

Catherine TROTTIER

Née le 10 avril 1971 à Tours

MIAP

Université Paul Valéry - Montpellier 3

UFR 6

Route de Mende

34199 Montpellier cedex 5

Tél. : (+33) [0]4 67 14 21 63

catherine.trottier@univ-montp3.fr

<http://www.univ-montp3.fr/miap/~trottier/>

IMAG

Université de Montpellier

UMR CNRS 5149

Case courrier 051

Place Eugène Bataillon

34090 Montpellier - France

Tél. : (+33) [0]4 67 14 41 64

catherine.trottier@umontpellier.fr

Cursus

- Depuis 2002 :

Maître de conférences à l'université Paul Valéry - Montpellier 3.

Rattachement recherche à l'IMAG (Institut Montpelliérain Alexander Grothendieck, anciennement I3M) - Université de Montpellier (UMR 5149).

- 1999-2002 :

Maître de conférences à l'université Charles de Gaulle - Lille 3.

Rattachement recherche au Gremars (Groupe de recherche : "Modélisation appliquée à la Recherche en Sciences Sociales") - Université Lille 3 (EA 2459).

- 1998-1999 :

Postdoctorat à l'Inria Rhône-Alpes - Projet IS2 (Inférence Statistique pour l'Industrie et la Santé) sur un CERD avec EDF.

- 1997-1998 :

ATER à mi-temps à l'ENSIMAG (École Nationale Supérieure d'Informatique et Mathématiques Appliquées de Grenoble).

- 1994-1997 :

Monitorat à l'université Joseph Fourier - Grenoble 1, U.F.R. Informatique et Mathématiques Appliquées.

Diplômes

- 2016 :

Habilitation à Diriger des Recherches de l'université de Montpellier.

Titre : *Voyage autour des modèles linéaires généralisés à la recherche de structures latentes.*

Jury : C. Biernacki, G. Celeux, Y. Guédon, C. Lavergne, D. Pommeret, J. Saracco.

- 1998 :

Doctorat de l'Institut National Polytechnique de Grenoble.

Titre : *Estimation dans les modèles linéaires généralisés à effets aléatoires.*

Jury : A. Antoniadis, M. Bonneu, J.-L. Foulley, C. Lavergne, J.-L. Soler.

Travail effectué au sein du Laboratoire de Modélisation et de Calcul (LMC - IMAG) et de l'Inria Rhône-Alpes.

Financement bourse MESR.

- 1994 :

D.E.A. de Mathématiques Appliquées de l'Université Joseph Fourier - Grenoble 1 - Department of Mathematics de l'Imperial College à Londres.

Diplôme d'ingénieur ENSIMAG (École Nationale Supérieure d'Informatique et Mathématiques Appliquées de Grenoble) de l'INPG (Institut National Polytechnique de Grenoble), option : Modélisation et Décision en Économie.

Diplôme international en mathématiques de l'Imperial College, Londres.

Activités d'enseignement

- De 1999 à 2002, à l'université Charles de Gaulle - Lille 3, mon activité d'enseignement s'est inscrite dans l'UFR de Psychologie, dont j'étais membre. J'étais en charge des cours en amphi et TD de statistique descriptive et inférentielle en Deug et Licence auprès des étudiants en psychologie. Nous avons reconstruit toutes les séances de TD en collaboration avec les psychologues qui animaient certains groupes de TD. Les contextes des exercices proposés s'appuyaient alors sur leurs travaux de recherche.
- Depuis 2002, après une mutation à l'université Paul Valéry - Montpellier 3, j'ai intégré le département MIAP (Mathématiques et Informatique Appliquées) de cette université LLASHS. Mes enseignements ont ainsi continué à s'adresser à des étudiants en sciences humaines et sociales : psychologie, sociologie, AES, sciences sociales, sciences du langage, géographie.
- J'ai été responsable d'enseignements à (très) gros effectifs, nécessitant une part importante de gestion des emplois du temps, de coordination des amphis et des nombreux groupes de TD, et de recrutement de vacataires ou de formation de doctorants SHS pour animer les séances de TD. En particulier, j'ai été responsable pendant 8 ans du cours de L1 en statistique descriptive à destination des étudiants des 3 filières : psychologie, sociologie et AES, représentant un peu moins de 1500 étudiants chaque année.
- Jusqu'en septembre 2014, en filière de psychologie nos enseignements se poursuivaient en L2 et L3, par de la statistique descriptive bivariée, de l'initiation aux probabilités et de l'inférence. J'ai été responsable du cours de L3 d'introduction aux probabilités pour 450 étudiants environ. À Montpellier malheureusement, le travail collaboratif avec les collègues psychologues ne s'est jamais avéré possible.
- En filière AES, j'assure depuis 4 ans un cours en Master 1 pour 350 étudiants environ et aussi avec un petit groupe d'apprentis au rythme particulier.

- Notre département MIAp (Mathématiques et Informatique Appliquées) a ouvert une filière MIASHS : la licence en 2014 et le master en alternance en 2016. Des nouveaux enseignements plus spécialisés ont ainsi été mis en place auxquels je participe. Je participe aussi au suivi de TER, stages et contrats d'alternance.

- En parallèle à cette activité régulière d'enseignement dans mon établissement, j'ai pu animer pendant quelques années :

- un cours-TD en L1 MASS de probabilités

- des cours, TP en maîtrise MASS de modèles linéaires

- des cours, TD et TP sous excel en école d'ingénieur ENSIA-SIARC (Industrie Agro-alimentaire des Régions Chaudes) de statistiques descriptives et inférentielles

- Depuis 2005, je co-anime un cours en master 2 Biostatistique (ex DEA) à l'université de Montpellier sur les thèmes : modèles linéaires avancés, modèles linéaires généralisés, modèles linéaires mixtes. Ce cours a été assuré sous diverses formes et avec divers volumes horaires, conjointement avec Christian Lavergne au début puis avec Frédéric Mortier (CIRAD) et depuis 3 ans avec Marie Denis (CIRAD). Il est désormais commun avec le master 2 MIND.

Le tableau ci-après donne une vue synthétique de ces enseignements. Mon volume d'enseignement annuel est approximativement de 240 heures TD.

- Depuis 2013, avec les collègues de 26ème section, nous avons mis en place des exercices sur la plateforme moodle. Les étudiants ont ainsi à leur disposition des exercices d'entraînement. J'ai eu plus particulièrement en charge les exercices concernant l'initiation aux probabilités pour les L3 de Psychologie.

L1	psycho / socio / AES Gestion d'enseignement très gros effectif	Statistique descriptive univariée	Amphi et TD	de 2002 à 2014 de 2005 à 2014
Deug1	MASS	Initiation aux probabilités	Cours et TD	de 2003 à 2005
Deug	Misashs	Statistique descriptive uni- et bi-variée	Cours et TD	de 2002 à 2004
Deug	psychologie	Statistique descriptive	Amphi et TD	de 1999 à 2002
L2	psychologie	Statistique descriptive bivariée	TD	de 2002 à 2010
L2	AES-misashs	Initiation aux probabilités	Cours et TD	de 2010 à 2014
L2	psycho/AES/sciences lang.	Option prépro. Initiation inférence	Cours, TD, TP	depuis 2015
L3	psychologie Gestion d'enseignement gros effectif	Initiation aux probabilités	Cours en amphi et TD	de 2004 à 2014
L3	psychologie	Inférence statistique	Cours et TD	de 2012 à 2014
L3	AES-misashs	Initiation au logiciel R	Cours et TP	de 2011 à 2014
Licence	psychologie	Statistique inférentielle	Amphi et TD	de 1999 à 2002
L3	psycho/AES/sciences lang.	Option prépro. Inférence avancée	Cours, TD, TP	depuis 2015
L3	MIASHS	Vraisemblance et inférence	Cours, TD, TP	depuis 2015
Maîtrise	MASS	Modèles linéaires	Cours, TD, TP	de 2002 à 2004
M1	IDS	Statistique desc et inf avec excel	Cours, TD, TP	de 2007 à 2009
M1	MIASHS	Données de panels	Cours TP	depuis 2016
École ingé.	ENSIAS-SIARC	Statistique desc et inf avec excel	Cours, TD, TP	de 2005 à 2007
M2	SOI	Statistique desc. et inf. avec R	Cours, TD, TP	2005-2006
M2	AES	Statistiques économiques	TP	2015-2016
M2	Biostatistique	Modèles linéaires mixtes	Cours, TD, TP	depuis 2010
M1&2	MIASHS	Régression logistique et log-linéaire	Cours, TP	depuis 2016

- **Sélection bayésienne de variable pour données longitudinales.**

C'est un travail qui démarre avec la thèse de Benjamin Heuclin et suite à son stage de master 2 au Cirad. Ce sujet de thèse en statistique appliquée a pour objectif de développer, dans le cadre des modèles de régression linéaire généralisée, des méthodes bayésiennes de sélection de variable en grande dimension adaptées à des données longitudinales qui permettront de sélectionner, selon l'âge de l'arbre, les variables explicatives qui agissent significativement sur son développement. Nous abordons notamment la question de la sélection d'effets aléatoires en grande dimension dans les modèles mixtes.

- **Construction supervisée de composantes prédictives dans les GLM multivariés.**

Avec Xavier Bry, nous travaillons depuis quelques années à la recherche de composantes, à partir d'un tableau X , ayant un bon pouvoir prédictif pour un ensemble de variables réponses Y . X contient de (très) nombreuses covariables quantitatives ou qualitatives avec redondance d'information et nécessitant de la régularisation. Y contient plusieurs variables à expliquer. Chacune est distribuée selon une loi appartenant à la famille exponentielle, ces lois pouvant être différentes d'une variable à l'autre. Dans nos applications, nous avons en particulier travaillé avec les lois Bernoulli, Binomiale et Poisson. La construction de composantes peut être vue comme une alternative aux méthodes classiques de régularisation par pénalisation. Elle tente d'extraire, dans l'observation multiple de covariables, le trait latent pertinent pour la modélisation de Y . Nous avons d'abord défini un critère à maximiser permettant à la fois de dégager les forces structurelles de X , et d'avoir un bon pouvoir de prédiction des variables dans Y . L'algorithme développé est disponible dans le package R *SCGLR*. Cette méthodologie a particulièrement été développée en lien avec le Cirad (Frédéric Mortier et Guillaume Cornu) pour modéliser, dans le cadre de la foresterie, des données de présence/absence ou d'abondance d'espèces d'arbres en forêt humide dans le bassin du Congo. Les covariables environnementales décrivent la topographie, la géologie, les pluies et des caractéristiques de végétation.

Jusqu'ici, cette méthode n'a toutefois concerné que les modèles à observations indépendantes. Or, dans de nombreuses situations concrètes, ces dernières sont interdépendantes : observations répétées et données multi-niveaux en écologie et sciences humaines - notamment en épidémiologie, données spatiales... Ainsi, ces travaux se poursuivent actuellement avec la thèse de Jocelyn Chauvet.

Par ailleurs, nous avons dernièrement intégré la possibilité d'un regroupement thématique des covariables et d'une recherche de composantes par thème. Une nouvelle version du package a été développée.

- **Combinaisons markoviennes et semi-markoviennes de modèles de régression.**

Nous avons co-encadré avec Yann Guédon et Christian Lavergne la thèse de Florence Chaubert (2005-2008) dont l'objet était l'analyse de données longitudinales structurées en phases successives, asynchrones entre individus. La démarche a consisté à développer une classe de modèles de Markov cachés combinant une (semi-)chaîne de Markov non-observable représentant la succession des phases, avec des modèles de régression associés à chaque état de la (semi-)chaîne de Markov, ceux-ci représentant dans chacune des phases la tendance, l'effet de covariables et l'hétérogénéité inter-individuelle. Les modèles de régression abordés ont été les GLM et les modèles linéaires mixtes.

L'introduction d'effets aléatoires induit un double niveau de structure cachée. Nous avons envisagé de considérer deux types d'effet aléatoire : l'effet aléatoire individuel et l'effet aléatoire temporel. Nous avons alors développé les méthodes d'estimation adéquates basées sur des algorithmes de type MCEM. Ces modèles ont été utilisés pour l'analyse de données de croissance d'arbres forestiers en fonction de covariables climatiques.

- **Modèles pour données catégorielles structurées hiérarchiquement.**

Avec la thèse de Jean Peyhardi (2010-2013), nous avons travaillé à la définition d'une nouvelle famille de modèles pour données catégorielles. Ce travail, en collaboration avec l'équipe Virtual Plants prend source dans la modélisation du développement et de la structure des plantes à l'échelle de l'organisme. Les données étudiées ici sont des séquences d'une variable catégorielle correspondant au type de production axillaire. Cette variable présente la particularité d'un ordre partiel, que la famille de modèles développée permet de prendre en compte. La nouvelle spécification d'un GLM pour données catégorielles par un triplet $(r; F; Z)$ permet d'englober un certain nombre de modèles existant dans la littérature, de pouvoir les comparer et d'en étudier diverses propriétés, en mettant en avant quelques indications quant au choix du ratio au regard de la nature de la réponse (nominale ou ordinale). Elle permet aussi de proposer de nouveaux modèles. L'intégration de ces modèles au niveau des nœuds d'un arbre de partitionnement, permet alors de traduire l'aspect hiérarchique, et nous a permis de définir des modèles linéaires généralisés partitionnés conditionnels (PCGLM), et de répondre ainsi à la modélisation de données partiellement ordonnées. Ces modèles ont été ensuite intégrés comme modèle d'observation sur des semi-chaînes de Markov décrivant la succession et les longueurs des zones de ramification.

Aujourd'hui diverses perspectives sont envisagées, notamment en abordant les questions de choix de modèle, mais aussi pour l'étude de méthodes de régularisation de ce type de modèles.

- **Données longitudinales de comptage.**

La problématique et les données amenées par le CIRAD concernent l'observation du nombre de nouvelles feuilles apparues dans des intervalles de temps successifs sur des plans de caféiers. L'analyse de ces données répétées de comptage a donné lieu à divers encadrements de stage de 2004 à 2007. Souhaitant modéliser les répétitions par l'introduction d'un effet aléatoire "individu", il s'est agi dans un premier temps d'étudier dans le cadre du modèle de Poisson à effet aléatoire la mise en place d'un critère de choix de modèle. Nous avons ensuite davantage travaillé à la prise en compte d'une covariable climatique, disposant aussi des relevés pluviométriques quotidiens sur la même période de temps. Ceci nous a amené d'une part à choisir des modèles de Poisson avec lien identité (plutôt que le lien canonique), et d'autre part à nous confronter au problème de la sous-dispersion dans le modèle de Poisson. C'est sur cette question de la sous-dispersion que Mariem Zouch, étudiante de l'école polytechnique de Tunisie, a effectué son projet de fin d'études. Dans la suite des travaux de Castillo et Pérez-Casany, nous nous sommes intéressés aux distributions de Poisson généralisées par pondération. Dans ces modèles, le type de pondération choisi doit vérifier un certain nombre de contraintes théoriques mais permet d'englober toutes les situations de sous- et sur- dispersion. Nous avons alors travaillé à introduire l'effet de la covariable climatique dans la modélisation et à mettre en place une méthode d'estimation des paramètres de ces modèles. Ce travail, dont nous avons élargi le cadre à la famille exponentielle bidimensionnelle s'est poursuivi au cours du stage de master 2

recherche de cette même étudiante soutenu en juin 2007.

- **Choix de modèle dans les modèles linéaires généralisés mixtes (GL2M) et modélisation par mélange de GL2M.**

Nous avons co-encadré avec Christian Lavergne la thèse de Marie-José Martinez de 2002 à 2006. Nous nous sommes intéressés d'une part à développer un critère simple de choix de modèle (type AIC, BIC ...) basé sur une pseudo log-vraisemblance et construit à partir du modèle linéarisé obtenu à convergence de la procédure d'estimation puisque l'expression de la log-vraisemblance marginale ne nous est pas directement accessible dans les GL2M. Cela a fait l'objet d'une étude numérique pour les modèles exponentiels à effets aléatoires dans le cadre de la modélisation de données répétées en fiabilité des logiciels. D'autre part, afin de prendre en compte différentes classes possibles d'individus, nous avons enrichi la modélisation en construisant des mélanges de GL2M. Nous avons mis en place deux méthodes d'estimation des paramètres dans ces modèles.

D'autres travaux viennent compléter ces thèmes de recherche. Ils concernent davantage des problématiques appliquées et ont plus de débouchés applicatifs que méthodologiques.

- Dans le domaine de la modélisation de l'architecture des plantes et dans le cadre d'une collaboration initiée en 2004 avec Pierre-Éric Lauri (INRA Montpellier - UMR AGAP (Amélioration Génétique et Adaptation des Plantes méditerranéennes et tropicales) – Équipe AFEF (Architecture et Fonctionnement des Espèces Fruitières)), plusieurs études ont été réalisées. Depuis les outils simples de la statistique jusqu'aux modèles linéaires généralisés mixtes, en passant par l'analyse de données (classification, ACP,...), les approches statistiques ont permis d'étudier différents aspects concernant l'arbre entier (caractérisation de la variabilité génétique des différentes variétés, effet de la longueur du rameau sur la probabilité de floraison, effet de manipulations expérimentales comme l'arcure sur l'architecture), ou la croissance du rameau (mécanismes déterminant certains aspects de la ramification - distribution et nature des rameaux).

Toujours en modélisation des plantes mais avec Jean-Baptiste Durand et Jean Peyhardi, nous avons pu nous intéresser plus récemment à la modélisation statistique de la production irrégulière d'arbres fruitiers dont la sélection de variétés régulières est souhaitée dès les premières années.

- En sciences du langage, les travaux de thèse d'Harold Guerrero (2004-2010) se sont attachés à caractériser la voix de l'enfant sourd appareillé et implanté cochléaire. Cette étude s'est réalisée à partir d'une analyse comparative acoustique et perceptive de la voix. Ces travaux ont mis en interaction divers acteurs et disciplines (médecins, linguistes, sciences de l'ingénieur). Nous avons travaillé à analyser les mesures réalisées au travers d'approches liées aux modèles linéaires généralisés.
- Réalisation d'une étude pour l'ICV (Institut Coopératif du Vin) (2006). À partir des données rassemblées par l'ICV, nous avons proposé avec Christian Lavergne une modélisation permettant d'une part de comparer les performances des bentonites commerciales pour la stabilisation protéique des vins et d'autre part d'optimiser le choix de la dose de bentonite utilisée en cours ou après fermenta-

tion alcoolique. Les résultats de cette étude ont été présentés au congrès de l'office international du vin.

- Réalisation d'une contre-expertise scientifique dans le cadre des affaires juridiques de l'INRA (2006).

Participation à projets financés

- Membre du projet CoForTips (2013-2016), "Forest of the Congo Basin : Resilience and Tipping Points", projet financé par Biodiversa ERaNET. Il a pour objectif d'aider à une meilleure gestion des forêts du bassin du Congo par l'élaboration de scénarios de biodiversité sur un horizon temporel de quelques décennies. Coordinateur : Claude Garcia.
- Membre du projet CoForChange (2009-2012), projet financé par Biodiversa ERaNET. Ses objectifs étaient de prévoir les impacts des changements globaux sur la structure et la composition floristique des peuplements forestiers dans le bassin du Congo, et de proposer des outils d'aide à la décision pour atténuer ces impacts. Coordinatrice : Sylvie Gourlet-Fleury.

Rayonnement international

↔ Invitations

- Invitation à Naples - University of Naples Federico II (Italie) du 26 au 27 octobre 2015 pour un exposé sur SCGLR dans le cadre du montage du projet ERC *SHAPE* (Statistical modelling of Human Perception and Evaluation), centré sur les données ordinales issues d'évaluation.
- Invitation à Washington – Georgetown University (États-Unis) du 23 mars au 4 avril 2015. Séminaire et travail sur des modèles de mélange de variables dans le cadre de SCGLR.
- Invitation à Piracicaba – ESALQ - Université de Sao Paulo (Brésil) du 5 au 9 mai 2012 pour donner un cours "Linear Mixed Models, theory and applications" dans le cadre de la conférence annuelle de la région brésilienne de l'International Biometric Society avec Frédéric Mortier.
- Invitation à Piracicaba – ESALQ - Université de Sao Paulo (Brésil) du 10 au 17 novembre 2009 pour donner un cours "Generalized Linear Mixed Models, theory and applications" à des chercheurs et doctorants (en collaboration avec Frédéric Mortier). Ce cours a été proposé à nouveau en juin 2011 au CIRAD – France.

↔ Expertise pour revues internationales : Biometrics, Statistics and Computing, Computational Statistics and Data Analysis, Revue d'Économie Politique, Journal de la Société Française de Statistique.

Participation à des groupes de travail

- Membre associé à l'équipe Virtual Plants (Inria-Cirad-Inra) (2005-2012) "Modélisation de la croissance des plantes. Des gènes au phénotype." Responsable : Christophe Godin. Le travail en collabo-

ration avec Yann Guédon s'inscrit dans ce cadre.

- Collaboration avec le CIRAD - UR "Biens et Services des Écosystèmes Forestiers tropicaux" (B&SEF) (depuis 2009). C'est dans ce cadre que s'inscrit la collaboration avec Frédéric Mortier.
- Participation à un groupe de réflexion (2010) sur le thème "Sélection de variables dans le modèle mixte en grande dimension" avec Frédéric Mortier (CIRAD), Franck Picard (CNRS - Lyon), Sophie Lambert-Lacroix (université de Grenoble), Nicolas Verzelen (INRA).
- Membre du groupe de travail MODEST (MOdélisation des DÉpendances STructurelles) depuis 2010.
- Membre du groupe de travail « modèles statistiques à structures cachées » (2004-2009)

Divers

- Invitation 1 mois en juin 2011 du professeur Clarice Demetrio, responsable du département de statistique ESALQ – USP, présidente de l'International Biometric Society. Invitation financée par l'université Montpellier 2 et la région Languedoc-Roussillon dans le cadre de son soutien à l'accueil de personnalités étrangères sur critère de prestige.
- Délégation CNRS 6 mois (2014-2015).
- Titulaire de la PES (2011-2015) et de la PEDR (2016-2020).
- Développement et mises à jour du package R *SCGLR*.

Co-encadrement d'activités de recherche _____

Thèses

- 2017 - ... : thèse de Benjamin Heuclin (bourse du ministère)
Sélection bayésienne de variable pour données longitudinales : application à la modélisation de la dynamique forestière et à l'amélioration génétique.
Co-encadrement (33%) avec Frédéric Mortier et Marie Denis.
Sujet : Comprendre l'influence des déterminants environnementaux, génétiques et leurs interactions sur les processus de croissance ou de survie d'espèces forestières est crucial pour prédire les effets des changements climatiques ou anthropiques sur les écosystèmes naturels ou plantés et ainsi adapter leurs modes de gestion. L'étude de la dynamique forestière se base sur des dispositifs expérimentaux qui permettent le suivi régulier de la croissance ou de la mortalité de chaque individu au cours du temps et l'acquisition d'un grand nombre de variables environnementales ou génétiques. L'analyse de tels dispositifs soulève différentes questions méthodologiques parmi lesquelles (i) la prise en compte des dépendances entre les différentes

mesures au cours du temps, (ii) la sélection des variables pertinentes alors qu'elles sont nombreuses et enfin (iii) l'estimation des effets des facteurs environnementaux et génétiques qui peuvent évoluer au cours du temps. Différentes approches sont envisagées. La première, dans le cadre des SSVS, repose sur l'élicitation de lois a priori des variables indicatrices et l'utilisation de champs de Markov ou prior d'Ising adaptés aux données longitudinales (Li, 2010). La seconde repose sur les modèles à coefficients aléatoires et une modélisation de la dynamique de ces coefficients au cours du temps (Verbeke, 2000).

- 2015 - ... : thèse de Jocelyn Chauvet (bourse du ministère)

Structures de dépendance complexes pour modèles à composantes supervisées.

Co-encadrement (50%) avec Xavier Bry.

Sujet : Une forte redondance des covariables cause de gros problèmes d'identifiabilité et, par conséquent, d'instabilité des coefficients dans les modèles de régression. Elle entraîne en outre, même lorsque l'estimation est possible, la quasi-impossibilité d'en interpréter les résultats. Dans les méthodes classiques de régularisation que sont Ridge et Lasso, un critère supplémentaire de pénalité sur la norme du vecteur des coefficients de régression est combiné à la vraisemblance. La pénalisation de cette norme permet de limiter la confusion des effets entre régresseurs redondants, mais ne facilite pas nécessairement l'interprétation du modèle estimé. Les méthodes de régression sur composantes principales facilitent quelque peu l'interprétation, mais n'optimisent guère la prédiction, dans la mesure où ces composantes ne sont pas supervisées par les variables à prédire. Nous avons proposé une autre technique de régularisation pour les modèles linéaires généralisés multivariés. Jusqu'ici, elle n'a toutefois concerné que les modèles à observations indépendantes. Or, dans de nombreuses situations concrètes, ces dernières sont interdépendantes : observations répétées et données multi-niveaux en écologie et sciences humaines - notamment en épidémiologie - séries temporelles, données spatiales... Il s'avère donc essentiel de procéder à l'extension de la méthode aux données à structure de dépendance complexe entre observations.

- 2010 - 2013 : thèse de Jean Peyhardi

A new GLM framework for analyzing categorical data. Application to plant structure and development. - Université Montpellier 2.

Co-encadrement (50%) avec Yann Guédon.

Résumé : This thesis aims at proposing a new class of GLMs for a hierarchically-structured categorical response variable such as a partially-ordered variable for instance. A first step consisted of clarifying differences and commonalities between GLMs for nominal and ordinal response variables. On this basis we introduced a new specification of GLM for categorical response variable, whether ordinal or nominal, based on three components : the ratio of probabilities r , the cumulative distribution function F and the design matrix Z . This framework allowed us to define a new family of models for nominal data, similar to the cumulative, sequential and adjacent families of models for ordinal data. Then we defined the class of partitioned conditional GLMs (PCGLMs) using directed trees and (r, F, Z) specification. In our biological context, data takes the form of multivariate sequences associating a categorical response variable (type of axillary production) with explanatory variables (e.g. internode length). In the semi-Markov switching partitioned conditional generalized linear models

(SMS-PCGLM) estimated on the basis of these sequences, the underlying semi-Markov chain represents both the succession and lengths of branching zones, while the PCGLMs represent the influence of growth explanatory variables on axillary productions within each branching zone. On the basis of these integrative statistical models, we showed that shoot growth influences specific branching events.

↔ Jean est actuellement maître de conférences à l'université de Montpellier - faculté de pharmacie.

- 2005 - 2008 : thèse de Florence Chaubert (bourse du ministère)

Modèles linéaires généralisés mixtes à structure markovienne cachée. - Université Montpellier 2.

Co-encadrement (33%) avec Christian Lavergne et Yann Guédon.

Résumé : Ce travail est consacré à l'étude des combinaisons markoviennes et semi-markoviennes de modèles de régression, i.e. des mélanges finis de modèles de régression avec dépendances (semi-)markoviennes. Cette famille de modèles statistiques permet l'analyse de données structurées en phases successives asynchrones entre individus, influencées par des covariables pouvant varier dans le temps et présentant une hétérogénéité inter-individuelle. L'algorithme d'inférence proposé pour les combinaisons (semi-)markoviennes de modèles linéaires généralisés est un algorithme du gradient EM. Pour les combinaisons (semi-)markoviennes de modèles linéaires mixtes, nous proposons des algorithmes de type MCEM où l'étape E se décompose en deux étapes de restauration conditionnelle : une pour les séquences d'états sachant les effets aléatoires (et les données observées) et une pour les effets aléatoires sachant les séquences d'états (et les données observées). Différentes méthodes de restauration conditionnelle sont présentées. Nous étudions deux types d'effets aléatoires : des effets aléatoires individuels et des effets aléatoires temporels. L'intérêt de cette famille de modèles est illustré par l'analyse de la croissance d'arbres forestiers en fonctions de facteurs climatiques. Ces modèles nous permettent d'identifier et de caractériser les trois principales composantes de la croissance (la composante ontogénique, la composante environnementale et la composante individuelle). Nous montrons que le poids de chaque composante varie en fonction de l'espèce et des interventions sylvicoles.

↔ Florence est actuellement statisticienne en Recherche et Développement chez Michelin - Clermont-Ferrand.

- 2004 - 2010 : thèse de Harold-Andrés Guerrero Lopez (non financée)

Caractérisation de la voix de l'enfant sourd appareillé et implanté cochléaire : approches acoustique et perceptuelle et proposition de modélisation. - Université Montpellier 3.

Co-encadrement (15%) avec Jacques Bres, Michel Mondain, Melissa Barkat-Defradas, Patrick Serraféro

Résumé : Cette thèse propose une analyse comparative acoustique et perceptive de la voix d'un effectif statistiquement fiable d'enfants sourds appareillés et implantés cochléaires. Peu de paramètres diffèrent de manière significative entre le groupe d'enfants sourds ayant été appareillés et implantés avant l'âge de trois ans, et le groupe d'enfants entendants. L'ensemble des résultats indiquent que la voix des enfants de notre étude ne présente pas les caractéristiques

traditionnellement retenues pour déterminer la voix pathologique. Par ailleurs, les caractéristiques de la voix des enfants implantés cochléaires sont sensiblement comparables à celles des enfants entendants. Fort de ces résultats expérimentaux, nous avons proposé un modèle «vibro-acoustique» de la régulation de la voix des enfants sourds «oralisés», et développé un simulateur numérique de la boucle audio-phonatoire.

↔ Harold est actuellement membre associé au laboratoire Modyco (UMR 7114) de l'université Paris-Ouest - Nanterre La Défense.

- 2002 - 2006 : thèse de Marie-José Martinez (bourse du ministère)

Modèles linéaires généralisés à effets aléatoires : contributions au choix de modèle et au modèle de mélange. - Université Montpellier 2.

Co-encadrement (50%) avec Christian Lavergne.

Résumé : Ce travail est consacré à l'étude des modèles linéaires généralisés à effets aléatoires (GL2M). Dans ces modèles, sous une hypothèse de distribution normale des effets aléatoires, la vraisemblance basée sur la distribution marginale du vecteur à expliquer n'est pas, en général, calculable de façon formelle. Dans la première partie de notre travail, nous revisitons différentes méthodes d'estimation non exactes par le biais d'approximations réalisées à différents niveaux selon les raisonnements. La deuxième partie est consacrée à la mise en place de critères de sélection de modèles au sein des GL2M. Nous revenons sur deux méthodes d'estimation nécessitant la construction de modèles linéarisés et nous proposons des critères basés sur la vraisemblance marginale calculée dans le modèle linéarisé obtenu à la convergence de la procédure d'estimation. La troisième et dernière partie s'inscrit dans le cadre des modèles de mélanges de GL2M. Les composants du mélange sont définis par des GL2M et traduisent différents états possibles des individus. Dans le cadre de la loi exponentielle, nous proposons une méthode d'estimation des paramètres du mélange basée sur une linéarisation spécifique à cette loi. Nous proposons ensuite une méthode plus générale puisque s'appliquant à un mélange de GL2M quelconques. Cette méthode s'appuie sur une étape de Metropolis-Hastings pour construire un algorithme de type MCEM. Les différentes méthodes développées sont testées par simulations.

↔ Marie-José est actuellement maître de conférences à l'université Pierre Mendès France - Grenoble (IUT2)

M2 Biostatistique et stage fin d'étude ingénieur

- 2015 : Jocelyn Chauvet *Extension de la méthode "Régression Linéaire Généralisée sur Composantes Supervisées" (SCGLR) aux données groupées.* Co-encadrement avec Xavier Bry.
- 2010 : Jean Peyhardi *Modèles markoviens à transitions régies par des modèles linéaires généralisés mixtes.* Co-encadrement avec Jean-Baptiste Durand et Yann Guédon.
- 2008 : Inès Nefzi (ingénieur) *Estimation de modèles linéaires mixtes et de modèles linéaires généralisés mixtes. Application en génétique quantitative : pousse annuelle et floraison des pommiers.* Co-encadrement avec Yann Guédon et Florence Chaubert.
- 2007 : Mariem Zouch *Loi de Poisson pondérée et famille exponentielle bidimensionnelle.* Co-encadrement avec Christian Lavergne.

- 2006 : Mariem Zouch (ingénieur) *Modèle de Poisson pondéré et sous-dispersion. Application à l'analyse de données longitudinales de comptage de croissance de caféiers*. Co-encadrement avec Yann Guédon.
- 2005 : Tessa Pierre-Joseph *Analyse de données longitudinales de comptage. Modélisation de l'effet de la pluviométrie sur la croissance de caféiers*. Co-encadrement avec Yann Guédon.
- 2004 : Lahcen Gourari *Méthodes d'analyse de données longitudinales de comptage appliquées à des suivis de croissance de caféiers*. Co-encadrement avec Benoit Liquet.

Participation à comité de suivi de thèse

- Alexandra Jestin (thèse suspendue au bout d'un an)
- Antoine Barbieri (soutenue 2016) *Contribution à l'analyse statistique de la qualité de vie en cancérologie*.
- Caroline Mwongera (soutenue 2012) *How smallholder farmers cope with climate variability. A case study of the Eastern slopes of Mount Kenya*.
- Olivier Taugourdeau (soutenue 2011) *Le sapin pectiné (Abies alba Mill., PINACEAE) en contexte méditerranéen : développement architectural et plasticité phénotypique*.
- Ciré Elimane Sall (soutenue 2009) *Modélisation spatiale hiérarchique bayésienne de l'apparentement génétique et de l'héritabilité en milieu naturel à l'aide de marqueurs moléculaires*.
- Carine Véra (soutenue 2004) *Modèles linéaires mixtes multiphasiques pour l'analyse de données longitudinales : Application à la croissance des plantes*.

Responsabilités collectives

- Depuis septembre 2016, directrice du département MIAp (Mathématiques et Informatique Appliquées aux sciences humaines et sociales) de l'UFR6 de l'université Paul Valéry - Montpellier 3 (18 membres : 13 EC, 3 PRAG, 2 PAST)
- Depuis novembre 2013, membre du conseil de l'UFR 6 (création de l'UFR en 2013).
- Membre des comités de sélection de l'université de Montpellier (2015), de l'université Paul Valéry - Montpellier 3 (2015), de l'université Montpellier 1 (2013), et membre des commissions de spécialistes de l'université Montpellier 3 (2007) et de l'université d'Aix-Marseille (suppléante) (2007).
- Membre du comité d'organisation des journées SFdS2016
- De 2008 à 2013, j'ai été en charge de la coordination des divers enseignements associés aux 2 sections CNU 25 et 26 à l'université Paul Valéry Montpellier 3 : gestion des 2500 heqTD d'enseignement, ventilation des services, recrutement des vacataires, coordination. J'ai reçu une prime pédagogique pour cela.
- De 2008 à 2011, coordination des actions du plan "Réussite en Licence" pour l'UFR 4 (ancienne

UFR du département MIAP jusqu'en 2013) de l'université Paul Valéry - Montpellier 3.

- Membre du conseil d'UMR I3M (2004-2007).

Liste des publications

Revue de statistique :

- 2017 Chauvet J., Trottier C., Bry X. "Component-based regularisation of generalised linear mixed models.", *soumis à Journal of Computational and Graphical Statistics*
- 2017 Bry X., Trottier C., Mortier F., Cornu G. "Supervised-component-based multivariate GLM for a thematic partitioning of the explanatory variables.", *en révision à Statistical Modelling*
- 2016 Peyhardi J., Trottier C., Guédon Y. "Partitioned conditional generalized linear models for categorical data.", *Statistical Modelling*, 16(4), pp 297-321
- 2015 Peyhardi J., Trottier C., Guédon Y. "A new specification of generalized linear models for categorical responses.", *Biometrika*, 102, pp. 889-906
- 2013 Bry X., Trottier C., Verron T., Mortier F. "Supervised Component Generalized Linear Regression using a PLS-extension of the Fisher scoring algorithm.", *Journal of Multivariate Analysis*, 119, pp. 47-60
- 2010 Chaubert - Pereira F., Guédon Y., Lavergne C., Trottier C. "Markov and Semi-Markov Switching Linear Mixed Models Used to Identify Forest Tree Growth Components.", *Biometrics*, 66(3), pp. 753-762
- 2009 Martinez M.-J., Lavergne C., Trottier C. "A mixture model-based approach to the clustering of exponential repeated data.", *Journal of Multivariate Analysis*, 100(9), pp. 1938-1951
- 2008 Lavergne C., Martinez M.-J., Trottier C. "Empirical Model Selection in Generalized Linear Mixed Effects Models.", *Computational Statistics*, 23 (1), pp. 99-110
- 2003 Diebolt J., Garrido M., Trottier C. "Improving extremal fit : a Bayesian regularization procedure.", *Reliability Engineering and System Safety*, 82, pp. 21-31
- 2000 Trottier C. "A quasi-score marginal approach in generalized linear mixed models.", *Statistics*, 33(4), pp. 291-308
- 2000 Lavergne C., Trottier C. "Sur l'estimation dans les modèles linéaires généralisés mixtes.", *Revue de Statistiques Appliquées*, XLVIII (1), pp. 45-63

Revue dans des domaines d'application :

- 2017 Trouillet R., Bry X., Trottier C., Briki W., Brouillet D. and Doron J. "Effect of the ordering of a simple digit span task on performances : estimation of the built-up of proactive interference and the memory of previous successes", *soumis à Cognitive Psychology*
- 2013 Durand J.B., Guitton B., Peyhardi J., Holtz Y., Guédon Y., Trottier C. and Costes E. "New insights for estimating the genetic value of segregating apple progenies for irregular bearing during the first years of tree production.", *Journal of Experimental Botany*, 64 (16), pp. 5099-5113
- 2013 Guerrero Lopez H.A., Mondain M., Amy de la Bretèque B., Serrafiero P., Trottier C., Barkat-Defradas M. "Acoustic, aerodynamic, and perceptual analyses of the voice of cochlear-implanted children.", *Journal of Voice*, 27(4), pp 523.e1-523.e17.
- 2013 Dambreville A., Lauri P.-E., Guédon Y., Trottier C., Normand F., "Deciphering structural and temporal interplays during the architectural development of mango trees.", *Journal of Experimental Botany*, 64(8), pp. 2467–2480
- 2011 Lauri P.-E., Gorza O., Cochard H., Martinez S., Celton J.-M., Ripetti V., Lartaud M., Bry X., Trottier C. and Costes E. "Genetic determinism of anatomical and hydraulic traits within an apple progeny.", *Plant, Cell and Environment*, 34, pp. 1276-1290
- 2010 Lauri P.-E., Kelner J. J., Trottier C., Costes E. "Insights into secondary growth in perennial plants : its unequal spatial and temporal dynamics in the apple (*Malus domestica*) is driven by architectural position and fruit load.", *Annals of Botany*, 105, pp. 607 – 616
- 2009 Normand F., Pambo Bello A. K., Trottier C. and Lauri P.-E. "Is axis position within tree architecture a determinant of axis morphology, branching, flowering and fruiting? An essay in mango.", *Annals of Botany*, 103, pp. 1325-1336
- 2008 Lauri P.-E., Bourdel G., Trottier C. and Cochard H. "Apple shoot architecture : evidence for strong variability of bud size and composition and hydraulics within a branching zone.", *New Phytologist*, 178, pp. 798-807
- 2007 Han H.-Hee., Coutand C., Cochard H., Trottier C. and Lauri P.-E. "Effects of shoot bending on lateral fate and hydraulics : invariant and changing traits across five apple genotypes.", *Journal of Experimental Botany*, 58(13), pp. 3537-3547
- 2006 Lauri P.-E., Maguylo K. and Trottier C. "Architecture and size relations : an essay in the apple (*Malus x domestica*) tree.", *American Journal of Botany*, 93(3), pp 357-368
- 2004 Lauri P.-E. and Trottier C. "Patterns of size and fate relationships of contiguous organs in the apple (*Malus domestica*) crown.", *New Phytologist*, 163, pp. 533-546

Conférences internationales :

- 2017 Chauvet J., Trottier C., Bry X. "Regularisation of Generalised Linear Mixed Models with autoregressive random effect", 32nd International Workshop on Statistical Modelling IWSM 2017, Groningen, Pays-Bas

- 2017 Bry X., Trottier C. "Supervised-Component Regression Methods", 37th Conference on Applied Statistics in Ireland CASI 2017, Mullingar, Irlande
- 2016 Chauvet J., Trottier C., Bry X., Mortier F. "Extension to mixed models of the Supervised Component-based Generalised Linear Regression.", 22nd International Conference on Computational Statistics, COMPSTAT 2016, Oviedo, Espagne
- 2015 Trottier C, Bry X., Mortier F., Cornu G., "A new component based approach of regularisation for multivariate generalised linear regression.", CLADAG2015, Cagliari, Italie, *sollicitée*.
- 2015 Bry X., Trottier C., Verron T., Mortier F., Cornu G., "Supervised Component-based Generalised Linear Regression with multiple explanatory themes : THEME-SCGLR.", ICCSDA, Istanbul, Turquie
- 2014 Bry X., Trottier C., Verron T., Mortier F., Cornu G., "Supervised Component-based Generalised Linear Regression with multiple explanatory themes : THEME-SCGLR.", ISC12- proceedings, Kermanshah, Iran
- 2014 Peyhardi J., Trottier C., and Guédon Y., "Partitioned conditional generalized linear models for categorical data.", 29th International Workshop on Statistical Modeling, Göttingen, Allemagne
- 2014 Mortier F., Bry X., Cornu G. and Trottier C. "SCGLR : a component-based multivariate regression method to model species distributions.", International Statistical Ecology Conference, Montpellier, France
- 2014 Bry X., Trottier C., Mortier F. and Verron T. "Extending SCGLR to multiple regressor groups : the Theme-SCGLR method.", 8th International Conference on Partial Least Squares and Related Methods, Montpellier, France
- 2013 Bry X., Trottier C., Mortier F. "Supervised component generalized linear regression.", 6th International Conference of the ERCIM WG on Computational and Methodological Statistics, Londres, Royaume-Uni
- 2013 Peyhardi J., Costes E., Caraglio Y., Lauri P.-E., Trottier C., Guédon Y., "Integrative models for analyzing jointly shoot growth and branching patterns.", 7th International Workshop on Functional Structural Plant Models, Saariselkä, Finlande
- 2013 Durand J.-B., Guitton B., Peyhardi J., Holtz Y. Guédon Y. Trottier C., Costes E., "Estimating the genetic value of F1 apple progenies for irregular bearing during first years of production.", 7th International Workshop on Functional Structural Plant Models, Saariselkä, Finlande
- 2013 Peyhardi J., Trottier C., and Guédon Y., "A unifying framework for specifying generalized linear models for categorical data.", 28th International Workshop on Statistical Modeling, Palermo, Italie
- 2012 Bry X., Trottier C., Verron T., Mortier F., "Supervised Component Generalized Linear Regression using a PLS-extension of the Fisher scoring algorithm.", COMPSTAT 2012, Limassol, Chypre
- 2012 Trottier C., Bry X., Mortier F., Verron T., "Component based Generalized Linear Regression using a PLS extended variant of the Fisher scoring algorithm.", 57a Reunião anual da RBRAS, Piracicaba, Brésil

- 2008 Chaubert-Pereira F., Guédon Y., Lavergne C., Trottier C. "Semi-Markov switching linear mixed model for analyzing forest tree development.", International Biometric Conference (IBC), Dublin, Irlande.
- 2008 Chaubert-Pereira F., Guédon Y., Lavergne C., Trottier C. "Estimating Markov and semi-Markov switching linear mixed models with individual-wise random effects.", COMPSTAT'2008, 18th Symposium of IASC, Volume II, pp. 11-18, Porto, Portugal
- 2007 Chaubert F., Caraglio Y., Lavergne C., Trottier C., Guédon Y. "A statistical model for analyzing jointly growth phases, the influence of environmental factors and inter-individual heterogeneity. Applications to forest trees." , 5th International Workshop on Functional-Structural Plant Models, p. 43, Napier, New Zealand
- 2006 Martinez M.-J., Lavergne C. and Trottier C., "Finite mixture model for exponential repeated data.", International conference on Statistical Latent Variable Models in the Health Sciences, Perugia, Italie.
- 2001 Garrido M., Diebolt J., Trottier C. "A Bayesian Regularization Procedure for a Better Extremal Fit.", Extremes 2001, International Symposium on Extreme Value Analysis : Theory and Practice, Leuven, Belgique
- 2000 Trottier C., Diebolt J. "Adjusting density functions for a better extremal fit.", Second International Conference on Mathematical Methods in Reliability, Bordeaux, France
- 1997 Lavergne C., Trottier C. "From a conditional to a marginal point of view in GL2M.", International Workshop on Statistical Modelling, Proceedings in Good Statistical Practice (1997), pp. 205-209, Biel/Bienne, Suisse

Communications :

- 2017 Chauvet J., Trottier C., Bry X. "Régularisation dans les Modèles Linéaires Généralisés Mixtes avec effet aléatoire autorégressif", Journées SFdS, Avignon, France
- 2016 Chauvet J., Trottier C., Bry X., Mortier F. "Extension de la régression linéaire généralisée sur composantes supervisées (SCGLR) aux données groupées", Journées SFdS, Montpellier, France
- 2015 Trottier C. "A new component-based approach of regularization for multivariate generalized linear regression." Naples, Italie.
- 2015 Trottier C., Bry X., Mortier F., Cornu G., Verron T. "Extension de la régression linéaire généralisée sur composantes supervisées à une partition thématique des régresseurs.", Journées SFdS, Lille, France.
- 2015 Trottier C. "A new component-based approach of regularization for multivariate generalized linear regression." Washington, États-Unis.
- 2014 Trottier C., Mortier F., Cornu G., Bry X. "SCGLR : un package R pour la régression linéaire généralisée sur composantes supervisées", 3èmes Rencontres R, Montpellier, France

- 2014 Peyhardi J., Trottier C., Guédon Y. "Modèles de références de régression multinomiale. Propriétés et applications en classification supervisée", Journées SFdS, Rennes, France.
- 2012 Peyhardi J., Trottier C., Guédon Y. "GLM multinomial hiérarchique", Journées SFdS, Bruxelles, Belgique.
- 2009 Guerrero H., Amy de la Breteque B., Mondain M., Serrafiero P., Trottier C., Barkat-Defradas M. "Étude descriptive préliminaire de la voix de l'enfant i+implanté cochléaire à partir des mesures aérodynamiques", 8èmes Rencontres Jeunes Chercheurs en Parole RJCP, Avignon, France.
- 2009 Guerrero H., Mondain M., Amy de la Breteque B., Serrafiero P., Trottier C., Barkat-Defradas M. "Identification et caractérisation acoustique et physiologique des troubles vocaux manifestés chez l'enfant implanté cochléaire", 3èmes Journées de Phonétique Clinique JPC3, Aix-en-Provence, France.
- 2008 Trottier C., Lavergne C., Guédon Y., Zouch M. "Weighted Poisson mixed model for underdispersed longitudinal count data", Journées SFdS, Ottawa, Canada.
- 2007 Chaubert F., Guédon Y., Lavergne C., Trottier C. "Combinaison markovienne de modèles linéaires généralisés pour données structurées en phases.", Journées SFdS, Angers, France.
- 2006 Martinez M.-J., Lavergne C., Trottier C. "Algorithme de type MCEM pour un modèle de mélange exponentiel mixte.", Journées SFdS, Clamart, France.
- 2005 Martinez M.-J., Lavergne C., Trottier C. "Mélange de modèles exponentiels à effets aléatoires.", Journées SFdS., Pau, France.
- 2004 Martinez M.-J., Lavergne C., Trottier C. "Modèles mixtes et loi exponentielle. Application à la fiabilité.", Journées SFdS, Montpellier, France.
- 1998 Trottier C., Lavergne C. "Une hétérogénéité dans les modèles linéaires généralisés mixtes", Journées SFdS, Rennes, France.
- 1997 Trottier C. "Les modèles linéaires généralisés mixtes : extension de la méthode GAR", Journées SFdS, Carcassonne, France.
- 1996 Trottier C., Lavergne C. "Estimation des composantes de la variance dans un GL2M", Journées SFdS, Québec, Canada.