

Radu Stefan STOICA

Dossier scientifique

Année 2016

Curriculum Vitae détaillé

Table des matières

1 Curriculum Vitae	3
1.1 Formation et parcours professionnel	3
1.2 Diplômes et quelques points clefs	4
2 Activité d'enseignement et d'encadrement	5
2.1 Service d'enseignement et encadrement d'étudiants à Lille	5
2.2 Encadrement doctoral et scientifique	5
2.3 Projet d'enseignement - lien avec la recherche : coopération Erasmus avec l'Université de Tartu - Estonie	8
3 Activité de recherche	9
3.1 Projet de recherche	9
3.2 Activité à l'Université de Lille	9
3.3 Perspectives - Nouvelle direction de recherche	11
3.4 Diffusion des travaux (rayonnement et vulgarisation)	12
3.5 Responsabilités scientifiques	14
3.6 Réalisation et diffusion de logiciels	15
4 Responsabilités collectives	15
5 Annexe A : liste complète des publications	17
6 Annexe B : plan du cours fait dans le cadre de la coopération Erasmus avec l'Université de Tartu - Estonie	22
7 Annexe C : bilan et projets futurs de mon activité de recherche	24
7.1 Questions concernant la modélisation de forme	24
7.1.1 Extension à l'infini, percolation et transition de phase	24
7.1.2 Processus de Cox, processus permanents et déterminants	25
7.2 Questions concernant la simulation MCMC	26
7.2.1 Méthodes classiques de simulation	26
7.2.2 Simulation exacte	27

7.3	Questions concernant l'inférence statistique	27
7.3.1	Détection conjointe de la forme et des paramètres	28
7.3.2	Simulation de la loi a posteriori	28
7.3.3	Choix de modèle	28
7.4	Projet : processus stochastiques et géométrie aléatoire	29
7.4.1	Modélisation spatio-temporelle	30
7.4.2	Simulation de processus stochastiques des formes	31
7.4.3	Statistiques pour les processus stochastiques des formes	31
7.4.4	Outils probabilistes et d'inférence statistique pour caractériser le chaos	32
8	Annexe D : personnes à contacter pour des lettres de recommandation	33
9	Annexe E : liste des publications transmises	35

1 Curriculum Vitae

Radu STOICA

Maître de conférences en Mathématiques :

Université de Lille - Laboratoire Paul Painlevé

Bâtiment M3, Bureau 316

59655 Villeneuve d'Ascq Cedex, France

tel : + 33 3 20 43 42 17

fax : + 33 3 20 43 67 74

e-mail : radu.stoica@math.univ-lille1.fr

web : www.math.univ-lille1.fr/~stoica/

Double nationalité : Française et Roumaine

Date de naissance : 24/09/1972

1.1 Formation et parcours professionnel

ÉTABLISSEMENTS français ou étrangers	FONCTIONS ET STATUTS (salarié, boursier, etc.)	DATES		OBSERVATIONS
		d'entrée en fonction en	de cessation de fonction	
Université de Lille (Lille - France)	salarié (maître de conférences)	01/09/2006		Laboratoire Paul Painlevé
INRA (Avignon - France)	salarié (post-doctorat)	01/12/2004	31/08/2006	Unité Biométrie
Université Jaume I (Castellon - Espagne)	boursier (post-doctorat)	01/03/2003	30/11/2004	Dpt. de Mathé- matiques
CWI (Amsterdam - Pays-Bas)	salarié (post-doctorat)	01/01/2001	31/12/2002	Stochastic Geometry Project
INRIA (Sophia Antipolis - France)	boursier (étudiant en thèse)	01/09/1997	31/12/2000	Projet Ariana
INRIA (Sophia Antipolis - France)	boursier (étudiant en DEA)	01/03/1997	31/08/1997	Projets Pastis/Ariana

1.2 Diplômes et quelques points clefs

Habilitation à diriger des recherches :

- Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées. Université de Lille, 2014. Garant : Yu. Davydov.

Doctorat :

- Processus ponctuels pour l'extraction de réseaux linéiques dans les images satellitaires et aériennes. Université de Nice-Sophia Antipolis, 2001. Directeurs : J. Zerubia et X. Descombes (INRIA - projet Ariana).

Autres diplômes :

- 1997 : DEA Signaux, Images et Communications, ENSEEIHT (Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Hydraulique et d'Informatique de Toulouse).
- 1997 : diplôme d'ingénieur en section spéciale en traitement du signal et des images à l'ENSEEIHT.
- 1996 : diplôme d'ingénieur en électronique et télécommunications à l'UPB (Université "POLITEHNICA" de Bucarest) comprenant plusieurs semestres d'études effectués à l'ESIEE Paris (Ecole Supérieure d'Ingénieurs en Electrotechnique et Electronique) dans le cadre de programmes d'échange TEMPUS.

Spécialité : Mathématiques - Probabilités et Statistique

Projet de recherche : modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées

- géométrie stochastique, processus spatiaux, simulation Monte Carlo, inférence statistique, statistiques spatiales

Réseaux scientifiques nationaux et internationaux :

- membre du GDR "Géométrie Stochastique" (responsables : P. Calka, D. Coupier, C.V. Tran)
- membre du GDR "Information Signal Image Vision" (responsable : L. Blanc-Ferraud)
- membre ISI International Astrostatistics Network (responsable : J. Hilbe)
- porteur action spécifique "Astrostatistique" (responsables : D. Fraix-Burnet, S. Girard)

Primes et délégations :

- délégation CNRS (un an) - l'Institut de Mécanique Céleste et Calcul des Ephémérides de l'Observatoire de Paris : 01/09/2013 - 31/08/2014
- PES/PEDR : 01/09/2007 - 31/08/2015

Langues : français/roumain (bilingue), anglais/espagnol(courant), catalan/néerlandais (notions).

Violons d'Ingres : lecture, judo (sportif licencié à la FFJDA), musique, cinéma.

2 Activité d'enseignement et d'encadrement

En septembre 2006, j'ai été recruté comme maître de conférence à l'Université de Lille. Depuis, j'ai enseigné à tous les niveaux, de la licence jusqu'au master. Le contenu des enseignements porte essentiellement sur les statistiques et les probabilités surtout quand l'enseignement a lieu au niveau master. En même temps, au niveau licence, j'ai aussi enseigné des mathématiques générales (analyse et algèbre). L'enseignement dispensé est soit théorique soit appliqué, en fonction de la formation où cette activité se déroule.

Mon activité d'enseignement inclut également l'encadrement des étudiants et le travail avec les doctorands.

Dans le cadre de mon activité de recherche, j'ai monté une coopération Erasmus entre l'Université Lille 1 et l'Université de Tartu en Estonie.

2.1 Service d'enseignement et encadrement d'étudiants à Lille

Enseignements en Master :

- M2 MA (Mathématiques Appliquées) : Cours de Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées
- M2 MA : Cours sur les Outils informatiques pour la statistique
- M2 MA : Cours sur les Processus ponctuels (en binôme avec Yu. Davydov)
- M2 ISN (Ingénierie Statistique et Numérique) : Cours d'Initiation aux méthodes de type Monte Carlo
- M2 ISN : Cours de Prédiction et séries chronologiques
- M2 ISN : TD/TP en Prédiction et séries chronologiques
- M2 IBVEM (Innovations en Biochnologies Végétales, Enzymatiques et Microbiennes) : Cours de Statistique
- M2 IBVEM : TP de Statistique
- M1 Agro (Agro - Alimentaire) : Cours de Statistique
- M1 Agro : TP de Statistique

Enseignements en Licence :

- L3 Mathématiques : TD de Statistique
- L3 MASS : TD sur les Tests statistiques et Modèles Linéaires
- L2 SVTE : Cours/TD d'Initiation à la Statistique
- L2 Informatique : TD d'Initiation aux Probabilités
- L1, L2 Télécom Lille : TD de Mathématiques
- L1 Parcours Aménagé SVTE : Cours/TD de Probabilités et Statistique

2.2 Encadrement doctoral et scientifique

Encadrement d'étudiants M2 MA (Mathématiques Appliquées)

- Antoine Hinge : modèle de cristallisation sur un réseau discret et mise en oeuvre numérique (co-encadrement avec Youri Davydov) (2014)

- Quentin Grimonprez : caractérisation morphologique et statistique de l'activité populationnelle d'une colonie d'abeilles (co-encadrement avec André Kretzschmar - INRA Avignon) (2011)

Suivi de projets de fin d'année en M2 ISN (Ingénierie Statistiques et Numérique)

- Mathieu Gérard - construction de grille de score (2013)
- Florent Lambert - assister les chargés d'études statistiques dans leur mission en gestion du risque des partenaires (2013)
- Céline Verhaeghe - étude sur des différentes bases des données afin de mieux qualifier leur clients et réalisation de score (2012)
- Anthony Cousien - application des modèles multi-états aux analyses de survie en dialyse (2010)
- Annie Decrequy - optimisation des opérations commerciales par comportement achats clients (2010)
- Benjamin Dhondt - réalisation d'un score d'octroi de crédit et étude de l'état du surendettement (2010)
- Yuze Li - méthodes statistiques appliquées à la vision par ordinateur (2010)
- Magali Lainé - classification des sites industriels du Nord-Pas-de-Calais potentiellement pollueurs de plomb (2009)
- Lin Wang - construction de modèle de durée de délai de financement (2009)
- Coralie Kaminski - construction d'une segmentation géomarketing (2008)
- Sébastien Prudhomme - développement d'un modèle de prévision de flux robuste (2008)
- Viviane Tahon - modélisation de la durée de vie client (2008)
- Cheikh Ba - modélisation de séries chronologiques (2007)

D'autres projets

- Issa Cherif Geraldo, Quentin Grimonprez, Julien Heiduk, Sarah Labbé : étude et implémentation de l'algorithme de simulation exacte CFTP (M2 ISN, 2012)
- Hugo Le Panse, étudiant en L3 Statistique et Informatique Décisionnelle à l'Université Toulouse 3 a effectué son projet de fin d'année dans l'Université Lille 1 : détection et analyse spatiale des cellules pour des images provenant de la spectrochimie (co-encadrement avec Julien Jacques et Cyril Ruckebusch, 2011)
- Yuze Li, Ayoung Oh : processus ponctuels marqués, modèles, simulation et inférence statistique (M1 ISN, 2009)

Travail avec des doctorants

Je dirige la thèse de Delphine Thomasson, depuis septembre 2015. Le sujet de sa thèse est "Modèles statistiques et population des débris spatiaux". Cette thèse est co-encadrée avec Florent Deleflie, astronome à l'IMCCE - Observatoire de Paris. Ce projet est co-financé par le CNES et l'Université de Lille. Cette thèse se déroule à la fois dans le Laboratoire Paul Painlevé et à l'IMCCE.

J'ai co-encadré avec Youri Davydov la thèse de Shuyan Liu. Le titre de sa thèse était "Lois stables et processus ponctuels : liens et estimation des paramètres". La composition du jury pour sa soutenance a été : P. Soulier, Yu. Davydov, J. M. Bardet, B. Garel, J. Segers, R. S. Stoica, A. Vienne. La soutenance a eu lieu le 10/12/2009 à Lille. Aujourd'hui Shuyan Liu est maître de conférences en mathématiques à l'Université Paris 1. Nous continuons à travailler ensemble sur des aspects pratiques et théoriques liés à l'application des lois stables pour l'étude des perturbations planétaires sur les dynamiques des comètes. Des articles issus de ces travaux ont été publiés dans des journaux et des actes de conférence [11,20,31].

Un autre article est en préparation.

Depuis janvier 2014 à l'IMCCE - Observatoire de Paris, je participe régulièrement, aux réunions de travail avec Irina Kovalenko (doctorante) et Daniel Hestroffer (son directeur de thèse). Mon rôle est d'apporter mes connaissances pour la partie probabilités et statistique de la thèse, notamment les aspects liés aux problèmes inverses et modélisation bayésienne. Le sujet de thèse est "Caractérisation orbitale et physique des astéroïdes binaires". Des articles communs ont été publiés dans des actes de conférence [11,31,32] et un article pour un journal à comité de lecture est en préparation.

A partir de février 2013, avec Fritz Kleinschroth (doctorant) et Frédéric Mortier (directeur de thèse) de CIRAD Montpellier, nous nous réunissons pour travailler sur le développement d'un modèle dynamique de réseau routier. Mon rôle est d'apporter mes connaissances pour la partie probabilités et statistiques de la thèse, notamment les aspects liés à la modélisation par processus ponctuels marqués. Ce travail inclut également Marie-Colette van Lieshout de CWI Amsterdam (Pays - Bas), ainsi que Sylvie Gourlet-Fleury et d'autres chercheurs de CIRAD. Un article commun a été soumis à la publication dans un journal à comité de lecture.

Dès juillet 2013 le doctorant Juhan Liivamägi a rejoint le groupe de chercheurs estoniens de l'Observatoire de Tartu (Elmo Tempel et Enn Saar), pour que l'on travaille ensemble sur l'étude de la distribution des filaments galactiques utilisant les processus ponctuels marqués. Dans cette nouvelle composition, notre groupe a publié un chapitre de livre et plusieurs articles [5,11,14,16]. En décembre 2015, le doctorant Rain Kipper de l'Observatoire de Tartu est venu travailler avec moi pendant un mois, avec une bourse de l'Ambassade de France en Estonie. Nous avons pu ainsi avancer sur nos projets en cours, en soumettant un article à la publication [13] et en étudiant la modélisation de la poussière galactique utilisant les champs aléatoires. La coopération avec les partenaires estoniens se poursuit depuis 2003.

Comités de thèse et jurys :

- Membre du jury de thèse de Jérôme Daquin (IMCCE - Observatoire de Paris – directeurs : F. Deleflie, J. Perez). Titre de la thèse : Essais sur les mouvements des corps artificiels résonants. Composition du jury : F. Deleflie, M. Guzzo, A. Lemaitre, J. Pelaez, J. Perez, A. Rossi, B. Sicardy, R. S. Stoica. La soutenance a eu lieu le 10/12/2015 à Paris.
- Membre du jury de thèse de Francisco Rodriguez Cortes (Université Jaume I, Castellon, Espagne – directeur : Jorge Mateu). Titre de la thèse : Modelling, estimation and applications of second-order spatio-temporal characteristics of point processes. Composition du jury : O. Cronnie, F. Montes, R. S. Stoica. La soutenance a eu lieu le 05/05/2014 à Castellon (Espagne). Francisco Rodriguez Cortes a effectué un séjour d'un mois à l'Université Lille 1 dans l'équipe Probabilités et Statistiques du Laboratoire Paul Painlevé. Le GDR "Géométrie Stochastique" a pris partiellement en charge ce séjour.
- Membre du jury de thèse de Julien Frouard (Observatoire de Paris – directeurs : Marc Fouchard, Alain Vienne). Titre de la thèse : Dynamique et stabilité des satellites irréguliers de Jupiter. Composition du jury : M. Fouchard, A. Lemaitre, C. D. Murray, B. Sicardy, B. Slawomir, R. S. Stoica, A. Vienne. La soutenance a eu lieu le 14/12/2010 à Paris.
- Membre du comité de thèse de Vera Georgescu (INRA Avignon - directeurs : André Kretzschmar, Rachid Senoussi, Samuel Soubeyrand). Titre de la thèse : Classification de données multivariées multitypes basée sur des modèles de mélange. Application à l'étude d'assemblages d'espèces en écologie. Comité de thèse : D. Agostini, J. N. Bacro, S. Soubeyrand, R. S. Stoica. La réunion du comité a eu lieu le 29/09/2008 en Avignon.

2.3 Projet d'enseignement - lien avec la recherche : coopération Erasmus avec l'Université de Tartu - Estonie

Après la publication des nombreux articles communs, ma coopération avec l'Estonie¹ nous a mené naturellement vers la formation. Nous avons décidé de mettre en place une coopération Erasmus entre l'Université de Lille (France) et l'Université de Tartu (Estonie), pour faciliter l'échange d'enseignants, de chercheurs et d'étudiants entre nos deux institutions.

Dans le cadre de cet accord, durant le mois de septembre 2014, j'ai enseigné un cours de 32h heures (4 semaines \times 2 séances \times 4 heures) à l'Université de Tartu. Le titre du cours est : "Mathematical statistics in physics and astronomy". Le cours s'adresse à des étudiants M2, doctorants, chercheurs débutants ou confirmés. Les étudiants proviennent ou bien des filières mathématiques et ou bien des filières physiques. La plupart des doctorants et des chercheurs sont des physiciens, des astronomes et des cosmologistes.

Cet enseignement se termine par un examen. Il aura lieu tous les deux ans. Cela nous permet d'avancer en parallèle notre activité commune de recherche et en fonction de cela, d'adapter l'enseignement en mathématiques. En effet, suite à l'activité de recherche commune, les questions mathématiques évoluent et ceci doit avoir une conséquence sur l'enseignement. Un premier objectif de ce projet d'enseignement est de fournir au public participant, un cadre probabiliste, amenant ainsi vers des réponses aux questions formulées, en utilisant l'inférence statistique.

Le plan du cours est donné dans l'Annexe B.

Un deuxième objectif de ce projet est de faire profiter mon institution, de l'expérience acquise à travers cette coopération Erasmus. Ainsi, ce cours a servi comme point d'ancrage pour l'enseignement "Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées" que j'ai effectué à Lille cette année dans le cadre du M2 Mathématiques Appliquées. Le contenu du cours a été adapté, afin que sa spécificité Mathématiques Appliquées soit prise en compte. Comme pour l'Estonie, j'ai bénéficié d'un public plus large que les étudiants concernés. A ce cours ont participé aussi des doctorants en mathématiques (trois étudiants) et en astronomie (deux étudiants) de Lille et de Namur (Belgique) (un étudiant), un post-doc en épidémiologie de Bruxelles (Belgique) et deux maîtres de conférences en mathématiques de Lille et de Paris, respectivement.

L'objectif à long terme de ce projet est de former et de faire interagir un groupe, un réseau de chercheurs, jeunes et confirmés, autour des mathématiques et des applications liés aux questions de mon projet de recherche.

1. voir le paragraphe *Activité de recherche* pour plus des détails.

3 Activité de recherche

3.1 Projet de recherche

En septembre 2006, j'ai été recruté à l'Université de Lille sur un projet de recherche que j'avais intitulé "Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées" [1]. Je me suis efforcé de rendre ce projet compatible avec l'environnement que m'offrait l'équipe "Probabilités et Statistiques" du Laboratoire P. Painlevé. Nouvel environnement, nouvelles questions, nouveaux horizons ont enrichi et approfondi mon projet de recherche.

C'est, de la même façon, sur la base de nouvelles collaborations et d'aides mutuelles que je vois mon activité de recherche future.

Dans les paragraphes qui suivent, je donne les grandes lignes de mes collaborations et de mes travaux pendant ma période lilloise. La liste complète des publications est présentée dans l'Annexe A. Le bilan et les projets futurs qui en découlent sont développées dans l'Annexe C.

3.2 Activité à l'Université de Lille

Ce projet se proposait de développer des outils probabilistes et statistiques pour pouvoir répondre aux questions du type "quelle est la forme qui se cache dans les données?". Plus précisément, je souhaitais me focaliser sur le développement des outils suivants :

- modélisation stochastique : qu'est-ce que la forme que l'on cherche ?
- simulation de type MCMC : comment mettre en oeuvre la forme ?
- inférence statistique : comment détecter la forme, quels sont les paramètres de la forme ?
- évaluation : la forme trouvée existe-t-elle vraiment, validation des résultats ?

Une fois en poste, j'ai commencé à mettre en place ce projet en suivant trois directions interdépendantes :

I - Partage du projet avec l'équipe "Probabilité et Statistiques". En 2007, le professeur Youri Davydov a créé le groupe de travail "Géométrie Stochastique". Je me suis impliqué tout de suite dans la vie de ce groupe qui est en lien direct avec mon projet de recherche. Peu-à-peu, des questions se sont concrétisées en publications [5,14,17,18,20].

II - Consolidation de mes contacts scientifiques en France et hors de France. Avant être recruté à Lille 1, j'ai travaillé comme post-doc au CWI Amsterdam (Pays-Bas), l'Université Jaume I Castellon (Espagne) et l'INRA Avignon. Pendant cette période, deux autres partenaires très importants ont été l'Observatoire de Tartu (Estonie) et l'Université de Valencia (Espagne). Une fois arrivé à Lille, nous avons continué travailler ensemble tout en cherchant faire bénéficier de ces contacts, l'équipe "Probabilité et Statistiques".

Marie-Colette van Lieshout de CWI Amsterdam a participé à deux des conférences sur la géométrie aléatoire organisées par le groupe de travail. Nous avons finalisé et publié un travail commencé pendant mon séjour au CWI [19]. Notre collaboration continue et actuellement nous travaillons sur l'application des statistiques descriptives de la forme et la construction d'un processus ponctuel marqué adapté pour la modélisation d'une dynamique d'évolution d'un réseau routier. Ce problème intéresse directement des chercheurs du CIRAD Montpellier (Frédéric Mortier, Sylvie Gourlet-Fleury) et c'est dans ce cadre que

je travaille avec le doctorant Fritz Kleinschroth. Un article a été soumis à la publication [10]

La coopération avec l'Université Jaume I continue et un article a été soumis à la publication [12]. Le sujet de ce travail porte sur la construction des nouvelles méthodologies pour l'estimation des paramètres des processus ponctuels marqués. Anne Philippe de l'Université de Nantes est aussi impliquée dans cette coopération.

La coopération avec l'INRA suit son chemin. Kien Kieu (INRA, Jouy-en-Josas) et Joel Chadoeuf (INRA, Avignon) ont participé à nos conférences sur la géométrie aléatoire. Naturellement, nous avons finalisé et publié les travaux commencés ensemble [15,22,24]. André Kretzschmar s'intéresse aux aspects géométriques des différentes structures bio-environnementale, comme l'évolution des colonies d'abeilles ou la fissuration du sol. En parallèle, Youri Davydov et Sergei Zuyev de l'Université de Chalmers (Suède) s'intéressent à des modèles stochastiques géométriques de croissance, et ils m'ont généreusement proposé travailler ensemble. L'encadrement de l'étudiant A. Hinge s'inscrit dans ce cadre. Personnellement, je souhaite appliquer ces modèles mathématiques aux problèmes provenant des sciences de la vie mentionnés précédemment.

La coopération avec Enn Saar de l'Observatoire de Tartu (Estonie) et Vicent Martinez de l'Université de Valencia (Espagne)² se développe et se concrétise en publications [5,11,13,14,16,21,23,31,34]. Nous étudions et construisons des modèles provenant de la géométrie aléatoire pour la détection et la caractérisation des structures formées par la distribution des galaxies dans l'Univers (agrégats, murs et filaments). L'équipe s'est agrandie en incorporant un brillant jeune scientifique, Elmo Tempel et les doctorants Juhan Liivamägi et Rain Kipper. Pour notre projet, nous avons bénéficié des aides financières de type BQR Lille 1 et de la part du GDR Géométrie Stochastique. Ces aides nous ont permis financer des visites des chercheurs et le symposium intitulé "Mathématiques pour la Cosmologie" qui a eu lieu à Lille. Nous avons reçu aussi le soutien de l'Ambassade de France en Estonie. Pendant l'année 2009, on nous a proposé la participation à l'action "Mois de la science française en Estonie". En décembre 2015, Rain Kipper a bénéficié d'une bourse de l'Ambassade de France en Estonie pour venir travailler avec moi à Lille.

Coopération Erasmus et formation d'étudiants. C'est sur la base de cette dynamique de recherche que ma coopération Erasmus a été lancée, et c'est sur ces thèmes de recherche que nous souhaitons lancer des thèses en co-tutelle.

III - Ouverture vers de nouveaux domaines d'application En arrivant à Lille j'ai lancé une coopération avec Alain Vienne et Marc Fouchard du Laboratoire d'Astronomie de Lille (LAL - IMCCE) et enseignants-chercheurs à l'Université Lille 1. Cette coopération s'est concrétisée dans des publications communes (astronomes et mathématiciens) issues des travaux effectués dans le cadre de la thèse de Shuyan Liu [11,20,31]. La coopération continue et un article est en cours de préparation.

Délégation CNRS (2013-14). Le CNRS m'a accordé un an de délégation au sein de l'Institut de Mécanique Céleste et Calcul des Ephémérides de l'Observatoire de Paris. Le thème général de cette délégation était le développement des outils probabilistes et statistiques pour l'étude des dynamiques chaotiques des corps célestes. Cette délégation a été le point de départ d'une coopération avec Daniel Hestroffer (IMCCE - Observatoire de Paris) et son étudiante en thèse, Irina Kovalenko, sur l'utilisation des problèmes inverses et la modélisation bayésienne pour la caractérisation orbitale et physique des astéroïdes binaires. Des articles ont été publiés dans des actes des conférences [11,31,32] et un article est en préparation.

2. E. Saar et V. Martinez sont les auteurs de l'ouvrage "Statistics for galaxy distributions" (Chapman & Hall, 2002).

En même temps avec l'astronome Florent Deleflie (LAL - IMCCE), nous avons lancé une thèse sur la modélisation des débris spatiaux. La thèse se propose d'utiliser des techniques de la statistique spatiale et des processus spatio-temporels pour caractériser et modéliser les nuages des débris autour de la Terre, formé par les satellites artificiels. Cette thèse est un partenariat entre le CNES (Juan Carlos Perez Dolado) et l'Université de Lille.

Astrostatistique. Les probabilités et les statistiques sont un cadre méthodologique naturel pour les nombreux problèmes provenant de l'astronomie et de la cosmologie. Sur cette base, un réseau international, "ISI International Astrostatistics Network" et un réseau local "Astrostatistique - France" ont été créés, afin que la synthèse et la communication entre mathématiques et astronomie puisse mieux se développer. J'ai tout de suite adhéré à ce mouvement scientifique, de par son caractère bénéfique pour les deux communautés. Mon implication dans l'organisation (et la participation à) des écoles d'été et des congrès internationaux s'inscrit aussi dans ces lignes.

3.3 Perspectives - Nouvelle direction de recherche

Nos travaux de recherche ont abouti à la construction d'une méthodologie mathématique pour détecter et caractériser la structure de données spatialisées. Cette construction procède en trois étapes. D'abord, un modèle de forme ou de structure est proposé à partir des données observées. Puis, une dynamique de simulation est construite en adéquation avec le modèle. Enfin, des procédures statistiques sont mises au point pour inférer les caractéristiques de la structure cachée et les paramètres du modèle. Chacune de ces étapes est attachée à un domaine particulier des probabilités et des statistiques. La modélisation repose sur les processus spatiaux markoviens, notamment sur les processus ponctuels marqués. La dynamique de simulation utilise les chaînes de Markov. L'inférence s'appuie sur l'analyse bayésienne, le recuit simulé, le maximum de vraisemblance, les tests. La synthèse de ces trois étapes se fait au confluent de trois domaines : la géométrie aléatoire, les chaînes de Markov et la statistique. Cette synthèse nous a permis d'aborder des applications concrètes en analyse d'image, en science de l'environnement et en astronomie, dans une relation d'interdépendance avec les mathématiques. Les détails concernant les différentes perspectives, à court et à long terme, de ce projet sont données dans [1].

Actuellement, avec Enn Saar (cosmologiste à l'Observatoire de Tartu, membre de l'Académie des Sciences de l'Estonie) et André Kretzschmar (biologiste à l'INRA Avignon, directeur de recherche), nous écrivons un livre qui présente des outils mathématiques permettant d'aborder une question fondamentale dans l'analyse des données spatialisées. Cette question est : quelle est la forme qui se cache dans les données ? Le livre accompagne cette présentation mathématique par des applications sur des données réelles provenant de l'astronomie, les sciences de l'environnement et l'analyse d'images. Nous ciblons un public d'étudiants et de chercheurs, jeunes et confirmés, provenant des domaines représentés par les trois auteurs.

La plupart des applications qui ont motivé notre activité de recherche jusqu'à maintenant, s'appuyaient seulement sur l'observation des données spatialisées à un moment de temps fixé. Cette situation a évolué. Aujourd'hui, pour ce type de données, nous disposons aussi de l'information temporelle. Dans ce cas, les données peuvent être vues comme des réalisations de processus stochastiques spatio-temporels. A temps fixé, ces processus peuvent être décrits par les outils existants de la géométrie stochastique et des statistiques spatiales, en apportant ainsi des éléments de réponses aux questions formulées. Cependant, l'intégration de ces éléments nécessite un mécanisme mathématiquement robuste. L'ajout de la dimension temporelle impose naturellement d'adapter des outils en provenance de la théorie des processus stochastiques à la description de dynamiques d'objets géométriques aléatoires. Cette idée a été lancée

par Youri Davydov dans le groupe de travail "Géométrie Stochastique" de l'Université de Lille. Et ceci constitue la direction de recherche majeure vers laquelle je souhaite m'orienter pour les années à venir.

3.4 Diffusion des travaux (rayonnement et vulgarisation)

Présentations dans des conférences :

- *ISI 2015 : 60th World Statistics Congress*, Rio de Janeiro, Brazil, 2015
- *Spatial Statistics : Emerging Patterns*, Avignon, France, 2015
- *Physique et Astrométrie des Objets du Systeme Solaire*, Hateg, Roumanie, 2015
- *IAU Symposium 306 : Statistical Challenges in 21st Cosmology*, Lisbonne, Portugal, 2014
- *Tracing the Cosmic Web*, Leiden, Pays-Bas, 2014
- *International Society for Bayesian Analysis World Meeting*, Kyoto, Japon, Juin, 2012
- *French-Danish Workshop on Spatial Statistics and Image Analysis in Biology*, Avignon, France, Mai 2012
- *Workshop on Stochastic Geometry and its Applications*, Rouen, France, March, 2012
- *Astronomie et Dynamique des Systèmes Gravitationnels*, Lille, France, Avril 2010
- *Mathematics and Astronomy*, Madrid, Espagne, Novembre 2009
- *Mécanisme de Transport des Comètes du Nuage de Oort*, Lille, France, Mai 2009
- *French-Danish Workshop on Spatial Statistics and Image Analysis in Biology*, Toulouse, France, Mai 2008
- *Mécanisme de Transport des Comètes du Nuage de Oort*, Lille, France, Avril 2008
- *Workshop on Spatial Point Process Modelling and its Applications*, Castellon, Espagne, Avril 2004
- *IEEE Workshop on Statistical Signal Processing*, Singapore, Août 2001
- *International Workshop on Bayesian Inference and Maximum Entropy Methods*, Gif-sur-Yvette, France, Juillet, 2000
- *International Conference on Spatial Statistics in the Agro-, Bio-, and Geosciences* (organisateur : J. Chadœuf, J. Møller, A. Penttinen et D. Stoyan), Freiberg, Allemagne, Juillet 2000
- *IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, Etats-Unis, Mai 1998.

Participations à d'autres manifestations scientifiques (écoles d'été, colloques) :

- intervenant : école d'été "Advanced Study School on Imaging for Medical Applications (SSIMA)" (organisateur : A. M. Bruckstein, E. Ovrei, I. Priescu) (invités : A. M. Bruckstein, D. Chetverikov, L. Cohen, D. Comaniciu, B. t. Haar Romeny, R. Kimmel, M. Miller, N. Sochen, L. Younes, J. Zerubia), Sinaia, Roumanie, Juin 2015
- organisateur et intervenant : école d'été "Ecole d'astrostatistique : méthodes et applications de la régression en astrophysique" (organisateur : D. Fraix-Burnet, D. Valls-Gabaud, J.-L. Starck, E. Slezak, T. Dudok de Wit, S. Girard, M. Clausel, G. Grégoire, P. Gratier, R. Stoica, J. Jacques), Annecy, France, Octobre 2013
- participant : Workshop on Stochastic Geometry and its Applications, Grenoble, France, Avril 2013
- intervenant : école d'été "La dynamique des systèmes gravitationnels : défis et perspectives" (organisateur : C. Froeschlé, Ch. Froeschlé et E. Lega), Aussois, France, Juillet 2009
- participant : école d'été "Stereology and Geometric Tomography" (organisateur : Eva B. Vedel-Jensen), Sandbjerg, Danemark, Juin 2002
- participant : "Meeting of Dutch Statisticians and Probabilists", Lunteren, Pays-Bas, Novembre 2001 et 2002

- participant : école d'été "Spatial Statistics and Computational Methods" (organisateurs : Martin B. Hansen, Jesper Møller), Aalborg, Danemark, Août 2001.

Visites et/ou séminaires :

- Université de Nice Sophia Antipolis, Laboratoire J. A. Dieudonné (T. Laloe et P. Reynaud - Bouret) : 01/03/2016
- Université de Lorraine - Institut Élie Cartan de Lorraine (R. Azais et A. Gégout - Petit) : 28/01/2016
- Université de Caen Basse-Normandie - Laboratoire de Mathématiques Nicolas Oresme (M. D. Biha et G. Levitt) : 12/01/2016
- Observatoire de Valencia et Observatoire de Javalambre, Espagne (V. Martinez et Mariano Moles) : 26/11/2015
- Université Paris 5, Groupe de Travail en Probabilités (A. Estrade et E. Saada) : 20/11/2015
- Journée SFDS "Méthodes et Logiciels" : traitement d'images (M. Clausel et E. Morand) : 23/10/2015
- Université d'Avignon, Séminaire de Statistique (D. Blanke et F. Bonneu) : 30/03/2015
- Université Paris 12, Groupe de Travail Probabilités (M. A. Asselah, J. Brémont, A. Le Ny) : 24/03/2015
- CWI Amsterdam, Pays - Bas (O. Cronie et M. N. M. van Lieshout) : 12-14/02/2014
- Université de Nantes - Laboratoire de Mathématiques Jean Leray (F. Lavancier et A. Philippe) : 15/11/2012
- Université "Politehnica" Bucarest - Faculté Electronique, Télécommunications et Technologie de l'Information (G. Brezeanu, D. Coltuc et M. Bodea) : 27/09/2012 et 31/10/2013
- Université de La Rochelle - Laboratoire Mathématiques Images Applications (M. Berthier et C. Choquet) : 26/01/2012
- IMCCE - Observatoire de Paris (F. Deleflie, D. Hestroffer)
- Observatoire de Toravere et Université de Tartu en Estonie (E. Saar, E. Tempel)
- Observatoire de Lille (M. Fouchard, A. Vienne)
- INRA Jouy-en-Josas (K. Adamczyk, K. Kieu et H. Monod)
- Université Rennes 2, Séminaire de Statistique de l'IRMAR (M. Emily) : 05-06/11/2009
- Université de Poitiers, Equipe Images, Couleur, Mouvement, Relief et Surfaces (O. Alata) : 07-08/02/2008
- DRASS Lille (C. Heyman) : 2007
- Telecom Lille 1 (M. Daoudi) : 01/06/2007
- Institut de Biologie de Lille (R. Blossey, L. Heliot, B. Vandebunder et D. Trinel)
- Observatoire de la Côte d'Azur (A. Bijaoui, E. Slezak) : 24/01/2006
- Groupe Spatial (organisateur : L. Bel), 2005
- Groupe de Travail "Milieux poreux : modèles et images" (organisateur : A. Estrade)
- Université Montpellier 2, Equipe de Probabilités et Statistiques (G. Biau) : 05/12/2005
- INRA Toulouse (département MIA) : 04/11/2005
- Séminaires de l'équipe Statistique et Modélisation Stochastique Grenoble (S. Degerine, C. D'Aubigny, F. Forbes) : 03/11/2005
- INRA Antibes (L. Lapchin et E. Lombaert) : 21/06/2005
- Université d'Avignon, Département de Mathématiques (A. Seeger) : 24/03/2005
- Centre de Geostatistique de l'Ecole de Mines de Paris (J.P. Chiles et C. Lantuejoul), 2003 et 2004
- INRA Avignon (R. Senoussi) : 18/10/2002
- ENS/INRIA (F. Baccelli) : 26/04/2002.

Invitation “Mois de la science française en Estonie” : Pendant la période 13 - 20 Mai 2009 j'ai été invité par l'Ambassade de France en Estonie pour participer au “Mois de la science française en Estonie”. Ma mission a été de faire une série de présentations pour le grand public et des spécialistes. L'Ambassade de France m'a également facilité la rencontre avec M. Priit Ennet de Vikerradio qui m'avait sollicité un interview. Cet interview a été ensuite diffusé à la radio estonienne.

Institut de Mécanique Céleste et Calcul des Ephémérides - Observatoire de Paris : depuis août 2012 j'y suis chercheur affilié

3.5 Responsabilités scientifiques

Editorial Board. Depuis janvier 2015, je suis membre du comité de rédaction du Spatial Statistics Journal. Ce comité est dirigé par Alfred Stein.

Reviewer : IEEE Transactions on Image Processing, Journal of Microscopy, Biometrical Journal, Annals of Statistics, Scandinavian Journal of Statistics, Annals of Applied Statistics, Journal of Agricultural, Biological and Environmental Statistics et Computational Statistics and Data Analysis.

Organisation de colloques :

- 47èmes Journées de Statistique de la SFdS 1/06-05/06/2015. Comité organisation : C. Biernacki, S. Dabo-Niang, A. Duhamel, M.-N. Nirel, C. Jamroz, S. Hertsoen, C. Rouillon, M. Romo, A. Amiri, N. Wicker, M.-F. Barme, E. Schmisser, E. Chazard, B. Thiam, T. Verdebout, B. Guedj, G. Castellan, R. S. Stoica, Y. Jung, K. Sahmer, L. Broze, (Lille, France).
- Stochastic Geometry Days 31/03-04/04/2014. Comité organisateurs : J.-C. Breton, D. Coupier, D. Dereudre, P. Heinrich, J. Kahn, R. S. Stoica, V. C. Tran, (Lille, France)
- Journées Mathématiques et Cosmologie 19/06-22/06/2012. Comité organisateurs : Yu. Davydov, R. S. Stoica, A. Vienne. Invités : M. Einasto, V. J. Martinez, E. Saar, E. Tempel, (Lille, France)
- Workshop on Stochastic Geometry and its Applications 28/03-30/04/2012. Comité organisateurs Rouen : J.-B. Bardet, P. Calka, N. Chenavier, M. El Machkouri, A. Rousselle, T. de la Rue. Comité organisateurs Lille : D. Coupier, Yu. Davydov, D. Dereudre, P. Heinrich, J. Kahn, R. S. Stoica, V. C. Tran, (Rouen, France)
- Journées Géométrie Aléatoire 30/03-01/04/2011. Comité organisateurs : D. Coupier, Yu. Davydov, D. Dereudre, P. Heinrich, J. Kahn, R. S. Stoica et V. C. Tran. Invités : F. Baccelli, P. Calka, J. Chadoeuf, A. Estrade, J. B. Gouère, K. Kieu, C. Lantuejoul, I. Molchanov, Ch. Thäle, M. N. M. van Lieshout, H. Zessin, S. Zuyev, (Lille, France)
- Mécanisme de transport des comètes du nuage d'Oort 05/05/ - 07/05/2009. Comité organisateurs : M. Fouchard, R. S. Stoica, A. Vienne. Invités : Yu. Davydov, C. Froeschlé, Ch. Froeschlé, S. Liu, H. Rickman, R. Tajeddine, G. B. Valsecchi, (Lille, France)
- Journées Géométrie Aléatoire 03/12 – 04/12/2008. Comité organisateurs : D. Coupier, Yu. Davydov, D. Dereudre, M. El Machkouri, P. Heinrich, R. S. Stoica et V. C. Tran. Invités : P. Calka, A. Illig, D.

Jeulin, K. Kieu, M. N. M. van Lieshout, J. Michel, S. Zuyev, (Lille, France)

- Mécanisme de transport des comètes du nuage d'Oort 28/04/ - 02/05/2008, Lille. Comité organisateurs : M. Fouchard, R. S. Stoica, A. Vienne. Invités : Yu. Davydov, C. Froeschlé, Ch. Froeschlé, S. Liu, H. Rickman, G. B. Valsecchi, (Lille, France)
- Workshop on Spatial Point Processes Modeling and its Applications 04/04 - 08/04/2004. Comité organisateurs : A. Baddeley, P. Gregori, J. Mateu, R. S. Stoica et D. Stoyan³, (Castellon, Espagne)

Edition de livres : suite à la conférence organisée à Castellon en 2004, deux livres ont été édités. Le premier est publié par Publicacions de la Universitat Jaume I et rassemble tous les travaux présentés pendant la conférence. Le deuxième livre est publié par Springer-Verlag et regroupe des études de cas concrets dans lesquels l'outil théorique fondamental est le cadre des processus ponctuels. La sélection des auteurs de ce deuxième livre s'était faite suite à la conférence organisée à Castellon. J'ai été l'un des éditeurs de ces deux livres, en tant que membre du comité organisateur de la conférence.

Organisation des séminaires : au cours de l'année 2002, j'ai été responsable de l'organisation de séminaires bi-mensuels au sein du département "Signal and Image Analysis" au CWI Amsterdam.

3.6 Réalisation et diffusion de logiciels

- La construction du processus ponctuel marqué, Bisous, et la coopération fructueuse avec les cosmologistes estoniens a abouti à la conception d'un logiciel pour la détection des filaments galactiques. La version actuelle du logiciel a été mise à la disposition de la communauté scientifique, afin que l'on puisse utiliser notre modèle et le comparer avec d'autres méthodes. Le logiciel est accessible à partir du lien : <https://github.com/etempel/Bisous>.

Ce logiciel a servi aussi à la création d'un catalogue qui cartographie les filaments galactiques dans notre Univers. Ce catalogue est accessible à partir du lien : <http://cosmodb.to.ee/>.

- MPPLIB est une librairie C++ qui sert à simuler des processus ponctuels marqués. J'ai participé à sa réalisation collective, essentiellement pendant le séjour post-doctoral effectué au CWI. Sa construction, orientée objet, permet son extension à d'autres modèles ou dynamiques de simulation. Son code (\approx 3400 lignes) est disponible sur simple demande. Ce logiciel est gratuit et peut être utilisé dans toute activité non-commerciale se déroulant dans le secteur académique.

4 Responsabilités collectives

Section 26 CNU :

- membre élu sur les listes SNESUP (suppléant) pour le mandat 2011 – 15

Université de Lille - UFR et UMR Mathématiques :

- membre du jury du domaine Mathématiques de l'Ecole Doctorale SPI de l'Université de Lille : comité des bourses doctorales, comité de suivi des thèses
- membre de la commission des services pour l'année 2015-16

3. Lien internet : <http://www.sppa.uji.es>

Vie sociale - UFR Mathématiques :

— membre fondateur de l'Amicale des personnels de l'UFR de Mathématiques – Lille 1"

5 Annexe A : liste complète des publications

Habilitation à diriger des recherches

[1] R. S. Stoica. *Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées*. Université Lille 1, 2014.

Thèse

[2] R. Stoica. *Processus ponctuels pour l'extraction des réseaux linéiques dans les images satellitaires et aériennes*. Université de Nice–Sophia Antipolis, 2001.

Livres édités

[3] A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan (eds.). *Case studies in spatial point process modeling*, Lecture Notes in Statistics by Springer Verlag, 2006.

[4] A. Baddeley, P. Gregori, J. Mateu, R. Stoica and D. Stoyan (eds.). *Spatial point process modelling and its applications*, Publicacions de la Universitat Jaume I, 2004.

Chapitres de livres

[5] R. S. Stoica, E. Tempel, L. J. Liivamägi, G. Castellan and E. Saar. Spatial patterns analysis in cosmology based on marked point processes. *Statistics for astrophysics. Methods and applications of the regression* (eds. D. Fraix-Burnet and D. Valls-Gabaud), European Astronomical Society Publication Series EDP Sciences, 2015.

[6] R. S. Stoica, F. Chatelain and M. Sigelle. Parametric inference for marked point processes in image analysis. *Stochastic Geometry for Image Analysis (Digital Signal and Image Processing series)*(ed. X. Descombes), John Wiley and Sons, 2012.

[7] R. S. Stoica, F. Chatelain and M. Sigelle. Inférence paramétriques pour les processus ponctuels marqués en analyse d'images. *Applications de la géométrie stochastique à l'analyse d'images (Série Signal et Image IC2)* (ed. X. Descombes), Editions Hermès Lavoisier, 2011.

[8] R. S. Stoica. Marked point processes for statistical and morphological analysis of astronomical data. *The European Physical Journal Special Topics*, 186 : 123 - 165, 2010.

[9] R. S. Stoica, X Descombes, M. N. M. van Lieshout and J. Zerubia. An application of marked point processes to the extraction of linear networks from images. *Spatial statistics through applications* (eds. J. Mateu and F. Montes), WIT Press, Southampton, 2002.

Articles soumis à la publication - disponibles sous la forme des rapports de recherche

[10] F. Kleinschroth, J. R. Healey, S. Gourlet-Fleury, F. Mortier and R. S. Stoica. Roadless space is greatly diminished by logging in intact forest landscapes of the Congo Basin. *Preprint ArXiv, arXiv :1602 :01436*,

2016.

[11] R. S. Stoica, S. Liu, L. J. Liivamägi, E. Saar, E. Tempel, F. Deleflie, M. Fouchard, D. Hestroffer, I. Kovalenko and A. Vienne. An integrative approach based on probabilistic modelling and statistical inference for morpho-statistical characterization of astronomical data. *Preprint ArXiv, arXiv :1510 :05553*, 2015.

[12] R. S. Stoica, A. Philippe, P. Gregori and J. Mateu. ABC Shadow algorithm : a tool for statistical analysis of spatial patterns. *Preprint ArXiv, arXiv :1507.04228*, 2015.

[13] E. Tempel, R. S. Stoica, R. Kipper and E. Saar. Bisous model – detecting filamentary pattern in point processes. *Soumis au Astronomy and Computing*, 2015.

Articles dans des revues internationales à comité de lecture

[14] E. Tempel, R. S. Stoica, E. Saar, V. J. Martinez, L. J. Liivamägi and G. Castellan. Detecting filamentary pattern in the cosmic web : a catalogue of filaments for the SDSS. *Monthly Notices of the Royal Astronomical Society*, 438 : 3465-3482, 2014.

[15] K. Kiêu, K. Adamczyk-Chauvat, H. Monod and R. S. Stoica. A completely random T-tessellation model and Gibbsian extensions. *Spatial Statistics*, 6 : 118-138, 2013.

[16] E. Tempel, R. S. Stoica and E. Saar. Evidence for spin alignment of spiral and elliptical galaxies in filaments. *Monthly Notices of the Royal Astronomical Society*, 428 : 1827-1836, 2013.

[17] P. Heinrich, R. S. Stoica and V. C. Tran. Level sets estimation and Vorob'ev expectation of random compact sets. *Spatial Statistics*, 2 : 47-61, 2012.

[18] M. El Machkouri and R. S. Stoica. Asymptotic normality of kernel estimates in a regression model for random fields. *Journal of Nonparametric Statistics*, 22 : 955 - 971, 2010.

[19] M. N. M. van Lieshout and R. S. Stoica. A note on pooling of labels in random fields. *Statistics and Probability Letters*, 80 : 1431 - 1436, 2010.

[20] R. S. Stoica, S. Liu, Yu. Davydov, M. Fouchard, A. Vienne and G. B. Valsecchi. Order statistics and heavy-tail distributions for planetary perturbations on Oort cloud comets. *Astronomy and Astrophysics*, 513 (A14) : 1-9, 2010.

[21] R. S. Stoica, V. J. Martinez and E. Saar. Filaments in observed and mock galaxy catalogues. *Astronomy and Astrophysics*, 510 (A38) : 1-12, 2010.

[22] L. Roques and R. S. Stoica. Species persistence decreases with habitat fragmentation : an analysis in periodic stochastic environments. *Journal of Mathematical Biology*, 55 : 189-205, 2007.

[23] R. S. Stoica, V. Martinez and E. Saar. A three dimensional object point process for cosmic filaments detection. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 56(4) : 459-477, 2007.

[24] R. S. Stoica, E. Gay and A. Kretzschmar. Cluster detection in spatial data based on Monte Carlo inference. *Biometrical Journal*, 49(4) : 505-519, 2007.

- [25] M. N. M. van Lieshout and R. S. Stoica. Perfect simulation for marked point processes. *Computational Statistics and Data Analysis*, 51 : 679-698, 2006.
- [26] R. S. Stoica, P. Gregori and J. Mateu. Simulated annealing and object point processes : tools for analysis of spatial patterns. *Stochastic Processes and their Applications*, 115 : 1860-1882, 2005.
- [27] R. S. Stoica, V. J. Martinez, J. Mateu and E. Saar. Detection of cosmic filaments using the Candy model. *Astronomy and Astrophysics*, 434 : 423-432, 2005.
- [28] R. Stoica, X. Descombes and J. Zerubia. A Gibbs point process for road extraction in remotely sensed images. *International Journal of Computer Vision*, 57(2) : 121-136, 2004.
- [29] M. N. M. van Lieshout and R. S. Stoica. The Candy model : properties and inference. *Statistica Neerlandica*, 57 : 1-30, 2003.
- [30] X. Descombes, R. Stoica, L. Garcin and J. Zerubia. A RJMCMC algorithm for object processes in image processing. *Monte Carlo Methods and Applications*, 7 : 149-156, 2001.

Articles dans des actes de conférence

- [31] R. S. Stoica, S. Liu, L. J. Liivamägi, E. Saar, E. Tempel, F. Deleflie, M. Fouchard, D. Hestroffer, I. Kovalenko and A. Vienne. Pattern detection and characterization for astronomical data through probabilistic modelling and statistical inference. *Proceedings of the 60th World Statistics Congress, ISI 2015*, Rio de Janeiro. A paraître.
- [32] I. D. Kovalenko, R. S. Stoica, N. V. Emelyanov, A. Doressoundiram and D. Hestroffer. Bayesian statistical approach to binary asteroid orbit determination. *Asteroids : New Observations, New Models. Proceedings IAU Symposium No. 318, 2015* (eds. S. Chesley, A. Morbidelli, R. Jedicke and D. Farnocchia). A paraître.
- [33] R. S. Stoica. Morpho-statistical characterization of the cosmic web using marked point processes. *Statistical Challenges in 21st Century Cosmology. Proceedings IAU Symposium No. 306* (eds. A. F. Heavens, J.-L. Starck and A. Krone-Martins), 2015.
- [34] E. Tempel and E. Saar and R. S. Stoica. Evidence for alignment of galaxies in filaments. *Proceedings of the Thirteenth Marcel Grossman Meeting on General Relativity* (eds. K. Rosquist, R. T. Jantzen and R. Ruffini), World Scientific, Singapore, 2013.
- [35] M. Emily and R. S. Stoica. Estimation Monte Carlo dans les processus ponctuels marqués en biologie tissulaire. *Journées des Statistiques*, Bordeaux, 2009.
- [36] P. Gregori, J. Mateu and R. S. Stoica. A marked point process for modelling three dimensional patterns. *Spatial point process modelling and its applications*, Publicacions de la Universitat Jaume I, 91-99, Castellon, Spain, 2004.
- [37] X. Descombes, M. N. M van Lieshout, R. S. Stoica and J. Zerubia. Parameter estimation by a Markov chain Monte Carlo technique for the Candy model. *Invited paper : Proc. IEEE Workshop Statistical*

Signal Processing, Singapore, 2001.

[38] R. Stoica, X. Descombes and J. Zerubia. Road extraction in remote sensed images using a stochastic geometry framework. *Proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods*, Gif-sur-Yvette, France, 2000.

[39] X. Descombes, R. Stoica and J. Zerubia. Two Markov point processes for simulating line networks. *Invited paper : Proceedings IEEE International Conference on Image Processing* Kobe, Japan, 1999.

[40] R. Stoica, J. Zerubia and J. M. Francos. Image retrieval and indexing : a hierarchical approach in computing the distance between textured images. *Proc. IEEE International Conference on Image Processing*, Chicago, 1998.

[41] R. Stoica, J. Zerubia and J. M. Francos. The two-dimensional Wold decomposition for segmentation and indexing in image libraries. *Proc. IEEE International Conference on Acoustics Speech and Signal Processing*, Seattle, 1998.

Rapports de recherche et pre-prints ayant déjà abouti à des publications

[42] E. Tempel, R. S. Stoica, E. Saar, V. J. Martinez, L. J. Liivamägi and G. Castellán. Detecting filamentary pattern in the cosmic web : a catalogue of filaments for the SDSS. *Preprint ArXiv, arXiv :1308.2533v1*, 2013.

[43] K. Kiêu, K. Adamczyk-Chauvat, H. Monod and R. S. Stoica. A completely random T-tessellation model and Gibbsian extensions. *Preprint HAL, hal-00785980*, 2013.

[44] E. Tempel, R. S. Stoica and E. Saar. Evidence for spin alignment of spiral and elliptical galaxies in filaments. *Preprint ArXiv, arXiv :1207.0068v1*, 2012.

[45] P. Heinrich, R. S. Stoica and V. C. Tran. Level sets estimation and Vorob'ev expectation of random compact sets. *Preprint HAL, hal-00495449*, 2010.

[46] M. N. M. van Lieshout and R. S. Stoica. A note on pooling of labels in random fields. *CWI Research report PNA-E0906*, 2009.

[47] R. S. Stoica, V. J. Martinez and E. Saar. Filaments in observed and mock galaxy catalogues. *Publications IRMA Lille, 69-IX*, 2009.

[48] R. S. Stoica, S. Liu, Yu. Davydov, M. Fouchard, A. Vienne and G. B. Valsecchi. Order statistics and heavy-tail distributions for planetary perturbations on Oort cloud comets. *Publications IRMA Lille, 69-VIII*, 2009.

[49] M. El Machkouri and R. S. Stoica. Asymptotic normality of kernel estimates in a regression model for random fields. *Publications IRMA Lille, 68-VIII*, 2008.

[50] R. S. Stoica, V. Martinez and E. Saar. A three dimensional object point process for cosmic filaments detection. *Research report n. 17*, Unité de Biometrie, INRA Avignon, 2006.

- [51] L. Roques and R. S. Stoica. Species persistence decreases with habitat fragmentation : an analysis in periodic stochastic environments. *Research report n. 23*, Unité de Biometrie, INRA Avignon, 2006.
- [52] R. S. Stoica and E. Gay. Cluster detection in spatial data using marked point processes. *Research report n.11*, *Unité de Biometrie*, INRA Avignon, 2005.
- [53] R. S. Stoica, P. Gregori and J. Mateu. Simulated annealing and object point processes : tools for analysis of spatial patterns. *Research report n. 69-2004*, University Jaume I Castellon, 2004.
- [54] R. S. Stoica, V. J. Martinez, J. Mateu and E. Saar. Detection of cosmic filaments using the Candy model. *Research report n. 68-2004*, University Jaume I Castellon, 2004.
- [55] M. N. M. van Lieshout and R. S. Stoica. Exact Metropolis-Hastings sampling for marked point processes using a C++ library. *CWI Research report PNA-0403*, 2004.
- [56] M. N. M. van Lieshout and R. S. Stoica. Perfect simulation for marked point processes. *CWI Research report PNA-0306*, 2003.
- [57] M. N. M. van Lieshout and R. S. Stoica. The Candy model revisited : Markov properties and inference. *CWI Research report PNA-0115*, 2001.
- [58] R. Stoica, X. Descombes and J. Zerubia. A Markov point process for road extraction in remote sensed images. *Rapport de recherche INRIA RR-3929*, 2000.
- [59] R. Stoica, J. Zerubia and J. M. Francos. Indexing and retrieval in multimedia libraries through parametric texture modeling using the 2D Wold decomposition. *Rapport de recherche INRIA RR-3594*, 1998.

Rapports techniques

- [60] A. G. Steenbeek, M. N. M. van Lieshout and R. S. Stoica. with contributions from P. Gregori and K. K. Berthelsen. MPPLIB, a C++ library for marked point processes. CWI, 2002-2003.

6 Annexe B : plan du cours fait dans le cadre de la coopération Erasmus avec l'Université de Tartu - Estonie

1. Title :

Mathematical statistics in physics and astronomy.

2. Annotation :

Spatial data are sets of observations made of elements having two components. The first component gives the coordinates where the observation took place. The second component, represented usually by a multi-dimensional real vector, represents the measures associated at the corresponding location.

Catalogues of celestial bodies and phenomena used in astronomy are natural examples of spatial data. A star catalogue contains the stars positions in the observed region and their attached characteristics. These characteristics may be the temperature, the mass, the radius and the luminosity of the considered star. A gamma-ray burst catalogue may contain the bursts locations in the observed regions to which the corresponding time occurrence and light curve is associated. A galaxy catalogue contains the galaxies positions together with attached characteristics such as brightness, mass and morphological type. Spatial data sets are encountered not only in astronomy and cosmology, but also in particle physics, environmental sciences, image analysis.

The spatial character of the data induces a strong morphological component to the possible answers that might be given to questions issued from the data analysis. This explains why the question almost always arising is what is the pattern hidden in the data ?

3. *Purpose of the course* : The aim of the course is to give to the student basic and advanced tools for performing statistical analysis and characterization of spatial data. These tools are stochastic modelling, MCMC simulation and statistical inference. The combination of all these tools allows the detection and the morpho-statistical characterization of the pattern hidden in the data.

4. *Learning outcomes* : Starting with concrete applications in cosmology, environmental sciences and image analysis, the student will learn about some very general concepts and techniques used in mathematical statistics and spatial data analysis. The course gives also the opportunity to practise some of these techniques on real and simulated data sets, under the R environment, using the spatstat library.

5. *Short description* : the course is structured in four main parts

I Spatial data examples : cosmology, environmental sciences, image analysis

II Stochastic modelling : marked point processes - definitions and tools (Palm distributions, Campbell-Mecke theorems, summary statistics), Poisson processes, Gibbs point processes.

Marked point processes are a mathematical tool that allows statistical inference from data representing random configurations of points having random characteristics. A random point with a random mark may represent a geometrical or more complex object. Hence, marked points processes allow the description of configurations or complex patterns made of interacting objects. The probabilistic character of these models allows a statistical characterisation of the modelled pattern. The geometrical properties of the objects together with their interactions enables a morphological description of the pattern. Therefore, marked point processes are able to describe a pattern from both points

of view, statistical and morphological.

III Monte Carlo simulation : spatial-and-death processes, Metropolis-Hastings dynamics, exact simulation.

The Markov chain Monte Carlo (MCMC) simulation is a general mathematical framework allowing the sampling of probability densities. The main philosophy of MCMC sampling methods is to simulate a Markov chain that has the equilibrium distribution the probability distribution to be sampled. The presentation of this part begins with an intuitive presentation of a MCMC sampler. Basic mathematical notions on general Markov chains are then given. The irreducibility, the Harris recurrence and the geometric ergodicity properties are also explained. Finally, several algorithms able to sample locally stable probability densities of marked point processes are shown.

IV Statistical Inference : parameter estimation (pseudo-likelihood and Monte Carlo likelihood based, complete and incomplete data), statistical tests, global optimization (simulated annealing), Bayesian analysis, pattern detection and results evaluation.

This part shows how the mathematical tools presented before, can be embedded in a "machinery" able to answer the question arisen by the course. Parameter estimation, pattern detection and results evaluation are presented here with the aim to give to the student a general statistical culture useful in solving problems coming from different application domains. These methods will be proved theoretically and practically.

7 Annexe C : bilan et projets futurs de mon activité de recherche

Cette dernière partie présente des conclusions et des perspectives qui découlent de mon activité de recherche⁴.

7.1 Questions concernant la modélisation de forme

7.1.1 Extension à l'infini, percolation et transition de phase

Les processus ponctuels que nous avons construits comportent des interactions entre les objets qui composent la forme. La loi de ces processus dépend des observations par un terme d'attache aux données. Ce terme introduit de la non stationnarité dans le modèle. Les interactions ne portent que sur les paires d'objets, et elles sont symétriques. Elles reposent sur la distance entre objets et leur marques. Le processus qui en résulte est localement stable et markovien au sens de Ripley-Kelly.

La stabilité locale garantit l'intégrabilité du modèle. C'est également une condition nécessaire à la convergence des dynamiques de Metropolis-Hastings. Le caractère markovien nous assure que les interactions sont de portée finie. Cela signifie qu'un objet n'interagit qu'avec les objets situés en deça d'un rayon fixe. Ce rayon dépend des paramètres du modèle. Le théorème de Hammersley-Clifford permet de caractériser ces processus par une mesure de Gibbs.

Les modèles sont définis dans une fenêtre d'observation finie $W \subset \mathbb{R}^d$. Nous souhaitons étudier l'existence de ces mesures quand le domaine d'observation s'étend vers \mathbb{R}^d . Notre motivation première est la cosmologie. En effet, la possibilité d'augmenter l'angle d'observation de notre univers justifie l'utilisation des modèles capables d'être étendus à l'infini.

Il existe déjà dans la littérature des résultats concernant l'extension des processus ponctuels markoviens. Nous souhaitons synthétiser les résultats existants et les adapter à notre cadre, où les données et la superposition des interactions jouent un rôle clef.

De point de vue méthodologique, une telle extension permettrait d'aller vers une fusion de deux courants d'utilisation des processus ponctuels. Le premier courant suppose que l'observation disponible peut s'étendre à l'infini et conduit à des résultats asymptotiques. Le deuxième courant suppose que l'observation n'est disponible que de façon locale et met en jeu des procédures d'inférence à partir de simulations numériques.

L'existence d'une mesure de Gibbs ne garantit pas son unicité. Une transition de phase signifie existence mais non unicité d'une mesure de Gibbs. Il est bien connu qu'une transition de phase se manifeste surtout quand les interactions du modèle sont fortes, ce qui est le cas de nos applications. Pour chaque modèle, il serait donc intéressant de déterminer la gamme de valeurs des paramètres pour laquelle une transition de phase a lieu. Des résultats concernant l'existence d'une transition de phase existent pour quelques modèles, mais les valeurs critiques exactes des paramètres sont inconnues. Nous souhaitons poursuivre ces pistes dans le cadre de nos modèles et nos applications.

Il se pourrait que la présence d'un terme d'attache aux données réduise les effets non désirés d'une transition de phase. Néanmoins, la connaissance des valeurs critiques de paramètres, pourrait largement

4. Cette partie est construite à partir du dernier chapitre de mon habilitation [1]. Ce chapitre contient également des références bibliographiques, que j'ai dû omettre ici, pour simplifier la lecture du présent document.

améliorer notre méthodologie. Nous pensons notamment à la construction de lois a priori judicieuses, quand un cadre bayésien est utilisé pour la détection de structures.

L'extension du domaine des observations demande l'extension des modèles. Il est par conséquent naturel de se demander si les structures détectées se prolongent à l'infini. Ce phénomène est connu sous le nom de percolation. Par exemple en cosmologie, la question se pose de l'existence des filaments galactiques infinis. La percolation continue a déjà été étudiée pour des schémas booléens, ainsi que pour des modèles de type Quermass. Nous souhaitons poursuivre ces voies, afin d'étudier le comportement de nos modèles d'un point de vue théorique, mais aussi par simulation, sous l'influence de paramètres critiques de percolation. Il est peut être possible d'établir la présence de la percolation dans un jeu de données par l'utilisation de procédures d'estimation des paramètres.

7.1.2 Processus de Cox, processus permanents et déterminants

Une fois les données observées, il y a deux manières d'aborder la détection des structures. La première suppose que la forme est la cause des observations. Cette idée se retrouve dans la construction de processus de Cox ou de Neymann-Scott. La deuxième suppose que la forme est conséquence des données observées. Cette idée se retrouve dans la construction de processus ponctuels marqués ou de champs de Markov.

La construction de formes par des processus ponctuels marqués permet de caractériser la forme d'une manière naturelle, à l'aide des statistiques exhaustives du modèle. La caractérisation de formes par des processus de type Cox se fait au moyen des fonctions d'intensité ou de covariance ou encore des mesures moments. Les processus de Cox permettent aussi d'effectuer des simulations conditionnelles une fois la structure connue.

Nous souhaitons rapprocher ces deux manières d'aborder la forme. Dans ce but, un premier pas serait de construire des modèles pour lesquels le terme d'attache aux données serait une loi de probabilité que l'on sait simuler conditionnellement quand la forme est connue.

Ces deux modélisations ont souvent leur loi accessible à une constante de normalisation près. Pour palier cet inconvénient, la modélisation par les processus permanents et déterminants se profile comme un choix possible. Les processus permanents peuvent être vus comme une classe particulière de processus de Cox. Ces processus simulent des populations de points qui ont tendance à s'agréger. Les processus déterminants s'obtiennent en suivant une démarche similaire à celle des processus permanents. Ces processus simulent des populations de points qui ont tendance à se repousser. L'intérêt de ces processus est que leur densité de probabilité est analytiquement accessible.

En conséquence, les processus permanents et déterminants peuvent s'avérer intéressants pour des problèmes de détection de formes. De plus, ces modèles pourraient créer un lien supplémentaire entre les champs aléatoires et les structures induites par les processus ponctuels. Plusieurs questions se posent : comment introduire des marques qui définiraient les objets composant la forme ? Est-il possible de construire des fonctions d'intensité et de covariance qui traduisent des propriétés géométriques de la forme ? Serait-il possible de prendre en compte des attractions et des répulsions simultanément ? La densité de probabilité est obtenue par la troncature du spectre d'une fonction de covariance. Cette approximation doit pouvoir être contrôlée. Cependant, il est important de connaître ses effets si l'inférence est effectuée à partir de cette densité de probabilité. Des procédures d'estimation de paramètres ont été proposées pour les processus déterminants. Là aussi, il serait intéressant de savoir dans quelle mesure les procédures proposées diminuent l'effet mentionné par (Baddeley et Silvermann, 1984).

7.2 Questions concernant la simulation MCMC

La simulation de lois de processus ponctuels marqués par itérations markoviennes est toujours un problème ouvert. En effet, cette approche comporte deux grandes familles de méthodes, la première famille, classique, garantissant la convergence à l'infini, la deuxième famille des méthodes parfaites fournissant avec précision le temps de convergence. La difficulté vient de ce que les méthodes classiques ont besoin d'approximer le temps de convergence, alors que les méthodes parfaites sont fortement dépendantes des paramètres du modèle.

Si l'on doit simuler un processus ponctuel de faibles interactions, et si les propriétés du modèle le permettent, nous pouvons utiliser des méthodes de simulation exacte. Dans ce cas, les temps de convergence sont raisonnables. Cependant, les structures émanent souvent de modèles à interactions fortes. De plus, la procédure de maximisation par recuit simulé requiert un renforcement des interactions de la loi simulée. Dans ce cas, les méthodes classiques de simulation sont plutôt utilisées.

Les perspectives sont ici d'améliorer ces différents aspects.

7.2.1 Méthodes classiques de simulation

Notre choix préféré pour simuler un processus ponctuel est la dynamique de Metropolis-Hastings. Ceci tient à ce qu'il a été possible de construire des algorithmes pour nos modèles qui leur sont bien adaptés (par l'utilisation de noyaux de transition spécifiques) et qui ont de bonnes propriétés de convergence.

Pour simuler un processus ponctuel marqué avec interactions, le point clef est la construction du noyau de transition. Ce noyau met en jeu trois types de transitions possibles : ajouter, enlever et modifier un objet. Si ces transitions s'effectuent en utilisant des lois d'instrumentation suffisamment simples, il est possible de montrer que la chaîne simulée par ce noyau est apériodique, irréductible, récurrente et ergodique. Une telle chaîne a comme unique distribution d'équilibre la loi du processus ponctuel étudié. Les échantillons ainsi obtenus permettent l'utilisation du théorème centrale limite. Il est souvent possible de proposer un noyau de transition un peu plus élaboré, qui aide le modèle à former des structures. Cependant, il n'est pas toujours facile de trouver les bonnes lois d'instrumentation. Ces lois doivent à la fois être adaptées au modèle et garantir toutes les propriétés de convergence de la chaîne simulée.

La construction de dynamiques de simulation qui produisent des échantillons utilisables pour l'inférence statistique est un problème qui se pose à chaque fois qu'un nouveau modèle est proposé.

Dans certaines situations, les dynamiques de Metropolis-Hastings adaptées peuvent être vues comme des dynamiques Monte Carlo à sauts réversibles (RJMCMC). Ces algorithmes ont un réel intérêt quand l'espace d'états n'est pas composé que d'états communicants récurrents. Les algorithmes RJMCMC ont reçu une attention particulière en traitement d'image. Ceci est dû à la capacité de ces dynamiques à effectuer des sauts entre sous-espaces de configuration de dimensions différentes. Elle est parfois confondue avec la capacité du modèle proposé de détecter le nombre d'objets contenus dans l'image. Cette idée est en partie justifiée. Cependant, nous sommes assez d'accord aujourd'hui avec le point de vue qui soutient que l'on rencontre plutôt rarement des preuves des propriétés de convergence des dynamiques RJMCMC. Une exception notable est (Cappé et al., 2003) qui établit de façon rigoureuse l'équivalence entre des dynamiques RJMCMC et les algorithmes de type naissance et mort.

Au sujet de la dynamique de simulation proposée pour simuler des modèles de pavages en T , nous souhaitons enrichir le noyau de transition de l'algorithme en cours pour garantir la récurrence et l'ergodicité de la chaîne simulée. Dans un premier temps, nous pensons essayer l'idée proposée par C. Lantuéjoul d'enlever une composante connexe entière. Cela permettrait d'adopter directement le cadre des dynamiques Metropolis-Hastings et d'appliquer les techniques de preuves de convergence déjà utilisées.

Les dynamiques construites peuvent être utilisés dans le cadre d'un algorithme de Gibbs, permettant ainsi l'échantillonnage de la loi conjointe de la forme et des paramètres. Dans notre cas, les paramètres sont spécifiés par une loi a priori, plaçant notre travail dans la perspective bayésienne. Nous souhaitons continuer dans cette direction en essayant d'intégrer à notre approche de nouvelles idées qui visent à améliorer les propriétés de l'algorithme de simulation.

Des stratégies particulières s'imposent quand le nombre d'objets de la configuration devient particulièrement important. Dans cette situation, la construction de noyaux de transition qui ajoutent et qui enlèvent plusieurs objets à la fois peut être envisagée. Il serait aussi peut être intéressant de s'inspirer des algorithmes de packing, même si leur construction s'appuie en général sur des principes différents de ceux des chaînes de Markov.

En général, le contrôle du comportement de ces dynamiques s'effectue en suivant l'évolution des séries chronologiques de statistiques exhaustives du modèle simulé. A ce propos, nous souhaitons enrichir notre arsenal de techniques en considérant comme point de départ les approches décrites dans (Robert et Casella, 2004).

7.2.2 Simulation exacte

Après la grande émulation de la fin des années 1990 et du début des années 2000, le nombre d'articles sur la simulation exacte est devenu moins important. Les explications données par certains sont la technicité du sujet, la difficulté d'y travailler avec des étudiants et le fait que le sujet soit à risque. Cependant, les perspectives que l'on peut déceler dans la littérature sont bien réelles.

Du point de vue de la simulation des processus ponctuels marqués avec application à la détection des structures, nous sommes intéressés par des méthodes capables d'échantillonner des modèles à interactions fortes. Cela signifie que nous avons besoin d'algorithmes flexibles pour simuler les modèles sur des gammes de paramètres les plus larges possibles.

Nous souhaitons poursuivre l'étude et l'adaptation théorique et pratique des algorithmes existants à notre problème. Un premier pas consisterait à regarder les performances de l'algorithme de Fill sur différents processus ponctuels marqués. Une deuxième idée est d'étudier l'applicabilité aux processus ponctuels des idées de simulation parfaite dans le sens direct. Finalement, une troisième piste serait de considérer les algorithmes de type "simulated tempering". Ces derniers algorithmes ont d'emblée un double avantage. D'abord, ils permettent de simuler des modèles à très basse température tout en réduisant les problèmes liés aux fortes corrélations entre échantillons et au faible mélange de la chaîne simulée. Ces algorithmes permettent aussi d'utiliser des algorithmes qui ont des noyaux de transition adaptés au modèle.

7.3 Questions concernant l'inférence statistique

La "machinerie" que nous avons construite suppose les paramètres du modèle connus. Pour les paramètres d'interaction cette condition a été relâchée en les modélisant par une loi a priori. Mais dans les

deux cas le problème de l'estimation des paramètres se pose.

7.3.1 Détection conjointe de la forme et des paramètres

Dans le cas de modèles à base de champs de Markov, des travaux ont abordé simultanément les problèmes de la détection (ou restauration) de la structure et de l'estimation des paramètres. Le principe de ces méthodes est connu plus souvent sous le nom de gradient stochastique. Un lien entre les méthodes basées sur le gradient stochastique et les algorithmes de type EM a été déjà établi, et le gradient stochastique a été appliqué à l'estimation des paramètres de processus ponctuels.

Nous souhaitons étudier ces approches pour essayer de les appliquer à notre cadre de détection de structures. Cela demandera de les adapter au cadre des processus ponctuels marqués. Un des points importants devrait être l'étude de la vitesse de convergence des dynamiques Monte Carlo mises en place pour simuler le modèle conditionnellement aux paramètres.

7.3.2 Simulation de la loi a posteriori

La simulation de la loi a posteriori $p(\theta|\mathbf{x})$, c'est-à-dire de la loi des paramètres conditionnellement à la forme observée est une question liée au problème précédent. En effet, nous sommes capables de simuler la loi de la forme conditionnellement aux paramètres du modèle, $p(\mathbf{x}|\theta)$. Si l'on est capable de simuler les deux lois conditionnelles, cela implique que l'on peut simuler la loi conjointe $p(\mathbf{x}, \theta)$, qui caractérise à la fois la forme et les paramètres.

La difficulté avec la simulation de la loi a posteriori est qu'elle met en jeu le calcul du rapport des constantes de normalisation. Ce calcul doit être fait à son tour par simulation. Une excellente idée qui élimine ce calcul par l'introduction des variables auxiliaires dans une dynamique de Metropolis-Hastings. L'incorporation dans le taux d'acceptation de la densité de probabilité de la variable auxiliaire simplifie son expression, au sens où le calcul des constantes de normalisation n'est plus nécessaire. De plus, le choix de variables auxiliaires est très large et très peu contraint par les hypothèses de travail.

Cette liberté de choix est à la fois le point fort et le point faible de la méthode. Du point de vue théorique, il est très facile de choisir une variable auxiliaire. De point de vue numérique, les choses sont un peu différentes, au sens où ce choix ne garantit pas forcément de bonnes propriétés de mélange pour la chaîne obtenue en incluant les variables auxiliaires. Concrètement, la chaîne peut rester bloquée dans le même état pendant de longs moments, chose qui nuit sérieusement à la qualité des échantillons obtenus. Les auteurs préviennent de la possibilité de cet effet non désiré sur certains modèles et proposent des heuristiques pour éviter le problème.

En s'inspirant de cette idée, nous avons proposé une méthode qui approxime la loi a posteriori d'une manière contrôlée. Les résultats numériques sont très prometteurs. Nous travaillons en ce moment pour trouver un encadrement théorique à la loi approximée. Un tel résultat permettrait de transposer notre idée à d'autres techniques d'échantillonnage approximé.

7.3.3 Choix de modèle

Valider statistiquement des résultats demande soit la connaissance à priori de la forme qui se cache dans les données, soit être capable de simuler un champ de données connaissant la forme qui s'y trouve. En

analyse d'image, la forme peut être considérée comme connue puisque on peut la voir. Cependant, si la structure de données est plus complexe, il devient plus difficile d'accéder à cette visualisation. Effectuer la démarche inverse, c'est-à-dire simuler des données en supposant connue la forme, peut se faire si l'on prend en compte la structure des données et le modèle de la forme. Ces considérations font que cette question reste pour le moment un problème ouvert.

L'échantillonnage de la loi a posteriori permet également effectuer des tests d'hypothèse en vue d'un choix de modèle.

Pour vérifier les résultats obtenus par notre méthode, nous avons proposé des tests de Monte Carlo basés sur les statistiques exhaustives du modèle. Sous l'hypothèse du modèle, ces tests ont montré qu'une structure était bien présente dans les données. Nous souhaitons poursuivre ce genre de tests en les enrichissant le plus possible pour pouvoir comparer des modèles.

L'intérêt pour ce problème a pris un essor particulier grâce à l'apparition de l'analyse par résidus des processus ponctuels. Les résidus sont définis comme la différence entre la somme d'une fonction aux points d'une configuration et l'intégrale de cette fonction sur tout le domaine. L'outil mathématique à la base de cette notion est la formule de Georgii-Nguyen-Zessin. Il existe également un lien fort entre cette méthode et l'estimation des paramètres par pseudo-vraisemblance. De point de vue applicatif, ce développement s'est concrétisé dans une série d'outils graphiques pour qualifier l'ajustement d'un modèle à un jeu de données. Cela facilite le choix d'un meilleur modèle.

Notre perspective dans cette direction est d'adapter ces idées pour notre problème de reconnaissance de formes. Cela revient d'abord à détecter la structure, puis à estimer ses paramètres et finalement calculer des résidus pour valider le modèle. Il y a indéniablement un lien fort avec les questions liées à l'estimation des paramètres. Techniquement, cela peut demander d'incorporer le terme d'attache aux données dans la mesure de référence. Certains outils existants pourraient aussi être adaptés en vue de valider le terme d'attache aux données, en plus du terme d'interaction.

7.4 Projet : processus stochastiques et géométrie aléatoire

Considérons quelques situations :

- un réseau routier évolue en fonction de l'activité humaine et économique qu'il connecte - sur une durée de plusieurs années des routes apparaissent et disparaissent ;
- les fissures au sol se propagent et se ramifient en formant un réseau - cette structure évolue au cours du temps jusqu'à atteindre une certaine stabilité en fonction de facteurs environnementaux
- lors de la collision entre deux satellites, des débris spatiaux se forment - au tout début, ces débris se déplacent dans l'espace d'une manière plutôt agrégé - puis le mouvement se stabilise, les débris étant repartis sur certaines orbites.

A chaque instant, notre approche de détection et de caractérisation des formes s'applique à chacune de ces situations en répondant à la question "quelle est la forme qui se cache dans les données?". Cependant, la présence de la dimension temporelle apporte plus d'information et permet de poser des nouvelles questions, comme par exemple :

- comment se développent les réseaux routiers en fonction de la croissance des villes ou la diminution des forêts ?
- quelle est la probabilité qu'une fracture géologique atteigne une certaine location ?

- y a-t-il une corrélation entre l'emplacement d'un gisement de minerais et la position de fractures géologiques ?
- quelle est la carte de probabilité d'impact d'un nouveau satellite avec les débris existants ?

Les situations décrites précédemment sont des exemples de processus stochastiques spatiotemporels. A temps fixé, ces processus peuvent être décrits par les outils existants de la géométrie stochastique et des statistiques spatiales, en apportant ainsi des éléments de réponses aux questions formulées. Cependant, l'intégration de ces éléments nécessite un mécanisme mathématiquement robuste. L'ajout de la dimension temporelle impose naturellement d'adapter des outils en provenance de la théorie des processus stochastiques à la description de dynamiques d'objets géométriques aléatoires. Cette idée a été lancée par Yu. Davydov dans le groupe de travail "Géométrie Stochastique" de l'Université de Lille. Et ceci constitue la direction de recherche majeure vers laquelle nous souhaitons nous orienter pour les années à venir.

Nous pensons démarrer sur cette nouvelle voie, en partant comme toujours d'applications et de questions concrètes. Pour répondre à ces questions, nous souhaitons construire des modèles, leur associer des dynamiques de simulation et bâtir des procédures d'inférence statistique. De cette manière, la relation bénéfique d'interdépendance entre les mathématiques et ses domaines d'application se trouve renforcée.

7.4.1 Modélisation spatio-temporelle

Les bases théoriques des processus spatiotemporels sont déjà décrites dans la littérature. Cependant, pour répondre aux questions pratiques il faut étudier des processus plus spécifiques. Nous avons actuellement deux familles de processus en vue : les processus ponctuels séquentiels et les processus de cristallisation.

Les processus ponctuels séquentiels sont des processus ponctuels pour lesquels compte l'ordre des points d'une configuration. Ils ont été introduits par M. N. M. van Lieshout. Ceci permet de s'affranchir des interactions symétriques et d'ajouter le temps comme caractéristique de la forme. Ces processus ont été appliqués en modélisation et analyse d'images vidéo. Une autre application concerne les processus de type absorption stochastique. Par rapport aux processus spatiaux de naissance et de mort, l'absorption stochastique n'est qu'un processus de naissance. En plaçant l'absorption stochastique dans le cadre des processus séquentiels, on obtient un modèle sur lequel le maximum de vraisemblance peut être utilisé pour en estimer les paramètres.

Les processus de cristallisation sont des processus germination-croissance qui évoluent dans le temps. A partir d'une configuration de points initiaux auxquels on a attaché un temps de début de croissance, des formes géométriques appelées grains se développent dans le temps en occupant l'espace environnant. Un point et son grain sont appelés cristal, et le mécanisme de développement du grain est appelée cristallisation. La réalisation d'un tel processus peut être vue comme un ensemble aléatoire, ou champ aléatoire, ou encore pavage si l'on s'intéresse à l'union des frontières des cristaux. Quant le point de vue champ aléatoire est adopté, sous des conditions assez générales portant sur la forme du cristal et sa vitesse de croissance, il est prouvé que ce champ est mélangeant. Ce résultat est appliqué pour estimer des paramètres du modèle.

Pour la suite, nous souhaitons étudier et appliquer ces processus à des problèmes concrets. Nous pensons que ces deux familles de processus offrent un cadre de modélisation commode pour introduire la composante temporelle dans nos problèmes de détection et de caractérisation des structures. Nous sommes

intéressés par la construction de statistiques spatio-temporelles qui permettent de décrire la forme observée. Dans le cas du processus de cristallisation, la considération de germes en interaction est une possible généralisation du modèle. Sur un plan théorique nous aimerions pouvoir établir des liens entre les processus séquentiels et les processus de cristallisation.

Au sujet de l'introduction du temps dans la modélisation d'une structure, il faut mentionner les excellents travaux effectués dans le groupe de recherche animé par E. B. Vedel-Jensen et ses co-auteurs concernant les processus spatio-temporels de Lévy. Ces travaux sont plutôt tournés vers les processus de type Cox, alors que les dynamiques associées relèvent des processus de Lévy. Ils constituent un complément de notre approche et pourraient aider, inspirer voire accompagner d'une manière harmonieuse la voie que nous souhaitons aborder.

La difficulté principale de l'introduction du temps est que dans un espace spatio-temporel on ne peut pas forcément appliquer les mêmes principes de modélisation que sur l'espace des formes. Une forme est vue comme une entité complexe construite à partir d'objets simples en interaction. Peut-on définir une structure dans le temps? Cette structure peut-elle être vue comme une entité composée d'objets simples? Si oui, quels sont ces objets simples? Ce sont des questions qui nous font tous un peu rêver. En même temps, il ne faut pas perdre de vue que des éléments de réponse à ces différents problèmes peuvent être apportés en considérant le temps comme une dimension supplémentaire de l'espace des formes.

7.4.2 Simulation de processus stochastiques des formes

L'ajout d'une dimension temporelle à la modélisation des structures amène à des changements dans les méthodes de simulation. D'autres méthodes, basées sur le filtrage particulaire, devraient aussi être étudiées.

Il est difficile de formuler des perspectives dans cette direction car la dimension temporelle pèse à la fois sur la définition du modèle et sur la complexité de la stratégie de simulation. Ce qui est clair, c'est que nous avons besoin de dynamiques de simulation capables de produire des échantillons de suffisamment bonne qualité pour pouvoir procéder à l'inférence statistique.

7.4.3 Statistiques pour les processus stochastiques des formes

L'utilisation des distributions de Palm et des capacités de Choquet comme outils de construction de statistiques descriptives de configurations de points est bien connue. Nous souhaitons compléter notre approche de détection de structures avec ce genre de statistiques. Cela nous permettrait d'étudier des corrélations entre formes différentes, notamment les formes construites avec des objets de plusieurs types. La généralité mathématique de cette approche donnerait la possibilité de comparer des résultats provenant de différentes méthodes de détection.

Il semble que cette manière de construire des statistiques exploratoires peut être étendue à un espace de formes enrichi d'une dimension temporelle supplémentaire. Comme nous souhaitons ajouter une dimension temporelle aux modèles de formes que nous mettrons en place, cette même dimension temporelle devrait être ajoutée aux statistiques décrivant la forme. Une telle construction doit être réalisée avec prudence et ne doit pas être une extension "automatique" des outils mathématiques : une telle extension risquerait de ne pas avoir de lien réel avec le phénomène étudié.

7.4.4 Outils probabilistes et d'inférence statistique pour caractériser le chaos

Le travail sur la dynamique des comètes se poursuit. En ce moment, nous appliquons notre méthodologie à des modèles plus simples sur lesquels on peut calculer un nombre important de grandeurs. Cela devrait nous permettre de simuler de nouvelles perturbations et de tester l'influence de ces perturbations sur la dynamique des comètes.

A long terme, nous souhaitons tester et consolider notre méthodologie pour qu'elle puisse appliquer à d'autres systèmes dynamiques perturbés.

8 Annexe D : personnes à contacter pour des lettres de recommandation

Youri Davydov
Université Lille 1
Laboratoire Paul Painlevé
59655 Villeneuve d'Ascq Cedex
Tel : + 33 3 20 43 45 75
Fax : + 33 3 20 43 67 74
Email : Youri.Davydov@math.univ-lille1.fr

Xavier Descombes
INRIA - Projet Ariana
2004, route des Lucioles
BP93 06902 Sophia Antipolis, Cedex France
Tel : + 33 4 92 38 76 63
Fax : + 33 4 92 38 76 43
Email : Xavier.Descombes@sophia.inria.fr

André Kretzschmar
Unité de Biométrie, INRA
Domaine St-Paul, site Agroparc 84 914
Avignon Cedex 9, France
Tel : + 33 4 32 72 21 75
Fax : + 33 4 32 72 21 82
Email : ktz@avignon.inra.fr

Marie-Colette van Lieshout
CWI - Department Signals and Images (PNA4)
Kruislaan 413
NL 1098 SJ Amsterdam, Pays-Bas
Tel : + 31 20 5924008
Fax : + 31 20 5924199
E-mail : colette@cwi.nl

Vicent Martinez
Observatori Astronòmic
Universitat de València Edificio de Institutos de Investigación de Paterna
(postal) Apartado de Correos 22085 E-46071 Valencia - Espagne
(courier) Polígono La Coma s/n 46980 Paterna - Valencia - Espagne
Tel : + 34 963 543484
Fax : + 34 963 54374
Email : vicent.martinez@uv.es

Jorge Mateu
Université Jaume I - Département de Mathématiques
Campus Riu Sec
E-12071 Castellon, Espagne
Tel : + 34 964 728391
Fax : + 34 964 728429
Email : mateu@mat.uji.es

Enn Saar
Tartu Observatorium
61602 Tõravere, Estonia
Tel : + 37 27 410 265
Fax : + 37 27 410 205
Email : saar@aai.ee

Alfred Stein
Université de Twente
PO Box 217, 7500 AE Enschede, Pays-Bas
Tel : +31 53 487 4552
Email : a.stein@utwente.nl

Alain Vienne
Université Lille 1
LAL-IMCCE Laboratoire d'Astronomie de Lille 1
1 Impasse de l'Observatoire,
59000 Lille France
Tel : + 33 3 20 52 44 24
Email : alain.vienne@univ-lille1.fr

9 Annexe E : liste des publications transmises

Habilitation à diriger des recherches

- R. S. Stoica. *Modélisation probabiliste et inférence statistique pour l'analyse des données spatialisées*. Université Lille 1, 2014.

Publications récentes

- R. S. Stoica, A. Philippe, P. Gregori and J. Mateu. ABC Shadow algorithm : a tool for statistical analysis of spatial patterns. *Preprint ArXiv, arXiv :1507.04228*, 2015.
- R. S. Stoica, E. Tempel, L. J. Liivamägi, G. Castellan and E. Saar. Spatial patterns analysis in cosmology based on marked point processes. *Statistics for astrophysics. Methods and applications of the regression* (eds. D. Fraix-Burnet and D. Valls-Gabaud), European Astronomical Society Publication Series EDP Sciences, 2015.
- K. Kiêu, K. Adamczyk-Chauvat, H. Monod and R. S. Stoica. A completely random T-tessellation model and Gibbsian extensions. *Spatial Statistics*, 6 : 118-138, 2013.
- P. Heinrich, R. S. Stoica and V. C. Tran. Level sets estimation and Vorob'ev expectation of random compact sets. *Spatial Statistics*, 2 : 47-61, 2012.
- M. El Machkouri and R. S. Stoica. Asymptotic normality of kernel estimates in a regression model for random fields. *Journal of Nonparametric Statistics*, 22 : 955 - 971, 2010.
- M. N. M. van Lieshout and R. S. Stoica. A note on pooling of labels in random fields. *Statistics and Probability Letters*, 80 : 1431 - 1436, 2010.
- R. S. Stoica. Marked point processes for statistical and morphological analysis of astronomical data. *The European Physical Journal Special Topics*, 186 : 123 - 165, 2010.
- R. S. Stoica, S. Liu, Yu. Davydov, M. Fouchard, A. Vienne and G. B. Valsecchi. Order statistics and heavy-tail distributions for planetary perturbations on Oort cloud comets. *Astronomy and Astrophysics*, 513 (A14) : 1-9, 2010.

Publications avant 2010

- L. Roques and R. S. Stoica. Species persistence decreases with habitat fragmentation : an analysis in periodic stochastic environments. *Journal of Mathematical Biology*, 55 : 189-205, 2007.
- R. S. Stoica, E. Gay and A. Kretschmar. Cluster detection in spatial data based on Monte Carlo inference. *Biometrical Journal*, 49(4) : 505-519, 2007.
- M. N. M. van Lieshout and R. S. Stoica. Perfect simulation for marked point processes. *Computational Statistics and Data Analysis*, 51 : 679-698, 2006.

- R. S. Stoica, P. Gregori and J. Mateu. Simulated annealing and object point processes : tools for analysis of spatial patterns. *Stochastic Processes and their Applications*, 115 : 1860-1882, 2005.