par un bruit gaussien multiplicatif.

**ACTION B.30. Etude des variations des processus stochastiques auto-similaires, théorèmes limites et calcul de Malliavin (X. Bardina et M. Jolis, Universitat Autònoma de Barcelona, I. Nourdin, Université Paris 6 LPMA, D. Nualart, University of Kansas, C. Tudor, SAMOS et F. Viens, Purdue University)**

Le travail de Peccati et Nualart (2004) « Central limit theorems for multiple Wiener-Itô integrals », qui donne un critère de convergence vers la loi normale pour les variables aléatoires vivant dans un chaos de Wiener d'ordre fixé, a déclenché un important flux de travaux portant sur l'application de ces résultats. Nous avons obtenu d'abord en collaboration avec G. Peccati un critère de convergence multidimensionnel. Ensuite, avec F. Viens, I. Nourdin ou D. Nualart, nous avons obtenu des théorèmes limites pour les variations d'ordre supérieur du mouvement brownien fractionnaire et de certains processus associés.

Nous avons également étudié des intégrales stochastiques multiples par rapport au fbm, ainsi que la convergence d'intégrales multiples.

2007 BARDINA, X. et TUDOR, C., The law of stochastic integrals with two independent fractional Brownian motions, *Boletín de la Sociedad Matemática Mexicana* 13(1).

2008 NOURDIN, I., NUALART, D. et TUDOR, C., Central and Non-Central limit theorems for weighted power variations of fractional Brownian motion, *soumis pour publication*, arXiv:0710.5639 et hal-00184057.

2008 TUDOR, C. et VIENS, F., Variations and estimates for the self-similarity order through Malliavin calculus, *soumis pour publication*, arXiv:0709.3896 et hal-00175730.

2008 BARDINA, X., JOLIS, M. et TUDOR, C., On the convergence of multiple Wiener-Itô integrals, arXiv:0712-3837 et hal-00200914.

**ACTION B.31. Equations stochastiques dirigées par le mouvement Brownien fractionnaire (R. Balan University of Ottawa, N. Eisenbaum et I. Nourdin, LPMA)**

**Université Paris 6, S. Tindel, Université de Nancy 1, C. Tudor, SAMOS et F. Viens, Purdue University)**

Le calcul stochastique par rapport au mouvement brownien fractionnaire a été suivi naturellement par l'étude des équations stochastiques ayant ce processus comme bruit. Nous avons traité des aspects variés, comme les propriétés d'une équation avec un bruit multiplicatif dirigé par un drap fractionnaire, ou le cas de la dimension infinie.

Dans le cas du mouvement brownien fractionnaire infini dimensionnel, nous avons prouvé des résultats (existence, unicité ou régularité de la solution) dans le cas du bruit linéaire additif.

2005 BOUFFOUSSI, B. et TUDOR, C., Kramers-Smolucchowsk approximation for stochastic equations with fbm. *Revue Roumaine de Mathématiques Pures et Appliquées*, 50 (2).

2006 NOURDIN, I. et TUDOR, C., Some linear stochastic fractional equations, *Stochastics and Stochastics Reports*, 78(2), p.51-65.

2007 BALAN, R. et TUDOR, C., The stochastic heat equation with a fractional colored noise: existence of the solution, *ALEA* (à paraître), arXiv mathPR/0703088 et hal-00134611.

**ACTION B.32. Analyse stochastique pour des processus fractionnaires généralisés (K. Bertin et S. Torres, Université de Valparaiso, Chili, K. Es-Sabaïy et C. Tudor SAMOS, M. Maejima, Keio University, Japon, D. Nualart, University of Kansas, Y. Ouknine)**

Cette partie contient quelques travaux ayant comme but le développement d'un calcul stochastique par rapport a des processus fractionnaires, autres que le mouvement brownien fractionnaire. Ce sont en général des extensions de celui-ci (par exemple le brownien fractionnaire infini dimensionnel ou les processus de Rosenblatt). Le processus de Rosenblatt a été défini dans les années 70-80 dans le contexte du Théorème de la Limite Non Centrale : en effet certaines fonctionnelles quadratiques du brownien fractionnaire convergent vers ce processus. Comme le processus de Rosenblatt est auto similaire et de mémoire longue, il a suscité de l'intérêt pour des applications diverses. Nous avons également regardé les processus de Bessel fractionnaires, les processus gaussiens avec dérive ou les processus d'Hermite.

2007 MAEJIMA, M. et TUDOR, C., Wiener integrals and Non Central limit theorem for Hermite processes, *Stochastic Analysis and Applications* 25(5), p. 1043-1056.

2008 TUDOR, C., Analysis of the Rosenblatt process, *ESAIM – Probability and Statistics*, 12, p. 230-257.

2008 TORRES, S. et TUDOR, C., Donsker theorem for the Rosenblatt process and a binary market model, arXiv:mathPR/0703085 et hal-00134608.

2008 BERTIN, K., TORRES, S. et TUDOR, C., Maximum likelihood estimators and random walks in long memory models, arXiv:0711.0513 et hal-00184842.

2008 ES-SEBAIY, K., NUALART, D., OUKNINE, Y. et TUDOR, C., Occupation densities for certain processes related to fractional Brownian motion, soumis pour publication, arXiv:0801.3314 et hal-00211827.

**ACTION B.33. Intégrales de Skorohod et martingales (K. Es-Selaiy et C.A. Tudor SAMOS, M. Thieullen et G. Peccati, Université Paris 6)**

Cette action comprend les travaux liés à l'intégrale de Skorohod sur les espaces de Wiener, Poisson ou Lévy. Les travaux listés ici étudient la relation entre cette intégrale de Skorohod qui est anticipante, c'est à dire que les intégrands ne sont pas forcément adaptés, et les martingales.

2006 PECCATI, G., THIEULLEN, M., TUDOR, C.. (2006), Martingale structure for Skorohod integral processes, *The Annals of Probability* 34(3), p. 1217-1239.

2007 PECCATI, G., TUDOR, C.A. (2007) Anticipating integrals and martingales on the Poisson space, *Random Operators and Stochastic Equations* 15(4), p. 327-352.

2008 ES-SABAIY, K., TUDOR, C.A., Lévy processes and Itô-Skorohod integrals, proceedings of the Skorohod Space Conference, Kyiv 2007, *Theory of Stochastic Processes*, à paraître.

**ACTION B 34: Analyse stochastique des équations de l'hydrodynamique perturbées par un bruit gaussien multiplicatif. (I. Chueshov, Kharkov National University, J. Duan, Illinois Institute of Technology, A. Millet SAMOS)**

Nous étudions diverses équations de l'hydrodynamique, telle que l'équation de Navier-Stokes couplée avec d'autres EDP décrivant l'évolution de la température ou d'un champ magnétique. Ces modèles sont utilisés pour modéliser la géophysique (plasma) ou l'océan, ainsi qu'en climatologie. Nous montrons que, sous des conditions de parabolicité stochastique, avec un bruit gaussien dont l'opérateur de covariance est à trace, ces solutions admettent une solution unique dans  $X=C([0,T];H) \cap L^2([0,T];V)$  lorsque la solution initiale a des moments d'ordre 4 dans H. Nous établissons des estimées a priori sur la solution. Sous des conditions plus restrictives sur le coefficient de diffusion, nous établissons un principe de grandes déviations dans X pour des solutions soumises à une petite perturbation. La méthode utilisée est basée sur la convergence faible de familles de solutions d'équations contrôlées stochastiques, c'est à dire telles que le bruit gaussien est translaté par un élément aléatoire de son espace auto-reproduisant.

2008 DUAN, J. et MILLET, A., Large deviations for the stochastic Boussinesq equation under random influence, à paraître dans *Stochastic Processes and their Applications*, arXiv :0802.1335 et hal-00250132 (prépublication de l'Institut Mittag Leffler).

2008 CHUESHOV, I. et MILLET, A, Stochastic 2D hydrodynamical type systems : well posedness and large deviations, soumis pour publication, arXiv 0807.1810 et hal-00295023.

**ACTION B 35 : Approximation des solutions ED backward doublement stochastiques (O. Aboura SAMOS)**

Dans sa thèse, O. Aboura étudie la vitesse de convergence de schémas de discrétisation de solutions d'ED backward doublement stochastiques introduites par Pardoux et Peng. La méthode est inspirée par les travaux de Zhang et repose sur une propriété de  $L^2$  régularité, mais les coefficients sont plus généraux et incluent le gradient de la solution. Les processus considérés sont fini-dimensionnels.

2008 ABOURA, O., Discretization of Backward doubly stochastic Differential Equations, *soumis pour publication*.

## 2 PRINCIPAUX RÉSEAUX FRANÇAIS, EUROPÉENS ET INTERNATIONAUX depuis 2005

Le SAMOS ou des membres du SAMOS participent aux réseaux et programmes internationaux suivants :

Intitulé	Nombre d'institutions partenaires	Intitulé du laboratoire coordinateur	Etablissement de rattachement du laboratoire coordinateur	Pays du laboratoire coordinateur	Nom et prénom du responsable scientifique dans l'unité	H/F	C pour Coordinateur ou P pour Partenaire	Date de début	Date de fin, le cas échéant
Coopération avec la Havane	2	CES	Université Paris 1	France	Cottrell Marie	F	C	1993	
STAFAV	4	Département Mathématique	Université Paris 11	France	Bardet Jean-Marc	H	P	2004	
BMF 2003-01345	2	Universitat de Barcelona	Universitat de Barcelona	Espagne	Millet Annie	F	P	2004	2008
PAI Picasso 07130NM	4	CES	Université Paris 1	France	Millet Annie	F	C	2004	2005
ISC-PIF« Institut des systèmes complexes Paris Ile-de-France »	19			France	Cottrell Marie	F	P	2005	
Semestre SPDE		Mittag-Leffler	Royal Swedish Academy of Sciences, Stockholm,	Suède	Millet Annie	F	P	2007	2008
CNRS/CONICYT 2008	2	CES	Université Paris 1	France	Tudor Ciprian	H	C	2008	2009
Coopération avec le Burkina-Faso	2	CES	Université Paris 1	France	Guyon Xavier	H	P	2008	
PAI Picasso (déposé avril 2008)	2	CES	Université Paris 1	France	Millet Annie	F	C	2009	2010
PAI Tassili (Déposé avril 2008)	2	CES	Université Paris 1	France	Cottrell Marie	F	C	2009	
AMSUD	5	Labo de Maths	Université Paris 11	France	Tudor Ciprian	H	P	2009	
Programme PSPD		Isaac Newton Institute	Isaac Newton Institute	Grande-Bretagne	Millet Annie	F	P	2010	

## II.3 LES ACTIVITES D'ANIMATION SCIENTIFIQUE 2005-2008

### II.3.1 Les colloques et conférences internationales organisées par des membres du SAMOS

Treize conférences et quatre sessions invitées ont été organisées ou co-organisées par des membres du SAMOS au cours de la dernière période quadriennale.

2008

Conférence *MASHS (Computational Methods for Modelling and leArning in Social and Human Sciences)*, 5-6 juin 2008, Créteil, organisée par le SAMOS et l'équipe ERUDITE de l'Université de Paris 12, 50 participants.

Responsables: **Patrice Gaubert et Madalina Olteanu.**

2008

Workshop à l'Université Galatasaray, Istanbul, dans le cadre d'une collaboration entre le CEE, l'Université Galatasaray (Istanbul) et le CEPN, Université Paris 12, « *Emploi et protection sociale : la Turquie parmi les pays aux frontières de l'Europe* », 31 mars-1<sup>er</sup> avril 2008.

Responsable : **Corinne Perraudin**

2008

Conférence « OR8, Havana 8<sup>th</sup> International Conference on Operations Research », organisée par le SAMOS en collaboration avec l'Université de la Havane et l'Université Humbolt (Allemagne), organisation d'une session de cinq exposés, 25-29 février 2008, 100 participants.

Responsable: **Marie Cottrell**

2008

Conférence « *Limit Theorems and Applications* », 14-16 janvier 2008, Paris (Centre Pierre Mendès France), 22 exposés, 80 participants

Responsables : **Jean-Marc Bardet et Ciprian Tudor**

2007

Conférence « *Stochastics Dynamics* », 11-12 juin 2007, Paris (Centre Pierre Mendès France), 13 exposés, 40 participants.

Responsables : **Annie Millet et Ciprian Tudor**

2007

Conférence *IWANN 2007(International Workshop on Artificial Neural Networks)*, juin 2007, à San Sebastien (Espagne). Organisation de deux sessions invitées : Analyse de données complexes.

Responsable : **Marie Cottrell**

2007

Workshop on « *Spatial Statistics* », 27 Avril 2007, Paris (Centre Pierre Mendès France), 6 invités étrangers, 70 participants

Responsable : **Cécile Hardouin**

2006

Congrès Bernoulli « *Stochastic Processes and their Applications* », organisation d'une session sur « *Stochastic Analysis for fractional processes* », 3 exposés, 20 participants

Responsable : **Ciprian Tudor**

2006

Journées MAS à Lille, septembre 2006, organisation d'une session sur « *EDP Stochastiques* », 5 exposés, 20 participants

Responsable: **Annie Millet**

2006

Organisation locale des deuxièmes journées du réseau MSTGA (Modélisation Spatio-Temporelle sur Graphe et Approximation, réseau du département Mathématiques et Informatique Appliquées de l'INRA), à l'Université Paris 1, Novembre 2006, 20 participants.

Responsable : **Cécile Hardouin**

2005

Conférence *WSOM'05* à Paris 1 Panthéon-Sorbonne en septembre 2005. Cette conférence a connu un grand succès avec plus de 140 participants. Une sélection des meilleurs papiers a été publiée (après un deuxième processus de referee par deux lecteurs) dans un numéro spécial de la revue *Neural Networks* édité par Marie Cottrell et Michel Verleysen.

Responsable: **Marie Cottrell**

2005

Conférence *ASMDA (Applied Stochastic Models and Data Analysis)*, mai 2005, à Brest, organisation d'une session invitée.

Responsable : **Marie Cottrell.**

2005

Conférence *STATDEP*, juin 2005 à ENSAE (co-organisé par l'ENSAE et le SAMOS)

Responsable : **Paul Doukhan**

2005-2006-2007-2008

Conférence « *ESANN* » à Bruges (Belgique). **Marie Cottrell** est membre du comité directeur (« steering committee »), depuis 1993.

**A venir :**

2008

Journées MAS à Rennes, août 2008, organisation d'une session invitée sur « *Statistique spatiale et spatio-temporelle* ».

Responsable: **Cécile Hardouin**

2008

Congrès Franco-Roumain de Mathématiques Appliquées, à Brasov (Roumanie), août 2008. Organisation d'une session sur les « *Processus stochastiques* ».

Responsable: **Ciprian Tudor**

### **3.2 Appartenance à des comités scientifiques de conférences**

#### **Marie Cottrell**

- Comités scientifiques de congrès récurrents : IWANN (International Workshop on Artificial Neural Networks) en Espagne tous les deux ans depuis 1991, WSOM tous les deux ans depuis 1997.
- Depuis 2000, comité scientifique et comité d'organisation des Conférences sur la Recherche Opérationnelle, qui se tiennent tous les deux ans à la Havane
- Membre du comité scientifique de la **conférence IJCNN 2005** qui s'est tenue à Montréal en juillet 2005.
- De 1997 à 2005, membre du comité scientifique des **rencontres internationales « Approches Connexionnistes en Sciences Economiques et de Gestion »**, qui se tenaient tous les ans en France, Espagne, Belgique.
- Depuis 2006, membre du comité scientifique de la **conférence MASHS** (Computational Methods for Modelling and leArning in Social and Human Sciences) qui lui a succédé. La dernière édition a eu lieu à Paris 12, en juin 2008 et a été co-organisée par le SAMOS et l'équipe ERUDITE de Paris 12.

#### **Patrice Gaubert**

- Depuis 2006 membre du comité scientifique de la **conférence MASHS** (Computational Methods for Modelling and leArning in Social and Human Sciences)
- Depuis 2000, comité scientifique et comité d'organisation des Conférences sur la Recherche Opérationnelle, qui se tiennent tous les deux ans à la Havane

#### **Patrick Letrémy**

- Depuis 2006 membre du comité scientifique de la **conférence MASHS** (Computational Methods for Modelling and leArning in Social and Human Sciences)

#### **Madalina Olteanu**

- Depuis 2008 membre du comité scientifique de la **conférence MASHS** (Computational Methods for Modelling and leArning in Social and Human Sciences)

### 3.3 Le séminaire du SAMOS 2005-2008

**Le Séminaire Samos « Probabilités, Statistiques et Réseaux de Neurones » est bimensuel.  
Responsables : Jean-Marc Bardet, Marie Cottrell, Annie Millet.**

**Programme du séminaire :**

**2005:**

- **Déc. 09, 11h00** : Régularité de convolutions stochastiques, *Szymon Peszat (Université de Cracovie et Paris 13)*
- **Nov. 25, 11h00** : Reconstruction des dynamiques des systèmes complexes, *Paul Bourgine (CREA, Ecole Polytechnique)*
- **Juin 17, 10h00** : Matinée spéciale de Calcul Stochastique au Samos, organisée par Ciprian Tudor,
- **Juin 10, 10h00** : Solutions of SDEs as saddle points of mini-max problems, *Istvan Gyongy (Edinburgh University, Grande Bretagne)*
- **Juin 03, 10h00** : Testing covariance stationarity for high frequencies panel data : a nonparametric approach based on the evolutionary spectral density, *Ibrahim Ahamada (Eurequa, Université Paris 1)*
- **Mai 27, 11h15** : Change-point detection in GARCH models : asymptotic and bootstrap tests, *Gilles Teyssière (Samos, Université Paris 1)*
- **Mai 27, 10h00** : Artificial Neural Network and Semiparametric Long-Memory ARCH, *Gilles Teyssière (Samos, Université Paris 1)*
- **Mai 20, 10h00** : Classification de courbes par apprentissage, *Olivier Winterberger (Samos, Université Paris 1)*
- **Avril 22, 10h00** : Connexité et analyse des données non linéaires, *Catherine Aaron (Samos, Université Paris 1)*
- **Avril 18, 10h00** : Recursive Self-Organizing Maps for Structures, *Barbara Hammer (Technical University of Clausthal, Allemagne)*
- **Avril 15, 10h00** : Dépendance faible : coefficients et applications statistiques, *Clémentine Prieur (INSA, Université Toulouse III)*
- **Avril 08, 10h00** : Sur le mouvement brownien bi-fractionnaire, *Ciprian Tudor (Samos, Université Paris 1)*
- **Mars 25, 11h15** : Un point de vue unifié pour les techniques de représentation euclidienne, *Elina Miret Barroso, Universidad de La Habana (Cuba)*
- **Mars 25, 10h00** : Rule Extraction from Prototype-based Classifiers, *Barbara Hammer (Technical University of Clausthal, Allemagne)*
- **Mars 11, 10h00** : Supervised Recursive Networks, *Barbara Hammer (Technical University of Clausthal, Allemagne)*
- **Mars 04, 10h00** : Relevance Determination in Learning Vector Quantization, *Barbara Hammer (Technical University of Clausthal, Allemagne)*
- **Fév. 18, 10h00** : Bornes et développements asymptotiques de la loi du maximum d'un champs aléatoire, *Jean-Marc Azaïs (Université Toulouse 3, Paul Sabatier)*
- **Fév. 04, 10h00** : Réseaux neuronaux et SVM à entrées fonctionnelles : une approche par régression inverse, *Nathalie Villa-Vialaneix (Université Toulouse 2, Le Mirail)*
- **Jan. 21, 11h00** : Sur les propriétés de beta-mélange des modèles hybrides à chaîne de Markov cachée, *Madalina Olteanu (Université Paris 1)*

**2006 :**

- **Déc. 15, 11h00** : Processus de naissance et de mort sur sur certains arbres aléatoires, *Jean-Marc Lasgouttes (INRIA Roquencourt), jean-marc.lasgouttes@inria.fr*
- **Déc. 01, 11h00** : Spatio-temporal modelling of epidemiological processes, *Carlo Gaetan (Université de Venizia, Italie), gaetan@unive.it*
- **Nov. 23, 00h00** : Journées Modélisation Spatio-Temporelle sur Graphe et Approximation,
- **Nov. 17, 11h00** : Rythmes complexes dans les systèmes multi-échelles, *Jean-Pierre Françoise (Université Paris 6)*
- **Nov. 10, 11h00** : Estimation adaptative de la densité pour des observations dépendantes, *Olivier Wintenberger (Université Paris 1), Olivier.Wintenberger@univ-paris1.fr*
- **Oct. 20, 11h00** : Approche MDL pour les chaînes de Markov cachées à émission gaussienne ou poissonnienne. Application à l'identification d'ordre., *Aurélien Garivier (Université Paris 11), Aurelien.Garivier@math.u-psud.fr*
- **Oct 13, 11h00** : Boosting Itéré pour la Detection de Donnees Aberrantes, *Jean-Michel Poggi (Université Paris 11-Orsay et Paris 5)*
- **Juin 23, 11h00** : Représentation temps-échelle pour la surveillance préventive du système du guidage d'un tramway sur pneumatiques « le Translohr », *Zahra Hamou Mamar (LIMOS, Université Clermont-Ferrand II)*
- **Juin 16, 11h00** : Cauchy problems with periodic controls, *Istvan Gyongy (Edinburgh University, Grande Bretagne)*
- **Juin 09, 11h00** : Dépendance faible, modèles et quelques utilisations, *Paul Doukhan (CREST et Samos-Matisse)*
- **Juin 02, 11h00** : Identification de l'architecture de perceptrons multicouches, *Joseph Rynkiewicz (Université Paris 1, Samos-Matisse)*
- **Mai 19, 11h00** : Distributions aléatoires en analyse d'images (Random distributions in image analysis), *Richard Emilion (Université d'Orléans)*
- **Avril 07, 12h00** : Dynamique de réseaux complexes : approximation de champ moyen et équations de corrélations, *Annick Lesne (LPTMC, Université Paris 6)*
- **Mars 31, 12h00** : Time-varying fractionnally integrated processes with discrete and continuous argument, *Donatas SURGAILIS (membre de l'Académie des Sciences de Lituanie)*
- **Mars 24, 12h00** : Random coefficient AR(1) process with heavy-tailed renewal-switching coefficient and heavy-tailed noise, *Donatas SURGAILIS (membre de l'Académie des Sciences de Lituanie)*
- **Mars 17, 11h00** : Information optimum vector quantization, *Thomas Villmann (Université de Leipzig)*
- **Mars 10, 11h00** : Prototype based fuzzy classification, *Thomas Villmann (Université de Leipzig)*
- **Mars 03, 12h00** : Application of topology preserving mapping using SOMs for medical data analysis, *Thomas Villmann (Université de Leipzig)*
- **Fév. 24, 11h00** : Dynamique de réseaux complexes : approximation de champ moyen et équations de corrélations, *Annick Lesne (LPTMC, Université Paris 6)*
- **Jan. 06, 11h00** : La parole naissant des interactions perceptuo-motrices : cadre théorique, données expérimentales et éléments de modélisation computationnelle, *Jean-Luc Schwartz (CNRS, I.N.P. Grenoble)*

2007 :

1. **Déc. 14, 11h00** : Distribution de cordes et formules de Rice, application aux milieux poreux, *Anne Estrade (Université Paris 5)*
2. **Nov. 30, 11h00** : Learning networks and curves, *Kevin Bleakley (Université Montpellier 2)*
3. **Nov. 23, 11h00** : Autour d'une technique de rééchantillonnage non paramétrique pour les champs aléatoires, *Lionel Truquet (CREST et Samos, Université Paris 1)*
4. **Oct. 18, 22h46** : Non-parametric Residual Variance Estimation in Supervised Learning with Applications, *Elia Liitiäinen (Helsinki University of Technology, Finlande)*
5. **Oct. 18, 22h43** : Modèle de mélange d'analyses en composantes principales robustes, *Michel Verleysen (U.C.L., Belgique)*
6. **Oct. 12, 11h00** : Modèles de mélange gaussien pour la classification des données de grande dimension, *Charles Bouveyron (Université Paris 1)*
7. **Juin 29, 11h05** : Stable and Unstable Manifolds for Stochastic Partial Differential Equations, *Jinqiao Duan (Illinois Institute of Technology, USA)*
8. **Juin 22, 11h00** : Conception de réseaux : quelques modèles et résultats., *Arnaud Knippel (LMI/INSA de Rouen, Département Génie Mathématique)*
9. **Juin 15, 12h00** : Exponential ergodicity for SPDEs, *Bohdan Maslowski (Académie des Sciences Tchèque), maslow@math.cas.cz*
10. **Juin 15, 11h00** : Estimateur de maximum de vraisemblance pour les EDS dirigées par un Brownien fractionnaire, *Frederi G. Viens (Purdue University, USA)*
11. **Juin 08, 11h00** : Le polyèdre des sous-graphes bipartis induits, conception de circuits VLSI et génomique, *Pierre Fouilhox (LIP6, Université Paris 6)*
12. **Juin 01, 11h00** : Multivariate Generalization of the Markov Switching Model, *Mohamad Khaled (EUREQua, CES, Université Paris 1), mokl79@yahoo.com*
13. **Avril 06, 11h45** : Borne "rayon-marge" sur l'erreur "leave-one-out" des SVM multi-classes, *Emmanuel Monfrini (Loria, INRIA Lorraine) monfrini@loria.fr*
14. **Avril 06, 11h00** : Les représentations multidimensionnelles doivent tenir compte du fléau de la dimension., *Sylvain Lespinats (INSERM unité U722), lespinats@bichat.inserm.fr*
15. **Mars 30, 12h00** : Propriétés asymptotiques et statistiques des modèles matriciels de dynamique des populations. Applications en foresterie., *Mélanie Zetlaoui (Université d'Evry), melanie.zetlaoui@univ-evry.fr*
16. **Mars 30, 11h00** : Comportement asymptotique de la distribution des pluies extrêmes en France, *Aurélié Muller (ENS Cachan Bretagne), muller@bretagne.ens-cachan.fr*
17. **Mars 23, 11h00** : Etude des configurations locales dans le modèle d'Ising, *David Coupier (Université Lille 1), david.coupier@math.univ-lille1.fr*
18. **Mars 16, 11h00** : Spatial Prediction for Massive Datasets, *Noel Cressie (Ohio State University, USA), ncressie@stat.ohio-state.edu*
19. **Mars 09, 11h00** : Estimation dans des modèles définis par équations différentielles ordinaires et stochastiques. Applications biostatistiques, *Sophie Donnet (Université Paris XI, Orsay), Sophie.Donnet@math.u-psud.fr*
20. **Fév. 23, 11h00** : Sélection de variables par information mutuelle et rééchantillonnage, *Michel Verleysen (Université catholique de Louvain, Belgique), verleysen@dice.ucl.ac.be*
21. **Fév. 09, 11h00** : Des techniques d'apprentissage statistique face à la complexité des données d'expression génomiques ou industrielles, *Philippe Besse (Université Toulouse III), besse@math.ups-tlse.fr*
22. **Fev. 02, 12h00** : Some experiences with Simulated Annealing in the solution of Statistical problems, *Carlos Bouza (Université La Havane, Cuba), bouza@matcom.uh.cu*

23. **Fev. 02, 11h00** : Sampling using Ranked Sets : some results in finite population inference, *Carlos Bouza (Université La Havane, Cuba), bouza@matcom.uh.cu*
24. **Jan. 26, 12h00** : Statistical Analysis of Signaling, *Carlos Bouza (Université La Havane, Cuba), bouza@matcom.uh.cu*
25. **Jan. 26, 11h00** : Statistical solutions to Stochastic Programs , *Carlos Bouza (Université La Havane, Cuba), bouza@matcom.uh.cu*
26. **Jan. 19, 11h00** : Mixed-integer rounding cutting planes for integer programming problems, *Sanjeeb Dash (Research Staff Member, New-York), sanjeebd@us.ibm.com*

#### 2008 :

- **Avril 18, 11h00** : Modèles à facteurs dynamiques, *Nathanaël Mayo (Samos et Exane BNP Paribas)*
- **Avril 11, 11h00** : Occupation densities for certain processes related to fractional Brownian motion, *Khalifa Es-Sebaiy (Samos, Université Paris 1)*
- **Avril 04, 11h00** : Bornes PAC-Bayésiennes pour des estimateurs par minimisation du risque empirique, *Pierre Alquier (Université Paris 7)*
- **Mars 28, 11h00** : Cartes auto-organisatrices pour données non vectorielles., *Fabrice Rossi (INRIA)*
- **Mars 21, 11h00**: To which class do known distributions of real valued infinitely divisible random variables belong?, *Makoto Maejima (Keio University, Japon)*
- **Mars 14, 11h30** : Division, *Paul Doukhan (CREST et Samos)*
- **Mars 07, 11h00** : The limits of nested subclasses of several classes of infinitely divisible distributions on  $\square$ , *Makoto Maejima (Keio University, Japon)*
- **Fev. 15, 11h00** : Classification methods for DNA barcoding, *Madalina Olteanu (Samos, Université Paris 1)*
- **Fév 08, 11h00** : Apprendre et prédire sur des populations différentes, *Julien Jacques (Laboratoire Paul Painlevé, Université Lille 1)*

### 3.4 Le séminaire Mathématiques des systèmes complexes (novembre 2006-mai 2008)

Ce séminaire est co-organisé par les équipes Marin Mersenne et Samos, en liaison avec l'Institut de Systèmes Complexes de Paris et Ile-de-France. Il se tient habituellement à l'Université Paris 1

#### 2006 :

- **Nov. 17, 11h00** : Rythmes complexes dans les systèmes multi-échelles, *Jean-Pierre Françoise (Université Paris 6)*
- **Oct. 13, 11h00** : Boosting Itéré pour la Détection de Données Aberrantes, *Jean-Michel Poggi (Université Paris 11-Orsay et Paris 5)*
- **Avril 07, 12h00** : Dynamique de réseaux complexes : approximation de champ moyen et équations de corrélations, *Annick Lesne (LPTMC, Université Paris 6)*
- **Fév. 24, 11h00** : Dynamique de réseaux complexes : approximation de champ moyen et équations de corrélations, *Annick Lesne (LPTMC, Université Paris 6)*
- **Jan. 06, 11h00** : La parole naissant des interactions perceptuo-motrices : cadre théorique, données expérimentales et éléments de modélisation computationnelle, *Jean-Luc Schwartz (CNRS, I.N.P. Grenoble)*

## 2007 :

- **Nov. 30, 11h00** : Learning networks and curves, *Kevin Bleakley (Université Montpellier 2)*
- **Oct. 12, 11h00** : Modèles de mélange gaussien pour la classification des données de grande dimension, *Charles Bouveyron (Université Paris 1)*
- **Juin 22, 11h00** : Conception de réseaux : quelques modèles et résultats., *Arnaud Knippel (LMI/INSA de Rouen, Département Génie Mathématique)*
- **Juin 08, 11h00** : Le polyèdre des sous-graphes bipartis induits, conception de circuits VLSI et génomique, *Pierre Fouilhoux (LIP6, Université Paris 6)*
- **Avril 06, 11h45** : Borne "rayon-marge" sur l'erreur "leave-one-out" des SVM multi-classes, *Emmanuel Monfrini (Loria, INRIA Lorraine) monfrini@loria.fr*
- **Avril 06, 11h00** : Les représentations multidimensionnelles doivent tenir compte du fléau de la dimension., *Sylvain Lespinats (INSERM unité U722), lespinats@bichat.inserm.fr*
- **Fév. 09, 11h00** : Des techniques d'apprentissage statistique face à la complexité des données d'expression génomiques ou industrielles, *Philippe Besse (Université Toulouse III), besse@math.ups-tlse.fr*
- **Jan. 19, 11h00** : Mixed-integer rounding cutting planes for integer programming problems, *Sanjeeb Dash (Research Staff Member, New-York), sanjeebd@us.ibm.com*

## 2008 :

- **Avril 18, 11h00** : Modèles à facteurs dynamiques, *Nathanaël Mayo (Samos et Exane BNP Paribas)*
- **Avril 04, 11h00** : Bornes PAC-Bayésiennes pour des estimateurs par minimisation du risque empirique, *Pierre Alquier (Université Paris 7)*
- **Mars 28, 11h00** : Cartes auto-organisatrices pour données non vectorielles., *Fabrice Rossi (INRIA)*
- **Fév. 15, 11h00** : Classification methods for DNA barcoding, *Madalina Olteanu (Samos, Université Paris 1)*

## Autres:

**Groupe de Travail « Aspects Fractals »** organisé par Ciprian Tudor (avec Ivan Nourdin et Giovanni Peccati) à l'Université de Paris 6 (Chevaleret). Ce groupe de travail se déroule une fois par mois et il a commencé en octobre 2006.

#### 4 LES CONVENTIONS DE RECHERCHE ET D'ETUDE 2005-2008

Au cours de la dernière période quadriennale, une douzaine de conventions de recherche et d'études et de contrat d'accompagnement de bourses CIFRE ont été signées avec des partenaires privés. De plus nous avons obtenu un contrat dans le cadre de l'ACI CNRS « Systèmes Complexes » et déposé un projet ANR Blanc en mathématiques.

##### Contrats industriels

Nom de l'entreprise	Laboratoire concerné	Responsable du contrat	Sujet	Date de début	Durée
GDF	CES	Cottrell Marie	Comportements de consommateurs	2003	2 ans
TRIALIS	SAMOS	Letrémy Patrick	Exploration de données	2005	6 mois
Informatique et Statistiques	SAMOS	Souchet Sandie	Séries chronologiques	2005	6 mois
GDF	SAMOS	Rynkiewicz Joseph	Test de modèle de prévision à court terme	2006	3 mois
GDF	SAMOS	Cottrell Marie	Modèle de prévision à court terme; faisabilité	2006	3 mois
NOEO	SAMOS	Olteanu Madalina	Veille sur les modèles de demande d'énergie	2007	1 mois
EXANE	SAMOS	Doukhan Paul	CIFRE (Nathanaël Mayo)	2007	3 ans
EDF	SAMOS	Girard Bernard	Explication du prix du gaz	2007	6 mois
STSI/CTSI	SAMOS	Vigneron Vincent	Surveillance des aéroports	2007	1 an
NOVACYT	SAMOS	Bouveyron Charles	Classifications d'images de frottis	2008	1 an + 3ans en cours de signature
SNECMA	CES	Cottrell Marie	Faisabilité – Cartographie de l'évolution de l'état du moteur	2008	4 mois + 3 ans en cours de signature
SNECMA	CES	Cottrell Marie	Faisabilité – Identifications de sources vibratoires	2008	4 mois + 3 ans en cours de signature
AUTOBIZZ	SAMOS	Bardet Jean-Marc	CIFRE (Solahona Dimby)	2008	3 ans

## **Contrats publics**

- 2004-2006

### **Programme « Système complexe en SHS »**

Equilibre dynamique, étude des trajectoires dans un système économique, le cas du marché du travail.

Resp. scientifique : Marie Cottrell

- 2008

### **Dépôt d'un projet ANR Blanc en Mathématiques**

TRACE : TRANSfert de Connaissance dans un Environnement changeant

Partenaires : Fabrice Rossi (INRIA), Jérôme Besnard (PDF Solutions SA), Jérôme Lacaille (SNECMA), Michel Verleysen (Machine Learning Group, Louvain-la-Neuve, Belgique)

Par ailleurs, il faut souligner le grand rayonnement des travaux plus particulièrement conduits dans le domaine des statistiques, de la modélisation et du traitement des données. En effet, les qualités remarquables de visualisation de données multidimensionnelles des algorithmes développés par le groupe SAMOS font qu'ils sont largement utilisés dans l'enseignement, notamment les M2 de Paris 1. Ainsi, les étudiants en font un usage très fréquent dans leur mémoire ou au cours de leurs stages. De même, plusieurs services de recherche en entreprise ont sollicité une formation pour l'utilisation de cet ensemble de programmes et de méthodes.

## **5 LA COMMUNICATION : LE SITE WEB DU SAMOS**

Depuis 12 ans, le SAMOS a mis en place un site internet (<http://samos.univ-paris1.fr>). Ce site est actuellement mis à jour par Yvonne Girard.

La page d'accueil présente l'actualité des activités principales du SAMOS et permet d'accéder à différentes rubriques parmi lesquelles : la présentation du projet scientifique de l'UMR, les publications et prépublications, la liste et les coordonnées des membres (y compris les membres associés), le programme du séminaire. Chaque membre du SAMOS dispose d'une page individuelle sur le site du MATISSE lui permettant de développer la présentation de ses travaux, de ses publications et de ses projets.

Depuis 2006, l'ensemble des publications, documents de travail et autres travaux sont déposés sur le serveur HAL.