

Curriculum vitae

LEONTE (ép. CIUPERCA) Gabriela

Mariée, deux enfants (11 et 16 ans).

Nationalité : française et roumaine.

Fonction actuelle :

Maître de conférences(HDR) en Statistiques, section 26, Université Lyon 1.

Qualification

aux fonctions de Professeurs des Universités, section 26 : 5 février 2010.

Adresse professionnelle :

Université Lyon 1, UMR 5208, Institut Camille Jordan,

Bat. Braconnier, 43, blvd du 11 novembre 1918, F - 69622 Villeurbanne, France,

Mail : Gabriela.Ciuperca@univ-lyon1.fr

Page web : <http://math.univ-lyon1.fr/~gciuperca/>

Cursus universitaire

- | | |
|-------------------------|---|
| 27 novembre 2009 | Habilitation à Diriger des Recherches, Discipline : Statistique,
Titre : " <i>Inférence statistique pour des modèles non-identifiables, Estimation du taux d'entropie, Applications des Statistiques</i> ",
Université Lyon 1. |
| 1992-1996 | Thèse de Doctorat, Discipline : Statistique Appliquée,
Titre : " <i>Modélisation du métabolisme de la glucose</i> ",
Mention "très honorable avec félicitations",
Institut National Agronomique Paris. Sous la direction du Professeur
Richard Tomassone. |
| 1991-1992 | DEA " <i>Modélisation Stochastique et Statistique</i> ", Mention "Assez Bien"
Université Paris XI Orsay. Mémoire sous la direction du Professeur Camille DUBY. |
| 1984-1988 | Maîtrise de Mathématique et d'Informatique, Mention "Très Bien"
Faculté de Mathématique, Université de Iasi, Roumanie. |

Carrière

- | | |
|-------------------------|--|
| à partir de 2001 | Maître de Conférences en Statistiques, Département de Mathématiques, Univ. Lyon 1, |
| 1998-2001 | Maître de Conférences en Statistiques, IUT STID de Poitiers, |
| 1997-1998 | ATER, Faculté de Sciences Economiques, Université Paris 13, |
| 1996-1997 | Post-doc Université Paris XI et AIRPARIF, |
| 1990-1991 | Maître Assistant en Informatique, Université Bacau, Roumanie. |
| 1988-1990 | Analyste-programmeur en Entreprise, Roumanie. |

Affiliations

- Membre *élue*, en août 2005, de l'*International Statistical Institut (ISI)*.
 - Membre de l'*ERCIM Working Group* (The European Research Consortium for Informatics and Mathematics)-*Mixture Models, High-dimensional statistics*.
-

Activité éditoriale

Associate Editor à la revue *Journal of Mathematics Research*.

Thèmes de recherche :

I. Statistique mathématique (estimations, tests d'hypothèse, simulations, applications données concrets :

- sélection automatique des variables, modèles en très grande dimension, méthodes LASSO
- modèles avec change-points
- modèles quantile, outliers, estimation robuste
- reconstitution des données manquantes
- modèles avec longue mémoire
- maximum de vraisemblance empirique
- mélanges de lois
- estimation des taux d'entropie pour des chaînes/processus de Markov

II. Applications des statistiques à la médecine (neurosciences, oncologie), biologie, industrie, transport, architecture, pollution de l'air. Parmi les méthodes statistiques utilisées (dans : thèse de doctorat, encadrement doctoral, collaborations, contrats)

- data mining, analyse de données (ACP, AFD, ACC, AFC), classification,
 - séries chronologiques
 - Chaînes de Markov, chaînes de Markov cachées
 - régressions : linéaires, non linéaires, logistiques, mixtes, censurées, PLS, sur les composantes principales
 - analyse de variance, analyse de covariance
 - plan d'expérience optimal
 - estimation/étude des paramètres d'un système d'équations différentielles non linéaires
-

Logiciels : R, SAS, Matlab.

Primes/distinctions

Bénéficiaire de la Prime d'Excellence Scientifique (PES), depuis octobre 2012, section 26.

Bénéficiaire de la Prime d'Encadrement Doctoral et de Recherche (PEDR) : octobre 2006-octobre 2010.

1 Publications, Contrats de Recherche, Encadrement, Expertise

1.1 Publications

1.1.1 Articles publiés, à paraître

Statistique mathématique (estimations, tests d'hypothèse, simulations)

1. Ciuperca G.(2014), "Model selection by LASSO methods in a change-point model", à paraître *Statistical Papers*, Vol. 55, No. 2, p. 349-374, 2014.
2. Ciuperca G.(2013), "Two tests for sequential detection of a change-point in a nonlinear model", *Journal of Statistical Planning and Inference*, Vol. 143, Issue 10, p. 1621-1834, 2013.
3. Ciuperca G.(2013), "Quantile regression in high-dimension with breaking", *Journal of Statistical Theory and Applications*, Vol.12, No.3, p. 288-305, 2013.
4. Ciuperca G.(2013), "Empirical likelihood for nonlinear model with missing responses", *Journal of Statistical Computation and Simulation*, Vol. 83, No. 4, p. 737-756, 2013.
5. Ciuperca G.(2012), "The S-estimator in change-point random model with long-memory", *Statistics*, Vol. 46, No. 6, p. 697-718, 2012.
6. Ciuperca G.(2011), "A general criterion to determine the number of change-points", *Statistics and Probability Letters*, Vol. 81, No. 8, p. 1267-1275, 2011.
7. Ciuperca G., Girardin V., Lhote L., (2011), "Computation and estimation of generalized entropy rates for denumerable Markov chains", *IEEE Transactions on Information Theory*, Vol. 57, No. 7, p. 4026-4034, 2011.
8. Ciuperca G.(2011), "Asymptotic behaviour of the LS estimator in a nonlinear model with long memory", *Journal of the Korean Statistical Society*, Vol. 40, p. 193-203, 2011.
9. Ciuperca G.(2011), "Penalized least absolute deviations estimation for nonlinear model with change-points", *Statistical Papers*, Vol. 52, No. 2, p. 371-390, 2011.
10. Ciuperca G.(2011), "Estimating nonlinear model with and without change-points by the LAD method", *Annals of the Institute of Statistical Mathematics*, Vol. 63, No. 4, p. 717-743, 2011.
11. Ciuperca G., Mercadier C.,(2010), "Semi-parametric estimation for heavy tailed distributions", *Extremes*, Vol. 13, No. 1, p. 55-87, 2010.
12. Ciuperca G.(2009), "The M-estimator in a multi-phase random regression model", *Statistics and Probability Letters*, Vol. 75, No. 5, p. 573-580, 2009.
13. Ciuperca G., Dapzol N.(2008), "Maximum Likelihood Estimator in a Multi-phase Random Regression Model", *Statistics*, Vol. 42, No. 4, p. 363-381, 2008.
14. Ciuperca G., Girardin V.(2007), "Estimation of the entropy rate of a countable Markov Chain", *Communications in Statistics- Theory and Methods*, Vol. 36, No. 14, p. 2493-2508, 2007.
15. Ciuperca G.(2004), "Maximum Likelihood Estimator in a two-phase Nonlinear Random Regression Model", *Statistics and Decision*, Vol. 22, No. 4, p. 335-349, 2004.
16. Ciuperca G., Ridolfi A., Idier J.(2003), "Penalized Maximum Likelihood Estimator for Normal Mixtures", *Scandinavian Journal of Statistics*, Vol. 30, No. 1, p. 45-59, 2003.
17. Ciuperca G.(2002), "Likelihood Ratio Statistic for Exponential Mixtures", *Annals of the Institute of Statistical Mathematics*, Vol. 54, No. 3, p. 585-594, 2002.
18. Ciuperca G.(1999), "Sur le test de maximum de vraisemblance pour le mélange de populations", *C.R.A.S.*, Série I, Tome 328, No. 4, Février 1999.

Applications des statistiques

19. Rouby C., Thomas-Danguin T., Vigouroux M., Ciuperca G., Jiang T., Alexanian J., Barges M., Gallice I., Demolis M., Degraix J.L., Sicard G.,(2011) "The Lyon Clinical Olfactory Test : validation and measurement of hyposmia and anosmia in healthy and demented populations", *International Journal of Otolaryngology*, Vol. 2011, 9 pages, DOI :10.1155/2011/203805, 2011.
20. Chambon V., Franck N., Koechlin E., Fakra E., Ciuperca G., Azorin J.-M., Farrer C.(2008), "The architecture of cognitive control in schizophrenia", *Brain*, 131, p. 962–970, 2008.
21. Plesa A., Ciuperca G., Louvet V., Pujo-Menjouet L., Génieys S., Dumontet C., Thomas X., Volpert V.(2006), "Diagnostics of the AML with immunophenotypical data", *Mathematical Modelling of Natural Phenomena*, Vol. 1 No. 2, p. 104–123, 2006.
22. Bel L., Bellanger L., Bonneau V, Ciuperca G., Dacunha-Castelle D., Deniau C., Ghattas B., Misiti M., Misiti Y., Oppenheim G., Poggi J.M., Tomassone R.(1999), "Éléments de comparaison de prévisions statistiques des pics d'ozone", *Revue de Statistique Appliquée*, XLVII(3), p. 7–25, 1999.
23. Bel L., Bellanger L., Ciuperca G., Dacunha-Castelle D., Gilibert E., Jakubowicz P., Oppenheim G., Tomassone R.(1998), "On Forecasting Ozone Episodes in the Paris Area", *Biometrical Letters*, Vol. 35, No. 1, p. 37–66, 1998.
24. Ciuperca G.(1998), "A method to treat the dynamical Statistical Models", *Journal of Biological Systems*, Vol. 6, No. 4, p. 357–375, 1998.
25. Ciuperca G.(1998), "Influence de la matrice de covariances dans des modèles décrits par un système d'équations différentielles", *Revue de Statistique Appliquée*, XLVI(2), p. 59–81, 1998.
26. Ciuperca G., Tomassone R., Flandrois J.P.(1996), "Use of mathematics and statistics in nutrition modelling", *Annales de zootechnie*, Vol. 45, p. 143-152, 1996.

1.1.2 Articles soumis ou en révision

27. Ciuperca G., "Adaptive LASSO model selection in a multiphase quantile regression", soumis.
28. Ciuperca G., Salloum Z. "Empirical likelihood test in a posteriori change-point nonlinear model", en révision.
29. Ciuperca G., "Estimation in a change-point nonlinear quantile model", en révision.
30. Ciuperca G., "Real time change-point detection in a model by adaptive LASSO and CUSUM", soumis.

1.2 Exposés

1.2.1 Conférences

1. "Réduction de dimension et sélection de modèle. Applications en génétique, industrie du pétrole", *Journée Scientifique 2014 FST, Lyon-Thématique BIG DATA, 25 juin 2014, Lyon, France.*
2. "Modèles avec change-points par LASSO adaptatif", *12e Colloque Franco-Roumain de Mathématiques Appliquées, 25-30 août 2014, Lyon, France.*
3. "Sélection de modèle par des méthodes de type LASSO dans un modèle avec change-points", *44èmes Journées de Statistique, 21-25 mai 2012, Bruxelles, Belgique.*

4. "Contribution au diagnostique de la leucémie LAM à l'aide d'une étude de données immunophénotypiques", *3ème Journée d'Emergence de l'Institut de Cancérologie Gustave-Roussy(Villejuif)*, 13 mai 2011. **Sur invitation.**
5. "Estimation robuste dans un modèle paramétrique avec rupture", *41èmes Journées de Statistique*, 25-29 mai 2009, Bordeaux, France.
6. "Weighted least Squares Estimation of the Extreme Value Index", *Joint Meeting of the Statistical Society of Canada and the Société Française de Statistique*, 25-28 mai 2008, Ottawa(Canada). Co-auteur avec C. Mercadier. **Sur invitation.**
7. "On the estimation of the entropy rate of finite Markov chains ", *XIth International Symposium on ASMDA (Applied Stochastic Models and Data Analysis)*, 17-20 mai 2005, Brest, France. Co-auteur avec V. Girardin.
8. "Estimateur du maximum de vraisemblance pénalisé pour des mélanges de densités dégénérées", *33èmes Journées de Statistique-A.S.U.*, 14-18 mai 2001, Nantes, France.
9. "Prévision de la concentration d'ozone dans la région parisienne", *XXIXèmes Journées de Statistique*, 26-30 mai 1997, Carcassonne, France.
10. "Use of Mathematics and Statistics in Nutrition Modelling", *IV th International Symposium on the nutrition of herbivores*, 8-9 septembre 1995, Paris. Co-auteur avec Tomassone R., Flandrois J.P.

1.2.2 Séminaires

1. "Quelques modèles statistiques en très grande dimension. Réduction de dimension et sélection de modèle", *Séminaire équipe ERIC, Université Lyon 2*, 17 février 2014.
2. "Sélection de modèle dans un modèle avec change-points", *Séminaire de statistiques, Institut Elie Cartan, Université de Lorraine (Nancy)*, 21 juin 2013.
3. "Inférence statistique dans un modèle de rupture : de l'indépendance à la longue mémoire", *Rencontres Statistiques Lyonnaises*, 4 février 2011.
4. "Inférence statistique dans un modèle de rupture : de l'indépendance à la longue mémoire", *Séminaire du LJK-Statistique Grenoble*, 14 octobre 2010.
5. "Inférence statistique dans un modèle de rupture", *Séminaire de Statistique I3M Montpellier*, 10 mai 2010.
6. "Estimation de mélanges gaussiens", *Séminaire de Probabilités, Université Lyon 1*, 8 novembre 2001.
7. "Maximum de vraisemblance pour des mélanges de densités", *Séminaire de Probabilités-Statistique, Université Grenoble 1*, 3 mai 2001.
8. "Maximum de vraisemblance pénalisé pour des mélanges gaussiens", *Groupe de travail de l'IUT - STID Paris V*, 23 mars 2000.
9. "Maximum de vraisemblance pour des mélanges de densités", *Séminaire de Probabilités, Université Lyon 1*, 18 janvier 2001.
10. "Test de vraisemblance pour des mélanges de populations, Prévision des pics de pollution", *Séminaire de Probabilités, Université Lyon 1*, 19 mai 1998.
11. "Test du maximum de vraisemblance pour des mélanges sur le paramètre de translation", *Séminaire de Probabilités et Statistique, Université Paris 13 Villetaneuse*, 14 janvier 1998.
12. "Prévision des pics de pollution d'ozone dans la région parisienne", *Séminaire à l'Institut Français du Pétrole*, 30 mai 1997.

13. "Méthodes statistiques pour la prévision des pointes de pollution d'ozone dans la région parisienne", *Séminaire de Statistique, Université de Marseille Luminy, 28 avril 1997.*
14. "La prévision de la pollution de l'ozone", *Journée "AIRPARIF-ORSAY", Exposé des travaux, 12 décembre 1996.*
15. "Modélisation statistique du métabolisme du glucose chez la chèvre", *Séminaire Biométrie, Université de Lyon 1, 20 novembre 1996.*
16. "Une méthode pour traiter les modèles statistiques paramétriques décrits par un système différentiel non linéaire", *Séminaire Institut National Agronomique Paris, 26 avril 1994.*

1.3 Participation à des contrats/groupes de recherche interdisciplinaires

J'ai collaboré avec des chercheurs ou avec des ingénieurs du milieu industriel d'horizons différents et sur des sujets variés (en ordre chronologique).

1. **Contrat Consultance entre G. Ciuperca et A. Perrut pour l'Université Lyon 1 et l'École Nationale d'Architecture de Lyon** dans les domaines : *des statistiques, de la modélisation statistique en très grande dimension, analyse de données et data mining.* A partir de janvier 2014.
2. Participation aux **Journées Maths-Entreprise** organisées par le GDR avec le même nom, à l'Université Lyon 1, entre *28 novembre et 2 décembre 2011.* J'ai supervisé deux groupes de doctorands travaillant sur des sujets proposés par 2 entreprises : *Rio Tinto Alcan-Rhône-Alpes* ("Anticiper la formation des défauts sur des anodes des cuves à l'électrolyse utilisées pour fabriquer l'aluminium"), *Oberthur* ("Points d'intérêt dans les courbes liées à un calcul cryptographique et implémentation en boîte blanche d'algorithmes cryptographiques").
3. **Partenariat industriel Université Lyon 1/ Volvo Truck** (*période avril-octobre 2010*) (G.Ciuperca et A. Perrut pour Univ. Lyon 1). Sujet : "Générateur des cycles par des méthodes statistiques". Application immédiate des résultats obtenus.
4. **GDR "Mascot NUM"** (Méthodes d'Analyse Stochastique pour les Codes et Traitements numériques). L'objectif principal de ce groupe de recherche (national) est de coordonner les efforts de recherche mathématique dans la modélisation et dans l'informatique. Ce GDR est mise en place *depuis le mois de mai 2008.*
5. **BQR "Diagnostique et modélisation des leucémies"** (*période 2005-2006*) basé sur la collaboration entre des mathématiciens de l'UMR 5208 (UCBL-ECL-INSA) et des médecins du service d'Hématologie Clinique (UCBL). A la suite de ce BQR, l'**article 15.** a été publié. (voir aussi Conférence **2.**).
6. **Contrat de recherche entre l'Université Paris XI Orsay et AIRPARIF** (Association de surveillance de la qualité de l'air en Ile de France) (*période septembre 1995 et août 1997*) sur la prévision des pointes de pollution dans la région parisienne. A la suite de ce contrat, les **articles 22.** et **23.** ont été publiés et surtout, la prévision de la pollution d'Ile de France est depuis opérationnelle.
7. Pendant ma thèse de doctorat, *entre 1994 et 1995,* j'ai participé à **deux courts contrats de recherche** sur la modélisation de la conservation dans l'**industrie agro-alimentaire** (voir les Parties 2.2.2 et 2.2.3)
8. Autres **collaborations avec des chercheurs non mathématiciens :**

- Médecins appartenant à l'UMR 5229-Lyon, INSERM-Université Paris 6, CHU-Marseille (**article 20.**).
- Chercheurs en neurosciences de l'Université Lyon 1, étude statistique sur un test clinique olfactif (**article 19.**).

1.4 Encadrement

J'ai privilégié l'encadrement des projets, stages, thèses, dans le domaine des applications des statistiques, sans oublier l'aspect théorique de la modélisation, sans lequel on risque d'utiliser un modèle inadéquat aux données et au problème posé.

1.4.1 Encadrement doctoral

1. Janvier 2003-avril 2006, Thèse en cotutelle : Université Lyon 1-INRETS/Bron.

J'ai codirigé, avec Hélène Tattegrain-Veste(INRETS) la thèse en Statistique de Nicolas DAPZOL "*Analyse de l'activité de conduite par les chaînes de Markov et les modèles de rupture multi-phase. Méthodologie et applications*". La thèse comportait des aspects théoriques sur des modèles avec change-points et des aspects appliqués pour modéliser la conduite d'un automobile. Par des techniques de modélisation complexes (change-points, Chaînes de Markov, semi-Markov cachées) on a d'abord modélisé, en utilisant une base de données d'apprentissage, des situations de conduite automobile. Ensuite, on a prévu les situations de conduite en temps réel ou à posteriori, avec le but principal de prévoir les décisions à risque du conducteur auto.

La thèse de N. Dapzol a reçu le prix : "*Young Researcher Seminar of the European Conference of Transport Research Institute*". Comme le nom l'indique, ce prix a été décerné à la Conférence : "European Conference of Transport Research Institute", La Haye, Pays-Bas, du 11 au 13 mai 2005.

Après sa thèse Nicolas Dapzol a été recruté en tant que Chargé de Recherche à INRETS. A la suite de ce co-encadrement, l'**article 13.** a été publié.

2. Depuis octobre 2012, je dirige la thèse de Zahraa SALLOUM (bourse allocataire de recherche) sur le sujet "*Détection de changements dans un modèle, avec un nombre faible ou très grand de variables*". La thèse comporte des aspects théoriques et numériques, pour tester en temps réel si un changement a eu lieu dans un modèle général non linéaire. La technique non-paramétrique du maximum de vraisemblance empirique est utilisée d'abord pour un modèle avec un nombre fixe de variables et ensuite avec un nombre très grand (qui converge à l'infini) de variables.

A la suite de cet encadrement l'**article 28.** a été soumis pour publication, en décembre 2013. Un deuxième article "Empirical likelihood confidence regions for the parameters of a two phases nonlinear model with and without missing response data" a été soumis par Z. Salloum. Un troisième article est en cours de rédaction.

1.4.2 Encadrement Mémoire Master Recherche (DEA)

1. Delphine RIOLI, élève de l'ENS-Lyon, DEA de Mathématiques(2003/2004), mémoire "*Vitesse de convergence dans le Théorème Central Limite pour des suites de variables aléatoires faiblement dépendantes*".

2. **Rima HADDAD, étudiante à l'ISFA LYON, M2R(2009/2010)**, mémoire "*Détection d'un changement dans un modèle à longue mémoire. Application à l'étude de l'inflation*".

1.4.3 Encadrement Master Professionnel et Stages Industriels

1. Depuis mon arrivée en 2001 à l'Université Lyon 1 j'assure, chaque année, l'**encadrement et le suivi** (visite en entreprise, conseils et suivi sur l'adéquation des méthodes statistiques utilisées, participation aux jury de soutenance) des **stages en entreprise des étudiants en deuxième année du Master Professionnalisant** : "Statistique, informatique et techniques numériques", du Département de Mathématiques, Université Lyon 1.

Les stages, de statistique appliquée, couvrent tous les domaines d'activité, par exemple :

- *biostatistique et santé* (BIOMERIEUX, SANOFI PASTEUR, INRA, HOSPICES CIVILS DE LYON, SOLADIS, A+A, ...)
- *transports* (RENAULT TRUCKS, INRETS, VEOLIA, ...)
- *hydrologie* (CEMAGREF, VEOLIA EAU, ...)
- *finances* (banques CREDIT AGRICOL, ...)
- *énergétique* (EDF)

2. **J'encadre des projets TER (Travaux d'Encadrement et de Recherche) de la première année du Master Professionnalisant "Master Ingénierie Mathématique" - Université Lyon 1** : année 2007/2008 : 5 TER, année 2008/2009 : 4 TER, année 2009/2010 : 4 TER, année 2010/2011 : 4 TER, année 2011/2012 : 3 TER (dont un sous forme de stage en entreprise : VEOLIA), année 2012/2013 : 4 TER, année 2013/2014 : 4 TER.

Quelques sujets proposés les dernières années :

- année 2013/2014 :
 - "*Classification supervisée par arbre de décision et réseau de neurones. Application au cancer du sein.*"
 - "*Introduction aux SVM (Support Vector Machines). Applications aux données médicales.*"
 - "*Modèles en très grande dimension avec des outliers. Théorie, simulations, applications.*"
 - *Ondelettes. Théorie et applications.*
- année 2012/2013 :
 - "*Modèle linéaire pour les données fonctionnelles. Applications.*"
 - "*Méthodes de type LASSO pour modélisation et sélection de variables en très grande dimension. Application au cancer de la prostate.*"
 - "*Régression pour des données de type catégorie. Applications sur des données : auto, médecine, économétrie.*"
 - "*Modèles linéaires mixtes pour des données longitudinales. Applications.*"
- année 2011/2012 :
 - "*Modèle logistique : binaire, polytomique, avec covariables dépendantes de temps. Applications.*"
 - "*Estimation non-paramétrique d'une densité. Estimation des données manquantes. Cas particulier : dans un modèle de régression linéaire par des méthodes semi-paramétriques.*"
 - "*Modélisation statistique de la consommation en carburant des bus. Modélisation de la conduite des bus.*" (Stage en entreprise)
- année 2010/2011 :
 - "*Méthodes de sélection de variables dans un modèle linéaire. Applications.*"
 - "*Méthodes de type CART. Applications.*"

- “*Quelques méthodes de classification. Applications*”.
 - “*Séries chronologiques avec applications*”.
 - année 2009/2010 :
 - “*Etude d’un modèle de régression linéaire avec données manquantes. Etude théorique, simulations, applications*”.
 - “*Modèles de panel linéaire avec applications en économétrie*”.
 - “*Modèles avec variable dépendante discrète. Applications en économétrie*”.
 - année 2008/2009 :
 - “*Méthodes statistiques pour des données de survie. Application au cancer*”.
 - “*Processus à longue mémoire ARFIMA. Application à la modélisation des ondes de l’océan*”.
 - “*Régression PLS. Applications*”.
 - “*Méthode CART. Application à la modélisation de la pollution de l’air*”.
 - “*Séries chronologiques avec applications en économétrie*”.
3. **2010/2011, projet de troisième année à Polytech Lyon** “*Modélisation de la pollution. Comparaison de méthodes*”. Les méthodes statistiques utilisées : modèles linéaires, analyse de données (factorielle), arbres de décision CART. La pollution de NO_2 modélisée est celle de la ville d’Oslo (Norvège).
 4. **2007/2008, encadrement de trois stages en entreprise pour des étudiants en deuxième année du Master Pro. de Neurosciences** “*Métrologie de la perception*” de l’Université Lyon 1. Les stages étaient dans le domaine des neurosciences : goût (DANONE), olfactif.
 5. **2001/2002, encadrement stage chez Renault-Véhicules Industriels**, d’un étudiant d’un étudiant en troisième année de l’Institut des Sciences et Techniques de l’Ingénieur de Lyon (ISTIL). Le stage a abouti à la modélisation, fonction du parcours effectué, de la consommation en carburant des bus.
 6. Encadrement **étudiants de l’IUT-STID Niort** (Université de Poitiers), pendant la période **septembre 1998- septembre 2001**, pendant laquelle j’ai exercé comme Maître de Conférences en Statistique :
 - j’ai encadré et suivi des étudiants de deuxième année en **stage d’entreprise** ;
 - j’ai **piloté des projets** pour les étudiants de deuxième année.

1.5 Autres activités scientifiques : Expertises et conseils scientifiques

1.5.1 Expertise internationale

1. Depuis *octobre 2011*, je fais partie de l'Editorial Board de la revue " *Journal of Mathematics Research*", publiée par Canadian Center of Sciences and Education.
Depuis 7 mars 2012 je suis aussi **Associate Editor** à la même revue.
2. Depuis *septembre 2010*, je suis **expert-évaluateur auprès des "Fonds de la Recherche Scientifique-FNRS à Bruxelles"**.
 - Pour 2012 j'ai évalué deux dossiers pour des bourses et mandat du FRS-FNRS.
 - Pour 2013 j'ai évalué deux dossiers (pour projet de recherche et pour concours de chargé de recherche).
 - Cette année, 2014, j'ai évalué deux dossiers (pour projet de recherche et pour concours de chargé de recherche).
3. **Referee** pour des revues internationales de statistique ou de modélisation :
 - Annals of the Institute of Statistical Mathematics : 2012 (1 article)
 - Canadian Journal of Statistics : 2014 (2 articles)
 - Communications in Statistics-Theory and Methods : 2005 (1 article), 2012 (2 articles), 2013 (1 article)
 - Communications in Statistics-Simulation and Computation : 2009 (1 article), 2011 (1 article)
 - Computational Statistics and Data Analysis : 2012 (1 article)
 - Electronic Journal of Statistics : 2013 (1 article)
 - Entropy : 2013 (1 article)
 - International Journal of Computer Mathematics : 2014 (1 article)
 - Journal of Statistical Computation and Simulation : 2014 (1 article)
 - Journal of Statistical Planning and Inference : 2009 (1 article), 2010 (1 article), 2014 (1 article)
 - Journal of Multivariate Analysis : 2010 (1 article), 2012 (1 article)
 - Journal of Nonparametric Statistics : 2010 (1 article)
 - Journal of Mathematics Research : 2011 (5 articles), 2012 (13 articles), 2013 (4 articles), 2014 (1 article)
 - Journal of the Royal Statistical Society : 2013 (1 article)
 - Journal of Statistical Computation and Simulation : 2012 (1 article), 2013 (1 article)
 - Pattern Recognition : 2011 (1 article)
 - Scandinavian Journal of Statistics : 2007 (1 article)
 - Selected Papers of SPE 2011 (the XIX Annual Congress of the Portuguese Statistical Society) : 2012 (1 article)
 - Statistical Papers : 2012 (2 articles)
 - Statistics : 2013 (1 article), 2014 (1 article).Je suis également reviewer pour *Mathematical Review*.

1.5.2 Expertise nationale

1. – Entre *septembre 2004 et septembre 2008*, j'ai été **membre de la Commission des Spécialistes**, section 25/26, Université Lyon 1.
 - Entre *septembre 2008 et septembre 2012*, j'ai fait partie du **Comité Consultatif** de l'Université Lyon 1, section 25/26.

- **Membre des Comités de Sélection McF** : 2 postes en 2009, 2 postes en 2010 dont un poste Chaire, 1 poste en 2011, 1 poste en 2012.
- 2. *Septembre 2010*, **membre du jury, en qualité d’expert, pour un concours externe d’Ingénieur de recherche CNRS.**
- 3. *27 octobre 2010*, **membre jury de thèse de doctorat** ”*Modélisation de l’incertitude sur les trajectoires d’avions*”, soutenue par Nobert FUEMKEU à INRETS/Université Lyon 1.
- 4. *Novembre 2010*, **expertise scientifique d’un dossier de bourse CIFRE.**

1.5.3 Autres

J’ai fait partie du Comité d’organisation pour ”*Le 12ème Colloque Franco-Roumain en Mathématiques Appliquées*”, Université de Lyon, 25-30 août 2014.

2 Présentation des thèmes de recherche

Mes recherches portent à la fois sur l'étude théoriques des modèles, avec simulations et applications et sur les applications des statistiques.

2.1 Statistique mathématique

2.1.1 Modèles en grande et très grande dimension, sélection automatique des variables

Supposons qu'on étudie une variable Y fonction d'un nombre très grand de variables X_1, \dots, X_p : $Y_i = \sum_{j=1}^p \beta_j X_{ji} + \varepsilon_i$, $i = 1, \dots, n$, avec la possibilité que le nombre de variables explicatives soit plus grand que le nombre d'observations ($n < p$). Le cas classique de régression ne fonctionne pas si $n < p$ (modèle en très grande dimension). Dans le cas $n > p$, même si on arrive à calculer les estimateurs pour les paramètres β_j , pour sélectionner les variables (étudier celles qui ont une influence significative sur Y), il faut passer par des tests d'hypothèse. Chose infaisable si p dépasse quelques dizaines (modèle en grande dimension) et surtout les tests d'hypothèse sont instables si les variables X_j sont corrélées entre elles. Une solution, très récente, pour faire de la sélection automatique (*propriétés oracle*) :

- les paramètres β_j qui ont une vraie valeur différente de zéro, auront un estimateur différent de zéro ;
- les paramètres β_j qui sont zéro (donc la variable X_j n'influe pas Y) seront directement estimés comme 0, sans passer par des tests d'hypothèse (avec une probabilité convergeant vers 1), est de minimiser, par rapport à $(\beta_j)_{1 \leq j \leq p}$, la fonction objectif

$$\sum_{i=1}^n (Y_i - \sum_{j=1}^p \beta_j X_{ji})^2 + \lambda_n \sum_{j=1}^p |\beta_j|.$$

Cette méthode s'appelle méthode LASSO. Mais, pour sélectionner automatiquement les variables, il faut que les données (X_{ji}) satisfassent certaines conditions.

Dans l'**article 27.**, j'ai proposé une méthode adaptative générale qui ne demande pas ces conditions :

$$(\hat{b}_n^*, \hat{\beta}_n^*) \equiv \arg \min_{(b, \beta)} \left(\sum_{i=1}^n \rho_\tau(Y_i - b - \mathbf{X}_i^t \beta) + \lambda_n \hat{\omega}_n^t |\beta| \right),$$

avec $\hat{\omega}_n = (\hat{\omega}_{n,1}, \dots, \hat{\omega}_{n,p}) \equiv \frac{1}{|\hat{\beta}_n|^g} = \left(\frac{1}{|\hat{\phi}_{n,1}|^g}, \dots, \frac{1}{|\hat{\phi}_{n,p}|^g} \right)$, $\hat{\beta}_n$ l'estimateur quantile. La fonction $\rho_\tau(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ est définie par $\rho_\tau(r) = r[\tau \mathbb{1}_{r>0} - (1-\tau) \mathbb{1}_{r \leq 0}]$. On démontre que l'estimateur proposé satisfait les propriétés oracle.

Des simulations numériques confirment la propriété de sélection automatique.

2.1.2 Modèles avec rupture (change-points) avec un nombre fixe et très grand de variables

Un phénomène/processus modélisé peut avoir à un certain moment un changement de comportement. La question qui se pose est de savoir si c'est un changement non significatif (dû aux aléas du phénomène), ou si, suite à des cause externes ou internes, le modèle a changé de forme.

Pour répondre à cette question il faut d'abord établir si l'étude est rétroactive ou séquentielle (en continu). Le modèle statistique, et surtout son approche, ne sont pas les mêmes dans les deux cas. D'autre part, fonction que les observations sont indépendantes, ou qu'elles dépendent de quelques observations du passé (modèle à courte mémoire) ou de tout le passé (modèle à longue mémoire) les résultats théoriques et donc leur application, sont différents. Alors, fonction du problème que l'on a, il faut choisir le "bon" modèle.

Remarquons, pas en dernier lieu, que la théorie classique d'inférence statistique pour des modèles paramétriques est construite sous l'hypothèse d'identifiabilité. Cette hypothèse n'est pas vérifiée par les modèles avec rupture. Une autre difficulté vient de la non régularité de la fonction à optimiser par rapport aux points de rupture. La plupart des modèles étudiés auparavant par d'autres auteurs, n'ont qu'un seul point de rupture et surtout les fonctions sur chaque segment sont linéaires. D'autre part, dans la littérature, il existe très peu d'études sur les modèles avec change-points en très grande dimension ou avec sélection automatique des variables.

1. Dans un modèle linéaire à *très grande dimension* (quelques centaines-milliers de variables), avec des points de rupture,

$$Y_i = \beta_1^t X_i \mathbb{1}_{1 < i \leq l_1} + \beta_2^t X_i \mathbb{1}_{l_1 < i \leq l_2} + \dots + \beta_{K+1}^t X_i \mathbb{1}_{l_K < i \leq n} + \varepsilon_i, \quad i = 1, \dots, n.$$

les propriétés asymptotiques ont été étudiées pour plusieurs estimateurs de type *LASSO*. Pour des estimateurs de type adaptatif, on prouve des propriétés de type oracle (le modèle détecte avec une probabilité proche de 1 les variables non significatives, sans passer par des tests d'hypothèse, même si le modèle contient un très grand nombre de variables), (**article 1.**). Mais, ce type d'estimateur peut être utilisé seulement quand le nombre d'observations est plus grand que le nombre de variables du modèle. Pour pallier cet inconvénient, dans l'**article 3.**, les méthodes de type *SCAD* et *LASSO-type* (pour un modèle médian) ont été étudiées. La méthode SCAD pose des problèmes numériques et la méthode LASSO-type est appropriée seulement pour des modèles particuliers. Alors, dans l'**article 27.**, on a considéré un *modèle adaptatif quantile*. Le résultat le plus intéressant est que l'estimation des change-points n'a aucune influence sur la sélection automatique des variables du modèle dans chaque phase. Pour chaque méthode d'estimation la vitesse de convergence des estimateurs et leur loi limite sont obtenues.

Dans l'**article 30.** la détection du changement est testée en temps réel en combinant une méthode adaptative LASSO et une statistique de test CUSUM. On obtient la loi asymptotique de la statistique de test, sous l'hypothèse que le modèle ne subit pas un changement, et on montre que cette statistique converge vers ∞ si il y a un changement.

2. Dans les **articles 13.** et **15.**, nous mettons en oeuvre la *méthode du maximum de vraisemblance* (MV) pour un modèle non linéaire avec les régresseurs aléatoires et avec un, respectivement, K points de rupture :

$$Y_i = h(\beta_1, X_i) \mathbb{1}_{1 < i \leq l_1} + h(\beta_2, X_i) \mathbb{1}_{l_1 < i \leq l_2} + \dots + h(\beta_{K+1}, X_i) \mathbb{1}_{l_K < i \leq n} + \varepsilon_i, \quad i = 1, \dots, n.$$

avec $h(\beta, X)$ une fonction, linéaire ou non linéaire en β , quelconque. On montre que les estimateurs par MV de $\beta_1, \dots, \beta_{K+1}$ et de l_1, \dots, l_K sont fortement consistants, avec une vitesse de convergence d'ordre n^{-1} pour les estimateurs des points de rupture et d'ordre $n^{-1/2}$ pour les estimateurs des paramètres de régression. Ils convergent en loi vers *argmax* de K processus de Poisson composés indépendants pour les premiers et vers un processus gaussien centré pour les seconds. Ces résultats ont été généralisés dans l'**article 12.**, en utilisant la méthode de *M-estimation*. Pour déterminer le nombre K de points de rupture, dans l'**article 6.** un critère général, consistant, est proposé.

On a appliqué les résultats obtenus pour détecter et estimer les changements dans un modèle compartimental (en médecine, par exemple, les différents échanges dans l'organisme humain sont ainsi modélisés) et dans un modèle de croissance (d'une plante, par exemple).

3. Si les erreurs d'un modèle présentent des *outliers* (*points extrêmes*), alors, les estimateurs considérés plus haut ont une grande variance. Dans ce cas il est préférable de considérer la méthode d'estimation *LAD* (*Least Absolute Deviation*). Par contre, l'étude théorique de ces estimateurs et le calcul numérique sont plus difficiles à cause de la non dérivabilité de la fonction objectif. Une étude théorique de cet estimateur est réalisée (consistance, vitesse de convergence, lois asymptotiques) dans l' **article 10**. Pour régulariser la fonction objectif, on considère une méthode LAD pénalisée dans l' **article 9**.

Des simulations sont réalisées pour des modèles de type exponentiel.

Parce que, assez souvent, dans les problèmes pratiques, on n'a aucune information concernant la loi des erreurs (ε_i), les méthodes présentées plus haut sont inappropriées. Alors, dans l'**article 29**., pour un taux quantile τ fixé, on a considéré un *modèle quantile*, et on a généralisé les résultats obtenus dans l'article 10.

Tous ces modèles ont été étudiés dans le cadre d'*observations indépendantes*. Dans des nombreuses applications (hydrologie, économétrie, ...), les régresseurs mais aussi les erreurs sont des *variables aléatoires à longue mémoire*.

4. Dans l'**article 5**., un modèle linéaire avec un seul change-point et avec des variables à longue mémoire est étudié :

$$Y_t = X_t \beta_1 \mathbb{1}_{1 \leq t \leq [n\pi]} + X_t \beta_2 \mathbb{1}_{[n\pi]+1 \leq t \leq n} + \varepsilon_t, \quad t = 1, \dots, n$$

avec $\pi \in (0, 1)$. Les vecteurs aléatoires X_t and ε_t sont Gaussiens, stationnaires, à long-mémoire. Soulignons que les propriétés théoriques des S-estimateurs sont totalement différentes de celles obtenues précédemment pour les modèles avec des variables aléatoires à courte mémoire ou indépendantes (par exemple, la vitesse de convergence dépend des paramètres de long mémoire et elle est plus lente).

Application à la modélisation du niveau de Nil. Dans le stage de R. Haddad, on a modélisé le taux d'échange du franc, par rapport au dollar, entre 1984 et 1985.

Des simulations ont été également réalisées pour comparer la méthode proposée avec d'autres méthodes de la littérature.

2.1.3 Modèles à longue mémoire

Dans l'**article 8**., j'ai trouvé la vitesse de convergence et la loi asymptotique de l'estimateur des moindres carrés des paramètres pour un modèle non linéaire de la forme : $Y_i = f(X_i; \beta) + \varepsilon_i$, $i = 1, \dots, n$. Les variables aléatoires X_i et ε_i sont Gaussiennes à longue mémoire. Les résultats obtenus dépendent des paramètres de longue mémoire et de l'espérance de la dérivée de la fonction de régression par rapport aux paramètres de régression. Des simulations confirment les résultats théoriques.

2.1.4 Modèle avec données manquantes

Pour un modèle non linéaire $Y_i = f(X_i; \beta) + \varepsilon_i$, $i = 1, \dots, n$, la variable réponse Y_i peut être manquante de manière aléatoire. Considérons alors, la variable aléatoire $\delta_i = 1$, si Y_i observée et $\delta_i = 0$, si Y_i manquante. Soit la fonction de probabilité de sélection $\pi(x) = \mathbb{P}[\delta = 1 | X = x]$. Les

paramètres β peuvent être estimés sur les données complètes, soit par moindres carrés $\hat{\beta}_{n,LS} = \arg \min_{\beta} \sum_{i=1}^n \delta_i [Y_i - f(X_i, \beta)]^2$, soit par moindres déviations $\hat{\beta}_{n,LAD} = \arg \min_{\beta} \sum_{i=1}^n \delta_i |Y_i - f(X_i, \beta)|$. Ces estimateurs permettent la construction de la statistique du maximum de vraisemblance empirique. On montre que cette statistique converge vers une loi de χ^2 , ce qui implique la construction des régions de confiance pour β . Les données manquantes sont reconstituées en utilisant une méthode semi-paramétrique et le théorème de Wilks est prouvé sur toutes les données reconstituées.

Application à la reconstitution des données ultrasoniques.

Une étude par simulations montre la performance de la méthode proposée, en termes de reconstitution des données manquantes et de qualité d'estimation, surtout si les données contiennent des outliers (points extrêmes). (**article 4.**)

2.1.5 Estimation des lois des valeurs extrêmes

Dans l'**article 11.** on propose un estimateur semi-paramétrique, qui généralise nombreux estimateurs existants, pour l'indice des valeurs extrêmes et pour le paramètre d'ordre 2 d'une loi des valeurs extrêmes. Des simulations ont été réalisées.

2.1.6 Mélanges de populations (densités)

Ce type de modèle statistique a de multiples applications pratiques : *biologie* (mélange de populations), *médecine* (diagnostic en fonction de différents tests cliniques), *génétique* (ex : dans le foie existent différentes classes de cellules qui ont une caractéristique différente de l'ADN), *chimie, imagerie* (ex : un objet composé de n pixels, chaque état x_i étant décrit par une couleur c_i , $i = 1, \dots, n$). Les c_i sont des réalisations i.i.d. d'un mélange de densités Gaussiennes) et *théorie du signal* (un signal peut être vu comme un mélange de plusieurs densités Gaussiennes).

Soit Y une variable aléatoire de densité un mélange de densités. Plus précisément, soit $\mathcal{F} = \{f_{\gamma}(x); \gamma \in \Gamma\}$ une famille de densités. Pour un p connu, l'ensemble des p -mélanges de densités de \mathcal{F} est : $\mathcal{G}_p := \{g_{\pi, \alpha} = \sum_{i=1}^p \pi_i f_{\gamma_i}(\cdot) / \pi = (\pi_1, \dots, \pi_p), \alpha = (\gamma_1, \dots, \gamma_p), \forall i = 1, \dots, p, \gamma_i \in \Gamma, 0 \leq \pi_i \leq 1, \sum_{i=1}^p \pi_i = 1\}$. Un problème important pour ce type de modèle, est de déterminer le nombre p de composantes; en fait, de tester si le modèle est un mélange de p -densités ou un mélange de q -densités, avec $q < p$. Dans le cas des modèles identifiables, le test du rapport de vraisemblance est une des solutions les plus utilisées. Pour le cas de mélanges de lois, la loi asymptotique du rapport de vraisemblance a été longtemps un problème ouvert, puisque sous l'hypothèse nulle, le modèle n'est pas identifiable et donc le théorème classique de χ^2 est inapplicable.

Dans l'**article 18.**, j'ai donné des conditions suffisantes, relativement faibles pour les densités f_{γ} , pour que l'ensemble des dérivées de la log-vraisemblance soit une classe de Donsker.

J'ai construit le premier exemple de *non-consistance du test de vraisemblance* pour un ensemble compact de paramètres (**article 17.**). Il s'agit de la famille des densités exponentielles : $f_{\gamma}(x) = e^{-(x-\gamma)} \mathbb{1}_{x>\gamma}$ pour $\gamma \in \Gamma = [0, G]$. Je montre que la statistique du maximum de vraisemblance vaut 0 avec une probabilité 1/2 et elle converge fortement vers $+\infty$ avec une probabilité 1/2. Une étude numérique est réalisée pour conforter le résultat théorique obtenu.

Conclusion : si on a un mélange de lois exponentielles, le test classique du maximum de vraisemblance est inapplicable!

Dans l'**article 16.**, on considère un mélange de p densités normales univariées. Si la variance

d'une des composantes est proche de zéro, la densité $g_{\pi,\alpha}$ peut dégénérer et donc l'estimateur du maximum de vraisemblance (MV) peut ne pas exister. Dans ce cas, l'algorithme EM dégénère. Pour palier à cet inconvénient, on a proposé un estimateur du MV pénalisé, qui est borné pour toute valeur de la variance, pour une taille d'échantillon fixée. Évidemment, l'algorithme EM pénalisé correspondant ne dégénère plus. Nous avons démontré que cet estimateur est fortement convergent, asymptotiquement efficace. Sa loi asymptotique est gaussienne centrée de matrice de covariance, l'inverse de l'information de Fisher. Les résultats obtenus sont généralisés pour un mélange de toute densité qui dégénère sur \mathbb{R} . Des simulations numériques mettent en évidence les avantages et les performances de l'estimateur proposé.

2.1.7 Estimation des taux d'entropie

Soit $\mathbf{X} = (X_i)_{1 \leq i \leq n}$ une suite aléatoire. Si X_i prend des valeurs dans un ensemble discret $E \subset \mathbb{R}$, l'entropie de Shannon au temps n de $(X_i)_{1 \leq i \leq n}$ est :

$$\mathbb{H}_n(\mathbf{X}) = - \sum_{(x_1, \dots, x_n) \in E^n} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] \log \mathbb{P}[X_1 = x_1, \dots, X_n = x_n]$$

Si la limite suivante $\lim_{n \rightarrow \infty} n^{-1} \mathbb{H}_n(\mathbf{X})$ existe, elle s'appelle *taux d'entropie (de Shannon)* et on la note $\mathbb{H}(\mathbf{X})$.

Il y a très peu de résultats dans la littérature concernant l'estimation de $\mathbb{H}(\mathbf{X})$ dans le cas d'une suite non i.i.d. Dans l'**article 14.**, nous avons considéré l'estimation de $\mathbb{H}(\mathbf{X})$ pour des chaînes de Markov. Soit une chaîne \mathbf{X} , d'espace d'états E de probabilités de transition $P = (P(i, j))_{i, j \in E}$ et de loi stationnaire $\pi = (\pi(i))_{i \in E}$. Le taux d'entropie de Shannon de la chaîne de Markov, s'il existe, est :

$$\mathbb{H}(\mathbf{X}) = - \sum_{i \in E} \pi(i) \sum_{j \in E} P(i, j) \log P(i, j)$$

Les estimateurs du taux d'entropie sont obtenus à partir des *estimateurs plug-in* pour les probabilités de transition et pour la loi stationnaire de la chaîne. On montre que l'estimateur du taux d'entropie est fortement consistant pour un espace d'états dénombrable. Si l'espace des états est fini et la loi de transition n'est pas uniforme, on montre la normalité asymptotique.

Ces résultats ont été généralisés dans l'**article 7.** où l'on étudie le taux d'entropie des suites de variables aléatoires dépendantes correspondant à des fonctionnelles d'entropie. On couvre ainsi les entropies : Shannon, Rényi, Tsallis, Sharma-Mittal, ... On montre que tous les taux d'entropie sont soit 0 soit infinis ; exceptés les taux d'entropie de Shannon et de Rényi.

Conclusion : si on veut mesurer l'agitation d'un système, on ne utilise pas ces entropies (sauf Shannon et Rényi) puisqu'elles sont soit 0 soit infinies !

2.2 Applications des statistiques

2.2.1 Modélisation du métabolisme du glucose

Ma thèse de doctorat a eu comme sujet la *modélisation par un système d'équations différentielles non linéaires, paramétriques* du métabolisme du glucose. Supposons que les variables d'état soient X_1, X_2, \dots, X_m . On connaît les valeurs de n variables ($n \leq m$) aux points (temps) t_1, t_2, \dots, t_N et de toutes les m variables au point t_0 . On suppose, en plus, que X_1, X_2, \dots, X_m satisfont au

système d'équations différentielles :

$$\begin{cases} \frac{dX_k(t)}{dt} = f_k(t, X; \theta), k = 1, 2, \dots, m \\ X(t_0) = X_0 \end{cases} \quad (1)$$

Etant donné un ensemble de valeurs y_{ij} , $i = 1, \dots, n$; $j = 1, \dots, N$ nous proposons d'estimer dans le paramètre θ , du modèle :

$$\mathbf{y}_{ij} = \mathbf{X}_i(\mathbf{t}_j; \theta^*) + \varepsilon_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, N \quad (2)$$

où $X_i(t_j; \theta^*)$ est la solution du système (1) pour la variable X_i au point t_j de paramètre θ^* . Le modèle statistique (2) est une régression multidimensionnelle non linéaire paramétrique, pour laquelle on ne connaît pas la forme analytique de la fonction non linéaire $X_i(t_j; \theta)$, mais, en échange, on peut calculer une estimation en chaque point t_j et pour un paramètre θ donné. On trouve les estimations des paramètres du système par maximum de vraisemblance et par moindres carrés. Des tests d'hypothèse montrent qu'on a le même modèle pour tous les individus traités avec glucose ce qui n'est plus vrai sous insuline. Les estimations obtenues sont en général très corrélées. Une *surparamétrisation* conduit à une instabilité des paramètres. J'ai mis en évidence de possibles relations linéaires entre les paramètres du modèle.

Le *plan d'expérience D_N -optimal*, c'est-à-dire le plan d'expérience qui minimise le déterminant de la matrice d'information, a été mis en évidence. Ce plan donne les instants où il faut faire les prises pour avoir plus d'information sur les paramètres. (**Articles 24., 25. et 26.**).

2.2.2 Modélisation des courbes d'impédance et d'activité acidifiante

Entre novembre 1994 et mars 1995 j'ai participé à un contrat de recherche, avec Prof. Richard Tomassone (de l'Institut National Agronomique de Paris) sur la *Modélisation des courbes d'impédance et d'activité acidifiante pour la fabrication d'un type de fromage, en différentes conditions*. Le sujet a été proposé par l'Institut Technique du Gruyère, La Roche sur Foron. Il s'agit de comparer l'influence des facteurs contrôlés sur ces courbes, dans le but de choisir un modèle approprié pour l'utilisation ultérieure. On a utilisé comme modèles statistiques : modèles non linéaires paramétriques, modèles linéaires généraux.

2.2.3 Modélisation de la conservation du hareng dans des conditions expérimentales différentes

Entre mars 1995 et juillet 1995 j'ai participé à un contrat de recherche, avec Prof. Richard Tomassone (de l'Institut National Agronomique de Paris) sur la *Modélisation de la conservation du hareng dans des conditions expérimentales différentes*. Le sujet a été proposé par le Centre d'Expérimentation et de Valorisation des Produits de la Mer, Boulogne-sur-mer. On a utilisé des **modèles non linéaires paramétriques** et des modèles linéaires généraux.

2.2.4 Modélisation de la leucémie

J'ai participé à un BQR basé sur la collaboration des mathématiciens de l'UMR 5208 et des médecins du Service d'Hématologie Clinique (Université Lyon 1 et HCL). Ce projet a fait partie d'un programme prioritaire régional "Cancer" et il a été consacré au développement des outils de diagnostic des leucémies aiguës myéloïdes. Des méthodes statistiques basées sur l'analyse

factorielle discriminante ont permis de prévoir certains types de leucémies en connaissant l'immunophénotypage. La régression logistique a permis de mieux comprendre les différents types de leucémie. D'autre part, sur la base des marqueurs immunophénotypiques, en utilisant des méthodes de classification non-supervisée (clustering), à la demande des médecins, on a proposé une nouvelle partition des leucémies. (**Article 21.**)

2.2.5 Etude sur la schizophrénie

Récemment j'ai collaboré avec un groupe de chercheurs/médecins sur une étude de la schizophrénie (**Article 20.**)

Des méthodes d'analyse de variance et de régression ont permis d'évaluer la performance cognitive, fonction du temps de réaction aux différents stimulus visuels, des malades et de comparer les résultats obtenus avec ceux obtenus pour les personnes non-schizophrènes.

2.2.6 Etude olfactive

J'ai participé à une étude sur les anosmiques, hyposmiques, concernant la sensibilité olfactive. (**article 19.**)

D'abord, par des modèles de covariances, on a étudié quels facteurs (âge, sexe, tabac) ont une influence sur les scores obtenus à la suite du test olfactif. Ensuite, pour chaque catégorie de personnes (normaux, anosmiques, hyposmiques) on a étudié, à partir de quelle âge les capacités olfactives commencent à se dégrader ? (par des tests de Students).

A la suite de cette étude on a également obtenu des critères clairs et simples pour le praticien qui vont lui permettre de diagnostiquer si le patient a des problèmes d'olfaction et quel est le degré d'affectation.

2.2.7 Architecture et autisme

Depuis janvier 2014, avec Anne Perrut (McF à l'Université Lyon 1), nous avons commencé la collaboration avec l'Ecole d'Architecture de Lyon. Dans un premier temps, nous étudions (par des méthodes d'analyse factorielle et des modèles linéaires) l'influence de l'architecture d'une pièce sur le comportement des autistes.

2.2.8 Générateur de cycles

Dans le cadre d'un partenariat industriel Université Lyon 1/Volvo Truck, par des méthodes de Chaînes de Markov cachées, séries chronologiques, on a construit un générateur de cycles (clause de confidentialité).

2.2.9 Modélisation de la pollution de l'air

Après ma thèse, j'ai fait un post-doc d'un an sur la modélisation des pics de pollution dans la région parisienne dans le cadre d'un contrat de recherche entre le Laboratoire de Probabi-

lités/Statistique de l'Université Paris XI et AIRPARIF. Dans le cadre de ce contrat, j'ai travaillé sur la mise en place effective du système de prévision des pointes de pollution par l'ozone et les oxydes d'azote dans la région parisienne, à court et à moyen terme. A court terme, la concentration journalière maximale d'ozone (de l'après-midi) est prévue à 6h du matin, en utilisant des données de pollution et données météo mesurées dans la région parisienne. A moyen terme, la prévision est faite un, voire plusieurs jours à l'avance. Pour prévoir la pollution, nous avons utilisé plusieurs méthodes statistiques : analyse de données, méthode de régression par arbre CART et des méthodes non-paramétriques. La méthode retenue est la non-paramétrique, les pics de pollution étant prévus en proportion de 90%. On considère (Y_n) le processus à prévoir et (X_n) un processus de variables exogènes. Dans ce cas précis, Y_n est la concentration maximale journalière d'ozone pour le jour n et X_n contient des variables météo (température, direction et vitesse de vent, ...). Le modèle statistique considéré est :

$$\begin{aligned} Y_{n+1} &= F(Y_n, X_n) + \varepsilon_{1n} \\ X_{n+1} &= G(X_n) + \varepsilon_{2n} \end{aligned}$$

avec (ε_{1n}) et (ε_{2n}) des bruits blancs indépendants. Les fonctions F et G sont inconnues et elles sont estimées en utilisant une méthode à noyau, en prenant le noyau Nadaraya-Watson.

Le système de prévision a été implanté à partir de l'été 1997 à AIRPARIF et depuis il est opérationnel (avec une amélioration apportée ultérieurement). Cette prévision de la pollution sert pour prévenir la population dans le cas d'un grand pic de pollution. Elle sert aussi aux décideurs politiques pour envisager des mesures qui vont défavoriser la formation de la pollution. (**Articles 22., 23.**).

3 Activités d'enseignement

3.1 Responsabilités

1. Membre élue au Conseil du Département de Mathématiques, Université Lyon 1 (depuis janvier 2014).
2. Coresponsable du parcours "Mathématiques pour la biologie et médecine : théorie et applications" du nouveau master recherche en Mathématiques appliquées (2013-2014).
3. Coresponsable des stages de fin d'étude du master Pro "Statistique, Informatique et Techniques Numériques(SITN)" de l'Université Lyon 1 (depuis 2001).
4. Participation à la rédaction de l'habilitation des masters de mathématiques appliquées ou de la Licence MASS concernant les cours de statistique.

J'ai enseigné et j'enseigne à un public varié d'étudiants, dans des filières mathématiques ou non mathématiques, de L1 jusqu'au M2.

3.2 Etudiants avec un parcours mathématique

1. **CM** : *DEA Mathématiques, Université Lyon 1, "Statistique Exploratoire" (2003/2004)* :
 - (a) introduction aux Statistiques ;
 - (b) méthodes factorielles : Analyse en Composantes Principales (ACP), Analyse Factorielle Discriminante (AFD), Analyse Factorielle des Correspondances (AFC), Analyse des Corrélations Canoniques (ACC) ;
 - (c) méthodes de classification : classification par partition, classification hiérarchique.
2. **CM/TD/TP** : *Master Pro 1 et 2 d'Ingénierie Mathématique(MIM), Université Lyon 1.* Jusqu'au 1er septembre 2011 la première année avait deux spécialisations : SITN(Statistique, Informatique et Techniques Numériques) et IR(Ingénierie des Risques).
 - (a) "Statistique paramétrique"(M1)(2009-2015), "Statistique inférentielle"(M2)(2001-2010), "Techniques Probabilités et Statistiques"(2001-2009) :
 - i. statistique descriptive ;
 - ii. notions d'échantillonnage, Théorème de Cochran ;
 - iii. théorie de l'estimation : méthodes d'estimation, propriétés des estimateurs, familles exponentielles, estimateur par intervalle ;
 - iv. théorie des tests d'hypothèse : *tests paramétriques* : lemme de Neyman-Pearson, test du rapport de vraisemblance et de Wald, tests de Student et de Fisher pour une ou deux populations ; *test non-paramétriques* : Théorème de Pearson, test de χ^2 , test de Kolmogorov-Smirnov, de la médiane, test de Spearman, test de Wilcoxon.
 - v. modèles linéaires
 - A. régression linéaire multiple ;
 - B. analyse de variance ;
 - vi. En TD/TP, nombreuses applications concrètes sont traitées, en utilisant les logiciels SAS et R.

- (b) "Analyse de données"(M2), (2001-2013). Depuis 2013, le cours a changé d'intitulé et de contenu, en "Data mining" :
- i. méthodes factorielles : ACP, AFD, AFC, ACC.
 - ii. construction d'un modèle sur un ensemble d'apprentissage et test sur un ensemble de test.
 - iii. évaluation des règles de décision.
 - iv. méthodes non-paramétriques d'estimation de la densité d'une loi et de la fonction de répartition.
 - v. data mining : régression PLS, méthode CART, méthodes de type LASSO.
- L'examen consiste d'un projet (individuel) sur des données réelles.
- (c) "Outils statistiques avancés"(M2)(2006-2015) : le contenu peut être différent d'une année à l'autre :
- i. séries temporelles, méthodes de Box-Jenkins, tests de stationnarité;
 - ii. modèles non-linéaires et algorithmes pour calculer les estimateurs;
 - iii. régression logistique et algorithmes pour le calcul des estimateurs;
 - iv. modèles mixtes;
 - v. modèles de censure; estimateur de Kaplan-Meier;
 - vi. méthodes de Jackknife et Bootstrap;
 - vii. modèles avec données manquantes.
- (d) "Méthodes des Monte-Carlo"(M2)(2001-2005)
- i. intégration par Monte-Carlo, Théorème de Rubinstein;
 - ii. méthode de Bootstrap;
 - iii. algorithme de Hastings-Metropolis.
- (e) "Logiciel SAS"(M2)(2012-2015)
- i. généralités;
 - ii. procédures simples;
 - iii. PROC GPLOT;
 - iv. PROC IML;
 - v. macros.
3. Chaque année je propose (et j'encadre) 3-4 sujets de **TER** pour les étudiants en M1 PRO de Mathématiques Appliquées. J'encadre également des **stages en entreprise** des étudiants en M2 PRO de la même formation. (Voir la Partie 1.4.3).
4. **CM/TD/TP** : *L1 et L2 MASS, Université Lyon 1,*
- (a) "Probabilités/Statistique"(L2)(2001-2007)
- i. introduction à la théorie de l'estimation et des tests d'hypothèse;
 - ii. modèles linéaires : régression et analyse de variance.
- (b) "Statistique descriptive"(L1)(2002/2003) :
- i. statistique descriptive unidimensionnelle : pour des variables quantitatives et qualitatives : résumés numériques (paramètres de position et de dispersion), représentations graphiques;
 - ii. statistique descriptive bidimensionnelle, distributions marginales et conditionnelles, coefficients de Spearman et Kendall.

3.3 Etudiants avec un parcours non-mathématique

1. **CM/TD/TP** : Master *Compétences Complémentaires en Statistique, Université Lyon 1* :
”Analyse de données” (2011/2012) :
 - (a) Analyse en Composantes Principales (ACP) ;
 - (b) Analyse Factorielle Discriminante (AFD) - classification supervisée ;
 - (c) Analyse Factorielle des Correspondances (AFC) ;
 - (d) Analyse des Corrélations Canoniques (ACC).
2. **CM/TD/TP** : *Master Pro M2 Neurosciences, Université Lyon 1*,
”Statistique appliquée aux neurosciences” (2007/2008) :
 - (a) introduction aux probabilités ;
 - (b) statistique descriptive ;
 - (c) notions d’estimation et tests d’hypothèse ;
 - (d) analyse de données : ACP et AFD.
3. **TD** : *DEUG II Sciences Economiques, Université Paris 13*,
”Statistique et Probabilités” (1997/1998) :
 - (a) variables aléatoires discrètes et continues ;
 - (b) estimateurs ponctuels et par intervalle ;
 - (c) tests d’hypothèse ;
 - (d) théorème limite centrale..
4. **TD** : *Ecole Ingénieurs Universitaire(Polytech), Université Lyon 1*,
”Introduction aux méthodes probabilistes et statistiques”(première année) (2008/2009) :
 - (a) modèle probabiliste, variables aléatoires discrètes, la loi Normale ;
 - (b) échantillons et lois d’échantillonnage ;
 - (c) estimation et test de la moyenne, tests de χ^2 .
5. **CM/TD/TP** : *IUT STID de Niort, Université Poitiers(1998/2001)*
Modèle linéaire :
 - (a) régression linéaire simple et multiple ;
 - (b) analyse de variance à un et à deux facteurs.

Logiciels utilisés en TP : SAS, R et Matlab.

Fait à Villeurbanne, le 24 novembre 2014