

# Table of Contents • Table des matières

Welcome • Bienvenue	2
Sponsors • Commanditaires	3
Organizers • Organisateurs	4
Exhibitors • Exposants	5
Language of Presentations • Langue des présentations	5
General Information • Information générale	6
Committees and Meetings • Comités et réunions	11
List of Sessions • Liste des sessions	15
Program • Programme	20
Abstracts • Résumés	64
Author Index • Index des auteurs	201
Author Index (per session) • Index des auteurs (par session)	207

# Welcome • Bienvenue

It is with great pleasure that the members of the Department of Mathematics and Statistics at Université Laval welcome you to the 38th Annual Meeting of the Statistical Society of Canada. We are delighted to host the SSC once again, whose last visit within our walls dates back to 1987. We are also glad to welcome the members of the *Association des statisticiennes et des statisticiens du Québec* who have enthusiastically accepted our invitation to hold their annual meeting jointly with ours.

Université Laval was the very first francophone university in America. Its origins date back to 1663, when Monseigneur de Laval founded the *Séminaire de Québec*. The university per se was officially created in 1852 when Queen Victoria signed the Royal Charter that gave the right to the *Séminaire de Québec* to “confer degrees” and “all rights, powers and privileges of a University”. Université Laval now has three campuses and its student population is approximately 44,000. The vast majority of its activities take place on the *cité universitaire*, which is located on a 190 hectare lot in Sainte-Foy. Aside from its numerous buildings of various architectural styles, plenty of green spaces can be found on campus, such as sports fields, wooded areas, hiking trails, bicycle paths and a botanical garden.

Of course, we strongly encourage you to take advantage of your stay in the area to go for a walk in the historical Old Québec district, a small piece of Europe only 7 km away from campus! Be it to admire the ramparts (declared UNESCO World Heritage Site), the Petit Champlain district, Château Frontenac, the St. Lawrence river or the Battle Fields of the Plains of Abraham, or to visit the numerous restaurants, pubs or museums that can be found there, a visit in the Old City can only enchant you.

C'est avec grand plaisir que les membres du Département de mathématiques et de statistique vous accueillent à l'Université Laval pour le 38<sup>e</sup> congrès annuel de la Société statistique du Canada. Nous sommes très heureux d'accueillir à nouveau la SSC, dont la dernière visite en nos murs remonte à 1987. Nous voulons aussi souhaiter la bienvenue aux membres de l'Association des statisticiennes et des statisticiens du Québec, qui ont accepté avec enthousiasme notre invitation à tenir leur rencontre annuelle avec nous.

L'Université Laval a été la toute première université francophone en Amérique. Ses origines remontent à 1663 alors que Monseigneur de Laval a fondé le Séminaire de Québec. L'université a elle-même été officiellement créée en 1852, quand la reine Victoria a signé la Charte royale qui donne au Séminaire de Québec le droit de « conférer des degrés » et « tous les droits, pouvoirs et privilèges d'Université ». L'Université Laval compte aujourd'hui trois campus et quelque 44 000 étudiants. La grande majorité des activités de l'Université Laval se déroulent sur la cité universitaire, localisée sur un terrain de 190 hectares à Sainte-Foy. En plus de ses nombreux pavillons de styles architecturaux variés, on retrouve sur le campus de nombreux espaces verts tels des terrains de sport, des boisés, des sentiers pédestres, des pistes cyclables et un jardin botanique.

Bien sûr, nous vous encourageons vivement à profiter de votre séjour dans la région pour aller marcher dans l'arrondissement historique du Vieux-Québec, un petit bout de l'Europe à seulement 7 km du campus ! Que ce soit pour admirer les fortifications (classées joyau du patrimoine mondial par l'UNESCO), le quartier Petit-Champlain, le Château Frontenac, le fleuve Saint-Laurent ou le parc des champs de bataille des plaines d'Abraham, ou pour visiter les nombreux pubs, restaurants ou musées qui s'y trouvent, un passage dans la Vieille Ville ne pourra que vous enchanter.

# Sponsors • Commanditaires

SSC 2010 would like to thank its numerous sponsors whose generous contribution has made this meeting possible:

SSC 2010 désire remercier ses nombreux commanditaires dont la généreuse contribution a rendu possible la tenue de ce congrès :

- Association des statisticiennes et des statisticiens du Québec
- Centre de recherches mathématiques
- Centre interuniversitaire québécois de statistiques sociales
- Chaire de recherche du Canada en échantillonnage statistique et en analyse de données
- Département de mathématiques et de statistique, Université Laval
- Faculté des sciences et de génie, Université Laval
- Fields Institute
- Institut de la statistique du Québec
- MITACS
- Pacific Institute for the Mathematical Sciences
- SAS
- Service de consultation statistique, Université Laval
- Zone, Coopérative de l'Université Laval

# Organizers • Organisateurs

## Local Arrangements Committee • Comité des arrangements locaux

**Chair • Président:** Thierry Duchesne

It would be impossible to organize an event of the size of the Annual Meeting of the SSC without the help of several individuals and organizations. I would first like to thank the members of the Local Arrangements Committee who generously devoted a large portion of their time to the organization of this meeting: Gaétan Daigle, Lajmi Lakhali-Chaieb, Sylvie Lambert and Louis-Paul Rivest. I must also thank our web master Sophie Baillargeon for her efficient maintenance of the conference web site. We benefited from a lot of support from many employees from other departments and divisions of the university: Guylaine Bélanger (Faculty of Forestry), Hélène Bissonnette (Faculty of Sciences and Engineering), Pierre Gauthier and Carole Nadeau (Faculty of Medicine), Martin Guay (Communications Direction), Robert Gosselin (University Theater), Christiane Masson (Finance Services), Nicole Nadeau (Public Affairs Division) and Nicolas Roy (Sodexo caterer).

Many members of the SSC also provided valuable help. Thanks to the organizers of previous meetings, in particular to Christian Léger and Matias Salibián-Barrera, for their numerous pieces of advice and to the members of the Executive Committee of the SSC (in particular to Duncan Murdoch, the Meetings Coordinator) for their suggestions and their help with the coordination between the local arrangements, the SSC office, the program committee and the members of the SSC. Thanks also to Gertrud Jeewanjee at the SSC office for all her work with the registrations and with the preparation of the conference bag contents.

An important element in the preparation of the Annual Meeting of the SSC is the translation of the abstracts and of the visual aids of the plenary sessions. Thanks to Félix Labrecque-Synnott for these translations, and to the Bilingualism Committee of the SSC (chaired by Cynthia Bocci) for its supervision of this work.

The program was prepared by the CRM. Particular thanks go to André Montpetit and Daniel Ouimet.

Il est impossible d'organiser un événement de l'envergure du congrès annuel de la SSC sans l'aide de nombreux individus et organismes. Je tiens tout d'abord à remercier les membres du comité des arrangements locaux qui ont généreusement accordé une grande partie de leur temps à l'organisation de ce congrès depuis de nombreux mois : Gaétan Daigle, Lajmi Lakhali-Chaieb, Sylvie Lambert et Louis-Paul Rivest. Je dois aussi remercier Sophie Baillargeon, notre webmestre, qui a toujours très efficacement tenu le site web du congrès à jour. Plusieurs employés d'autres départements et divisions de l'université nous ont aussi apporté beaucoup de soutien : Carole Nadeau et Pierre Gauthier (faculté de médecine), Nicole Nadeau (direction des affaires publiques), Martin Guay (direction des communications), Hélène Bissonnette (faculté des sciences et de génie), Robert Gosselin (théâtre de la cité universitaire), Guylaine Bélanger (faculté de foresterie), Nicolas Roy (traiteur Sodexo) et Christiane Masson (service des finances).

Plusieurs membres de la SSC nous ont également beaucoup aidés. Merci aux organisateurs des congrès antérieurs, en particulier à Christian Léger et Matias Salibián-Barrera, pour leurs nombreux conseils et aux membres de l'exécutif de la SSC (en particulier à Duncan Murdoch, le coordinateur des congrès) pour leurs suggestions et leur aide dans la coordination entre les arrangements locaux, le bureau de la SSC, le comité du programme et les membres de la SSC. Merci aussi à Gertrud Jeewanjee du bureau de la SSC pour tout son travail au niveau des inscriptions et de la préparation de la trousse des congressistes.

Un élément important de la préparation du congrès de la SSC est la traduction des résumés et des supports visuels des conférences plénières. Merci à Félix Labrecque-Synnott pour ces traductions et au comité du bilinguisme de la SSC (présidé par Cynthia Bocci) pour en avoir assuré la supervision.

Le programme a été préparé par le CRM. Un merci particulier à André Montpetit et à Daniel Ouimet.

## Program Committee • Comité du programme

**Chair • Président:** Christian Léger, *Université de Montréal*

**Members • Membres:**

**Biostatistics Section • Groupe de biostatistique**

Robert Platt, *McGill University*

**Business and Industrial Statistics Section • Groupe de statistique industrielle et de gestion**

Smiley Cheng, *University of Manitoba*

**Probability Section • Groupe de probabilité**

Murray Burke, *University of Calgary*

**Survey Methods Section • Groupe de méthodologie d'enquête**

Steven Thompson, *Simon Fraser University*

**Association des statisticiennes et statisticiens du Québec**

Martin Rioux, *Geyser Statistique*

**Concentration of Sessions in Actuarial Science • Concentration de séances en actuariat**

José Garrido, *Concordia University*

## Exhibitors • Exposants

Displays for examination and purchase will be available in the hall of the Ferdinand-Vandry building. Confirmed exhibitors at the time this program went to press are: CRC Press, Minitab, Nelson, SAS, Springer, SPSS, W.H. Freeman & Company and Wiley.

Les kiosques des exposants seront situés dans le hall du pavillon Ferdinand-Vandry. Au moment de mettre ce programme sous presse, les exposants suivants ont confirmé leur présence : CRC Press, Minitab, Nelson, SAS, Springer, SPSS, W.H. Freeman & Company et Wiley.

## Language of Presentations • Langue des présentations

At the time that they submitted their abstract, speakers were asked to provide the language in which they intend to give their oral presentation as well as the language of their visual aids. Icons are used to provide this information for each paper. For the oral presentation, we have used the icons , , whereas ,  indicate the language of the visual aids. The letter inside identifies the language: E for English and F for French. Please note that the visual aids for the plenary talks will be provided in both languages.

Lorsque les conférenciers ont soumis leur résumé, ils ont spécifié la langue dans laquelle ils comptaient faire leur présentation orale, ainsi que la langue du support visuel. À titre informatif, nous avons inclus cette information à l'aide d'icônes pour chaque présentation. Pour la présentation orale nous avons utilisé les icônes , , alors que ,  indiquent le support visuel. La lettre à l'intérieur identifie la langue : F pour français et E pour anglais (English). Veuillez noter que le support visuel des conférences plénières sera présenté dans les deux langues.

# General Information • Information générale

## Registration • Inscription

The registration desk will be located in the hall of the Ferdinand-Vandry building. The registration schedule is as follows:

- Sunday May 23: 8:00–18:00
- Monday May 24: 7:00–17:00
- Tuesday May 25: 8:00–17:00
- Wednesday May 26: 8:00–17:00

Please note that the registration desk is located in the same building as the Sunday workshops, **but at more than a 10 minute walk from the building of the opening sessions on Monday and Tuesday.**

Le comptoir d'inscription sera situé dans le hall du pavillon Ferdinand-Vandry. L'horaire des inscriptions sera le suivant :

- Dimanche 23 mai : 8h à 18h
- Lundi 24 mai : 7h à 17h
- Mardi 25 mai : 8h à 17h
- Mercredi 26 mai : 8h à 17h

Veillez noter que le comptoir d'inscription est situé dans le même pavillon que les ateliers du dimanche, **mais à plus de 10 minutes de marche des séances d'ouverture du lundi et du mardi.**

## Directions • Emplacement

Campus maps are available in the conference bag and **on the back cover of this program**. The Université Laval Campus is located in the Sainte-Foy district of Québec City, 4 km away from the bridges and 7 km away from Old Québec. Underground and outside parking are available at the Ferdinand-Vandry building, for which parking tickets can be purchased in one of the pay and display machines (circled “P” on the campus map) or at the Parking Office (Ernest-Lemieux building).

For those staying at the Sainte-Foy hotels, it is possible to walk to campus (about 2.5 km) or to take one of the buses of the *Réseau de transport de la Capitale* (RTC) (route 11, about 25 minutes total travel time). Those who choose to walk can, if they wish, cover more than one third of the itinerary inside the Laurier, Place de la Cité and Place Sainte-Foy shopping malls.

Those who choose to stay at the Palace Royal hotel have to allow about 30 minutes to cover the distance (6.5 km) by bus. Routes 7, 11, 800 and 801 all leave from the *Terminus d'Youville* (located just behind the hotel) and travel westbound with stops on or close to the university campus. The RTC's web site offers a tool that gives detailed itineraries for all of its bus routes (<http://www.rtcquebec.ca/Trajecto/tabid/296/Default.aspx>, click on “Trajecto”, in French only).

Cash payment (\$2.60, no change given) is accepted on buses. Bus tickets (\$2.50) or day passes (\$6.70) can also be purchased in most convenience stores, grocery stores, pharmacies, etc.

Des cartes du campus sont disponibles dans le sac du congressiste et **à l'endos de ce programme**. Le campus de l'Université Laval est situé dans l'arrondissement Sainte-Foy de Québec, à 4 km des ponts et à 7 km du Vieux-Québec. Un stationnement sous-terrain de même qu'un stationnement extérieur sont disponibles au pavillon Ferdinand-Vandry, pour lesquels il faut se procurer un billet de stationnement dans un des horodateurs (voir les « P » encadrés sur la carte du campus) ou au bureau du stationnement (pavillon Ernest-Lemieux).

Pour ceux qui logent aux hôtels de Sainte-Foy, il est possible de se rendre sur le campus en marchant (environ 2,5 km) ou en prenant les autobus du Réseau de transport de la Capitale (RTC) (parcours numéro 11, environ 25 minutes au total). Ceux qui choisissent de marcher peuvent, s'ils le désirent, faire plus du tiers du trajet à l'intérieur des centres commerciaux Laurier, Place de la Cité et Place Sainte-Foy.

Ceux qui auront choisi l'hôtel Palace Royal doivent prévoir environ 30 minutes pour le trajet (6,5 km) en autobus. Les autobus 7, 11, 800 et 801 qui partent tous du Terminus d'Youville (situé derrière l'hôtel), vont tous vers l'ouest et font tous au moins un arrêt sur ou près du campus de l'Université Laval. Le site web du RTC offre un outil donnant des itinéraires détaillés pour chacun de ses parcours d'autobus (<http://www.rtcquebec.ca/Trajecto/tabid/296/Default.aspx>, cliquer sur « Trajecto »).

Les paiements en argent (2,60\$, on ne vous rend pas la monnaie) sont acceptés dans les autobus. Vous pouvez aussi acheter des billets d'autobus (2,50 \$) ou des passes journalières (6,70 \$) dans la plupart des dépanneurs, épiceries, pharmacies, etc.

## Rooms • Salles

All scientific activities, coffee breaks and annual general meetings will take place in the rooms and the hall of the Ferdinand-Vandry building (VAN, number 9 on campus map), except for the plenary sessions on Monday (am and pm) and Tuesday (am) that will be in the University Theatre (Palasis-Prince building, PPR, number 6 on campus map) and on Tuesday (pm) and Wednesday (am and pm) that will be in the auditorium 1112 of the Adrien-Pouliot building (POU, number 12 on campus map). Committee meetings will take place in the rooms 2320 and 2330 of the Gene-H.-Kruger building (KRU, number 32 on campus map). The welcome reception and the barbecue will take place in the Atrium of the Charles-De Koninck building (DKN, number 13 on campus map).

Toutes les activités scientifiques de même que les pauses-café et les assemblées générales annuelles se dérouleront dans les salles et dans le hall du pavillon Ferdinand-Vandry (VAN, pavillon numéro 9 sur la carte), à l'exception des conférences plénières du lundi (am et pm) et du mardi (am) qui auront lieu au Théâtre de la cité universitaire (pavillon Palasis-Prince, PPR, numéro 6 sur la carte) et du mardi (pm) et mercredi (am et pm) qui auront lieu à l'auditorium 1112 du pavillon Adrien-Pouliot (POU, numéro 12 sur la carte). Les réunions des différents comités auront lieu dans les salles 2320 et 2330 du pavillon Gene-H.-Kruger (KRU, numéro 32 sur la carte). Le coquetel de bienvenue et le barbecue se dérouleront à l'Atrium du pavillon Charles-De Koninck (DKN, numéro 13 sur la carte).

## Workshops • Ateliers

The three workshops organized by the four sections will take place on Sunday May 23 from 9:00 to 16:00 in the Ferdinand-Vandry building (VAN). The Biostatistics Section workshop will be in room VAN 1245A, the Survey Methods Section workshop will be in room VAN 2811A and the joint workshop of the Business and Industrial Statistics and Probability sections will be in room VAN 2809A.

Les trois ateliers organisés par les quatre groupes auront lieu le dimanche 23 mai de 9h à 16h au pavillon Ferdinand-Vandry (VAN). L'atelier du groupe de biostatistique aura lieu dans la salle VAN 1245A, l'atelier du groupe de méthodologie d'enquête aura lieu dans la salle VAN 2811A et l'atelier conjoint des groupes de probabilité et de statistique industrielle et de gestion aura lieu dans la salle VAN2809A.

## Poster Session • Séance d'affichage

All poster and case studies sessions will take place in the hall of the Ferdinand-Vandry (VAN) building either on Monday May 24 or Tuesday May 25 between 8:30 and 18:00. Participants will be able to interact with authors during the second session of the afternoon as they are expected to be by their poster during that time period.

Toutes les séances d'affichage et les études de cas se dérouleront dans le hall du pavillon Ferdinand-Vandry (VAN) soit le lundi 24 ou mardi 25 mai de 8h30 à 18h. Les congressistes pourront interagir avec les auteurs durant la seconde séance de l'après-midi, moment où les auteurs sont invités à être présents près de leur affiche.

## **Panel Discussion on the Role, Benefits, and Importance of a National Statistics Institute • Groupe de discussion sur le rôle, les bénéfices et l'importance d'un Institut national en statistique**

**Sunday, May 23 • Dimanche 23 mai, 16:15, VAN 1245A**

A panel discussion will consider the importance and support of statistical research for Canada, and the need for a national statistical sciences institute will be reassessed. Broad input is encouraged at this discussion; there will be ample time for engagement by the community after the panelists provide their viewpoints on these issues. Both large-scale visionary remarks as well as suggestions for smaller ways we can support each other and the pivotal role of statistical science in Canada are welcome.

Un groupe de discussion étudiera l'importance et le support de la recherche en statistique au Canada, et les besoins pour un institut national sur les sciences statistiques seront réévalués. Comme une vaste contribution à la discussion est encouragée, la communauté aura amplement le temps de s'exprimer quand les panélistes auront affirmé leur point de vue sur ces questions. Tant les remarques visionnaires à grande échelle que les suggestions de petits gestes de support mutuel ou du rôle pivot de la statistique au Canada sont les bienvenues.

## **Information on 2010 NSERC Discovery Grant Competition • Information sur le concours 2010 des subventions à la découverte du CRSNG**

**Monday, May 24 • Lundi 24 mai, 18:10, VAN 1245A**

Representatives from NSERC will make a presentation to inform researchers on the peer review process for NSERC Discovery Grants. Summary statistics will also be provided on the 2010 Competition. All researchers are welcome to attend.

Des représentants du CRSNG feront une présentation afin d'informer les chercheurs sur le processus d'évaluation par les pairs des subventions à la découverte du CRSNG. Des statistiques résumant le concours de 2010 seront fournies. Tous les chercheurs sont les bienvenus.

## **Welcome Reception • Réception de bienvenue**

We cordially invite all conference participants to a welcome reception in the Atrium of the Charles-De Koninck building on Sunday May 23rd from 18:00 to 20:30. Coupons for free drinks are available in the conference bags.

Nous invitons cordialement tous les participants à la conférence à une réception de bienvenue à l'Atrium du pavillon Charles-De Koninck le dimanche 23 mai de 18h à 20h30. Des coupons donnant droit à des consommations gratuites se trouvent dans le sac du congressiste.

## **Barbecue**

A barbecue is organized in the Atrium of the Charles-De Koninck building on Monday May 24th from 18:30 until 21:00. Graduate students and other conference participants who have purchased tickets for this event will find these tickets in their conference bag.

Un barbecue est organisé à l'Atrium du pavillon Charles-De Koninck le lundi 24 mai de 18h30 à 21h. Tous les étudiants gradués ainsi que les autres participants qui ont acheté des billets pour cet événement les trouveront dans le sac du congressiste.

## **Banquet**

The conference banquet will take place at Hotel Plaza Québec (3031 boulevard Laurier) in the evening of Tuesday May 25. A cocktail with cash bar will start at 18:30 and dinner service will start at 20:15. All conference participants who have not declined their participation to the banquet will find a colored ticket in their conference bag. They must bring this banquet ticket

with them and put it on their table to indicate their choice of menu. Note to musicians: A piano will be available if you feel like entertaining the crowd!

Though it is possible to walk from the Ferdinand-Vandry building to Hotel Plaza Québec (2.5 km), transportation by bus from this building to the Hotel Plaza Québec will be offered to conference participants. There will be several departures between 18:15 and 19:30 to allow participants to attend the SSC's Annual General Meeting. Buses will also be available to bring participants back from Hotel Plaza Québec to the university residences and to Hotel Palace Royal at the end of the banquet near 23:00.

Le banquet du congrès se tiendra à l'hôtel Plaza Québec (3031, boulevard Laurier) en soirée le mardi 25 mai. Un coquetel avec bar payant débutera à 18h30 et le souper sera servi à partir de 20h15. Tous les participants au congrès qui n'ont pas décliné leur participation au banquet retrouveront un billet de couleur qu'ils devront déposer sur leur table pour indiquer leur choix de menu. Avis aux musiciens : un piano sera disponible si vous désirez égayer la foule !

Bien qu'il soit possible de marcher du pavillon Ferdinand-Vandry à l'hôtel Plaza Québec (2,5 km), un transport en autobus de ce pavillon jusqu'à l'hôtel Plaza Québec sera offert aux congressistes. Plusieurs départs auront lieu entre 18h15 et 19h30 afin de permettre aux participants d'assister à l'assemblée générale annuelle de la SSC. Des autobus seront également disponibles pour transporter les congressistes de l'hôtel Plaza Québec aux résidences et à l'hôtel Palace Royal à la fin du banquet vers 23h.

## **Dinner of the Caucus for Women in Statistics and the Committee on Women in Statistics • Souper du Caucus des femmes en statistique et du comité de la promotion de la femme en statistique**

Please join the Caucus for Women in Statistics and the Committee on Women in Statistics of the SSC for an informal get-together over dinner on Monday, May 24, at 18:00, at the *Chez Greco Restaurant*:

2300 chemin Sainte-Foy (Sainte-Foy Road), Phone: 418-652-0319, Web: [http://www.chezgreco.com/index\\_en.html](http://www.chezgreco.com/index_en.html).

The average dinner entrée costs between \$15.00 and \$26.00, but, on Mondays, most dinner entrées on the "À la carte" menu are half price. Chez Greco Restaurant accepts cash, debit and credit cards.

How to get there: It is within walking distance of Université Laval (about 5 minutes). From Université Laval, take the *avenue des Sciences Humaines* (des Sciences Humaines Av.) all the way to chemin Sainte-Foy (Sainte-Foy Road). It will be on your left on chemin Sainte-Foy. All people (both men and women) are invited and students are especially welcome.

Nous vous invitons à rejoindre le Caucus des femmes en statistique et le Comité de la promotion de la femme en statistique de la SSC, pour un souper informel, le lundi 24 mai, à 18h, au restaurant Chez Greco : 2300 chemin Sainte-Foy, tél. : 418-652-0319, Site web : <http://www.chezgreco.com/>.

En soirée, le plat principal coûte en moyenne entre 15 \$ et 26 \$, cependant, les lundi, la plupart des plats principaux du menu à la carte sont à moitié prix. Le restaurant Chez Greco accepte l'argent comptant, ainsi que les cartes de débit et crédit.

Pour s'y rendre : On peut le faire à pied de l'Université Laval (environ 5 minutes). De l'Université Laval, prenez l'Avenue des Sciences Humaines, jusqu'au Chemin Sainte-Foy. Ce sera à votre gauche, sur le Chemin Sainte-Foy. Tout le monde est invité (hommes et femmes) et les étudiants et les étudiantes sont tout particulièrement bienvenus.

## **The SSC Committee on New Investigators • Le comité des nouveaux chercheurs de la SSC**

This year the SSC New Investigators Committee has planned two events: one formal (the panel discussion) and one informal (a beer party). The panel discussion (session 10D) will take place on Tuesday, May 25 at 16:25 and will feature Paul Cabilio, Acadia University, Kelly Zou, Pfizer Inc. & Harvard University and Nancy Reid, University of Toronto who will provide advises on managing your first steps after graduation, i.e., career paths job search, interviews, grants, paper submission etc. Please, send your list of questions to [ygl@math.uwaterloo.ca](mailto:ygl@math.uwaterloo.ca).

The very informal beer party will take place on Sunday, May 23 at 20:00 at the *Café au temps perdu*, located at 867 avenue Myrand (Phone: 418-681-5601).

Cette année le comité des nouveaux chercheurs de la SSC a planifié deux événements : un formel (panel) et un informel (un pot de bière). Le panel (séance 10D) aura lieu le mardi 25 mai à 16h25 et mettra en vedette Paul Cabilio de l'Université Acadia,

Kelly Zou de Pfizer Inc. et de l'Université Harvard et Nancy Reid de l'Université de Toronto qui donneront des conseils sur la gestion des premières étapes qui suivent la graduation : recherche de carrières et d'emplois, entrevues, subventions, soumission d'articles, etc. S'il vous plaît envoyez vos questions à [ygl@math.uwaterloo.ca](mailto:ygl@math.uwaterloo.ca)

Le très informel pot de bière aura lieu le dimanche 23 mai à 20h au Café au temps perdu, situé au 867, avenue Myrand (tél : 418-681-5601).

## **Internet Access • Accès internet**

All conference participants will find a user name and a password to access the wireless network `ulaval-inv` for the duration of the conference in their conference bag. This network broadcasts a signal that is available in most rooms and public spaces of most buildings on campus. Moreover, access to computer labs in the Alexandre-Vachon and Adrien-Pouliot building will be given to the conference participants. More information is available in the conference bag.

Tous les participants inscrits au congrès trouveront dans le sac du congressiste un nom d'utilisateur, un mot de passe et des instructions leur permettant d'avoir accès au réseau internet sans fil `ulaval-inv` pour la durée du congrès. Ce réseau diffuse un signal disponible dans la très grande majorité des salles et des espaces publics de la plupart des pavillons du campus. De plus, des laboratoires informatiques des pavillons Alexandre-Vachon et Adrien-Pouliot seront mis à la disposition des participants. Plus d'information est disponible dans le sac des congressistes.

## **Food on campus and nearby • Nourriture sur le campus et à proximité**

Please see the detailed information on food services available on the campus and its vicinity in the conference bag.

Prière de consulter les feuillets d'information détaillée sur les services de vente de nourriture disponibles sur le campus et dans les environs inclus dans le sac du congressiste.

## **Discounts at various tourist attractions • Rabais pour certaines attractions touristiques**

A 10% discount on the cruises offered by *Croisières AML* (<http://www.croisieresaml.com/en/index.php>) is available to conference participants. Simply show the coupon in your conference bag as well as a proof of participation to the conference (e.g., name tag or receipt) when purchasing your cruise ticket. A \$2 discount is also available at the *Musée de la civilisation* upon presentation of your name tag.

Un rabais de 10 % sur les croisières offertes par Croisières AML (<http://www.croisieresaml.com/fr/index.php>) est disponible pour les participants au congrès. Simplement présenter le coupon qui se trouve dans votre sac du congressiste de même qu'une preuve de participation au congrès (par exemple votre cocarde ou votre reçu) à l'achat de votre billet de croisière. Un rabais de 2 \$ est également disponible au Musée de la civilisation sur présentation de votre cocarde.

# Committees and Meetings

## Comités et réunions

VAN rooms = Ferdinand-Vandry building (number 9 on campus map), KRU rooms = Gene-H.-Kruger building (number 32 on campus map), PPR rooms = Palasis-Prince building (number 6 on campus map), POU rooms = Adrien-Pouliot building (number 12 on campus map).

**Lunch boxes will be provided for all lunch time meetings in the KRU 2320/KRU 2330 room.**

Les salles VAN sont dans le pavillon Ferdinand-Vandry (numéro 9 sur la carte du campus). Les salles KRU sont dans le pavillon Gene-H.-Kruger (numéro 32 sur la carte du campus). Les salles PPR sont dans le pavillon Palasis-Prince (numéro 6 sur la carte du campus). Les salles POU sont dans le pavillon Adrien-Pouliot (numéro 12 sur la carte du campus).

**Des boîtes à lunch seront fournies pour toutes les réunions de comités se déroulant dans la salle KRU 2320/KRU 2330.**

### Calendar Listing • Horaire journalier

#### Saturday May 22 • Samedi 22 mai

18:00–22:00	Executive Committee • <i>Comité exécutif</i>	Hôtel Plaza Québec (Bach room/salle Bach)
-------------	--	--

#### Sunday, May 23 • Dimanche 23 mai

9:00–10:30	Finance Committee • <i>Comité des finances</i>	KRU 2320/KRU 2330
10:30–12:00	Publications Committee • <i>Comité des publications</i>	KRU 2320/KRU 2330
10:30–12:00	Department Heads Meeting • <i>Rencontre des directeurs de départements</i>	VAN 1853
12:00–16:00	Board of Directors • <i>Conseil d'administration</i>	KRU 2320/KRU 2330
16:15–17:45	Panel Discussion, National Institute in Statistics • <i>Groupe de discussion, Institut national en statistique</i>	VAN 1245A

#### Monday, May 24 • Lundi 24 mai

11:55–13:15	Probability Section (Section Officers) • <i>Groupe de probabilité (directeurs du groupe)</i>	KRU 2320/KRU 2330
	Biostatistics Section (Section Officers) • <i>Groupe de biostatistique (directeurs du groupe)</i>	KRU 2320/KRU 2330
	Business and Industrial Statistics Section (Section Officers) • <i>Groupe de statistique industrielle et de gestion (directeurs du groupe)</i>	KRU 2320/KRU 2330
	Survey Methods Section (Section Officers) • <i>Groupe des méthodes d'enquête (directeurs du groupe)</i>	KRU 2320/KRU 2330
	Committee on Women in Statistics • <i>Comité sur les femmes en statistique</i>	KRU 2320/KRU 2330
	Statistical Education Committee • <i>Comité d'éducation en statistique</i>	KRU 2320/KRU 2330
11:55–20:00	Award for Case Studies and Data Analysis Committee • <i>Comité du prix pour les études de cas et l'analyse de données</i>	VAN 1853

18:10–19:10	Biostatistics Section (General Meeting) • <i>Groupe de biostatistique (assemblée générale)</i>	VAN 2289A
	Business and Industrial Statistics Section (General Meeting) • <i>Groupe de statistique industrielle et de gestion (assemblée générale)</i>	VAN 2291A
	Probability Section (General Meeting) • <i>Groupe de probabilité (assemblée générale)</i>	VAN 2811A
	Survey Methods Section (General Meeting) • <i>Groupe des méthodes d'enquête (assemblée générale)</i>	VAN 1203A
	Association des statisticiennes et des statisticiens du Québec (General Meeting) • <i>Association des statisticiennes et des statisticiens du Québec (assemblée générale)</i>	VAN 2809A
	Information on 2010 NSERC Competition • <i>Information sur la compétition du CRSNG 2010</i>	VAN 1245A

### Tuesday, May 25 • Mardi 25 mai

10:25–18:00	Award for Case Studies in Data Analysis Committee • <i>Comité du prix pour les études de cas et l'analyse de données</i>	VAN 1853
11:55–13:15	Bilingualism Committee • <i>Comité du bilinguisme</i>	KRU 2320/KRU 2330
	Public Relations Committee • <i>Comité des relations publiques</i>	KRU 2320/KRU 2330
	NICDS • <i>INSDC</i>	KRU 2320/KRU 2330
	CJS Editorial Board • <i>Comité de rédaction de la RCS</i>	KRU 2320/KRU 2330
	Accreditation General Meeting • <i>Assemblée générale des membres accrédités</i>	VAN 1871
18:00–19:30	SSC Annual General Meeting • <i>Assemblée générale annuelle de la SSC</i>	VAN 1245A
	Student Presentation Award Committee • <i>Comité du prix pour les présentations étudiantes</i>	VAN 1853

### Wednesday, May 26 • Mercredi 26 mai

12:00–13:20	Accreditation Committee • <i>Comité sur l'accréditation</i>	KRU 2320/KRU 2330
	Liaison	KRU 2320/KRU 2330
	Program Committee • <i>Comité du programme</i>	KRU 2320/KRU 2330
	Probability Section (Section Officers) • <i>Groupe de probabilité (directeurs du groupe)</i>	KRU 2320/KRU 2330
	Biostatistics Section (Section Officers) • <i>Groupe de biostatistique (directeurs du groupe)</i>	KRU 2320/KRU 2330
	Business and Industrial Statistics Section (Section Officers) • <i>Groupe de statistique industrielle et de gestion (directeurs du groupe)</i>	KRU 2320/KRU 2330
	Awards Committee • <i>Comité des prix</i>	KRU 2320/KRU 2330
16:15–17:45	Board of Directors • <i>Conseil d'administration</i>	KRU 2320/KRU 2330
17:45	Executive Committee • <i>Comité exécutif</i>	KRU 2320/KRU 2330

## Alphabetical Listing • Liste alphabétique

*Lunch boxes* will be provided for all lunch time meetings in the KRU 2320/KRU 2330 room.

*Des boîtes à lunch* seront fournies pour toutes les réunions de comités se déroulant dans la salle KRU 2320/KRU 2330 le midi.

### Accreditation Committee • *Comité sur l'accréditation*

Wednesday, May 26 • mercredi 26 mai, 12:00–13:20, KRU 2320/KRU 2330

### Accreditation General Meeting • *Assemblée générale des membres accrédités*

Tuesday, May 25 • mardi 25 juin, 11:55–13:15, VAN 1871

### Association des statisticiennes et des statisticiens du Québec General Meeting • *Assemblée générale de l'Association des statisticiennes et des statisticiens du Québec*

Monday, May 24 • lundi 24 mai, 18:10–19:10, VAN 2809A

**Award for Case Studies and Data Analysis Committee • Comité du prix pour les études de cas et l'analyse de données**

Monday, May 24 • lundi 24 mai, 11:55–20:00, VAN 1853

Tuesday, May 25 • mardi 25 mai, 10:25–18:00, VAN 1853

**Awards Committee • Comité des prix**

Wednesday, May 26 • mercredi 26 mai, 12:00–13:20, KRU 2320/KRU 2330

**Bilingualism Committee • Comité du bilinguisme**

Tuesday, May 25 • mardi 25 mai, 11:55–13:15, KRU 2320/KRU 2330

**Biostatistics Section General Meeting • Groupe de biostatistique (assemblée générale)**

Monday, May 24 • lundi 24 mai, 18:10–19:10, VAN 2289A

**Biostatistics Section (Section Officers) • Groupe de biostatistique (directeurs du groupe)**

Monday, May 24 • lundi 24 mai, 11:55–13:15, KRU 2320/KRU 2330

Wednesday, May 26 • mercredi 26 mai, 12:00–13:20, KRU 2320/KRU 2330

**Board of Directors • Conseil d'administration**

Sunday, May 23 • dimanche 23 mai, 12:00–16:00, KRU 2320/KRU 2330

Wednesday, May 26 • mercredi 26 mai, 16:15–17:45, KRU 2320/KRU 2330

**Business and Industrial Statistics Section General Meeting • Groupe de statistique industrielle et de gestion (assemblée générale)**

Monday, May 24 • lundi 24 mai, 18:10–19:10, VAN 2291A

**Business and Industrial Statistics Section (Section Officers) • Groupe de statistique industrielle et de gestion (directeurs du groupe)**

Monday, May 24 • lundi 24 mai, 11:55–13:15, KRU 2320/KRU 2330

Wednesday, May 26 • mercredi 26 mai, 12:00–13:20, KRU 2320/KRU 2330

**CJS Editorial Board • Comité de rédaction de la RCS**

Tuesday, May 25 • mardi 25 mai, 11:55–13:15, KRU 2320/KRU 2330

**Committee on Women in Statistics • Comité sur les femmes en statistique**

Monday, May 24 • lundi 24 mai, 11:55–13:15, KRU 2320/KRU 2330

**Department Heads Meeting • Rencontre des directeurs de département**

Sunday, May 23 • dimanche 23 mai, 10:30–12:00, VAN 1853

**Executive Committee • Comité exécutif**

Saturday May 22 • samedi 22 mai, 18:00–22:00, Hôtel Plaza Québec (Bach room / salle Bach)

Wednesday May 26 • mercredi 26 mai, 17:45, KRU 2320/KRU 2330

**Finance Committee • Comité des finances**

Sunday, May 23 • dimanche 23 mai, 9:00–10:30, KRU 2320/KRU 2330

**Information on 2010 NSERC Competition • Information sur la compétition du CRSNG 2010**

Monday, May 24 • lundi 24 mai, 18:10–19:10, VAN 1245A

**Liaison**

Wednesday, May 26 • mercredi 26 mai, 12:00–13:20, KRU 2320/KRU 2330

**NICDS • INSDC**

Tuesday, May 25 • mardi 25 mai, 11:55–13:15, KRU 2320/KRU 2330

**Panel discussion, National Institute in Statistics • Groupe de discussion, Institut national en statistique**

Sunday, May 23 • dimanche 23 mai, 16:15–17:45, VAN 1245A

**Probability Section General Meeting • Groupe de probabilité (assemblée générale)**

Monday, May 24 • lundi 24 mai, 18:10–19:10, VAN 2811A

**Probability Section (Section Officers) • Groupe de probabilité (directeurs du groupe)**

Monday, May 24 • lundi 24 mai, 11:55–13:15, KRU 2320/KRU 2330

Wednesday, May 26 • mercredi 26 mai, 12:00–13:20, KRU 2320/KRU 2330

**Program Committee • Comité du programme**

Wednesday, May 26 • mercredi 26 mai, 12:00–13:20, KRU 2320/KRU 2330

**Public Relations Committee • Comité des relations publiques**

Tuesday, May 25 • mardi 25 mai, 11:55–13:15, KRU 2320/KRU 2330

**Publications Committee • *Comité des publications***

Sunday, May 23 • dimanche 23 mai, 10:30–12:00, KRU 2320/KRU 2330

**SSC General Meeting • *Assemblée générale de la SSC***

Tuesday, May 25 • mardi 25 mai, 18:00–19:30, VAN 1245A

**Statistical Education Committee • *Comité d'éducation en statistique***

Monday, May 24 • lundi 24 mai, 11:55–13:15, KRU 2320/KRU 2330

**Student Presentation Award Committee • *Comité du prix pour les présentations étudiantes***

Tuesday, May 25 • mardi 25 mai, 18:00–19:30, VAN 1853

**Survey Methods Section General Meeting • *Groupe des méthodes d'enquête (assemblée générale)***

Monday, May 24 • lundi 24 mai, 18:10–19:10, VAN 1203A

**Survey Methods Section (Section Officers) • *Groupe des méthodes d'enquête (directeurs du groupe)***

Monday, May 24 • lundi 24 mai, 11:55–13:15, KRU 2320/KRU 2330

# List of Sessions • Liste des sessions

VAN rooms = Ferdinand-Vandry building (number 9 on campus map), PPR rooms = Palasis-Prince building (number 6 on campus map), POU rooms = Adrien-Pouliot building (number 12 on campus map).

Les salles VAN sont dans le pavillon Ferdinand-Vandry (numéro 9 sur la carte du campus). Les salles PPR sont dans le pavillon Palasis-Prince (numéro 6 sur la carte du campus). Les salles POU sont dans le pavillon Adrien-Pouliot (numéro 12 sur la carte du campus).

## Sunday May 23 • Dimanche 23 mai

### 09:00–16:00

- VAN 1245A Workshop of the Biostatistics Section “Draw your Assumptions before your Conclusions: Graphs for Causal Inference”  
*Atelier du Groupe de biostatistique « Dessinez vos hypothèses avant vos conclusions : graphiques pour l’inférence causale »*
- VAN 2809A Workshop of the Business and Industrial Statistics & Probability Sections “Copulas: Introduction to the Theory, Implementation in R, with Applications in Finance and Insurance”  
*Atelier des Groupes de statistique industrielle et de gestion et de probabilité « Copules : Introduction à la théorie, la mise en œuvre en R, avec des applications en finance et en assurance »*
- VAN 2811A Workshop of the Survey Methods Section “Small Area Estimation”  
*Atelier du Groupe des méthodes d’enquête « Estimation pour petits domaines »*

### 16:15–17:45

- VAN 1245A Session 00A Panel Discussion on the Role, Benefits, and Importance of a National Statistics Institute  
*Panel sur le rôle, les bénéfices et l’importance d’un Institut national en statistique*

## Monday May 24 • Lundi 24 mai

### 08:15–08:30

- PPR Session 01A Inaugural Session  
*Séance inaugurale*

### 08:30–09:45

- PPR Session 01B SSC Presidential Invited Address  
*Allocution de l’invité du président de la SSC*

### 10:25–11:55

- VAN 1245A Session 02A Introduction to some Topics in Statistical Data Mining  
*Introduction à quelques sujets en forage statistique de données*
- VAN 2289A Session 02B Probability and Finance  
*Probabilité et finance*
- VAN 2291A Session 02C Statistics and Science  
*Statistique et science*
- VAN 2809A Session 02D Innovative Tools in Teaching Statistics  
*Outils innovateurs pour l’enseignement de la statistique*
- VAN 2811A Session 02E Living in Canada: Developing a Household Panel Survey in Canada  
*Vivre au Canada : Développer une enquête par panel sur les ménages au Canada*

VAN 1203	Session 02F	Randomized Trials, Meta-Analysis, and Data Analysis <i>Essais randomisés, méta-analyse et analyse de données</i>
VAN 1211	Session 02G	Case-Control, Cohort Studies, and Other Trials <i>Études cas-témoins, de cohortes et autres essais</i>
VAN1871	Session 02H	Clustering, Soccer, and the Scientific Method <i>Classification, soccer et la méthode scientifique</i>
<b>10:25–11:40</b>		
VAN 1883	Session 02I	Mathematical Statistics <i>Statistique mathématique</i>
<b>13:15–14:15</b>		
PPR	Session 03A	SSC Impact Award Address <i>Allocution du récipiendaire du prix pour impact de la SSC</i>
<b>14:30–16:00</b>		
VAN 1245A	Session 04A	Statisticians in the Fight against Cancer <i>Statisticiens dans la bataille contre le cancer</i>
VAN 2289A	Session 04B	Large (Industrial-scale) Computer Experiments for Business and Industry <i>Grandes expériences par ordinateur (à échelle industrielle) pour l'industrie et la gestion</i>
VAN 2291A	Session 04C	Optimization and Applications in MCMC <i>Optimisation et applications des algorithmes Monte Carlo par chaînes de Markov</i>
VAN 2809A	Session 04D	Panel Discussion on Opportunities & Challenges in Interdisciplinary Research and Training <i>Panel sur les opportunités et les défis de la formation et la recherche interdisciplinaire</i>
VAN 2811A	Session 04E	Methods and Applications in Cluster Analysis <i>Méthodes et applications de la classification automatique</i>
VAN 1203	Session 04F	Methodologies for Count Data and Ecological Data <i>Méthodologies pour données de dénombrement et données écologiques</i>
VAN1211	Session 04G	Methodology for Survival Analysis <i>Méthodologie pour l'analyse de survie</i>
VAN 1871	Session 04H	Computational Methods in Finance <i>Méthodologie computationnelle en finance</i>
<b>14:30–15:45</b>		
VAN 1883	Session 04I	Distribution Theory and Copulas <i>Théorie des distributions et copules</i>
<b>14:30–16:00</b>		
VAN Hall	Session 04J	Poster Session 1 <i>Séance d'affichage 1</i>
VAN Hall	Session 04K	Case Study 1 “The Effects of Photo-interpreted Variables in the Estimation of Stand-level Merchantable Volumes in the Province of Québec” <i>Étude de cas 1 « L'effet de variables photo-interprétées sur l'estimation du volume marchand à l'échelle du peuplement dans la province du Québec »</i>
VAN Hall	Session 04L	Case Study 2 “Metabolism of Bradykinin and Endogenous des-Arg <sup>9</sup> -Bradykinin in Human Plasma: Contribution to the Pathophysiology of Angioedema Associated with ACE Inhibitors” <i>Étude de cas 2 « Métabolisme de la bradykinine et de la des-Arg<sup>9</sup>-bradykinine endogène dans le plasma humain : contribution à la physiopathologie de l'angio-œdème associé aux inhibiteurs de l'enzyme de conversion de l'angiotensine »</i>
<b>16:30–17:30</b>		
VAN 1245A	Session 05A	Special Invited Session of the Survey Methods Section—Cell Phone Surveys <i>Allocution de l'invité du président du groupe des méthodes d'enquête — Enquêtes à partir des cellulaires</i>

**16:30–18:00**

- VAN 2289A Session 05B Methods for Longitudinal and Spatial Categorical Data  
*Méthodes pour des données catégorielles longitudinales et spatiales*
- VAN 2291A Session 05C Probability, Regression, and Time Series  
*Probabilité, régression et séries chronologiques*
- VAN 2809A Session 05D What is the Role of a Statistical Association in Insuring Rigour in the Publication of Statistical Results in Media?  
*Quel est le rôle d'une association de statisticiens pour assurer la rigueur dans la publication médiatique de résultats statistiques ?*
- VAN 2811A Session 05E Financial Modelling in Actuarial Sciences  
*Modèles financiers en actuariat*
- VAN 1203 Session 05F Biostatistics  
*Biostatistique*
- VAN 1211 Session 05G Regression Methodology in Biostatistics and Neuronal Data  
*Méthodologie de régression en biostatistique et données neuronales*
- VAN 1871 Session 05H Hypothesis Testing and Inference  
*Tests d'hypothèses et inférence*

**18:10–**

- VAN 1245A Information on 2010 NSERC Discovery Grant Competition  
*Information sur le concours 2010 des subventions à la découverte du CRSNG*

**Tuesday May 25 • Mardi 25 mai****08:30–09:45**

- PPR Session 06A Gold Medal Award Address  
*Allocution de la récipiendaire de la médaille d'or*

**10:25–11:55**

- VAN 1245A Session 07A The Biology, Technology, and Statistical Modelling of High-throughput Genomics Data  
*La biologie, la technologie et la modélisation statistique de données génomiques à haut débit*
- VAN 2289A Session 07B New Developments in Modelling Using Survey Data  
*Nouveaux développements en modélisation de données d'enquêtes*
- VAN 2291A Session 07C Quality Improvement  
*Amélioration de la qualité*
- VAN 2809A Session 07D Applications of Statistics in Hydrology, Climatology, and Meteorology  
*Statistique appliquée à l'hydrologie, la climatologie et la météorologie*
- VAN 2811A Session 07E Ruin Theory  
*Théorie de la ruine*
- VAN 1203 Session 07F Methodology for Hazard Rate Estimation and Censored Data  
*Méthodologie pour l'estimation du taux de risque et les données censurées*
- VAN 1211 Session 07G Shrinkage Estimation and Bayesian Methodology  
*Estimation par rétrécissement et méthodologie bayésienne*

**10:25–11:40**

- VAN 1871 Session 07H Density Estimation and Robust Regression  
*Estimation de la densité et régression robuste*

**10:25–11:55**

- VAN 1883 Session 07I Time Series, Spatio-temporal Data, and Statistical Education  
*Séries chronologiques, données spatio-temporelles et l'enseignement de la statistique*

**13:15–14:15**

- POU 1112 Session 08A CRM–SSC Prize in Statistics Address  
*Allocution de la récipiendaire du prix CRM-SSC*

**14:25–15:55**

- VAN 1245A Session 09A Statistical Methods for Life History Analysis  
*Méthodes statistiques pour l'analyse de l'historique de vie*
- VAN 2289A Session 09B Recent Advances in Modelling Long Range Dependent and Nonlinear Time Series, with Applications to Business, Econometrics, and Environmental Sciences  
*Développements récents dans la modélisation de la dépendance à long terme et de séries chronologiques non linéaires, avec applications en gestion, économétrie et sciences environnementales*
- VAN 2291A Session 09C Quality in Surveys  
*Qualité dans les enquêtes*
- VAN 2809A Session 09D Stochastic Dynamics in Biology  
*Dynamique stochastique en biologie*
- VAN 2811A Session 09E Privacy Legislation—Implications for Canadian Statisticians  
*Législations sur la vie privée — Conséquences pour les statisticiens canadiens*
- VAN 1203 Session 09F Binary or Discrete Data  
*Données binaires ou discrètes*
- VAN 1211 Session 09G Risk Theory and Probability Theory in Finance  
*Théorie du risque et probabilité en finance*
- VAN 1871 Session 09H Nonresponse, Imputation, and Bootstrapping  
*Non-réponse, imputation et bootstrap*

**14:25–15:40**

- VAN 1883 Session 09I Quality Control and Design of Experiments  
*Contrôle de qualité et planification d'expériences*

**14:25–15:55**

- VAN Hall Session 09J Poster Session 2  
*Séance d'affichage 2*

**16:25–17:25**

- VAN 1245A Session 10A Isobel Loutit Invited Address of the Business and Industrial Statistics Section Special: Quality Technology in the High-tech Age  
*Allocution Isobel Loutit du Groupe de statistique industrielle et de gestion : La technologie en qualité à l'ère de la haute technologie*

**16:25–17:55**

- VAN 2289A Session 10B Analysis of Next Generation Sequencing Data for Functional Genomics  
*Analyse de données de séquençage de la prochaine génération en génomique fonctionnelle*
- VAN 2291A Session 10C Stochastic Models in Population Genetics  
*Modèles stochastiques en génétique des populations*
- VAN 2809A Session 10D Panel Discussion on Problems and Issues Faced by New Researchers  
*Panel sur les problèmes et questions auxquels font face les nouveaux chercheurs*
- VAN 2811A Session 10E Analysis of Multilevel Models for Complex Survey Data  
*Analyse de modèles multi-niveaux pour des données d'enquêtes complexes*

**16:25–17:40**

- VAN 1203 Session 10F Missing or Incomplete Observations  
*Observations manquantes ou incomplètes*
- VAN 1211 Session 10G Longitudinal Data and Mixed Effects  
*Données longitudinales et effets mixtes*

**16:25–17:55**

- VAN 1871 Session 10H Methodological Issues in Designing Surveys  
*Questions méthodologiques dans la planification d'enquêtes*

**Wednesday May 26 • Mercredi 26 mai****08:30–09:15**

POU 1112 Session 11A Pierre Robillard Award Address  
*Allocution de la récipiendaire du prix Pierre-Robillard*

**09:15–10:00**

POU 1112 Session 11B *Canadian Journal of Statistics* Award Address  
*Allocution des récipiendaires du prix de la Revue canadienne de statistique*

**10:30–12:00**

VAN 1245A Session 12A Data and Models on the Move: Functional Data Analysis  
*Données et modèles en déplacement : Analyse de données fonctionnelles*

VAN 2289A Session 12B Spatial Statistics and the Environment  
*Statistique spatiale et l'environnement*

VAN 2291A Session 12C Statistics in Conservation  
*Statistique en préservation*

VAN 2809A Session 12D Dependence Modelling in Risk Theory  
*Modèles de dépendance en théorie du risque*

VAN 2811A Session 12E Genomics and Association Studies  
*Génomique et études d'association*

**10:30–11:45**

VAN 1203 Session 12F Probability  
*Probabilité*

VAN 1211 Session 12G Estimation in Survey Sampling  
*Estimation en échantillonnage*

**10:30–12:00**

VAN 1871 Session 12H Times Series and Climate  
*Séries chronologiques et le climat*

**13:20–14:20**

POU 1112 Session 13A The Lasso  
*Le lasso*

**14:30–16:00**

VAN 1245A Session 14A Response-biased Sampling and Missing Covariates  
*Échantillonnage biaisé par les réponses et covariables manquantes*

VAN 2289A Session 14B Computer Experiments  
*Expérimentation numérique*

VAN 2291A Session 14C Network Sampling  
*Échantillonnage par réseau*

VAN 2809A Session 14D What can we learn from each other about teaching statistics?  
*Que peut-on apprendre les uns des autres concernant l'enseignement de la statistique ?*

VAN 2811A Session 14E Statistical Modelling in Actuarial Sciences  
*Modèles statistiques en actuariat*

**14:30–15:45**

VAN 1203 Session 14F Mixed Effects  
*Effets mixtes*

**14:30–16:00**

VAN 1211 Session 14G Data Analysis in Biostatistics  
*Analyse de données en biostatistique*

**14:30–15:45**

VAN 1871 Session 14H Regression and Data Mining  
*Régression et forage de données*

# Program • Programme

## Dimanche 23 mai • Sunday, May 23

---

**09:00 – 16:00**

**VAN 1245A**

Workshop of the Biostatistics Section “Draw your Assumptions before your Conclusions: Graphs for Causal Inference”

Atelier du Groupe de biostatistique « Dessinez vos hypothèses avant vos conclusions : graphiques pour l’inférence causale »

Leader • Animateur: Miguel A. HERNÁN, *Harvard School of Public Health*

**09:00 – 16:00**

**VAN 2809A**

Workshop of the Business and Industrial Statistics & Probability Sections “Copulas: Introduction to the Theory, Implementation in R, with Applications in Finance and Insurance”

Atelier des Groupes de statistique industrielle et de gestion et de probabilité « Copules : Introduction à la théorie, la mise en œuvre en R, avec des applications en finance et en assurance »

Leaders • Animateurs: Christian GENEST, *Université Laval* & Johanna NEŠLEHOVÁ, *McGill University*

**09:00 – 16:00**

**VAN 2811A**

Workshop of the Survey Methods Section “Small Area Estimation”

Atelier du Groupe des méthodes d’enquête : « Estimation pour petits domaines »

Leader • Animateur: Pascal ARDILLY, *Institut national de la statistique et des études économiques, France*

**16:15 – 17:45 Session 00A**

**VAN 1245A**

Panel Discussion on the Role, Benefits, and Importance of a National Statistics Institute

Panel sur le rôle, les bénéfices et l’importance d’un Institut national en statistique

Invited Panel Discussion • Panel sur invitation

Chair • Président: Richard LOCKHART, *Simon Fraser University*

Organizers • Responsables: Charmaine DEAN & Richard LOCKHART, *Simon Fraser University*

– François BELLAVANCE, *HEC-Montréal*  

– Hugh CHIPMAN, *Acadia University*  

- Gail IVANOFF, *Université d'Ottawa*  
- Mary LESPERANCE, *University of Victoria*  
- Louis-Paul RIVEST, *Université Laval*  

## Lundi 24 mai • Monday, May 24

---

### 08:15 – 08:30 Session 01A

PPR

Inaugural Session

Séance inaugurale

Special Session • Conférence spéciale

Special Guest Speakers • Conférenciers invités de marque

Chair • Président: Thierry DUCHESNE, *Université Laval*Organizers • Responsables: Thierry DUCHESNE, *Université Laval* & BOVAS ABRAHAM, *University of Waterloo*08:15 Edwin BOURGET, *Vice-recteur à la recherche et à la création, Université Laval*08:20 BOVAS ABRAHAM, *University of Waterloo*, Thierry DUCHESNE, *Université Laval*

### 08:30 – 09:45 Session 01B

PPR

SSC Presidential Invited Address

Allocution de l'invité du président de la SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: BOVAS ABRAHAM, *University of Waterloo*– Robert ENGLE, *NYU Stern School of Business*  

Long Term Skewness and Systemic Risk • Asymétrie à long terme et risque systémique [MS-1]

### 09:45 – 10:25 Pause-café • Coffee Break

VAN Hall

### 10:25 – 11:55 Session 02A

VAN 1245A

Introduction to some Topics in Statistical Data Mining

Introduction à quelques sujets en forage statistique de données

Introductory Overview Lecture • Conférence de vue d'ensemble

Organizer and Chair • Responsable et président: Christian LÉGER, *Université de Montréal*– Robert TIBSHIRANI, *Stanford University*  

Topics in Modern Statistical Learning and Data Mining • Sujets en apprentissage statistique et forage de données modernes [MS-2]

### 10:25 – 11:55 Session 02B

VAN 2289A

Probability and Finance

Probabilité et finance

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Probability Section*Session parrainée par le *Groupe de probabilité*Organizer and Chair • Responsable et président: Rafal KULIK, *University of Ottawa*

- 10:25 Miklós CSÖRGŐ, *Carleton University*    
Random Walking on a Comb • Marche aléatoire alvéolée [MS-3]
- 10:55 BRUNO RÉMILLARD, *HEC-Montréal*, Sylvain RUBENTHALER, *Université de Nice-Sophia Antipolis*   
  
Optimal Hedging in Discrete and Continuous Time • Réplication optimale en temps discret et en temps continu [MS-4]
- 11:25 Anatoliy SWISHCHUK, *University of Calgary*    
Stochastic Modelling and Pricing of Energy Markets' Contracts with Local Stochastic Delayed and Jumped Volatilities • Modélisation stochastique et fixation de prix de contrats dans les marchés de l'énergie avec volatilités stochastiques locales de sauts et de délais [MS-5]

**10:25 – 11:55 Session 02C****VAN 2291A**

## Statistics and Science

## Statistique et science

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Agnes M. HERZBERG, *Queen's University*

- 10:25 Michel CHRÉTIEN, *Institut de recherche de l'Hôpital d'Ottawa*    
Viral Pandemics: Academics Filling the Gap • Les pandémies virales : le monde académique à la rescousse [MS-6]
- 10:55 Haile DEBAS, *University of California, San Francisco*    
Global Health Challenges and Opportunities in the 21st Century • Défis globaux et possibilités à exploiter en santé au 21e siècle [MS-7]
- 11:25 Peter KAVANAGH, *Canadian Broadcasting Corporation*    
Why Journalists Can't Count, Statisticians Can't Read and whether it Matters • Pourquoi les journalistes ne peuvent pas compter, les statisticiens ne peuvent pas lire, et est-ce que ça importe? [MS-8]

**10:25 – 11:55 Session 02D****VAN 2809A**

## Innovative Tools in Teaching Statistics

## Outils innovateurs pour l'enseignement de la statistique

Invited Paper Session • Présentations sur invitation

Chair • Présidente: Zeny MATEO, *University of Manitoba*Organizers • Responsables: Zeny MATEO & Saumen MANDAL, *University of Manitoba*

- 10:25 Paul COREY, *University of Toronto*    
Games and Stories in the Teaching of Statistics • Jeux et histoires en enseignement de la statistique [MS-9]
- 10:55 David LOEWEN, *University of Manitoba*    
Electronic Homework: Does Student Learning Improve if Questions Have More Detailed Steps? • Devoirs électroniques : l'apprentissage des étudiants est-il amélioré si les questions ont des étapes plus détaillées? [MS-10]
- 11:25 Larry WELDON, *Simon Fraser University*  

Strategies for Teaching an Enduring Knowledge of Statistics • Stratégies permettant d'inculquer une connaissance durable de la statistique [MS-11]

### 10:25 – 11:55 Session 02E

VAN 2811A

Living in Canada: Developing a Household Panel Survey in Canada

Vivre au Canada : Développer une enquête par panel sur les ménages au Canada

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Inter-university Centre for Social Statistics*

Séance parrainée par le *Centre interuniversitaire québécois de statistiques sociales*

Chair • Président: Jean POIRIER, *CIQSS*

Organizers • Responsables: Jean POIRIER, *CIQSS* & Céline LE BOURDAIS, *McGill University & CIQSS*

10:25 Andrew HEISZ, *Statistics Canada*  

A Summary of Content Evaluations for the Living in Canada Survey Pilot • Un résumé des évaluations de contenu de l'enquête Vivre au Canada [MS-12]

10:55 Christian NADEAU, Marianne GOSSEN & Caroline CAUCHON, *Statistique Canada*  

An Evaluation of the Canadian Household Panel Survey Methodology • Une évaluation de la méthodologie de l'Enquête par panel auprès des ménages canadiens [MS-13]

11:25 Raynald LÉTOURNEAU, Magalie BROCHU, Mélanie RAYMOND & Julie TOUSIGNANT, *Ressources humaines et Développement des compétences Canada*  

Longitudinal Data to Support Adult and Family Policy Development • Les données longitudinales et le développement de politiques sur les adultes et les familles [MS-14]

### 10:25 – 11:55 Session 02F

VAN 1203

Randomized Trials, Meta-Analysis, and Data Analysis

Essais randomisés, méta-analyse et analyse de données

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*

Séance parrainée par le *Groupe de biostatistique*

Chair • Président: James HANLEY, *McGill University*

10:25 Michael ROTONDI, *University of Western Ontario*  

Meta-Regression and Cluster Randomized Trials: Theory and Practice • Métarégression et essais randomisés par grappes : théorie et pratique [MS-15]

10:40 Julia TALEBAN & Guang Yong ZOU, *University of Western Ontario*  

Confidence Intervals for Common Effect Measures in Cluster Randomization Trials • Intervalles de confiance pour mesures d'effets communs en essais randomisés par grappes [MS-16]

10:55 Depeng JIANG, *St. Michael's Hospital, University of Toronto*  

Statistical Power and Effect Size Estimation in Intervention Research under Population Heterogeneity • Puissance statistique et estimation de la taille de l'effet dans la recherche en intervention sous hétérogénéité de la population [MS-17]

11:10 Russell STEELE, *McGill University & Jewish General Hospital*, Franco CARNEVALE, *Montreal Children's Hospital & McGill University*, Lisa CHAN, *McGill University & Jewish General Hospital*, Mary Ellen McDONALD, *Montreal Children's Hospital & McGill University*, Javad SHAHIDI, *McGill*

University, Ian SHRIER, *Jewish General Hospital & McGill University*  

Bayesian Models for Meta-Analysis of Randomized Control Trials: Do We Know What People are Thinking? • Modèles bayésiens pour méta-analyse d'essais de contrôle randomisés : savons-nous ce que pensent les gens ? [MS-18]

11:25 Li Hua YUE, Duncan MURDOCH & Wenqing HE, *University of Western Ontario*  

Cost-efficient Variable Selection: A Case Study • Sélection efficace de variables : une étude de cas [MS-19]

11:40 Nathalie SAVARD, *Université Laval et INSPQ*, Louis-Paul RIVEST & Patrick LEVALLOIS, *Université Laval*  

Association Between Environmental Characteristics and the Occurrence of Small Gestational Aged Birthweight in Quebec • Association entre le milieu et le faible poids de naissance pour l'âge gestationnel au Québec [MS-20]

## 10:25 – 11:55 Session 02G

VAN 1211

Case-Control, Cohort Studies, and Other Trials

Études cas-témoins, de cohortes et autres essais

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*  
Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Shelley BULL, *University of Toronto*

10:25 Marco CARONE, *Johns Hopkins Bloomberg School of Public Health*  

Nonparametric Lifetime Risk Estimation using Prevalent Cohort Survival Data • Estimation non-paramétrique du risque à vie à partir de données de survie provenant d'une cohorte prévalente [MS-21]

10:40 Michael McISAAC & Richard COOK, *University of Waterloo*  

Efficient Two-phase Selection of Patients for Biomarker Studies in Chronic Disease • Sélection efficace de patients à deux phases pour études à biomarqueurs de maladies chroniques [MS-22]

10:55 Russell SHINOHARA, Constantine FRANGAKIS, Elizabeth PLATZ & Konstantinos TSILIDIS, *Johns Hopkins University*  

Estimating Effects by Combining Instrumental Variables with Case-control Designs • Estimation d'effets par le jumelage de variables instrumentales et de plans d'expérience cas-témoin [MS-23]

11:10 Yanqing YI, *Memorial University of Newfoundland*, Xikui WANG, *University of Manitoba*  

Comparison of Score, Likelihood ratio and Wald Tests for Response Adaptive Designs • Comparaison des tests du score, du rapport de vraisemblance et de Wald pour plans adaptatifs à la réponse [MS-24]

11:25 Ying-Pin CHEN, *Lewis and Clark College*  

Which design is more powerful? Ehrenfest Urn versus Biased Coin • Quel plan est le plus puissant ? Urne d'Ehrenfest ou pièce biaisée [MS-25]

11:40 Juxin LIU, *University of Saskatchewan*, Paul GUSTAFSON, *University of British Columbia*, Nicola CHERRY & Igor BURSTYN, *University of Alberta*  

Bayesian Analysis of a Matched Case-control Study with a Misclassified Exposure Variable • Analyse bayésienne d'une étude cas-témoins appariée avec une variable d'exposition mal classée [MS-26]

**10:25 – 11:55 Session 02H****VAN 1871**Clustering, Soccer, and the Scientific Method  
Classification, soccer et la méthode scientifique

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Nadia GHAZZALI, *Université Laval*10:25 Jeffrey ANDREWS & Paul McNICHOLAS, *University of Guelph*  

Model-based Clustering via tCLUST • Classification à base de modèles au moyen de tCLUST [MS-27]

10:40 Vahid PARTOVI NIA, *McGill University*, Anthony DAVISON, *Ecole Polytechnique Fédérale de Lausanne*  

Bayesian High-dimensional Clustering with Variable Selection • Classification bayésienne en hautes dimensions avec sélection de variables [MS-28]

10:55 Laura SANGALLI, Piercesare SECCHI, Simone VANTINI & Valeria VITELLI, *Politecnico di Milano, Italy*  

Classification of Functional Data: Curve Clustering when Curves are Misaligned • Classification des données fonctionnelles : classification de courbes lorsque les courbes sont désalignées [MS-29]

11:10 Yu XIA, *University of Guelph*  

Constrained Clustering by a Global Optimization Method • Classification sous contraintes au moyen d'une méthode d'optimisation globale [MS-30]

11:25 David BRILLINGER, *University of California, Berkeley*  

A Review of Statistical Modelling and Data Analyses for Soccer • Une revue de la modélisation statistique et de l'analyse de données pour le soccer [MS-31]

11:40 G. ARTHUR MIHRAM, *Princeton, NJ*, Danielle MIHRAM, *University of Southern California*  

A Statistician's Corrective for Editor [Wired Magazine] Anderson's "Data Deluge Makes the Scientific Method Obsolete" • La rectification d'un statisticien au sujet de l'éditorial "Le Déluge de données rend la méthode scientifique obsolète", d'Anderson [dans Wired Magazine]. [MS-32]

**10:25 – 11:40 Session 02I****VAN 1883**Mathematical Statistics  
Statistique mathématique

Contributed Paper Session • Séance de communications libres

Chair • Président: Éric MARCHAND, *Université de Sherbrooke*10:25 Mikelis BICKIS, *University of Saskatchewan*, Xiao Bo LI, *Saskatchewan Assessment Management*  

Calibrating P-values using Splines. • Calibrage des valeurs p au moyen de splines [MS-33]

10:40 Fuqi CHEN & Sévérien NKURUNZIZA, *University of Windsor*  

Equivariance and Generalized Inference in Two-sample Location-scale Families • Équivariance et inférence généralisée pour les lois bivariées de position-échelle [MS-34]

10:55 Katherine DAVIES, *University of Manitoba*, N. BALAKRISHNAN, *McMaster University*, Jerome P. KEATING, *University of Texas at San Antonio*, Robert L. MASON, *Southwest Research Institute*  

Simultaneous Closeness among Order Statistics to Population Quantiles • Proximité de l'ensemble des statistiques d'ordre aux quantiles de population [MS-35]

11:10 Debaraj SEN, Yogendra P. CHAUBEY & Murari SINGH, *Concordia University*  

Distribution of Coefficient of Variation: a Transformation in a New Perspective • Distribution du coefficient de variation : une transformation sous une nouvelle perspective [MS-36]

11:25 Ximing XU, *University of Toronto*  

On the Robustness of Composite Likelihood Inference • De la robustesse de l'inférence par vraisemblance composite [MS-37]

## 11:55 – 13:15 Dîner • Lunch

### 13:15 – 14:15 Session 03A

PPR

SSC Impact Award Address

Allocution du récipiendaire du prix pour impact de la SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Jerry LAWLESS, *University of Waterloo*

– Rick ROUTLEDGE, *Simon Fraser University*  

Preserving Wild Pacific Salmon • La préservation du saumon sauvage du Pacifique [MS-38]

### 14:30 – 16:00 Session 04A

VAN 1245A

Statisticians in the Fight against Cancer

Statisticiens dans la bataille contre le cancer

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Judy-Anne CHAPMAN, *NCIC Clinical Trials Group/Queen's University*

Organizer • Responsable: Lehana THABANE, *McMaster University*

14:30 Karen KOPCIUK, *Alberta Health Services – Cancer Care*, Yun-Hee CHOI, *University of Western Ontario*, Pat PARFREY & Jane GREEN, *Memorial University*, Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute*  

Statistical Methods and Study Designs for Evaluating Cancer Risk in Families Carrying a Known Genetic Mutation • Méthodes statistiques et plans d'enquête pour évaluer le risque associé au cancer dans les familles porteuses d'une certaine mutation génétique [MS-39]

15:00 Patrick BROWN & Lutong ZHOU, *Cancer Care Ontario*  

Small-area Spatial Variation in Smoking Incidence using Survey Data • Variation spatiale en petits domaines des taux de tabagisme pour des données d'enquête [MS-40]

15:30 Gregory POND, *Ontario Clinical Oncology Group, Department of Oncology, McMaster University*  

Modern Cancer Clinical Trials: Recent Statistical Advancements and Controversies • Essais cliniques oncologiques modernes : percées et controverses statistiques récentes [MS-41]

### 14:30 – 16:00 Session 04B

VAN 2289A

Large (Industrial-scale) Computer Experiments for Business and Industry

Grandes expériences par ordinateur (à échelle industrielle) pour l'industrie et la gestion

## Invited Paper Session • Présentations sur invitation

Session sponsored by the *Business and Industrial Statistics Section*Session parrainée par le *Groupe de statistique industrielle et de gestion*Organizer and Chair • Responsable et président: Will WELCH, *University of British Columbia*14:30 Andrew BOOKER, *Boeing*  

Challenges in Industrial Applications of Computer Experiments • Défis en applications industrielles d'expériences informatiques [MS-42]

15:00 Anthony O'HAGAN & Jeremy OAKLEY, *University of Sheffield*  

Emulating a Computationally Intensive Simulation Model of Disease to Aid Decision-making on Drug Development • Émulation d'un modèle lourd de simulation d'une maladie pour aider à la prise de décision en développement de médicaments [MS-43]

15:30 Derek BINGHAM, *Simon Fraser University*, James HOLLOWAY & Bruce FRYXELL, *University of Michigan*  

Predictive Modelling of a Radiative Shock Physics • Modélisation prédictive de la physique de chocs radiatifs [MS-44]

**14:30 – 16:00 Session 04C****VAN 2291A**

## Optimization and Applications in MCMC

## Optimisation et applications des algorithmes Monte-Carlo par chaînes de Markov

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Mylène BÉDARD, *Université de Montréal*14:30 Radu CRAIU & Yan BAI, *University of Toronto*, Antonio DI NARZO, *University of Bologna*  

Divide and Conquer: A Mixture-Based Approach to Regional Adaptation for MCMC • Diviser pour régner : un algorithme régional et adaptatif pour l'échantillonnage Monte-Carlo avec des chaînes de Markov [MS-45]

15:00 Jeffrey ROSENTHAL, *University of Toronto*  

Optimising and Adapting the Metropolis Algorithm • Comment optimiser et adapter l'algorithme Metropolis [MS-46]

15:30 Simon GUILLOTTE, *University of Prince-Edward Island*, François PERRON, *Université de Montréal*, Johan SEGERS, *Université Catholique de Louvain*  

A Bayesian Sieves Model for Inference on Bivariate Extremes • Un modèle de tamis bayésien pour l'inférence sur les lois de valeurs extrêmes bivariées [MS-47]

**14:30 – 16:00 Session 04D****VAN 2809A**

## Panel Discussion on Opportunities and Challenges in Interdisciplinary Research and Training

## Panel sur les opportunités et les défis de la formation et la recherche interdisciplinaire

Invited Panel Discussion • Panel sur invitation

Organizer and Chair • Responsable et présidente: Charmaine DEAN, *Simon Fraser University*– Laurie AINSWORTH, *Simon Fraser University*  – Shelley BULL, *University of Toronto*  – Richard COOK, *University of Waterloo*  

- John PETKAU, *University of British Columbia*  
- Francis ZWIERS, *Environment Canada*  

**14:30 – 16:00 Session 04E****VAN 2811A**

Methods and Applications in Cluster Analysis

Méthodes et applications de la classification automatique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Guohua YAN, *University of New Brunswick, Fredericton*14:30 Ruben ZAMAR, *University of British Columbia*  

Clustering using Linear Patterns • Classification au moyen de patrons linéaires [MS-48]

15:00 Rong ZHU, *McMaster University*, Abdel EL-SHAARAWI, *National Water Research Institute, Environment Canada*  

Model Clustering and Applications • Classification à partir de modèles et applications [MS-49]

15:30 Peng ZHANG, *University of Alberta*, Xiaogang WANG, *York University*, Peter SONG, *University of Michigan*  

Clustering Categorical Data Based On Distance Vectors • Classification de données catégorielles à partir de vecteurs de distance [MS-50]

**14:30 – 15:30 Session 04F****VAN 1203**

Methodologies for Count Data and Ecological Data

Méthodologies pour données de dénombrement et données écologiques

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*Session parrainée par le *Groupe de biostatistique*Chair • Présidente: Sylvia ESTERBY, *University of British Columbia, Okanagan*14:30 Yunqi Ji & Zhaozhi FAN, *Memorial University*  

Analysis of Longitudinal Count Data with Measurement Error and the Application in Epidemiology • Analyse de données de dénombrement longitudinales avec erreurs de mesure et application en épidémiologie [MS-51]

14:45 Xiaoqin XIONG & Joel DUBIN, *University of Waterloo*  

A Regression Tree Based Method for Longitudinal Count Data • Une méthode à base d'arbres de régression pour des données longitudinales de dénombrement [MS-52]

15:00 Tasneem ZAIHRA, *University of New Brunswick—Saint John*, Sudhir PAUL, *University of Windsor*  

Variance Function in Semiparametric Analysis of Count Data • La fonction de variance dans l'analyse semi-paramétrique de données de dénombrement [MS-53]

15:15 Henrik STRYHN & Charles CARAGUEL, *University of Prince Edward Island*  

Evaluation of Diagnostic Tests by a Latent Class Model with Three Latent Classes • Evaluation de tests diagnostiques avec le modèle de classes latentes incluant trois classes [MS-54]

**14:30 – 16:00 Session 04G****VAN 1211**

## Methodology for Survival Analysis

## Méthodologie pour l'analyse de survie

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*Session parrainée par le *Groupe de biostatistique*Chair • Président: JASON NIELSEN, *Carleton University*14:30 Theresa CONROY & Mei Ling HUANG, *Brock University*  

On Estimation of Extreme Values in Survival Analysis • Estimation de valeurs extrêmes en analyse de survie [MS-55]

14:45 Mohammad DEGHAN & Thierry DUCHESNE, *Université Laval*  

A Comparison of the Performance of Some Non-parametric Estimators of the Conditional Survival Function with Interval-censored Data • Une comparaison de certains estimateurs non paramétriques de la fonction de survie conditionnelle avec données censurées par intervalle [MS-56]

15:00 Jianrong WU, *St Jude Children's Research Hospital*  

Nonparametric Confidence Intervals for the Difference of Median Failure Times • Intervalles de confiance non paramétriques pour la différence de médianes de temps de défaillance [MS-57]

15:15 Cecilia COTTON, *University of Waterloo*, Patrick HEAGERTY, *University of Washington*  *G*-estimation for the Comparison for Dynamic Treatment Regimens • *G*-estimation afin de comparer des traitements dynamiques [MS-58]15:30 Yongling XIAO, Michal ABRAHAMOWICZ & Erica MOODIE, *McGill University*  

Accuracy of Conventional and Marginal Structural Cox Model Estimators: A Simulation Study • Précision des estimateurs pour un modèle de Cox marginal structurel et conventionnel : une étude de simulation [MS-59]

15:45 Yingwei PENG & Jianfeng XU, *Queen's University*  

Proportional Density Mixture Cure Model • Modèle de mélange proportionnel de densités avec fraction non à risque [MS-60]

**14:30 – 16:00 Session 04H****VAN 1871**

## Computational Methods in Finance

## Méthodologie computationnelle en finance

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Business and Industrial Statistics Section*Session parrainée par le *Groupe de statistique industrielle et de gestion*Chair • Présidente: Gail IVANOFF, *Université d'Ottawa*14:30 Kin Hung (Felix) KAN & Mark REESOR, *University of Western Ontario*  

Correcting the Bias in Least-Squares Monte Carlo Estimators of American Option Values • Corriger le biais dans l'estimation aux moindres carrés de la valeur d'options américaines obtenus par Monte-Carlo [MS-61]

14:45 Amaan MEHRABIAN & David STEPHENS, *McGill University*  

Stochastic Volatility Models with Leverage Effect, Heavy-tailed Errors, and Jump Components. • Modèles de volatilité stochastique avec effet de levier, erreurs à ailes relevées et sauts. [MS-62]

- 15:00 Zhongxian MEN, *University of Waterloo*    
 Multivariate Stochastic Volatility Models: A Slice Sampler Within Gibbs Approach • Modèles de volatilité stochastique multivariés : une approche basée sur un échantillonneur par tranches dans Gibbs [MS-63]
- 15:15 François WATIER & René FERLAND, *Université du Québec à Montréal*    
 Mean-variance Portfolio Optimization in a Market with Stochastic Correlations • Gestion moyenne-variance dans un marché avec corrélations stochastiques [MS-64]
- 15:30 Hui ZHAO & Paul MARRIOTT, *University of Waterloo*    
 Variational Bayesian Learning for Regime-switching Model • Apprentissage bayésien variationnel pour modèles à changement de régime [MS-65]
- 15:45 Jean-Philippe BOUCHER, *Université du Québec à Montréal*    
 Using Count Distributions to Assess the Impact of Graduate Studies on the Birth Rate for Students • Utilisation des distributions de comptage pour évaluer l'impact des études supérieures sur le taux de natalité des étudiants [MS-66]

**14:30 – 15:45 Session 04I****VAN 1883**

## Distribution Theory and Copulas

## Théorie des distributions et copules

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Johanna NEŠLEHOVÁ, *McGill University*

- 14:30 Maciej AUGUSTYNIAK & LOUIS DORAY, *Université de Montréal*    
 Inference for a Leptokurtic Symmetric Family of Distributions Represented by the Difference of Two Gamma Variates • Inférence pour une famille de distributions symétriques et leptokurtiques représentée par la différence de deux variables aléatoires gamma [MS-67]
- 14:45 Holly FALSETTI & Mei Ling HUANG, *Brock University*    
 On Tail Index Estimation of a Heavy Tailed Distribution • Estimation de l'indice caudal d'une distribution à ailes relevées [MS-68]
- 15:00 Elif ACAR, Radu CRAIU & Fang YAO, *University of Toronto*    
 Nonparametric Covariate Adjustment in Conditional Copulas • Ajustement non paramétrique des covariables pour des copules conditionnelles [MS-69]
- 15:15 Ralf WILKE, *University of Nottingham*, Simon Lo, *Lingnan University, Hong Kong*    
 Identifiability and Estimation of the Sign of a Covariate Effect in the Competing Risks Model • Identifiabilité et estimation du signe d'un effet covariable dans un modèle à risques compétitifs [MS-70]
- 15:30 Yildiz YILMAZ, *Samuel Lunenfeld Research Institute*, Jerald LAWLESS, *University of Waterloo*    
 Comparison of Semiparametric Maximum Likelihood Estimation and Two-Stage Semiparametric Estimation in Copula Models • Comparaison de l'estimation du maximum de vraisemblance semi-paramétrique et de l'estimation semi-paramétrique en deux étapes pour des modèles de copules [MS-71]

**14:30 – 16:00 Session 04J****VAN Hall**

## Poster Session 1

## Séance d'affichage 1

Contributed Poster Session • Séance d'affichage libre

– Alice DRAGOMIR & Jean-François ANGERS, *Université de Montréal*, Jean-Eric TARRIDE, *McMaster University*, Sylvie PERREAULT, *Université de Montréal*  

Validation of a Micro-simulation Monte-Carlo Markov Model for Schizophrenia • Validation d'un modèle de Markov pour la schizophrénie avec microsimulations de Monte-Carlo [MS-72]

– Man-Kee Maggie CHU, *University of Western Ontario*  

Application of the Expectation-Maximization (EM) Algorithm for Mixture Models • Application de l'algorithme espérance-maximisation (EM) aux modèles de mélanges [MS-73]

– Abdulkadir HUSSEIN, *University of Windsor*, Hassen MUTTLAK & Essam AL-SAWI, *King Fahd University of Petroleum and Minerals*  

Increasing the Efficiency of Group Sequential Methods by Auxiliary Information • Augmenter l'efficacité de méthodes séquentielles en groupes au moyen d'information auxiliaire [MS-74]

– Ryan BROWNE, *University of Guelph*, Paul McNICHOLAS, *University of Guelph*  

Mixtures of Latent Class Analyzers for Incomplete Categorical Data • Mélanges d'analyseurs de classes latentes pour données catégorielles incomplètes [MS-75]

– Weibin JIANG, Hao YU & Reg KULPERGER, *University of Western Ontario*  

Identifiability and Stationarity Conditions for General Multivariate GARCH Models • Conditions d'identifiabilité et de stationnarité pour modèles GARCH généraux multivariés [MS-76]

– Michael McDONALD & Zeny FENG, *University of Guelph*  

Simultaneous Association Test on Multiple Traits • Test d'association simultané pour de multiples traits [MS-77]

– Xiaolei CHEN, Paul CABILIO & Ying ZHANG, *Acadia University*  

Kendall's Tau and Sign Test as Tests for Trend in Time Series Data • Le Tau de Kendall et le test des signes en tant que tests de tendance en analyse de séries chronologiques [MS-78]

– Daria MARTCHENKO, Rob DEARDON & Paul McNICHOLAS, *University of Guelph*  

Designing Experiments to Assess the Space-time Dynamics of Plant Diseases • Planification d'expériences afin d'évaluer la dynamique spatio-temporelle de maladies des plantes [MS-79]

**14:30 – 16:00 Session 04K****VAN Hall**

Case Study 1 “The Effects of Photo-interpreted Variables in the Estimation of Stand-level Merchantable Volumes in the Province of Québec”

Étude de cas 1 « L'effet de variables photo-interprétées sur l'estimation du volume marchand à l'échelle du peuplement dans la province du Québec »

Case Study • Étude de cas

Chair • Président: Kevin KEEN, *University of Northern British Columbia*

Organizers • Responsables: Kevin KEEN, *University of Northern British Columbia*, Gaétan DAIGLE & Louis-Paul RIVEST, *Université Laval*

– Fanny HARVEY & Bastien FERLAND-RAYMOND, *Université Laval*  

- Bing XIA, Chunlin WANG, Emmanuel THOMPSON & Niroshan WITHANAGE, *University of Calgary*  
- Zeinab MASHREGHI, Audrey BÉLIVEAU & Valérie PATENAUDE, *Université de Montréal*  
- Chunyi WANG, *University of Toronto*  
- Bakri ABDALLA, Kyle BALKISSOON, Chelsia BARAL & Herlini Salim ALI, *McMaster University*  
- Christina ASZTALOS & Jessica BLACKWOOD-BECKFORD, *McMaster University*  

**14:30 – 16:00 Session 04L****VAN Hall**

Case Study 2 “Metabolism of Bradykinin and Endogenous des-Arg<sup>9</sup>-Bradykinin in Human Plasma: Contribution to the Pathophysiology of Angioedema Associated with ACE Inhibitors”

Étude de cas 2 « Métabolisme de la bradykinine et de la des-Arg<sup>9</sup>-bradykinine endogène dans le plasma humain : contribution à la physiopathologie de l'angio-œdème associé aux inhibiteurs de l'enzyme de conversion de l'angiotensine »

Case Study • Étude de cas

Chair • Président: Kevin KEEN, *University of Northern British Columbia*

Organizers • Responsables: Kevin KEEN, *University of Northern British Columbia* & Miguel CHAGNON, *Université de Montréal*

- Nianping HU & Wenbin LI, *Queen's University*  
- Rena MANN, Ryan STONE & Linghong LU, *University of Victoria*  
- Wei (Becky) LIN, Hong LU & Ximing XU, *University of Toronto*  
- Xiaoyin ZHONG & Jun (Jason) CHEN, *University of British Columbia*  
- Amidu RAIFU & Zhihui Amy LIU, *McMaster University*  
- Michael (Jack) DAVIS & Emelie GUSTAFSSON, *University of British Columbia, Okanagan*  
- Corinne RIDDELL, Eric CORMIER & Yitian (Sky) LIANG, *University of British Columbia*  
- Eric FU & Kevin USHEY, *University of British Columbia*  
- Ashley BONNER, Faye DE LEON, Abdul QURESHI & Saqib ABBASI, *McMaster University*  

**16:00 – 16:30 Pause-café • Coffee Break****VAN Hall****16:30 – 17:30 Session 05A****VAN 1245A**

Special Invited Session of the Survey Methods Section—Cell Phone Surveys

Allocution de l'invité du président du groupe des méthodes d'enquête — Enquêtes à partir des cellulaires

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Survey Methods Section*

Session parrainée par le *Groupe des méthodes d'enquête*

Organizer and Chair • Responsable et président: Steven THOMPSON, *Simon Fraser University*

– J Michael BRICK, *Westat*  

Dual Frame Landline and Cell Phone Surveys • Enquêtes à partir d'une base de sondage duale de téléphones terrestres et cellulaires [MS-80]

### 16:30 – 18:00 Session 05B

VAN 2289A

Methods for Longitudinal and Spatial Categorical Data

Méthodes pour des données catégorielles longitudinales et spatiales

Invited Paper Session • Présentations sur invitation

Chair • Président: John PETKAU, *University of British Columbia*

Organizer • Responsable: Rachel ALTMAN, *Simon Fraser University*

16:30 Yanbing ZHENG, *University of Kentucky*, Brian AUKEMA, *Canadian Forest Service, Natural Resources Canada and University of Northern British Columbia*  

Spatio-temporal Modeling of Outbreaks of Mountain Pine Beetle • Modélisation spatio-temporelle d'infestations de dendroctone du pin ponderosa [MS-81]

17:00 Farouk NATHOO, *University of Victoria*  

Joint Spatial Modeling of Recurrent Infection and Growth • Modélisation spatiale conjointe de récurrences d'infections et de croissance [MS-82]

17:30 R. Jeanette O'HARA HINES, *University of Waterloo*, W.G.S. HINES, *University of Guelph*  

An Appraisal of Methods for the Analysis of Longitudinal Categorical Data with MAR Drop-outs • Une évaluation des méthodes d'analyse de données catégorielles longitudinales avec attrition aléatoire [MS-83]

### 16:30 – 18:00 Session 05C

VAN 2291A

Probability, Regression, and Time Series

Probabilité, régression et séries chronologiques

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Probability and Business and Industrial Statistics Sections*

Session parrainée par les *Groupes de probabilité et de statistique industrielle et de gestion*

Chair • Président: Hao YU, *University of Western Ontario*

Organizers • Responsables: Rafal KULIK, *University of Ottawa* & Hao YU, *University of Western Ontario*

16:30 Reg KULPERGER, Hao YU & Weibin JIANG, *University of Western Ontario*  

Multivariate GARCH : Some Foundations • GARCH multivarié : quelques fondements [MS-84]

17:00 Wei WU, *University of Chicago*  

Second Order Inference of Time Series • Inférence de second ordre pour séries chronologiques [MS-85]

17:30 Yuliya MARTSYNYUK, *University of Manitoba*  

Functional Central Limit Theorems in Linear Regression Models • Théorèmes limites centraux fonctionnels pour modèles de régression linéaires [MS-86]

**16:30 – 18:00 Session 05D****VAN 2809A**

What is the Role of a Statistical Association in Insuring Rigour in the Publication of Statistical Results in Media?

Quel est le rôle d'une association de statisticiens pour assurer la rigueur dans la publication médiatique de résultats statistiques ?

Invited Panel Discussion • Panel sur invitation

Session sponsored by the *Association des statisticiennes et statisticiens du Québec*

Session parrainée par l'*Association des statisticiennes et statisticiens du Québec*

Organizer and Chair • Responsable et présidente: Nathalie MADORE, *Régie des rentes du Québec*

– Michel SAULNIER, *Président du chapitre du Québec de l'Association de la recherche et de l'intelligence marketing*  

– Lise MILLETTE, *Presse canadienne et membre du CA de la Fédération professionnelle des journalistes du Québec*  

– Pierre-Paul NOREAU, *Directeur de l'éditorial, Journal Le Soleil*  

**16:30 – 18:00 Session 05E****VAN 2811A**

Financial Modelling in Actuarial Sciences

Modèles financiers en actuariat

Invited Paper Session • Présentations sur invitation

Chair • Présidente: Hélène COSSETTE, *Université Laval*

Organizer • Responsable: Gary PARKER, *Simon Fraser University*

16:30 Patrice GAILLARDETZ, *Concordia University*  

Pricing and Hedging Equity-indexed Annuities Using Risk Measures • Fixation de prix et répliation d'annuités indexées aux capitaux propres au moyen de mesures de risques [MS-87]

17:00 Marianito RODRIGO, *ITAM, Mexico*, Rogemar MAMON, *University of Western Ontario*  

A Generating Function Approach in the Calibration of Vasicek and CIR Bond Price Models • Une approche à base de fonctions génératrices pour calibrer les modèles de Vasicek et CIR pour le prix d'obligations [MS-88]

17:30 Ken Seng TAN, *University of Waterloo*, Chengguo WENG, *Towson University*  

Risk Measure Based Approach to Optimal Reinsurance • Une approche basée sur les mesures de risque pour réassurance optimale [MS-89]

**16:30 – 18:00 Session 05F****VAN 1203**

Biostatistics

Biostatistique

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Erica MOODIE, *McGill University*

16:30 Ashkan ERTEFAIE & David STEPHENS, *McGill University*  

On the Efficiency of Propensity Score Matching • De l'efficacité de l'appariement des scores de propension [MS-90]

- 16:45 Lisa LIX & Tolulope SAJOBI, *University of Saskatchewan*    
 Testing Multiple Outcome Variables with a Priori Weights in the Presence of Non-normality and Variance Heterogeneity • Tester de multiples variables réponses avec des poids a priori en présence de non-normalité et d'hétérogénéité de la variance [MS-91]
- 17:00 Eleanor PULLENAYEGUM, Jean-Eric TARRIDE, Feng XIE & Daria O'REILLY, *McMaster University*    
 Analysis of Health Utility Data • Analyse de données d'utilité de la santé [MS-92]
- 17:15 Tolulope SAJOBI, Lisa LIX, William LAVERTY & Longhai LI, *University of Saskatchewan*    
 Discriminant Analysis for Repeated Measures Data: Effects of Covariance Structure on Bias and Error in Discriminant Function Coefficients • Analyse discriminante pour données de mesures répétées : effets de la structure de covariance sur le biais et l'erreur dans les coefficients de la fonction discriminante [MS-93]
- 17:30 Candemir CIGSAR & Jerald LAWLESS, *University of Waterloo*    
 Carryover Effects Testing in Heterogeneous Recurrent Event Processes • Tester pour les effets résiduels en processus d'événements récurrents hétérogènes [MS-94]
- 17:45 Yin LI & Keumhee Chough CARRIERE, *University of Alberta*    
 An Approach to Improving Goodness of Fit for Social Network Models • Une approche afin d'améliorer l'adéquation de modèles de réseaux sociaux [MS-95]

## 16:30 – 18:00 Session 05G

VAN 1211

### Regression Methodology in Biostatistics and Neuronal Data

### Méthodologie de régression en biostatistique et données neuronales

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Chair • Président: Russell STEELE, *McGill University*

- 16:30 Nabil CHANNOUF, *Groupe d'études et de recherche en analyse des décisions*, Brenda MACGIBBON, *UQAM and GERAD*, Marc FREDETTE, *HEC and GERAD*    
 Correlation Effects on Sample Size Calculations for Poisson and ZIP Regression Models • Effet de la corrélation sur la détermination de la taille d'un échantillon dans les modèles de régression de Poisson et de Poisson gonflé à zéro [MS-96]
- 16:45 Laura DUMITRESCU, *University of Ottawa*    
 Estimation of a Longitudinal Linear Structural Error-in-variables Model • Estimation d'un modèle longitudinal linéaire structurel à erreurs dans les variables [MS-97]
- 17:00 Olga MELNICHOUK, Salomon MINKIN, Lisa MARTIN & Norman BOYD, *Ontario Cancer Institute*    
 Power of Logistic Regression with Measurement Error in Predictor Variable and Varying Number of Observations per Subject • Puissance de la régression logistique avec des erreurs de mesure dans la variable prédictive et un nombre variable d'observations par sujet [MS-98]
- 17:15 Benjamin RICH, *McGill University*, Erica MOODIE, *McGill University*, David STEPHENS, *McGill University*    
 Model Checking with Residuals for  $G$ -estimation of Optimal Dynamic Treatment Regimes • Vérification de modèle avec résidus pour  $g$ -estimation de régimes de traitement dynamiques optimaux [MS-99]

17:30 Shojaeddin CHENOURI, *University of Waterloo*, Sam BEHSETA, *California State University, Fullerton*  
 ⓘ ⓘ

Comparison of Two Populations of Curves with an Application in Neuronal Data Analysis • Comparaison de deux populations de courbes avec une application en analyse de données neuronales [MS-100]

17:45 Qing GUO, Hall GEOFF & Eleanor PULLENAYEGUM, *McMaster University* ⓘ ⓘ

A Critical Review of Structural Equation Modelling in fMRI Studies of Effective Connectivity • Revue critique de la modélisation par équations structurelles pour la connectivité effective en études IRMf [MS-101]

## 16:30 – 18:00 Session 05H

VAN 1871

### Hypothesis Testing and Inference

#### Tests d'hypothèses et inférence

Contributed Paper Session • Séance de communications libres

Chair • Président: Serge PROVOST, *University of Western Ontario*

16:30 Félix LABRECQUE-SYNNOTT & Jean-François ANGERS, *Université de Montréal* ⓘ ⓘ

Estimation and Testing For Continuous Zero-modified Models • Estimation et tests pour modèles continus modifiés à zéro [MS-102]

16:45 Denis LAROCQUE, *HEC Montréal*, Jaakko NEVALAINEN, *University of Turku*, Hannu OJA, *University of Tampere* ⓘ ⓘ

Multivariate Nonparametric Two-sample Tests for Mixed Outcomes • Tests non paramétriques multivariés pour comparer deux échantillons avec des variables réponses de type mixtes [MS-103]

17:00 Xiaoqing NIU & Pengfei LI, *University of Alberta* ⓘ ⓘ

Hypothesis Testing in Multivariate Mixture Models: the EM approach • Tests d'hypothèses pour des modèles de mélanges multivariés : l'approche EM [MS-104]

17:15 Jesper RYDÉN & Sven Erick ALM, *Uppsala University* ⓘ ⓘ

The Effect of Interaction and Rounding Error in Two-way ANOVA: An Example on the Impact of Testing for Normality • Effet de l'interaction et l'erreur d'arrondi dans le modèle de l'analyse de la variance à deux facteurs : l'exemple des tests de normalité [MS-105]

17:30 Ying YAN & Changbao WU, *University of Waterloo* ⓘ ⓘ

Empirical Likelihood Inference for Two-sample Problems with Missing Data • Inférence par vraisemblance empirique pour problèmes à deux échantillons avec données manquantes [MS-106]

17:45 Jianan PENG, *Acadia University*, Hubert CHEN, *National Cheng-Kung University*, Jiancheng JIANG, *University of North Carolina at Charlotte* ⓘ ⓘ

Interval Hypotheses for Comparing Several Treatments with a Control • Hypothèses par intervalles pour comparer plusieurs traitements à un contrôle [MS-107]

## 18:10 –

VAN 1245A

### Information on 2010 NSERC Discovery Grant Competition

### Information sur le concours 2010 des subventions à la découverte du CRSNG

## Mardi 25 mai • Tuesday, May 25

---

### 08:30 – 09:45 Session 06A

PPR

Gold Medal Award Address

Allocution de la récipiendaire de la médaille d'or

Special Session • Conférence spéciale

Organizer and Chair • Responsable et présidente: Charmaine DEAN, *Simon Fraser University*– Nancy REID, *University of Toronto*  

Some Thoughts on the Theory of Statistics • Quelques idées sur la théorie de la statistique [MS-108]

### 09:45 – 10:25 Pause-café • Coffee Break

VAN Hall

### 10:25 – 11:55 Session 07A

VAN 1245A

The Biology, Technology, and Statistical Modelling of High-throughput Genomics Data

La biologie, la technologie et la modélisation statistique de données génomiques à haut débit

Introductory Overview Lecture • Conférence de vue d'ensemble

Chair • Président: Raphaël GOTTARDO, *Institut de recherches cliniques de Montréal*Organizer • Responsable: Christian LÉGER, *Université de Montréal*– Naomi ALTMAN, *Penn State University*  

The Biology, Technology and Statistical Modelling of High-throughput Genomics Data • La biologie, la technologie et la modélisation statistique de données génomiques à haut débit [MS-109]

### 10:25 – 11:55 Session 07B

VAN 2289A

New Developments in Modelling Using Survey Data

Nouveaux développements en modélisation de données d'enquêtes

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Survey Methods Section*Session parrainée par le *Groupe des méthodes d'enquête*Organizer and Chair • Responsable et présidente: Lenka MACH, *Statistics Canada*10:25 François VERRET & Suzelle GIROUX, *Statistique Canada*  

Pooling Respondents' Blood Serum for Analysis in the Canadian Health Measures Survey • Regroupement du sérum sanguin de répondants pour analyse dans l'Enquête canadienne sur les mesures de la santé [MS-110]

10:50 Anne-Sophie CHAREST, *Carnegie-Mellon Statistics Department*, Jean-François BEAUMONT, *Statistique Canada*  

Bootstrap Variance Estimation of Model Parameters with Survey Data • Méthode bootstrap pour l'estimation de la variance de paramètres de modèle avec des données d'enquêtes [MS-111]

11:15 Richard VALLIANT, *Universities of Michigan and Maryland*  

Linear Regression Diagnostics for Survey Data • Diagnostics de régression linéaire pour données d'enquête [MS-112]

11:40 Discussant • Modératrice: Georgia ROBERTS, *Statistics Canada*

## 10:25 – 11:55 Session 07C

VAN 2291A

Quality Improvement

Amélioration de la qualité

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Business and Industrial Statistics Section*

Session parrainée par le *Groupe de statistique industrielle et de gestion*

Organizer and Chair • Responsable et président: Gemai CHEN, *University of Calgary*

10:25 Su-Fen YANG, *National Chengchi University, Taiwan*, Smiley W CHENG, *University of Manitoba*  

New Nonparametric Control Charts • Nouvelles cartes de contrôle non paramétriques [MS-113]

10:55 John BREWSTER, *University of Manitoba*  

Some Issues in the Design and Analysis of Industrial split-plot Experiments • Quelques problèmes en planification et analyse d'expériences industrielles avec subdivision des parcelles [MS-114]

11:25 Geoff VINING & Laura FREEMAN, *Virginia Tech*  

Issues in Planning Experiments for Reliability Data • Problèmes en planification d'expériences pour données de fiabilité [MS-115]

## 10:25 – 11:55 Session 07D

VAN 2809A

Applications of Statistics in Hydrology, Climatology, and Meteorology

Statistique appliquée à l'hydrologie, la climatologie et la météorologie

Invited Paper Session • Présentations sur invitation

Chair • Présidente: Anne-Catherine FAVRE, *Université Laval*

Organizers • Responsables: Anne-Catherine FAVRE, *Université Laval* & Jean-François ANGERS, *Université de Montréal*

10:25 Luc PERREAU & James MERLEAU, *Institut de recherche d'Hydro-Québec*, Guillaume ÉVIN, *Université Laval*  

Mixtures of Normal, Gamma and Gumbel Distributions: An Application to Peaks and Volumes of Spring Floods • Mélanges de lois de probabilités normales, gamma et Gumbel : une application aux données de pointes et de volumes de crues printanières [MS-116]

10:55 Vincent FORTIN, *Environnement Canada*, Juan Alberto VELAZQUEZ, François ANCTIL & Anne-Catherine FAVRE, *Université Laval*, Richard TURCOTTE, *Centre d'Expertise Hydrique du Québec*, Heri KUSWANTO, *Université Laval*, Nguyen Bao TRINH, *Institut National de la Recherche Scientifique*  

Evaluation of the Canadian Meteorological Ensemble Prediction System for Short-term Hydrological Forecasting • Evaluation du système de prévision météorologique d'ensemble canadien pour la prévision hydrologique à courte échéance [MS-117]

11:25 Eric PARENT & Jacques BERNIER, *AgroParisTech*  

Some Bayesian Investigations about the Bivariate Gamma Model with Applications to Environmental Data • Quelques investigations bayésiennes autour du modèle gamma bivarié avec applications environnementales [MS-118]

### 10:25 – 11:55 Session 07E

**VAN 2811A**

Ruin Theory

Théorie de la ruine

Invited Paper Session • Présentations sur invitation

Chair • Président: Étienne MARCEAU, *Université Laval*

Organizer • Responsable: Steve DREKIC, *University of Waterloo*

10:25 Andrei BADESCU, *University of Toronto*  

Vector-valued Martingales in Risk Theory • Martingales vectorielles en théorie du risque [MS-119]

10:55 Steve DREKIC, *University of Waterloo*  

A Threshold-based Risk Process with a Waiting Period to Pay Dividends • Un processus de risque basé sur un seuil avec un délai pour le paiement des dividendes [MS-120]

11:25 Manuel MORALES, *University of Montreal*  

On the Ruin Problem for Two New Levy Insurance Risk Processes • Du problème de la ruine pour deux nouveaux processus du risque d'assurance de Levy [MS-121]

### 10:25 – 11:40 Session 07F

**VAN 1203**

Methodology for Hazard Rate Estimation and Censored Data

Méthodologie pour l'estimation du taux de risque et les données censurées

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Cynthia STRUTHERS, *University of Waterloo*

10:25 Bin DONG & David E. MATTHEWS, *University of Waterloo*  

Empirical Likelihood Estimation of a Covariate-adjusted Cumulative Hazard Ratio • Estimation par vraisemblance empirique d'un rapport de risques cumulatifs ajusté par covariables [MS-122]

10:40 Na LEI & Wenqing HE, *University of Western Ontario*  

Weakly Parametric PH Regression Model for Multivariate Survival Data • Modèles faiblement paramétriques de régression à risques proportionnels pour données de survie multivariées [MS-123]

10:55 Longyang WU & Richard COOK, *University of Waterloo*  

Estimation of Treatment Effects based on Composite Endpoints • Estimation d'effets de traitement à partir de points terminaux composites [MS-124]

11:10 Jun LI & Yogendra P. CHAUBEY, *Concordia University*, Pranab SEN, *University of North Carolina*  

Some New Smooth Density Estimators for Length Biased Data • Quelques nouveaux estimateurs lisses de densités pour données biaisées en longueur [MS-125]

- 11:25 Héla ROMDHANI, M'Hamed Lajmi LAKHAL CHAIEB & Louis-Paul RIVEST, *Université Laval*    
 A Copula Model for Bivariate Observations with Below Detection Limits • Une copule pour données bidimensionnelles avec un seuil de détection [MS-126]

**10:25 – 11:40 Session 07G****VAN 1211**

## Shrinkage Estimation and Bayesian Methodology

## Estimation par rétrécissement et méthodologie bayésienne

Contributed Paper Session • Séance de communications libres

Chair • Président: Simon GUILLOTTE, *University of Prince Edward Island*

- 10:25 Saber FALLAHPOUR & S. Ejaz AHMED, *University of Windsor*    
 Shrinkage Estimation Strategy in Non-linear Models • Stratégie d'estimation par rétrécissement pour modèles non linéaires [MS-127]
- 10:40 Sévérin NKURUNZIZA, *University of Windsor*    
 Shrinkage Methods in Some Multivariate Dynamical Models • Méthodes par rétrécissement dans certains modèles dynamiques multidimensionnels [MS-128]
- 10:55 Mohamed TAMAZIGHT & Éric MARCHAND, *Université de Sherbrooke*    
 Bayesian Confidence Intervals using Balanced Loss Functions • Intervalles de confiance bayésiens à l'aide de fonctions de perte pondérées [MS-129]
- 11:10 Longhai LI, *University of Saskatchewan*    
 Are Bayesian Inferences Weak for Wasserman's Example? • Les inférences bayésiennes sont-elles inefficaces pour l'exemple de Wasserman ? [MS-130]
- 11:25 Othmane KORTBI & Éric MARCHAND, *Université de Sherbrooke*    
 On the Performance of Truncated Linear Estimators for Estimating a Restricted Multivariate Normal Mean • Sur la performance d'estimateurs linéaires tronqués pour estimer la moyenne d'une loi multinormale sous contraintes [MS-131]

**10:25 – 11:40 Session 07H****VAN 1871**

## Density Estimation and Robust Regression

## Estimation de la densité et régression robuste

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Katherine DAVIES, *University of Manitoba*

- 10:25 Alexandre LEBLANC, *University of Manitoba*    
 On the Boundary Properties of Bernstein Polynomial Estimators • Comportement des estimateurs de Bernstein aux frontières de l'intervalle unitaire [MS-132]
- 10:40 Serge PROVOST, *University of Western Ontario*, Min JIANG, *Statistics Canada*    
 On the Determination of the Polynomial Component of Certain Density Estimates • Sur la détermination de la composante polynomiale de certains estimateurs de fonctions de densité [MS-133]
- 10:55 Mark WOLTERS & John BRAUN, *University of Western Ontario*    
 Using a Linear Combination of Adjustments to Enforce Shape Constraints on a Density Estimate • Utilisation d'une combinaison linéaire d'ajustements afin de faire respecter des contraintes de forme à un estimateur de densité [MS-134]

- 11:10 Enayetur RAHEEM & S. Ejaz AHMED, *University of Windsor*    
Shrinkage  $M$ -Estimation in Partially Linear Models •  $M$ -estimation par rétrécissement dans des modèles partiellement linéaires [MS-135]
- 11:25 Matias SALIBIAN-BARRERA, *University of British Columbia*, Azadeh ALIMADAD, *Simon Fraser University*    
An Outlier-robust Fit for Generalized Additive Models with Applications to Outbreak Detection • Ajustement robuste aux valeurs aberrantes de modèles additifs généralisés avec applications à la détection d'épidémies [MS-136]

**10:25 – 11:55 Session 07I****VAN 1883**

## Time Series, Spatio-temporal Data, and Statistical Education

## Séries chronologiques, données spatio-temporelles et l'enseignement de la statistique

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Bethany WHITE, *University of Western Ontario*

- 10:25 Bei CHEN & Yulia GEL, *University of Waterloo*, N. BALAKRISHNA, *Cochin University of Science and Technology*, BOVAS ABRAHAM, *University of Waterloo*    
Forecasting ARCH/GARCH Processes using Sieve Bootstrap • Prévission de processus ARCH/GARCH via le bootstrap par tamis [MS-137]
- 10:40 Lengyi HAN & John BRAUN, *University of Western Ontario*    
Calibrating the Block Bootstrap for Time Series • Calibration du bootstrap par blocs pour séries chronologiques [MS-138]
- 10:55 Ameer MANCEUR & Pierre DUTILLEUL, *McGill University*    
Maximum Likelihood Estimation for Separable Variance-covariance Structures in 2D and 3D • Estimation par maximum de vraisemblance de structures de variance-covariance séparables en 2D et 3D [MS-139]
- 11:10 Katrina ROGERS-STEWART, Patrick FARRELL & JASON NIELSEN, *Carleton University*    
Spatio-temporal Estimates as Predictors in a Hierarchical Model • Estimateurs spatio-temporels comme prédicteurs dans un modèle hiérarchique [MS-140]
- 11:25 AnneMichele MILLAR, *Mount Saint Vincent University*, Candace SCHAU, *CS Consultants, LLC*    
Student Attitudes to Statistics: Survey Results and Discussion • Attitudes étudiantes par rapport à la statistique : résultats d'enquête et discussion [MS-141]
- 11:40 Tasneem ZAIHRA, *University of New Brunswick—Saint John*, Hasan RIZVI    
Teaching Statistics Utilizing Web 2.0 • Enseigner la statistique avec le Web 2.0 [MS-142]

**11:55 – 13:15 Dîner • Lunch****13:15 – 14:15 Session 08A****POU 1112**

## CRM–SSC Prize in Statistics Address

## Allocution de la récipiendaire du prix CRM-SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Louis-Paul RIVEST, *Université Laval*

– Grace YI, *University of Waterloo*  

Statistical Analysis of “Imperfect” Data: Handling Missing Observations and Measurement Error • Analyse statistique de données “imparfaites” : Tenir compte d’observations manquantes et d’erreurs de mesure [MS-143]

### 14:25 – 15:55 Session 09A

**VAN 1245A**

Statistical Methods for Life History Analysis

Méthodes statistiques pour l’analyse de l’historique de vie

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Organizer and Chair • Responsable et président: Richard COOK, *University of Waterloo*

14:25 Per Kragh ANDERSEN, *University of Copenhagen, Denmark*, Giuliana CORTESE, *Università degli studi di Padova, Italia*  

Competing Risks and Time-dependent Covariates • Risques concurrents et covariables dépendant du temps [MS-144]

14:55 Niels KEIDING & Oluf HANSEN, *University of Copenhagen, Denmark*, Rémy SLAMA, *Institut Albert Bonniot/Inserm U823, Grenoble, France*  

Analysis of Time-to-pregnancy from Current Duration Data • Analyse du temps d’attente avant une grossesse à partir de durées d’attentes actuelles [MS-145]

15:25 Daniel FAREWELL, *Cardiff University*  

The Cox Model via Generalized Estimating Equations • Le modèle de Cox par des équations d’estimation généralisées [MS-146]

### 14:25 – 15:55 Session 09B

**VAN 2289A**

Recent Advances in Modelling Long Range Dependent and Nonlinear Time Series, with Applications to Business, Econometrics, and Environmental Sciences

Développements récents dans la modélisation de la dépendance à long terme et de séries chronologiques non linéaires, avec applications en gestion, économétrie et sciences environnementales

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Yulia GEL, *University of Waterloo*

14:25 Nalini RAVISHANKER, *University of Connecticut, Storrs*, Jeffrey PAI, *University of Manitoba*  

Fast Bayesian Estimation for Sub-Gaussian Stable VARFIMA Processes via a Multivariate Preconditioned Conjugate Gradient Algorithm • Estimation bayésienne rapide pour processus stables VARFIMA sous-gaussiens au moyen d’un algorithme de gradient conjugué préconditionné multivarié [MS-147]

14:55 Melody GHAHRAMANI, *University of Winnipeg*, Aerambamoorthy THAVANESWARAN, *University of Manitoba*  

On the Combined Estimating Functions Method with Applications in Finance • De la méthode des fonctions estimatrices combinées avec applications en finance [MS-148]

15:25 Gemai CHEN, *University of Calgary*  

A New Quantile Function Based Model for Modeling Price Behaviours in Financial Markets • Un

nouveau modèle à base de fonctions quantiles pour modéliser le comportement des prix en marchés financiers [MS-149]

### 14:25 – 15:55 Session 09C

VAN 2291A

Quality in Surveys

Qualité dans les enquêtes

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Survey Methods Section*

Session parrainée par le *Groupe des méthodes d'enquête*

Chair • Présidente: Georgia ROBERTS, *Statistics Canada*

Organizer • Responsable: Laurie REEDMAN, *Statistics Canada*

14:25 France LAPOINTE, Sylvain VÉGIARD & Louise BOURQUE, *Institut de la statistique du Québec*  

The Compendium of Recommended Survey Practices of the Institut de la statistique du Québec •  
Le Recueil de bonnes pratiques dans les enquêtes de l'Institut de la statistique du Québec [MS-150]

14:55 Maureen KELLY, *Canadian Institute for Health Information*  

Quality in Continuing Care: How new CIHI Data and Indicators are Supporting Health Care •  
Qualité en soins de suivi : comment les nouvelles données et indicateurs ICIS soutiennent les soins  
de santé [MS-151]

15:25 Xinye (Hannah) YANG & Caroline PELLETIER, *Statistics Canada*  

Improving the Quality of Automated Coding at Statistics Canada • Améliorer la qualité du codage  
automatisé à Statistique Canada [MS-152]

### 14:25 – 15:55 Session 09D

VAN 2809A

Stochastic Dynamics in Biology

Dynamique stochastique en biologie

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Probability Section*

Session parrainée par le *Groupe de probabilité*

Organizer and Chair • Responsable et présidente: Lea POPOVIC, *Concordia University*

14:25 Nicolas LANCHIER & Jared NEUFER, *Arizona State University*  

Geometric Properties of the Majority Rule Model • Propriétés géométriques du modèle de la loi  
majoritaire [MS-153]

14:55 John MAYBERRY & Rick DURRETT, *Cornell University*  

Traveling Waves of Selective Sweeps • Vagues voyageuses de balayages sélectifs [MS-154]

15:25 Anita WINTER, *TU München*, Andreas GREVEN, *Universität Erlangen-Nürnberg*, Peter PFAFFELHUBER, *Freiburg*, Anja STURM, *Universität Göttingen*, Iljana ZÄHLE, *Universität Erlangen-Nürnberg*  
 

A Multitype Branching Model with Local Self-regulation • Un modèle de branchement multitype  
avec auto-régulation locale [MS-155]

**14:25 – 15:55 Session 09E****VAN 2811A**

## Privacy Legislation—Implications for Canadian Statisticians

## Législations sur la vie privée — Conséquences pour les statisticiens canadiens

Invited Panel Discussion • Panel sur invitation

Session sponsored by the *Accreditation Committee*Session parrainée par le *Comité d'accréditation*Organizer and Chair • Responsable et présidente: Mary LESPERANCE, *University of Victoria*14:25 Don WILLISON, *Ontario Agency for Health Protection and Promotion*  

What Should you Have in Hand when you Knock on the Door of your Research Ethics Board? • Que devriez-vous avoir en main avant de frapper à la porte de votre comité d'éthique? [MS-156]

14:45 Christian WHALEN, *Office of the Ombudsman, Government of New Brunswick*  

Data, Privacy and the Law • Données, vie privée et loi [MS-157]

15:05 Philip GILES, *Statistics Canada*  

Statistics Canada's Policies and Practices Related to Privacy, Security and Confidentiality • Politiques et pratiques de Statistique Canada liées à la vie privée, à la sécurité et à la confidentialité [MS-158]

15:25 Discussant • Modérateur: Kevin KEEN, *University of Northern British Columbia*

15:45 Floor discussion • Discussion plénière

**14:25 – 15:55 Session 09F****VAN 1203**

## Binary or Discrete Data

## Données binaires ou discrètes

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*Session parrainée par le *Groupe de biostatistique*Chair • Présidente: Karen Kopciuk, *Alberta Health Services – Cancer Care*14:25 Blache Paul AKPOUE & Jean-François ANGERS, *Université de Montréal*  

Modelling of Negatively Correlated Multivariate Discrete Data • Modélisation de données discrètes multidimensionnelles corrélées négativement [MS-159]

14:40 Zhijian CHEN, Grace YI & Changbao WU, *University of Waterloo*  

Estimating Functions for Correlated Binary Data Analysis with Misclassified Responses • Fonctions estimatrices pour l'analyse de données binaires corrélées avec mauvaise classification de la réponse [MS-160]

14:55 Gerald LBOVIC, *University of Toronto*, Patrick BROWN, *Cancer Care Ontario*, George TOMLINSON, *University Health Network*, James STAFFORD, *University of Toronto*  

Estimating Time-varying Intensity Matrices in Continuous Time Multi-state Markov Models Using Local Likelihood Methods • Estimation des matrices d'intensité dépendantes du temps dans des modèles de Markov à multiétats à temps continu par méthodes de vraisemblance locale [MS-161]

15:10 Krishna SAHA, *Central Connecticut State University*  

Tests for Proportions in the Clustered Binary Data, with Applications to Toxicological Data • Test de proportions pour données binaires en grappes avec application à des données de toxicologie [MS-162]

- 15:25 Xuemao ZHANG & Sudhir PAUL, *University of Windsor*    
 Estimation of Regression Effects in Longitudinal Binary Data • Estimation des effets de régression pour des données binaires longitudinales [MS-163]
- 15:40 David STEPHENS, *McGill University*, Martin CROWDER, *Imperial College London*    
 Markov Chain Monte Carlo for Markov Chain Macro Data • Chaînes de Markov Monte-Carlo pour macro données markoviennes [MS-164]

**14:25 – 15:55 Session 09G****VAN 1211****Risk Theory and Probability Theory in Finance****Théorie du risque et probabilité en finance**

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Probability and Business and Industrial Statistics Sections*Session parrainée par les *Groupes de probabilité et de statistique industrielle et de gestion*Chair • Président: Andrei BADESCU, *University of Toronto*

- 14:25 Yingying GAN & Xiaoming LIU, *University of Western Ontario*    
 Model Mortality Index with Markovian Drift Under the Lee-Carter Framework • Modélisation de l'indice de mortalité avec dérive markovienne dans le cadre du modèle de Lee-Carter [MS-165]
- 14:40 Ya Fang WANG, *Concordia University*, Ghislain LÉVEILLÉ, *Université Laval*, José GARRIDO, *Concordia University*    
 The Distribution of the Compound Sums and Risk Measures • Distribution de sommes composées et mesures de risque [MS-166]
- 14:55 Ben Salah ZIED & Romuald Hervé MOMEYA, *Université de Montréal*    
 The Minimal Entropy Martingale Measure (MEMM) for a Markov-modulated Lévy Model • La mesure d'entropie martingale minimale (MEMM) dans un modèle de Lévy à modulations markoviennes [MS-167]
- 15:10 Frédéric GODIN & Manuel MORALES, *Université de Montréal*    
 Derivatives of the European Carbon Market • Les produits dérivés du marché du carbone européen [MS-168]
- 15:25 Romuald Hervé MOMEYA & Manuel MORALES, *Université de Montréal*    
 On the Price of Risk of the Underlying Markov Chain in an Exponential Regime-Switching Lévy model • La prise en compte du risque sous-jacent à une chaîne de Markov dans un modèle de Lévy avec changements de régime [MS-169]
- 15:40 Ramin OKHRATI, *Concordia University*, Alejandro BALBÁS, *Universidad Carlos III de Madrid, Spain*, José GARRIDO, *Concordia University*    
 Risk Measures and Arbitrage Opportunities • Mesures de risque et opportunités d'arbitrage [MS-170]

**14:25 – 15:55 Session 09H****VAN 1871****Nonresponse, Imputation, and Bootstrapping****Non-réponse, imputation et bootstrap**

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Survey Methods Section*Session parrainée par le *Groupe des méthodes d'enquête*

Chair • Président: Wilson LU, *Acadia University*

- 14:25 James BRENNAN & Jack LOTHIAN, *Statistics Canada*    
Integrating Quality Indicators in Complex Surveys • Intégrer des indicateurs de qualité dans des enquêtes complexes [MS-171]
- 14:40 François VERRET, *Statistique Canada*    
Calibrating Complex Survey Weights to Adjust for Refusals to Record Linkage • Calibration de poids d'enquête complexes pour traiter le refus à l'appariement des fichiers [MS-172]
- 14:55 Ivan CARRILLO-GARCIA, *Statistics Canada*, Milorad KOVACEVIC, *United Nations Development Programme*, Changbao WU, *University of Waterloo*    
A Comparison of Reweighting and Imputation Approaches for Missing Responses in Longitudinal Surveys • Une comparaison des approches de repondération et d'imputation pour réponses manquantes dans des enquêtes longitudinales [MS-173]
- 15:10 Audrey BÉLIVEAU & David HAZIZA, *Université de Montréal*, Myriam MAUMY, *Université de Strasbourg*    
Nonparametric Estimation of Response Probabilities in Survey Sampling: An Empirical Investigation • Estimation non paramétrique des probabilités de réponse dans les enquêtes : une étude empirique [MS-174]
- 15:25 Lenka MACH, Jean-François BEAUMONT, Keven BOSHA & Yves LAFORTUNE, *Statistics Canada*    
Mean Bootstrap Variance Estimation for Survey Data • Estimation de la variance du bootstrap moyen pour les données d'enquêtes [MS-175]
- 15:40 Zeinab MASHREGHI, David HAZIZA & Christian LÉGER, *Université de Montréal*    
Bootstrap Variance Estimation in the Presence of Imputed Data • L'estimateur bootstrap de la variance en présence de données imputées [MS-176]

## 14:25 – 15:40 Session 09I

**VAN 1883**

### Quality Control and Design of Experiments

### Contrôle de qualité et planification d'expériences

Contributed Paper Session • Séance de communications libres

Chair • Président: Tom LOUGHIN, *Simon Fraser University, Surrey*

- 14:25 Jaiwook BAIK, *Korea National Open University*    
Current Practices in Reliability Improvement and Assurance Tests in Korea • Pratiques actuelles pour l'amélioration de la fiabilité et les tests d'assurance en Corée [MS-177]
- 14:40 Chithran VASUDEVAN & Asokan VARIYATH, *Memorial University of Newfoundland*    
Empirical Likelihood Based Information Criteria for Change Point Problems • Critères d'information basés sur la vraisemblance empirique pour problèmes de point de rupture [MS-178]
- 14:55 Xiaoli SHANG & Xiaojian XU, *Brock University*    
Optimal Designs for Fourier Regression Models • Plans optimaux pour modèles de régression de Fourier [MS-179]
- 15:10 Wai Kong YUEN & Xiaojian XU, *Brock University*    
Optimal Implementations for Robust Designs with Densities • Implémentations optimales pour plans robustes avec densités [MS-180]

15:25 Philip BERTRAND, *Consultant Statistician*  

The Importance of Experimental Design • L'importance de la planification d'expériences [MS-181]

## 14:25 – 15:55 Session 09J

VAN Hall

Poster Session 2

Séance d'affichage 2

Contributed Poster Session • Séance d'affichage libre

– Abbie GARDNER, Rob DEARDON & Gerarda DARLINGTON, *University of Guelph*  

Goodness-of-fit Measures for Individual-level Models for Infectious Disease • Mesures de qualité de l'ajustement pour modèles de maladies infectieuses au niveau de l'individu [MS-182]

– Chandima KARUNANAYAKE, Donna RENNIE & Punam PAHWA, *University of Saskatchewan*, Yue CHEN, *University of Ottawa*, James DOSMAN, *University of Saskatchewan*  

Relationship Between Lung Function and Hypertension among Rural Canadians using Fractional Polynomials • Relation entre fonction pulmonaire et hypertension en milieu rural au moyen de polynômes fractionnels [MS-183]

– Lee CHEL HEE & Lisa LIX, *University of Saskatchewan*  

Heterogeneity Control of Capture-recapture • Contrôle de l'hétérogénéité en capture-recapture [MS-184]

– Irene VRBIK, Rob DEARDON & Feng ZENY, *University of Guelph*, Braun JOHN, *University of Western Ontario*  

Modelling the Spatio-temporal Dynamics of Fire Spread • Modéliser la dynamique spatio-temporelle de la propagation d'incendies [MS-185]

– Lorna DEETH & Rob DEARDON, *University of Guelph*  

Finite Mixtures of Individual-level Models for Infectious Diseases • Mélanges finis de modèles au niveau de l'individu pour maladies infectieuses [MS-186]

– Marie FOREST & Fabrice LARRIBE, *Université du Québec à Montréal*  

Fine Genetic Mapping of Polygenic Traits by the Coalescent Process with Recombination • Cartographe génétique fine de caractères polygéniques par le processus de coalescence avec recombinaison [MS-187]

– Ye LI & Patrick BROWN, *University of Toronto*  

Continuous Model in Disease Mapping with Aggregated Data • Modèle continu en cartographie des maladies avec données agrégées [MS-188]

– Malorie CHABOT-BLANCHET & Geneviève LEFEBVRE, *Université du Québec à Montréal*  

Bayesian Model Averaging to Estimate Causal Effects • Moyennage bayésien pour l'estimation d'effets causaux [MS-189]

– Selva SELVARATNAM & Alwell OYET, *Memorial University*  

Wavelet Designs for Nonparametric Regression Models with Autocorrelated Errors • Plans à base d'ondelettes pour modèles de régression non paramétriques à erreurs autocorrélées [MS-190]

– Dandan LUO & Peng ZHANG, *University of Alberta*  

Efficient Estimation for Patient-specific Rates of Disease Progression Using Nonnormal Linear Mixed Models for Multiple Characteristics • Estimation efficace de taux de progression de maladies spécifiques aux patients au moyen de modèles linéaires mixtes non normaux pour de multiples caractéristiques [MS-191]

– Zeny FENG, Xiaojian YANG & Paul McNICHOLAS, *University of Guelph*  

Genome Selection for Predicting a Quantitative Trait • Sélection de génome pour la prédiction d'un trait quantitatif [MS-192]

– Rhonda ROSYCHUK, *University of Alberta*  

Implementation of a Cost Quotation Tool for Statistical Consulting in an Academic Medical Setting: Preliminary Results • Implémentation d'un outil d'estimation des coûts de consultation statistique dans un cadre médico-académique : résultats préliminaires [MS-193]

– Ryan BROWNE & Judy-Anne CHAPMAN, *NCIC Clinical Trials Group, Queen's University*  

SAS General Parametric Step-wise Regression Macro • Macro SAS de régression générale paramétrique pas-à-pas [MS-194]

– Matthew SPARLING & Paul McNICHOLAS, *University of Guelph*  

Mixtures of Gaussian and Uniform Distributions for Model-based Clustering • Mélanges de distributions gaussiennes et uniformes pour classification à base de modèles [MS-195]

– Zhang LIN & Rob DEARDON, *University of Guelph*  

Assessing the Fit of Epidemic Models over Time and Space • Évaluer l'adéquation de modèles épidémiques à travers le temps et l'espace [MS-196]

– Sanjeena SUBEDI & Paul McNICHOLAS, *University of Guelph*  

Parameter Estimation for Model-based Clustering of Longitudinal Data • Estimation paramétrique en classification à base de modèles pour données longitudinales [MS-197]

– Brian FRANZAK & Paul McNICHOLAS, *University of Guelph*  

Model-based Clustering Using Mixtures of Gamma Distributions • Classification à base de modèles utilisant des mélanges de distributions gamma [MS-198]

– Swarna WEERASINGHE, *Dalhousie University*, Bruce SMITH, *Dalhousie university*, Sam STEWART, *Dalhousie University*  

Periodic Autoregressive Moving Average (PARMA) Models: Analyses of Ozone Levels in Nova Scotia • Modèles autorégressifs périodiques à moyenne mobile (MAPMM) : analyse de niveaux d'ozone en Nouvelle-Écosse [MS-199]

– Hedy JIANG, *Cancer Care Ontario & McMaster University*, Patrick BROWN, *Cancer Care Ontario & University of Toronto*  

A Bayesian Evaluation of Breast Cancer Screening • Une évaluation bayésienne du dépistage du cancer du sein [MS-200]

– Ian McLEOD & Changjiang XU, *University of Western Ontario*  

Generalized Information Criteria for Model Selection with Penalized Maximum Likelihood Estimation • Un critère d'information généralisé pour la sélection d'un modèle avec l'estimation par maximum de vraisemblance pénalisée [MS-201]

– Mateen SHAIKH, Paul McNICHOLAS & Anthony F. DESMOND, *University of Guelph*  

A Pseudo-EM Algorithm for Clustering Incomplete Longitudinal Data • Un algorithme pseudo-EM pour la classification de données longitudinales incomplètes [MS-202]

**15:55 – 16:25 Pause-café • Coffee Break**

**VAN Hall**

**16:25 – 17:25 Session 10A****VAN 1245A**

Isobel Loutit Invited Address of the Business and Industrial Statistics Section Special:  
Quality Technology in the High-tech Age

Allocution Isobel Loutit du Groupe de statistique industrielle et de gestion : La technologie  
en qualité à l'ère de la haute technologie

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Business and Industrial Statistics Section*

Session parrainée par le *Groupe de statistique industrielle et de gestion*

Organizer and Chair • Responsable et président: Smiley CHENG, *University of Manitoba*

– C. F. Jeff WU, *Georgia Inst of Technology*  

Quality Technology in the High-tech Age • La gestion de la qualité à l'ère de la haute technologie  
[MS-203]

**16:25 – 17:55 Session 10B****VAN 2289A**

Analysis of Next Generation Sequencing Data for Functional Genomics

Analyse de données de séquençage de la prochaine génération en génomique fonctionnelle

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Naomi ALTMAN, *Pennsylvania State University*

16:25 Raphaël GOTTARDO, *Institut de recherches cliniques de Montréal (IRCM)*  

PICS: Probabilistic Inference for ChIP-seq • PICS : Inférence probabilistique pour ChIP-Seq  
[MS-204]

16:55 John MARIONI, *University of Chicago*  

Statistical Issues in the Analysis of RNA-seq Data • Problèmes statistiques dans l'analyse de données d'ARN-seq [MS-205]

17:25 Russell ZARETZKI & Michael GILCHRIST, *University of Tennessee*, William BRIGGS, *Methodist Hospital, Brooklyn N.Y.*, Artin ARMAGAN, *Duke University*  

Bias Correction and Bayesian Analysis of Digital Gene Expression Libraries • Correction du biais et analyse bayésienne de bibliothèques digitales d'expression génétique [MS-206]

**16:25 – 17:55 Session 10C****VAN 2291A**

Stochastic Models in Population Genetics

Modèles stochastiques en génétique des populations

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Probability Section*

Session parrainée par le *Groupe de probabilité*

Organizer and Chair • Responsable et président: Sabin LESSARD, *Université de Montréal*

16:25 John WAKELEY & Sohini RAMACHANDRAN, *Harvard University*  

Sources of Randomness in Genetic Ancestry • Sources de variation aléatoire en ascendance génétique [MS-207]

16:47 François ROUSSET, *CNRS & Université Montpellier 2, France*, Raphaël LEBLOIS, *Museum National d'Histoire Naturelle, Paris, France*  

Likelihood-based Inference of Dispersal Rates from Patterns of Genetic Diversity • Méthodes de

vraisemblance pour inférer sur des taux de dispersion à partir de patrons de diversité génétique [MS-208]

17:10 Martin MÖHLE, *University of Tübingen*  

Coalescent Processes without Proper Frequencies and Applications to the Poisson-Dirichlet Coalescents • Processus coalescents sans fréquences propres et applications aux coalescents Poisson-Dirichlet [MS-209]

17:33 Robert GRIFFITHS & ALISON ETHERIDGE, *Oxford University*, Jay TAYLOR, *Arizona State University*  

A Coalescent Dual Process in a Cannings Model with Genic Selection, and the Lambda Coalescent Limit • Un processus dual de coalescence dans un modèle de Cannings avec sélection génétique et la limite de coalescence Lambda [MS-210]

### 16:25 – 17:55 Session 10D

VAN 2809A

Panel Discussion on Problems and Issues Faced by New Researchers

Panel sur les problèmes et questions auxquels font face les nouveaux chercheurs

Invited Panel Discussion • Panel sur invitation

Session sponsored by the *Committee on New Investigators*  
Session parrainée par le *Comité des nouveaux chercheurs*

Organizer and Chair • Responsable et présidente: Yulia GEL, *University of Waterloo*

– Paul CABILLO, *Acadia University*  

– Kelly ZOU, *Pfizer Inc. & Harvard University*  

– Nancy REID, *University of Toronto*  

### 16:25 – 17:55 Session 10E

VAN 2811A

Analysis of Multilevel Models for Complex Survey Data

Analyse de modèles multiniveaux pour des données d'enquêtes complexes

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Zilin WANG, *Wilfrid Laurier University*

16:25 Danny PFEFFERMANN, *Hebrew University & University of Southampton*  

Estimation of Variance Components for Multilevel Models • Estimation de composantes de la variance pour modèles multiniveaux [MS-211]

This talk will be given by Zilin WANG • Cette conférence sera présentée par Zilin WANG.

16:55 Wilson LU, *Acadia University*, Zilin WANG, *Wilfrid Laurier University*, Mary THOMPSON, *University of Waterloo*  

Estimation of Empirical Distributions of Variance Component Estimators for Multilevel Models • Estimation des distributions empiriques des estimateurs des composantes de la variance pour modèles multiniveaux [MS-212]

17:25 J. N. K. RAO, *Carleton University*  

A Weighted Estimating Equations Approach to Inference for Two-level Models from Survey Data • Une approche d'inférence par équations estimatrices pondérées pour modèles à deux niveaux et données d'enquêtes [MS-213]

**16:25 – 17:40 Session 10F****VAN 1203**

## Missing or Incomplete Observations

## Observations manquantes ou incomplètes

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*Session parrainée par le *Groupe de biostatistique*Chair • Présidente: Lisa LIX, *University of Saskatchewan*16:25 Jinhui MA, Noori AKHTAR-DANESH, Lisa DOLOVICH & Lehana THABANE, *McMaster University*  

Imputation Strategies for Missing Binary Outcomes in Cluster Randomized Trials • Stratégies d'imputation pour réponses binaires manquantes en essais cliniques randomisés par grappes [MS-214]

16:40 Michael REGIER, *McGill University*  

A Unified Approach for Handling the Combined Problem of Missing Data and Measurement Error • Une approche unifiée pour traiter le problème de données manquantes combinées à des erreurs de mesure [MS-215]

16:55 Hua SHEN & Richard COOK, *University of Waterloo*  

Analysis of Left-truncated and Right-censored Survival Times with Incomplete Covariate Data • Analyse de données de survie tronquées à gauche et censurées à droite avec covariables manquantes [MS-216]

17:10 Yang ZHAO & Meng LIU, *University of Regina*  

Regression Model with Covariates Missing in Non-monotonic Patterns • Modèle de régression avec covariables manquantes de façon non-monotone [MS-217]

17:25 Dianliang DENG & Ting ZHANG, *University of Regina*  

Modelling the Lagged Effect for Incomplete Data in Tumour Xenograft Experiments • Modélisation de l'effet de rémanence pour données incomplètes en expériences de xénogreffes [MS-218]

**16:25 – 17:40 Session 10G****VAN 1211**

## Longitudinal Data and Mixed Effects

## Données longitudinales et effets mixtes

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*Session parrainée par le *Groupe de biostatistique*Chair • Présidente: Julie HORROCKS, *University of Guelph*16:25 Taraneh ABARIN, *Samuel Lunenfeld Research Institute*, He LI & Liqun WANG, *University of Manitoba*, Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute*, Craig PENNELL, *University of Western Australia*  

Linear Mixed Effects Models for Longitudinal Studies with Measurement Error on Covariates and Response • Modèles à effets mixtes linéaires pour études longitudinales avec erreurs de mesures dans les covariables et la réponse [MS-219]

16:40 Tanvir QUADIR & S. Ejaz AHMED, *University of Windsor*  

Improving the Estimation of Fixed Effects in the Linear Mixed Model for Longitudinal Data • Améliorer l'estimation d'effets fixes dans un modèle linéaire mixte pour données longitudinales [MS-220]

- 16:55 Guohua YAN, Renjun MA & M. Tariqul HASAN, *University of New Brunswick*    
 Generalized Linear Models with Nonparametric Random Effects for Longitudinal Data • Modèles linéaires généralisés avec effets aléatoires non paramétriques pour données longitudinales [MS-221]
- 17:10 Haocheng LI & Grace YI, *University of Waterloo*    
 A Composite Likelihood Approach for Longitudinal Data with Nonignorable Non-monotone Missing Observations in Both Response and Covariates • Une approche par vraisemblance composite pour données longitudinales avec données manquantes non ignorables et non monotones dans la réponse et les covariables [MS-222]
- 17:25 Erica MOODIE & David STEPHENS, *McGill University*    
 Marginal Structural Models for Competing Risks • Modèles structuraux marginaux pour risques compétitifs [MS-223]

**16:25 – 17:55 Session 10H****VAN 1871**

## Methodological Issues in Designing Surveys

## Questions méthodologiques dans la planification d'enquêtes

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Survey Methods Section*Session parrainée par le *Groupe des méthodes d'enquête*Chair • Présidente: Cynthia Bocci, *Statistique Canada*

- 16:25 Sébastien LABELLE-BLANCHET, Claude GIRARD & Martin HAMEL, *Statistique Canada*    
 Methodological Issues of a Survey on Industrial Processes • Enjeux méthodologiques d'une enquête portant sur les processus industriels [MS-224]
- 16:40 Jeannine CLAVEAU, Xinye (Hannah) YANG & Laurie REEDMAN, *Statistique Canada*    
 Business Survey Data Collection Research at Statistics Canada • La recherche sur la collecte de données des enquêtes-entreprises à Statistique Canada [MS-225]
- 16:55 JAVIER OYARZUN & ERIC PELLETIER, *Statistique Canada*    
 Redesign of Research and Development in the Canadian Industry Survey: Combining Tax and Survey Data • Remaniement de l'Enquête sur la recherche et développement dans l'industrie canadienne : Combiner des données fiscales et d'enquêtes [MS-226]
- 17:10 Mamadou DIALLO, Marie-Claude DUVAL & Steven THOMAS, *Statistique Canada*    
 The Methodological Challenges of the Survey on Living with Chronic Diseases in Canada (SLCDC) • Les défis méthodologiques de l'enquête sur les personnes ayant une maladie chronique au Canada (EPMCC) [MS-227]
- 17:25 Eric HORTOP, *Statistics Canada*    
 Multidimensional Tabular Data Suppression in the Canadian Cancer Registry • Suppression des données dans les tableaux multidimensionnels du Registre canadien du cancer [MS-228]
- 17:40 Cindy UBARTAS, *Statistique Canada*    
 Evaluation of Telephone Interview Collection in First Nations Communities: 2009 Pilot Study • Évaluation de la collecte par entrevue téléphonique dans les communautés des Premières Nations : étude pilote de 2009 [MS-229]

## Mercredi 26 mai • Wednesday, May 26

---

### 08:30 – 09:15 Session 11A

POU 1112

Pierre Robillard Award Address

Allocution de la récipiendaire du prix Pierre-Robillard

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Jean-François ANGERS, *Université de Montréal*

– Qian ZHOU, *Harvard School of Public Health*, Peter SONG, *University of Michigan*, Mary THOMPSON, *University of Waterloo*  

Information Matrices in Estimating Function Approach: Test for Model Misspecification and Model Selection • Matrices d'information dans l'approche des fonctions d'estimation : Test pour inadéquation du modèle et sélection de modèle [MS-230]

### 09:15 – 10:00 Session 11B

POU 1112

*Canadian Journal of Statistics* Award Address

Allocution des récipiendaires du prix de la *Revue canadienne de statistique*

Special Session • Conférence spéciale

Organizer and Chair • Responsable et présidente: Mary LESPERANCE, *University of Victoria*

– Piotr KOKOSZKA, *Utah State University*, Lajos HORVÁTH, *University of Utah*, Matthew REIMHERR, *University of Chicago*  

Two Sample Inference in Functional Linear Models • Inférence pour deux échantillons dans des modèles linéaires fonctionnels [MS-231]

### 10:00 – 10:30 Pause-café • Coffee Break

VAN Hall

### 10:30 – 12:00 Session 12A

VAN 1245A

Data and Models on the Move: Functional Data Analysis

Données et modèles en déplacement : Analyse de données fonctionnelles

Introductory Overview Lecture • Conférence de vue d'ensemble

Chair • Président: Duncan MURDOCH, *University of Western Ontario*

Organizer • Responsable: Christian LÉGER, *Université de Montréal*

– Jim RAMSAY, *McGill University*  

Data and Models on the Move: Functional Data Analysis • Données et modèles en mouvement : l'analyse de données fonctionnelles [MS-232]

### 10:30 – 12:00 Session 12B

VAN 2289A

Spatial Statistics and the Environment

Statistique spatiale et l'environnement

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Organizer and Chair • Responsable et président: Patrick BROWN, *Cancer Care Ontario*

10:30 Olaf BERKE, *University of Guelph*  

Geographic Epidemiology: Finding Spatial Patterns in Regional Count Data and the Effect of Diagnostic Misclassification Bias • Épidémiologie géographique : chercher des structures spatiales dans des données régionales de dénombrement et l'effet d'un biais de mauvaise classification diagnostique [MS-233]

11:15 Raquel MENEZES, *University of Minho, Portugal*, Peter DIGGLE, *Lancaster University & Johns Hopkins University School of Public Health*, Ting-Li SU, *Lancaster University*  

Geostatistics, Preferential Sampling and Environmental Monitoring • Géostatistique, échantillonnage préférentiel et surveillance environnementale [MS-234]

### 10:30 – 12:00 Session 12C

VAN 2291A

Statistics in Conservation

Statistique en préservation

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Committee on Women in Statistics and the Caucus for Women*

Session parrainée par le *Comité de promotion de la femme en statistique et le Caucus des femmes*

Organizer and Chair • Responsable et présidente: Joanna MILLS FLEMMING, *Dalhousie University*

10:30 Eva CANTONI, *University of Geneva, Switzerland*, Joanna MILLS FLEMMING, *Dalhousie University*, Alan WELSH, *The Australian National University, Canberra*  

A Random Effects Hurdle Model for Shark Counts • Un modèle de type « haie » avec effets aléatoires pour le comptage de requins [MS-235]

11:00 Julie HORROCKS, *University of Guelph*, David HAMILTON & Hal WHITEHEAD, *Dalhousie University*  

Estimating Abundance of Animals from Correlated Binary Data • Estimation de l'abondance d'animaux à partir de données binaires corrélées [MS-236]

11:30 Laurie AINSWORTH & Charmaine DEAN, *Simon Fraser University*  

Zero-inflated Spatial Models for Environmental Data: Clustered Zeros versus Spatial Outliers • Modèles spatiaux à surreprésentation de zéros pour données environnementales : zéros en grappe contre valeurs spatialement aberrantes [MS-237]

### 10:30 – 12:00 Session 12D

VAN 2809A

Dependence Modelling in Risk Theory

Modèles de dépendance en théorie du risque

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Business and Industrial Statistics Section*

Session parrainée par le *Groupe de statistique industrielle et de gestion*

Chair • Président: Steve DREKIC, *University of Waterloo*

Organizer • Responsable: Étienne MARCEAU, *Université Laval*

10:30 Hélène COSSETTE & Étienne MARCEAU, *Université Laval*, Véronique MAUME-DESCHAMPS, *Université Lyon 1 (France)*  

Discrete-time Risk Models Based on Time Series for Count Random Variables • Modèles de risque

en temps discret basés sur les séries temporelles pour des variables aléatoires de dénombrement [MS-238]

11:00 Étienne MARCEAU, Hélène COSSETTE & Mélina MAILHOT, *Université Laval*  

TVaR-based Capital Allocation with Dependence • Allocation de capital basé sur la mesure TVaR en supposant la dépendance [MS-239]

11:30 Johanna NEŠLEHOVÁ, *McGill University*  

Archimedean Dependence and Beyond • Au-delà de la dépendance archimédienne [MS-240]

## 10:30 – 12:00 Session 12E

VAN 2811A

Genomics and Association Studies

Génomique et études d'association

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Chair • Président: Angelo CANTY, *McMaster University*

10:30 Naomi ALTMAN, Qingyu WANG, Vishesh KARWA & Aleksandra SLAVKOVIC, *Penn State University*  

Resolving Isoform Expression using RNA-seq data • Résoudre l'expression isoforme au moyen de données d'ARN-seq [MS-241]

10:45 Ali GHODSI & Hadi ZARKOUB, *University of Waterloo*  

Gene Selection for Microarray Data via Hilbert–Schmidt Independence Criterion • Sélection de gènes pour données de micropuces au moyen du critère d'indépendance d'Hilbert-Schmidt [MS-242]

11:00 Kyle RAYMOND & Thomas SCHEIKE, *University of Copenhagen*  

Assessing the Effects of Haplotype Match for Competing Risks with Unphased Genotype Data • Évaluer les effets de la compatibilité haplotype sur les risques compétitifs avec des données génotypiques déphasées [MS-243]

11:15 Juan XIONG & Wenqing HE, *University of Western Ontario*  

Predicting of Survival Time by Combining Mismeasured Gene Expression Data from Different Platforms • Prévoir le temps de survie en combinant des données d'expression génétique mal mesurées provenant de différentes plateformes [MS-244]

11:30 Alexandre BUREAU, *Université Laval*, Jordie CROTEAU, *Centre de recherche Université Laval – Robert-Giffard*, Aurélie LABBE, *McGill University*, Chantal MÉRETTE, *Université Laval*  

Latent Class Model with Familial Dependence to Address Heterogeneity in Complex Diseases: Adapting the Approach to Family-based Association Studies • Modèle de classes latentes avec dépendance familiale pour surmonter l'hétérogénéité des maladies complexes : adaptation de l'approche aux études d'association dans les familles [MS-245]

11:45 Marie-Pierre SYLVESTRE, *Samuel Lunenfeld Research Institute*, Angelo J. CANTY, *McMaster University*, Daryl WAGGOTT, *Samuel Lunenfeld Research Institute*, John D. BRUNZELL, *University of Washington*, Andrew P. BORIGHT, Andrew D. PATERSON & Shelley BULL, *University of Toronto*  

Semiparametric Mixed Models for Multivariate Profile Traits in Genome-wide Association Studies of Complex Diseases: An Application to Lipid Density Gradient Profiles • Modèles mixtes semi-paramétriques pour l'analyse de profils multivariés dans les études d'association pangénomiques de maladies complexes : une application aux profils lipidiques [MS-246]

**10:30 – 11:45 Session 12F****VAN 1203**

## Probability

## Probabilité

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Probability Section*Session parrainée par le *Groupe de probabilité*Chair • Président: Jean-François RENAUD, *University of Waterloo*10:30 Djilali AIT Aoudia, *Université de Montréal*, Éric Marchand, *Université de Sherbrooke*  

On the Number of Runs for Bernoulli Arrays • La distribution du nombre de suites de succès de longueur deux associées à un tableau de variables aléatoires Bernoulli [MS-247]

10:45 Luai AL Labadi, *University of Ottawa*, Mahmoud Zarepour, *Associate Professor*  

Exact and Approximate Expressions of the Kolmogorov and Lévy Distances between a Dirichlet Process and Another Continuous Probability Measure with some Applications • Expressions exactes et approximatives des distances de Kolmogorov et de Lévy entre un processus de Dirichlet et une autre mesure de probabilité continue, avec quelques applications [MS-248]

11:00 Amir Kermany & Sabin Lessard, *Université de Montréal*  

Fixation Probability of a Gamete under Weak Selection and the Effect of Recombination Modifiers on the Probability of Fixation • Probabilité de fixation d'un gamète sous sélection faible et effet de modificateurs du taux de recombinaison [MS-249]

11:15 Jeffrey Picka, *University of New Brunswick*  

Statistical Assessment of Models for Powder Flow • Évaluation statistique de modèles d'écoulement des poudres [MS-250]

11:30 Djilali AIT Aoudia & François Perron, *Université de Montréal*  

A Time-dependent Version of Pólya Urn • Une version dépendante du temps de l'urne de Pólya [MS-251]

**10:30 – 11:45 Session 12G****VAN 1211**

## Estimation in Survey Sampling

## Estimation en échantillonnage

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Survey Methods Section*Session parrainée par le *Groupe des méthodes d'enquête*Chair • Président: Cindy Ubartas, *Statistique Canada*10:30 Catalin Dochitoiu & Stuart Pursey, *Statistics Canada*  

Calendarization Methods for Reference Period Adjustments with Applications to Monthly Surveys • Méthodes de calendarisation pour ajustements à la période de référence avec applications aux enquêtes mensuelles [MS-252]

10:45 Valery Dongmo Jiongo, David Haziza & Pierre Duchesne, *Université de Montréal*  

Robust Small Area Estimation • Estimation robuste pour petites régions [MS-253]

11:00 M. Hidiroglou & François Verret, *Statistics Canada*, J. N. K. Rao, *Carleton University*  

Small Area Estimation under Informative Sampling • Estimation pour petites régions sous un plan de sondage informatif [MS-254]

11:15 Chen XU, *University of British Columbia*  

Penalized Likelihood Methods for Variable Selection in Analysis of Survey Data • Méthodes de sélection de variables basées sur la vraisemblance pénalisée en analyse de données d'enquête [MS-255]

11:30 Dagmar MARIACA HAJDUCEK & Jerald F. LAWLESS, *University of Waterloo*  

Duration Analysis of Survey Data using Inverse Probability of Censoring Weights (IPCW) • Analyse de durée pour données d'enquête au moyen de la pondération par probabilité inverse de censure [MS-256]

## 10:30 – 12:00 Session 12H

**VAN 1871**

Times Series and Climate

Séries chronologiques et le climat

Contributed Paper Session • Séance de communications libres

Chair • Président: Pierre DUCHESNE, *Université de Montréal*

10:30 Azadeh MOGHADERI, Patrick FLANDRIN & Pierre BORGNAT, *Ecole Normale Supérieure de Lyon, France*  

Trend Estimation by Means of the Empirical Mode Decomposition • Estimation de tendances par les moyennes de la décomposition des modes empiriques [MS-257]

10:45 Taoufik BOUEZMARNI, *McGill University*, Abderrahim TAAMOUTI, *Universidad Carlos III de Madrid*  

Nonparametric Estimation for Causality Measures • Estimation non paramétrique des mesures de causalité [MS-258]

11:00 Karim RAHIM & David THOMSON, *Queen's University*  

Multitaper Spectral Estimation Package in R • Package R d'estimation spectrale à fenêtres de pondération multiples [MS-259]

11:15 Azadeh MOGHADERI, *Ecole Normale Supérieure de Lyon*, Glen TAKAHARA & David THOMSON, *Queen's University*  

A Graphical Test for Stationary and Uniformly Modulated Processes • Un test graphique pour processus stationnaires et uniformément modulés [MS-260]

11:30 Lilia RAMIREZ RAMIREZ, Ziming HUANG & Yulia GEL, *University of Waterloo*  

Probabilistic Weather Forecasting using the Geostatistical Output Perturbation Method on Hierarchical Bayesian Space-time Models • Prévisions météorologiques probabilistes utilisant la méthode de perturbation géostatistique de modèles espace-temps bayésiens hiérarchiques [MS-261]

11:45 David THOMSON, *Queen's University*  

Edgeworth Expansions for Analyzing Frequency Modulation in Climate Time Series • Développements d'Edgeworth afin d'analyser la modulation de fréquences en séries chronologiques climato-logiques [MS-262]

## 12:00 – 13:20 Dîner • Lunch

**13:20 – 14:20 Session 13A****POU 1112**

The Lasso

Le lasso

Special Session • Conférence spéciale

Chair • Président: Hugh CHIPMAN, *Acadia University*Organizer • Responsable: Christian LÉGER, *Université de Montréal*– Robert TIBSHIRANI, *Stanford University*  

The Lasso: Some Novel Algorithms and Applications • Le lasso : quelques algorithmes et applications originaux [MS-263]

**14:30 – 16:00 Session 14A****VAN 1245A**

Response-biased Sampling and Missing Covariates

Échantillonnage biaisé par les réponses et covariables manquantes

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Biostatistics Section*Session parrainée par le *Groupe de biostatistique*Organizer and Chair • Responsable et président: Don McLEISH, *University of Waterloo*14:30 John NEUHAUS, *University of California, San Francisco*, Alastair SCOTT & Chris WILD, *University of Auckland*, Charles McCULLOCH, *University of California, San Francisco*  

A General Method to Analyze Data from Outcome Dependent Cluster Sampling Designs • Une méthode générale pour analyser des données de plans d'échantillonnage en grappes dépendantes de la réponse [MS-264]

15:00 Chris WILD, Alan LEE & Alastair SCOTT, *University of Auckland*  

Efficient Estimation in Multiphase Case-control Studies • Estimation efficace en études cas-témoins à phases multiples [MS-265]

15:30 Richard COOK, *University of Waterloo*  

Estimating Treatment Effects for Episodic Interventions Based on Response-dependent Observation • Estimation des effets de traitement pour interventions épisodiques à partir d'observations dépendantes de la réponse [MS-266]

**14:30 – 16:00 Session 14B****VAN 2289A**

Computer Experiments

Expérimentation numérique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Pritam RANJAN, *Acadia University*14:30 Thomas SANTNER, *Ohio State University*  

Simultaneous Determination of Tuning and Calibration Parameters for Computer Experiments • Détermination simultanée de paramètres de réglage et de calibrage pour des expériences informatiques [MS-267]

- 14:52 Hugh CHIPMAN, Weiwei WANG & Pritam RANJAN, *Acadia University*    
 Sequential Optimization of a Computer Model: What's so Special about Gaussian Processes Anyway? • Optimisation séquentielle d'un modèle informatique : pourquoi les processus gaussiens? [MS-268]
- 15:15 Earl LAWRENCE, Dave HIGDON, Katrin HEITMANN & Salman HABIB, *Los Alamos National Laboratory*, Martin WHITE, *University of California, Berkeley*, Christian WAGNER, *Astrophysikalisches Institut Potsdam*    
 Emulating the Nonlinear Matter Power Spectrum for the Universe • Émuler le spectre de puissance non linéaire de la matière pour l'univers [MS-269]
- 15:37 Leanna HOUSE, *Virginia Tech*    
 Reinforcing Reification with Application to a Rainfall-runoff Computer Model • Renforcement de la réification avec applications à un modèle informatique de précipitation-débit [MS-270]

**14:30 – 16:00 Session 14C****VAN 2291A**

## Network Sampling

## Échantillonnage par réseau

Invited Paper Session • Présentations sur invitation

Session sponsored by the *Survey Methods Section*Session parrainée par le *Groupe des méthodes d'enquête*Organizer and Chair • Responsable et président: Steven THOMPSON, *Simon Fraser University*

- 14:30 Pierre LAVALLÉE, *Statistique Canada*, Louis-Paul RIVEST, *Université Laval*    
 Capture-recapture Sampling and Indirect Sampling • L'échantillonnage capture-recapture et l'échantillonnage indirect [MS-271]
- 15:00 Art POON, *BC Centre for Excellence in HIV/AIDS*, Sergei L. KOSAKOVSKY POND, *University of California at San Diego*, Simon D. W. FROST, *University of Cambridge*    
 Estimating Mixing Between Subpopulations using Respondent Driven Sampling • Estimation d'homophilie entre sous-populations avec échantillonnage dirigé par les répondants [MS-272]
- 15:30 Mike KWANISAI, *National Opinion Research Center*, Steven K. THOMPSON, *Simon Fraser University*    
 Bayes Estimation with Adaptive Web Sampling • Estimation bayésienne sous échantillonnage adaptatif en toile [MS-273]

**14:30 – 16:00 Session 14D****VAN 2809A**

## What can we learn from each other about teaching statistics?

## Que peut-on apprendre les uns des autres concernant l'enseignement de la statistique ?

Invited Paper Session • Présentations sur invitation

Chair • Présidente: Alison GIBBS, *University of Toronto*Organizers • Responsables: Alison GIBBS, *University of Toronto* & Bethany WHITE, *University of Western Ontario*

- 14:30 Robert DELMAS, Joan GARFIELD & Andrew ZIEFFLER, *University of Minnesota*    
 A Radical Approach to Teaching Introductory Statistics: The CATALST Project • Une approche radicale à l'enseignement de la statistique élémentaire : le projet CATALST [MS-274]
- 15:00 Dennis PEARL, *Ohio State University*    
 CAUSE: Building Partnerships to Advance College Level Statistics Education • CAUSE : bâtir des partenariats pour l'évolution de l'éducation statistique universitaire [MS-275]

15:30 Bethany WHITE, *University of Western Ontario*  

The Canadian Statistics Educator Network • Le Réseau canadien des éducateurs en statistique  
[MS-276]

### 14:30 – 16:00 Session 14E

**VAN 2811A**

Statistical Modelling in Actuarial Sciences

Modèles statistiques en actuariat

Invited Paper Session • Présentations sur invitation

Chair • Président: Manuel MORALES, *Université de Montréal*

Organizer • Responsable: José GARRIDO, *Concordia University*

14:30 Jean-Philippe BOUCHER, *Université du Québec à Montréal*, Montserrat GUILLÉN, *Université de Barcelone*  

Semiparametric Approaches to Model the Heterogeneity of Count Data • Approches semi-paramétriques dans la modélisation de l'hétérogénéité du nombre de réclamations [MS-277]

15:00 Christian GENEST, *Université Laval*, Yvan KOJADINOVIC, *University of Auckland*, Johanna NEŠLEHOVÁ, *McGill University*, Jun YAN, *University of Connecticut at Storrs*  

Goodness-of-fit Tests for Bivariate Extreme-value Copulas • Tests d'adéquation pour des copules de valeurs extrêmes bivariées [MS-278]

15:30 José GARRIDO, *Concordia University*, Ana DEBÓN, *U. Politécnica de Valencia, Spain*  

Mortality Improvement; an Actuarial Perspective • Amélioration de la mortalité ; une perspective actuarielle [MS-279]

### 14:30 – 15:45 Session 14F

**VAN 1203**

Mixed Effects

Effets mixtes

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Chair • Président: Alexandre BUREAU, *Université Laval*

14:30 Wei LIU, *York University*, Lang WU, *University of British Columbia*  

A Semiparametric Nonlinear Mixed-effects Model with Non-ignorable Missing Data and Measurement Errors for HIV Viral Data • Un modèle semi-paramétrique non linéaire à effets mixtes pour données virales VIH à erreurs de mesures et données manquantes non ignorables [MS-280]

14:45 Yan Yan WU, Georges MONETTE & Augustine WONG, *York University*  

An Accurate Method for the Test of Variance Component of Random Effects in Linear Mixed Models • Une méthode précise pour tester la composante de la variance des effets aléatoires de modèles linéaires mixtes [MS-281]

15:00 Ahlem HAJJEM, François BELLAVANCE & Denis LAROCQUE, *HEC Montréal*  

Mixed Effects Trees for Clustered Data • Arbres à effets mixtes pour données en grappe [MS-282]

15:15 J LOREDO-OSTI, *Memorial University*  

Bootstrap Test for Mixed Linear Models • Test bootstrap pour modèles linéaires mixtes [MS-283]

15:30 Mahmoud TORABI & Rhonda ROSYCHUK, *University of Alberta*  

Spatial Modeling of Disease Mapping of Rates Based on Putative Sources • Modélisation spatiale de taux d'infection à partir de sources putatives [MS-284]

## 14:30 – 16:00 Session 14G

VAN 1211

Data Analysis in Biostatistics

Analyse de données en biostatistique

Contributed Paper Session • Séance de communications libres

Session sponsored by the *Biostatistics Section*

Session parrainée par le *Groupe de biostatistique*

Chair • Présidente: Jeanette O'HARA HINES, *University of Waterloo*

14:30 James HANLEY, *McGill University*  

Underestimation of Mortality Reductions in Cancer Screening Studies: The ERSPC as a Case Study • Sous-estimation des réductions de la mortalité dans les études de dépistage du cancer : le ERSPC comme exemple [MS-285]

14:45 Élodie SAMSON, *Centre de recherche en cancérologie de l'Université Laval, Québec*, Wei XU, *Biostatistics Department, Princess Margaret Hospital, Toronto*, Isabelle BAIRATI & François MEYER, *Centre de recherche en cancérologie de l'Université Laval, Québec*, Azad ABUL KALAM, *Applied Molecular Oncology, Princess Margaret Hospital, Toronto*, Geoffrey LIU, *Medical Biophysics Department, Ontario Cancer Institute/Princess Margaret Hospital, Toronto*  

Stratified Cox Proportional Hazards Models to Identify Genetic Prognostic Markers in Head and Neck Cancer Based on Different Genetic Models • Ajustement de modèles de Cox stratifiés pour identifier des marqueurs génétiques pronostiques chez les patients atteints d'un cancer otorhinolaryngologique selon différents modèles génétiques [MS-286]

15:00 Dena SCHANZER, *Public Health Agency of Canada*, Jason GILMORE, *Statistics Canada*, Hui ZHENG, *Public Health Agency of Canada*  

Statistical Estimates of Absenteeism Attributable to Influenza from the Labour Force Survey • Estimations statistiques de l'absentéisme attribuable à la grippe basée sur l'Enquête sur la population active [MS-287]

15:15 David TOLUSSO & Sheilah HOGG-JOHNSON, *Institute for Work & Health*  

The Use of Mixture Models to Examine Workplace Disability in Ontario • L'utilisation de modèles de mélanges pour examiner l'invalidité au travail en Ontario [MS-288]

15:30 Delbert STEWART & David STEWART, *Brogan Inc.*  

MDCConnect Physician Perspectives and Segmentation Research Study • Perspectives sur MDCConnect et étude de segmentation [MS-289]

15:45 Muhammad ISLAM, *Grant MacEwan University*  

Analysis of Incomplete and Irregularly Spaced Blood Sugar Response Data from a Longitudinal Study • Analyse de données de mesures glycémiques incomplètes et irrégulièrement espacées dans une étude longitudinale [MS-290]

**14:30 – 15:45 Session 14H****VAN 1871**

## Regression and Data Mining

## Régression et forage de données

Contributed Paper Session • Séance de communications libres

Chair • Président: Alexandre LEBLANC, *University of Manitoba*14:30 François BELLAVANCE, Abdessamad DINE & Denis LAROCQUE, *HEC Montréal*  

Prediction of Mixed Outcomes using a Weighted Combination of Univariate and Multivariate Random Forests • Prédiction de variables réponses multivariées mixtes avec une combinaison de forêts aléatoires univariées et multivariées [MS-291]

14:45 Shabnam CHITSAZ & S. Ejaz AHMED, *University of Windsor*  

Improving the Performance of Intercept Parameter Estimation in the Multivariate Reduce-rank Regression Model • Améliorer la performance de l'estimation du paramètre d'ordonnée à l'origine dans le modèle de régression de rang réduit multivarié [MS-292]

15:00 Abbas KHALILI, *McGill University*  

Feature Selection in Mixture Regression Models with Diverging Number of Parameters • Sélection de caractéristiques en mélanges de modèles de régression à nombres de paramètres divergents [MS-293]

15:15 Xuewen LU, *University of Calgary*  

Direct Estimation for Density-weighted Average Derivatives when Responses are Missing at Random • Estimation directe de dérivées moyennes pondérées par les densités en présence de réponses aléatoirement manquantes [MS-294]

15:30 Xu (Sunny) WANG, *St. Francis Xavier University*, Hugh CHIPMAN, *Acadia University*  

Mining Drug Data via Mixtures and Related Computational Issues • Forage de données de médicaments au moyen de mélanges et problèmes informatiques liés [MS-295]

## Abstracts • Résumés

**Session 00A      Dimanche 23 mai • Sunday, May 23, 16:15 – 17:45      VAN 1245A**

**Panel Discussion on the Role, Benefits, and Importance of a National Statistics Institute  
Panel sur le rôle, les bénéfices et l'importance d'un Institut national en statistique**

This session aims at bringing the statistical sciences community together to discuss the construction of a National Statistical Sciences Institute. Our pilot institute, the National Institute for Complex Data Structure, accomplished a number of outstanding things, opening many doors for interdisciplinary research with statistical leadership. Many members of our community support the idea of a national institute and there is a danger that the goal may fade away if nothing is done.

This panel discussion will consider the importance and support of statistical research for Canada, and the need for an institute will be reassessed. We will consider what challenges are faced by statistical researchers in Canada; how statistics advances Canadian science and technology and public policy; how Canadian statistical science impacts the global statistical community; our linkages with other Canadian scientific communities, including mathematics and operations research; what sorts of structures and resources are required to advance Canadian statistical research in a substantial way; what would be the impact of an inability to procure such resources.

Broad input is encouraged at this discussion; there will be ample time for engagement by the community after the panellists provide their viewpoints on these issues. Both large-scale visionary remarks as well as suggestions for smaller ways we can support each other and the pivotal role of statistical science in Canada are welcome.

Cette séance vise à rassembler la communauté scientifique statistique afin de discuter de la fondation d'un Institut national des sciences statistiques. Notre institut pilote, l'Institut national sur les structures de données complexes, a accompli plusieurs choses remarquables et a ouvert plusieurs portes pour la recherche interdisciplinaire avec une direction statistique. Plusieurs membres de notre communauté soutiennent l'idée d'un institut national et il y a un danger que ce but nous échappe si rien n'est fait.

Ce débat considérera l'importance et le soutien de la recherche statistique pour le Canada, et le besoin d'un institut sera réévalué. Nous considérerons les défis auxquels font face les chercheurs en statistique au Canada; comment la statistique fait progresser la science, la technologie et la politique publique canadiennes; comment la science statistique canadienne influence la communauté statistique internationale; nos liens aux autres communautés scientifiques canadiennes, incluant les communautés mathématiques et de recherche opérationnelle; quelles sortes de structures et de ressources sont nécessaires afin de faire progresser significativement la recherche statistique canadienne; quel serait l'impact d'une incapacité d'obtenir ces ressources.

Des contributions générales à cette discussion sont encouragées; il y aura amplement de temps pour entendre la communauté après que les panélistes aient donné leur point de vue sur ces questions. Les remarques visionnaires sont autant le bienvenue que les suggestions plus modestes sur les façons de se soutenir entre nous et de soutenir le rôle central de la science statistique au Canada.

**Session 01A      Lundi 24 mai • Monday, May 24, 08:15 – 08:30      PPR**

**Inaugural Session  
Séance inaugurale**

**Session 01B                      Lundi 24 mai • Monday, May 24, 08:30 – 09:45                      PPR**

SSC Presidential Invited Address  
 Allocution de l'invité du président de la SSC

[MS-1]

**Long Term Skewness and Systemic Risk**  
**Asymétrie à long terme et risque systémique**

Robert ENGLE, *NYU Stern School of Business*

This talk shows how the asymmetric volatility models widely used in financial econometrics imply long term negative skewness in aggregate market returns which can be interpreted as systemic risk much like what we have recently observed. The economic underpinnings of this asymmetry are developed and portfolio implications are suggested based on hedge portfolios.

Cet exposé montre comment les modèles à volatilité asymétrique souvent utilisés en économétrie financière impliquent une asymétrie négative à long terme dans les rendements agrégés des marchés pouvant être interprétée comme un risque systémique semblable à ce que nous avons récemment observé. Les fondements économiques de cette asymétrie sont développés et des conséquences pour les portefeuilles sont suggérées à partir de portefeuilles de couvertures.

**Session 02A                      Lundi 24 mai • Monday, May 24, 10:25 – 11:55                      VAN 1245A**

Introduction to some Topics in Statistical Data Mining  
 Introduction à quelques sujets en forage statistique de données

[MS-2]

**Topics in Modern Statistical Learning and Data Mining**  
**Sujets en apprentissage statistique et forage de données modernes**

Robert TIBSHIRANI, *Stanford University*

In this talk I will give an introduction to modern supervised learning techniques, including linear models, generalized additive models, neural networks, classification trees, random forests, boosting, support-vector machines and kernel methods. The main reference is the 2nd edition of the text “The Elements of Statistical Learning” (2009) by Hastie, Tibshirani and Friedman.

Dans cet exposé, je vais donner une introduction aux techniques modernes d'apprentissage supervisé, incluant les modèles linéaires, les modèles additifs généralisés, les réseaux de neurones, les arbres de classification, les forêts aléatoires, le boosting, les machines à vecteurs de support, et les méthodes à noyaux. La principale référence est la 2e édition du livre *The Elements of Statistical Learning* (2009) de Hastie, Tibshirani et Friedman.

**Session 02B      Lundi 24 mai • Monday, May 24, 10:25 – 11:55      VAN 2289A**  
**Probability and Finance**  
**Probabilité et finance**

[MS-3]

**Random Walking on a Comb**  
**Marche aléatoire alvéolée**

Miklós CsÖRGÖ, *Carleton University*

Consider a simple symmetric random walk on the 2-dimensional comb lattice  $\mathbf{C}^2$  that is obtained from  $\mathbf{Z}^2$  by removing all horizontal edges off the  $x$ -axis. A strong approximation result will be presented for this random walk for the sake of studying its path behaviour, such as joint weak convergence and Strassen type LIL of its two components, as well as their marginal Hirsch type behaviour. Time permitting, the local time process of this random walk will also be studied via strong approximations. This talk is based on my collaboration with Endre Csaki, Antonia Foldes and Pal Revesz.

Considérons une marche aléatoire simple symétrique sur le réseau alvéolé  $\mathbf{C}^2$  obtenu à partir de  $\mathbf{Z}^2$  en enlevant toutes les arrêtes horizontales de l'axes des  $x$ . Une approximation pour cette marche aléatoire sera présentée afin d'étudier son comportement, notamment la faible convergence conjointe et la LLI de type Strassen de ses deux composantes, ainsi que leur comportement marginal de type Hirsch. Si le temps le permet, le processus de temps local de cette marche aléatoire sera aussi étudié au moyen d'approximations fortes. Cet exposé est basé sur ma collaboration avec Endre Csaki, Antonia Foldes et Pal Revesz.

[MS-4]

**Optimal Hedging in Discrete and Continuous Time**  
**Réplication optimale en temps discret et en temps continu**

Bruno RÉMILLARD, *HEC-Montréal*, Sylvain RUBENTHALER, *Université de Nice-Sophia Antipolis*

In this talk we give the optimal solution of the hedging problem in discrete time by minimizing the mean square hedging error, when the underlying assets are multidimensional. We also find explicit expressions for the optimal hedging problem in continuous time when the underlying assets are modeled by a regime-switching geometric Lévy process. It is also shown that the continuous time solution can be approximated by discrete time Hidden Markov models processes.

Dans cet exposé, nous donnons la solution optimale du problème du portefeuille de réplication en temps discret lorsque l'on désire minimiser l'erreur quadratique moyenne entre la valeur finale du portefeuille et la valeur intrinsèque d'une option, lorsqu'il y a plusieurs actifs sous-jacents. Nous trouvons aussi la solution en temps continu de ce problème de réplication lorsque les actifs sous-jacents sont des processus de Lévy avec changement de régimes. Nous montrons aussi que la solution optimale en temps continu peut être approchée par celle en temps discret lorsque les actifs sont modélisés par des chaînes de Markov avec états cachés.

[MS-5]

**Stochastic Modelling and Pricing of Energy Markets' Contracts with Local Stochastic Delayed and Jumped Volatilities**

**Modélisation stochastique et fixation de prix de contrats dans les marchés de l'énergie avec volatilités stochastiques locales de sauts et de délais**

Anatoliy SWISHCHUK, *University of Calgary*

In this talk we study stochastic modelling and pricing of electricity, gas and temperature mar-

Dans cet exposé nous étudions la modélisation stochastique et la fixation des prix de l'électricité, du gaz et de contrats

kets' contracts with delay and jumps, modelled by independent increments processes. The spot price models are based on a sum of non-Gaussian Ornstein–Uhlenbeck processes (including geometric and arithmetic models), describing the long- and short-term fluctuations of the spot dynamics. The models include jumps not only in the spot dynamics but also in the stochastic volatility to describe typical features like spikes of energy spot prices and jumps in the volatility. The basic products are spot, futures and forward contracts, swaps and options written on these.

sur la température avec des sauts et délais, modélisés par des processus incrémentaux indépendants. Les modèles de prix au comptant sont basés sur une somme de processus d'Ornstein-Uhlenbeck non gaussiens (incluant des modèles géométriques et arithmétiques) décrivant les fluctuations à court et long terme de la dynamique au comptant. Les modèles comprennent des sauts non seulement dans la dynamique au comptant, mais aussi dans la volatilité stochastique décrivant des caractéristiques typiques comme les pics des prix d'énergie au comptant et les sauts de volatilité. Les produits de base sont les opérations au comptant, les contrats à terme, les swaps et les options sur ceux-ci.

## Session 02C      Lundi 24 mai • Monday, May 24, 10:25 – 11:55      VAN 2291A

### Statistics and Science Statistique et science

#### [MS-6]

#### **Viral Pandemics: Academics Filling the Gap**

#### **Les pandémies virales : le monde académique à la rescousse**

Michel CHRÉTIEN, *Institut de recherche de l'Hôpital d'Ottawa*

Infectious diseases kill 17 million people each year. Viral infections can cross the globe in days. The recent A/H1N1 pandemic demonstrates clearly the danger.

Antivirals are an absolute necessity. In this field, pharma have little financial incentives. That has produced the “10/90 gap”, i.e., spending only 10% in research on diseases affecting 90% of the world population. Academics had to come to the rescue and bridge that gap. They have invented a new drug development model that could decrease the cost by as much as 80% compared to the pharmaceutical industry.

Les maladies infectieuses font 17 millions de victimes par année. Des infections virales peuvent couvrir la planète en quelques jours. La récente pandémie A/H1N1 en est une démonstration éclatante.

Les antiviraux sont nécessaires. C'est un champ où les Pharma ont la « grippe ». En n'allouant que 10 % des crédits de recherche pour les maladies qui touchent 90 % de la population, elles ont provoqué le « vide 10/90 ». Pour combler ce vide, le monde académique a inventé un nouveau modèle dont les médicaments pourront être 80 % moins chers que ceux des boîtes pharmaceutiques.

#### [MS-7]

#### **Global Health Challenges and Opportunities in the 21st Century**

#### **Défis globaux et possibilités à exploiter en santé au 21e siècle**

Haile DEBAS, *University of California, San Francisco*

This presentation will discuss the complex global health challenges of the 21st century and the opportunities that may help to develop effective strategies to address them. The challenges (pandemics; microbial resistance; poverty, disparity and lack of access to health care; the impact of human migra-

Cette présentation discutera de défis globaux complexes en santé au 21<sup>e</sup> siècle et des possibilités à saisir pour aider à développer des stratégies efficaces pour les surmonter. Les défis (pandémies, résistance microbienne, pauvreté, inégalité et inaccessibilité des soins de santé, l'impact de la migration humaine et des changements climatiques, une crise

tion and climate change; a global workforce crisis, etc.) are enormous. However, significant opportunities exist to address them. These include the enthusiasm and passion of students and faculty, the generosity of the global community, the wide use of interdisciplinary approach to solve complex problems, and scientific and technological innovations.

globale de la main-d'oeuvre, etc.) sont énormes. Cependant, il existe des possibilités importantes pour les surmonter. Celles-ci incluent l'enthousiasme et la passion des étudiants et professeurs, la générosité de la communauté planétaire, l'utilisation répandue de l'approche interdisciplinaire pour résoudre des problèmes complexes, et l'innovation scientifique et technologique.

**[MS-8]**

**Why Journalists Can't Count, Statisticians Can't Read and whether it Matters**

**Pourquoi les journalistes ne peuvent pas compter, les statisticiens ne peuvent pas lire, et est-ce que ça importe ?**

Peter KAVANAGH, *Canadian Broadcasting Corporation*

This presentation will address the true Two Culture split of the 21st Century, the rise in the importance and near ubiquity of statistical measurements and the possibility that numerical representations of reality will displace other more traditional forms of narrative construction of human meaning while at the same time the media and journalists lose interest in and the ability to interpret numbers. In terms of general society this creates the near perfect storm conditions for confusion, consternation and controversy. A world relying on numeracy, informed by the innumerate is the description of a dystopia of alarming dimensions.

Cette présentation portera sur la réelle séparation des deux cultures au 21<sup>e</sup> siècle : d'une part, la montée en importance et la quasi-omniprésence de mesures statistiques et la possibilité que les représentations numériques de la réalité supplantent d'autres formes plus traditionnelles de constructions narratives de l'existence humaine et, d'autre part, une perte d'intérêt et de capacité de la part des médias et des journalistes pour l'interprétation de chiffres. Pour la société en général, ceci crée des conditions parfaites pour la confusion, la consternation et la controverse. Un monde s'appuyant sur les chiffres et informés par le achiffrés constitue une dystopie de dimension alarmante.

**Session 02D      Lundi 24 mai • Monday, May 24, 10:25 – 11:55      VAN 2809A**

**Innovative Tools in Teaching Statistics**

**Outils innovateurs pour l'enseignement de la statistique**

**[MS-9]**

**Games and Stories in the Teaching of Statistics**

**Jeux et histoires en enseignement de la statistique**

Paul COREY, *University of Toronto*

The teacher lucky enough to be with an exceptionally bright and enthusiastic group of students in the health and life sciences must appreciate that their first love is not statistics. In order to convey the essential concepts of statistics and show their relevance to data analysis they must use a minimum of formulae and embed the message in a collection of games and stories. The purpose of this talk is to share with my fellow teachers some stories and games that I have found to be very useful and to attract from the audience other examples of such gamesmanship.

L'enseignant ayant la chance d'encadrer un groupe exceptionnellement brillant et enthousiaste d'étudiants en sciences de la santé doit considérer le fait que leur passion n'est pas la statistique. Afin de transmettre les concepts essentiels de la statistique et de montrer leur pertinence en analyse de données, il faut donc utiliser un minimum de formules et emboîter les concepts dans une série d'histoires et de jeux. Le but de cet exposé est de partager avec mes collègues enseignants quelques jeux et histoires qui m'ont été très utiles et d'obtenir du public d'autres exemples de tels jeux.

**[MS-10]****Electronic Homework: Does Student Learning Improve if Questions Have More Detailed Steps?****Devoirs électroniques : l'apprentissage des étudiants est-il amélioré si les questions ont des étapes plus détaillées ?**David LOEWEN, *University of Manitoba*

Online homework systems can provide many pedagogical advantages for students. In multi-section introductory statistics courses at the University of Manitoba, the on-line homework delivery system WebAssign is used for all assignments completed outside of class. This study, conducted as a randomized block design experiment in the second course, investigates the effect of giving students the opportunity to answer intermediate steps in solving homework problems. The treatment group received the assignments with the intermediate steps. Each group was the treatment group on two of the four questions tested. Some benefits were detected, but not to the degree hoped.

Les systèmes de devoirs en ligne peuvent comporter plusieurs avantages pédagogiques pour les étudiants. Dans les cours d'introduction à la statistique enseignés dans plusieurs sections à l'Université du Manitoba, le système de devoirs en ligne WebAssign est utilisé pour tous les travaux effectués à l'extérieur de la classe. Cette étude, une expérience à blocs aléatoires, s'est servie du deuxième cours pour étudier l'effet de permettre aux étudiants de répondre aux étapes intermédiaires lors de la résolution de problèmes. Le groupe traitement reçoit les devoirs incluant les étapes intermédiaires. Chaque groupe d'étudiants était groupe traitement pour deux des quatre questions formant l'étude. Un effet bénéfique mais moins important qu'espéré a été détecté.

**[MS-11]****Strategies for Teaching an Enduring Knowledge of Statistics****Stratégies permettant d'inculquer une connaissance durable de la statistique**Larry WELDON, *Simon Fraser University*

The futility of teaching students to master calculations without conceptual understanding is widely acknowledged. However, there are incentives for both instructors and students to focus on calculations. In this paper strategies are suggested for encouraging students to gain a conceptual understanding of statistics: open book tests and exams, elimination of the traditional textbook, emphasis on verbalization and visualization, use of open-ended case studies, and an emphasis on process rather than product. It is argued that these strategies are likely to result in a more useful knowledge of statistics than more traditional approaches. Some evidence from the author's experience is provided.

On sait pertinemment combien il est futile d'enseigner aux étudiants à maîtriser un calcul sans en comprendre les concepts sous-jacents. Cependant, mettre l'accent sur les calculs comporte des avantages pour les professeurs et les étudiants. Nous suggérons diverses stratégies pour encourager les étudiants à développer une compréhension conceptuelle de la statistique : tests et examens à livre ouvert, élimination des manuels traditionnels, importance accrue de la verbalisation et de la visualisation, emploi d'études de cas ouvertes, et accent sur la démarche plutôt que le résultat. À la lumière de notre expérience, nous soutenons que ces stratégies mèneront à une connaissance de la statistique plus utile que les approches traditionnelles.

**Session 02E      Lundi 24 mai • Monday, May 24, 10:25 – 11:55      VAN 2811A**

Living in Canada: Developing a Household Panel Survey in Canada

Vivre au Canada : Développer une enquête par panel sur les ménages au Canada

**[MS-12]****A Summary of Content Evaluations for the Living in Canada Survey Pilot****Un résumé des évaluations de contenu de l'enquête Vivre au Canada**Andrew HEISZ, *Statistics Canada*

In 2006, a conference on longitudinal surveys hosted by Statistics Canada concluded that Canada lacks a longitudinal survey which collects information on multiple subjects such as family, human capital, labour health. Following this conference, funds were received to support a pilot survey for a new Canadian Household Panel Survey (CHPS-Pilot). Consultations on the design and content were held with academic and policy experts in 2007 and 2008, and a pilot survey was conducted in the fall of 2008. This paper highlights results from a series of evaluations of content from the CHPS-Pilot.

En 2006, un congrès sur les enquêtes longitudinales organisé par Statistique Canada a permis de conclure que le Canada n'a pas d'enquête longitudinale lui permettant de recueillir des renseignements sur de multiples sujets tels que la famille, le capital humain, la santé, et le travail. Suite à ce congrès, du financement a été accordé afin de soutenir l'enquête pilote pour la nouvelle Enquête par panel auprès des ménages canadiens (EPMC-Pilote). Des consultations concernant le plan et le contenu ont eu lieu avec des experts universitaires et politiques en 2007 et 2008, et une enquête pilote a été menée à l'automne 2008. Cet article met en évidence ces résultats d'une série d'évaluations de contenu de l'EPMC-Pilote.

**[MS-13]****An Evaluation of the Canadian Household Panel Survey Methodology****Une évaluation de la méthodologie de l'Enquête par panel auprès des ménages canadiens**Christian NADEAU, Marianne GOSSEN & Caroline CAUCHON, *Statistique Canada*

In the fall of 2008, Statistics Canada conducted a pilot survey in order to evaluate the content and methodology of a long duration household panel survey. The Canadian Household Panel Survey (CHPS), which collect data on multiple aspects of individuals' life, was conducted on about 2,000 households. It allowed testing the collection methodology mainly characterized by the use of non-proxy personal interviews of every adult member in the selected households. This paper provides a description of the main features of the CHPS methodology and focuses on the analysis of the impact of the collection methodology and effort on nonresponse.

Statistique Canada a mené, à l'automne 2008, une enquête pilote afin d'évaluer le contenu et la méthodologie d'une enquête par panel de longue durée auprès des ménages. L'Enquête par panel auprès des ménages canadiens (EPMC), qui vise à recueillir des données sur divers aspects de la vie des individus, a été menée auprès d'environ 2 000 ménages. Elle a notamment permis la mise à l'essai de la méthodologie de collecte dont une des principales caractéristiques consiste à n'utiliser que des interviews personnelles sans procuration auprès de chacun des membres adultes des ménages sélectionnés. Cet article présente une description des principales caractéristiques de la méthodologie de l'EPMC et s'intéresse plus particulièrement à l'analyse de l'impact de la méthodologie et de l'effort de collecte sur la non-réponse.

[MS-14]

**Longitudinal Data to Support Adult and Family Policy Development****Les données longitudinales et le développement de politiques sur les adultes et les familles**Raynald LÉTOURNEAU, Magalie BROCHU, Mélanie RAYMOND & Julie TOUSIGNANT, *Ressources humaines et Développement des compétences Canada*

Longitudinal data can enrich policy development with information on how individuals' pasts influence their current lives and how their situations can change over time. This presentation makes the case for the relevance of longitudinal data for policy development. It then explores longitudinal data needed to inform policy development in three areas: learning, labour market participation, and family life. The presentation concludes by examining some of the desirable features of the new longitudinal data, such as the capacity to reflect emerging needs and to offer a holistic perspective.

Les données longitudinales et le développement de politiques sur les adultes et les familles Les données longitudinales peuvent contribuer au développement de politiques en montrant comment le passé influence la vie actuelle de chaque individu et comment l'état de cette personne peut changer au fil des ans. Cette présentation illustre la pertinence des données longitudinales pour le développement de politiques. Elle discute ensuite des données longitudinales requises pour appuyer le développement de politiques dans trois domaines : apprentissage, participation au marché du travail et vie familiale. Cette présentation se termine par l'examen de caractéristiques clés des nouvelles données longitudinales, en particulier de refléter les besoins émergents et une perspective holistique.

**Session 02F****Lundi 24 mai • Monday, May 24, 10:25 – 11:55****VAN 1203****Randomized Trials, Meta-Analysis, and Data Analysis****Essais randomisés, méta-analyse et analyse de données**

[MS-15]

**Meta-Regression and Cluster Randomized Trials: Theory and Practice****Métarégression et essais randomisés par grappes : théorie et pratique**Michael ROTONDI, *University of Western Ontario*

Cluster randomized trials are increasingly popular in health research as investigators recognize their breadth of application. In light of their popularity, methods for the synthesis of their results become increasingly important. This presentation is focused on the meta-regression of cluster randomized trials alone, and in combination with their individually randomized counterparts, while emphasizing the required modifications to traditional techniques.

Issues such as the homogeneity of treatment effects and the exploration of sources of heterogeneity through meta-regression will be discussed. Applications include the investigation of the protective effect of vitamin A supplementation on reducing child mortality in the developing world.

Les essais randomisés par grappes gagnent en popularité en recherche en santé, car les chercheurs en reconnaissent leur vaste potentiel d'application. Les méthodes de synthèse des résultats de tels essais sont donc de plus en plus importantes. Cette présentation est centrée sur la méta-régression d'essais randomisés par grappes seuls et en combinaison avec leur contrepartie randomisée individuellement, et met l'emphase sur les modifications devant être apportées aux techniques traditionnelles.

Des questions telles l'homogénéité des effets de traitement et l'exploration de sources d'hétérogénéité par méta-régression seront abordées. Les applications incluent l'étude de l'effet préventif de suppléments de vitamine A en réduction de la mortalité infantile dans les pays en voie de développement.

**[MS-16]****Confidence Intervals for Common Effect Measures in Cluster Randomization Trials  
Intervalles de confiance pour mesures d'effets communs en essais randomisés par grappes**Julia TALEBAN & Guang Yong ZOU, *University of Western Ontario*

Randomization by intact social units are sometimes more convenient than the random allocation of individuals to interventions. Such designs are referred to as cluster randomization trials. However, statistical inferences are complicated when the unit of inference differs from that of randomization. Here we present confidence intervals for several effect measures in cluster randomization trials with continuous outcomes. The essence is to recover variance estimates from confidence limits for parameters that form the parameters of interest such as the lognormal mean and the standardized mean difference. We show that the resulting procedures perform better than existing procedures using a simulation study.

La randomisation par unités sociales intactes est souvent plus commode que l'allocation aléatoire d'individus entre différentes interventions. De tels plans sont dits essais randomisés par grappes. Cependant, les inférences statistiques sont plus complexes lorsque l'unité d'inférence diffère de l'unité de randomisation. Nous présentons ici des intervalles de confiance pour plusieurs mesures d'effets en essais randomisés par grappes avec réponses continues. Essentiellement, nous retrouvons les estimés de la variance à partir des bornes de confiance pour les paramètres d'intérêt telles la moyenne lognormale et la différence moyenne standardisée. Nous montrons que les procédures résultantes surpassent les procédures existantes à l'aide d'une étude de simulation.

**[MS-17]****Statistical Power and Effect Size Estimation in Intervention Research under Population Heterogeneity  
Puissance statistique et estimation de la taille de l'effet dans la recherche en intervention sous hétérogénéité de la population**Depeng JIANG, *St. Michael's Hospital, University of Toronto*

In this study, we focus on the influence of population heterogeneity on statistical power in design and evaluation of intervention programs. We conduct Monte Carlo simulations to demonstrate how statistical power changes if the data generated from a mixture of two subpopulations is misspecified as one-class model. The simulation study also shows how the statistical power under population heterogeneity depends on the quality of classification of participants to the subpopulations, which, in turn, is affected by what covariates are used and how they are associated with outcomes and subpopulation membership.

Dans cette étude, nous nous concentrons sur l'influence de l'hétérogénéité de la population sur la puissance statistique dans la conception et l'évaluation de programmes d'intervention. Nous effectuons des simulations de Monte-Carlo pour démontrer comment la puissance statistique change si les données générées à partir d'un mélange de deux sous-populations sont mal spécifiées comme faisant partie d'une seule classe. Les simulations montrent également comment la puissance statistique varie en présence d'hétérogénéité de la population et dépend de l'efficacité de la classification des participants aux sous-populations, qui, à son tour, est influencée par les covariables utilisées et comment ces dernières sont associées aux réponses et l'appartenance aux sous-populations.

**[MS-18]****Bayesian Models for Meta-Analysis of Randomized Control Trials: Do We Know What People are Thinking?  
Modèles bayésiens pour méta-analyse d'essais de contrôle randomisés : savons-nous ce que pensent les gens ?**

Russell STEELE, *McGill University & Jewish General Hospital*, Franco CARNEVALE, *Montreal Children's Hospital & McGill University*, Lisa CHAN, *McGill University & Jewish General Hospital*, Mary Ellen McDONALD, *Montreal Children's Hospital & McGill University*, Javad SHAHIDI, *McGill University*, Ian SHRIER, *Jewish General Hospital & McGill University*

Bayesian methods for meta-analysis have become quite common, particularly in the analysis of multi-centre data from medical randomized control trials (RCTs). However, there has been little research on the question of whether current methods (in particular, for model selection and prior specification) produce results that concur with the actual beliefs of people interpreting the data from published meta-analyses. In this talk, we will present the methods and preliminary results from a mixed methods study using both quantitative and qualitative methods that tries to resolve differences between subjective Bayesian posterior distributions and what people state as their beliefs.

Les méthodes bayésiennes de méta-analyse sont devenues plus répandues, surtout en analyse de données multicentres d'essais contrôles randomisés (ECR). Cependant, peu de travaux ont porté sur l'accord entre les résultats obtenus par les méthodes actuelles (en particulier, en sélection de modèles et spécification des a priori) et l'opinion actuelle des gens interprétant les données de méta-analyses publiées. Dans cet exposé, nous allons présenter les méthodes et résultats préliminaires d'une étude employant plusieurs méthodes quantitatives et qualitatives visant à résoudre les conflits entre les distributions a posteriori subjectives et ce que les gens affirment réellement penser.

**[MS-19]****Cost-efficient Variable Selection: A Case Study  
Sélection efficace de variables : une étude de cas**

Li Hua YUE, Duncan MURDOCH & Wenqing HE, *University of Western Ontario*

A study was conducted in Southwestern Ontario to investigate Assertive Community Treatment (ACT) for patients with severe mental illness (SMI). Our role in the ACT project is to assess what cost-efficient factors influence outcomes of clients with SMI receiving care from ACT. There are two types of cost of collecting the data. The first is monetary cost, and the second is the level of difficulty to get the data for a potential predictor. We analyze the complicated cost structure, apply the Branching LARS method we have developed previously to the variable selection process, and present the resulting cost-efficient models.

Une étude a été menée dans le sud-ouest de l'Ontario afin d'étudier le suivi intensif dans le milieu (SIM) pour les patients atteints de troubles mentaux graves (TMG). Notre rôle dans le projet SIM est d'évaluer quels facteurs rentables influencent la réponse de clients atteints de TMG qui reçoivent des soins SIM. Il y a deux types de coûts associés à la collecte des données. Le premier est un coût monétaire, et le second est le niveau de difficulté lié à l'obtention de données pour un prédicteur potentiel. Nous analysons la structure de coûts complexe, appliquons la méthode LARS à branches (que nous avons développée antérieurement) au processus de sélection de variables, et présentons le modèle rentable résultant.

**[MS-20]****Association Between Environmental Characteristics and the Occurrence of Small Gestational Aged Birthweight in Quebec****Association entre le milieu et le faible poids de naissance pour l'âge gestationnel au Québec**Nathalie SAVARD, *Université Laval et INSPQ*, Louis-Paul RIVEST & Patrick LEVALLOIS, *Université Laval*

We explore associations between environmental characteristics and small gestational aged (SGA) birthweight. The study is conducted at the Quebec's Community Local Services Centers (CLSC) level. The proportion of SGA birthweight for each CLSC is calculated with Quebec's birth data. CLSC environmental characteristics come from socioeconomic census and aggregated Statistics Canada survey data. Questions from the Canadian Community Health Survey concerning life habits are used. Factorial analysis enables us to group the environmental variables; regression modelling then relates environmental characteristics and SGA birthweight.

Cet exposé explore si certaines caractéristiques du milieu de vie sont associées au faible poids de naissance. L'étude est réalisée au niveau des CLSC du Québec; la proportion de faible poids pour chacun est calculée à partir du fichier des naissances. Le milieu est décrit à l'aide de données socioéconomiques de recensement et de données agrégées d'enquête de Statistique Canada. Les questions d'habitudes de vie des Enquêtes sur la santé dans les collectivités canadiennes sont utilisées. Des analyses factorielles permettent de regrouper les variables de milieu; des modèles de régression les mettent ensuite en relation avec les faibles poids.

**Session 02G****Lundi 24 mai • Monday, May 24, 10:25 – 11:55****VAN 1211****Case-Control, Cohort Studies, and Other Trials****Études cas-témoins, de cohortes et autres essais****[MS-21]****Nonparametric Lifetime Risk Estimation using Prevalent Cohort Survival Data****Estimation non-paramétrique du risque à vie à partir de données de survie provenant d'une cohorte prévalente**Marco CARONE, *Johns Hopkins Bloomberg School of Public Health*

Lifetime risk, the proportion of a population which will ever experience some disease of interest, is a common epidemiologic measure of disease risk. Incident cohort designs constitute the gold-standard in studying lifetime risk. However, due to material constraints, data are often collected from prevalent cohort studies whereby individuals are recruited cross-sectionally and followed forward in time. We derive a nonparametric estimator of lifetime risk accounting for loss to follow-up in the context of prevalent cohort studies, apply our methodology to data from the Canadian Study of Health and Aging, and describe the risk of dementia in the Canadian elderly population.

Le risque à vie, la proportion d'une population éventuellement atteinte d'une maladie donnée, est une mesure épidémiologique de risque courante. L'étude du risque à vie est idéalement faite à l'aide de cohortes incidentes. Toutefois, en raison de contraintes matérielles, des données sont souvent récoltées à partir de cohortes prévalentes, où des individus échantillonnés transversalement sont suivis. Nous présentons un estimateur non-paramétrique du risque à vie dans le contexte d'études de cohortes prévalentes incluant des patients perdus de vue, appliquons notre méthodologie aux données provenant de l'Étude canadienne sur la santé et le vieillissement, et décrivons le risque à vie de démence dans la population âgée du Canada.

**[MS-22]****Efficient Two-phase Selection of Patients for Biomarker Studies in Chronic Disease****Sélection efficace de patients à deux phases pour études à biomarqueurs de maladies chroniques**Michael McISAAC & Richard COOK, *University of Waterloo*

In cohort studies of chronic disease processes, budgetary constraints often make it impossible to process all stored serum samples. Instead, a subsample must be selected for processing to obtain biomarker data. Inexpensive baseline and follow-up information can be used to inform the selection process and optimize the efficiency of resultant estimators. We describe how efficiency varies according to the two-phase sampling scheme adopted. We show how these results can be used to improve estimators of the effect of biomarkers on outcomes. A study of the relation between biomarkers of inflammation and disease progression in psoriatic arthritis is used for illustration.

En études cohortes de processus de maladies chroniques, des contraintes budgétaires rendent impossible l'analyse de tous les échantillons de sérum recueillis. Un sous-échantillon doit plutôt être choisi pour analyse afin d'obtenir des données de biomarqueurs. Des informations peu coûteuses recueillies au début de l'étude et lors du suivi peuvent être utilisées afin d'informer le processus de sélection et d'optimiser l'efficacité des estimateurs résultants. Nous décrivons comment l'efficacité varie selon le plan d'échantillonnage en deux étapes choisi. Nous montrons comment ces résultats peuvent être utilisés afin d'améliorer les estimateurs des effets des biomarqueurs sur la réponse. Une étude sur la relation entre biomarqueurs d'inflammation et progression de rhumatismes psoriasiques est utilisée à des fins illustratives.

**[MS-23]****Estimating Effects by Combining Instrumental Variables with Case-control Designs****Estimation d'effets par le jumelage de variables instrumentales et de plans d'expérience cas-témoin**Russell SHINOHARA, Constantine FRANGAKIS, Elizabeth PLATZ & Konstantinos TSILIDIS, *Johns Hopkins University*

The instrumental variables framework is commonly used for the estimation of causal effects from cohort samples. As the use of Mendelian randomization studies is increasing and the cost of genotyping and gene expression data can be high, the analysis of data gathered from more cost-effective sampling designs is of prime interest. We show that the standard analysis does not appropriately estimate the causal effects of interest when the instrumental variables design is combined with the case-control design. We also propose a method for estimating the causal effects in such combined designs. We illustrate the method with a study in oncology.

Les variables instrumentales sont fréquemment utilisées afin d'estimer des effets causaux. Étant donné l'essor récent d'études de randomisation mendélienne et le coût prohibitif du génotypage et des données d'expression génétique, il est d'intérêt de considérer l'analyse de données provenant de plans d'expériences moins coûteux. Nous démontrons que l'utilisation naïve de variables instrumentales peut engendrer une estimation erronée des effets causaux lorsque les observations proviennent d'une étude cas-témoin jumelée à un plan de variables instrumentales. Nous présentons par la suite une méthode d'estimation appropriée des effets causaux dans de telles circonstances et illustrons notre méthode à l'aide de données oncologiques.

**[MS-24]****Comparison of Score, Likelihood ratio and Wald Tests for Response Adaptive Designs****Comparaison des tests du score, du rapport de vraisemblance et de Wald pour plans adaptatifs à la réponse**Yanqing Yi, *Memorial University of Newfoundland*, Xikui WANG, *University of Manitoba*

The score test statistic has been generalized to the data collected from response adaptive designs and a generalized score statistic which includes more

La statistique de test de score a été généralisée pour traiter des données provenant de plans adaptatifs à la réponse, et une statistique de score généralisée comportant plus

information on the adaptation of the treatment allocation is introduced. Simulation studies are conducted on the comparison of the powers of the Wald, the score, the generalized score and the likelihood ratio statistics. The overall power of the Wald statistic is better than the score, generalized score and likelihood ratio statistics. Simulation results also indicate that the power is sensitive to the types of response adaptive designs.

d'information sur l'adaptation de la répartition du traitement est présentée. Des études de simulation sont menées afin de comparer la puissance des statistiques de Wald, de score, de score généralisé et du rapport de vraisemblances. La puissance globale de la statistique de Wald est la plus élevée. Les résultats de simulation indiquent aussi que la puissance est sensible au type de plan.

#### [MS-25]

##### **Which design is more powerful? Ehrenfest Urn versus Biased Coin Quel plan est le plus puissant ? Urne d'Ehrenfest ou pièce biaisée**

Ying-Pin CHEN, *Lewis and Clark College*

In a sequentially randomized trial where prospective comparisons of two or more treatments are studied, the power of an inference procedure depends on the number of assignments of each treatment, and hence on the design of randomization that allocates treatments. We analyze the power functions of some standard inference procedures under both the biased coin design and the Ehrenfest urn design. Our work shows that the Ehrenfest urn design is uniformly more powerful than the biased coin design.

En essais randomisés séquentiellement où des comparaisons prospectives d'au moins deux traitements sont étudiées, la puissance d'une procédure inférentielle dépend du nombre de sujets par traitement, et donc du plan de randomisation. Nous analysons les fonctions de puissance de quelques méthodes d'inférence usuelles sous le plan de la pièce biaisée et le plan d'urne d'Ehrenfest. Nos travaux montrent que le plan d'urne d'Ehrenfest est uniformément plus puissant que le plan de la pièce biaisée.

#### [MS-26]

##### **Bayesian Analysis of a Matched Case-control Study with a Misclassified Exposure Variable Analyse bayésienne d'une étude cas-témoins appariée avec une variable d'exposition mal classée**

Juxin LIU, *University of Saskatchewan*, Paul GUSTAFSON, *University of British Columbia*, Nicola CHERRY & Igor BURSTYN, *University of Alberta*

We propose a Bayesian adjustment for the misclassification of a binary exposure variable in a matched case-control study. The method admits a priori knowledge about both the misclassification parameters and the exposure-disease association. The standard Dirichlet prior distribution for a multinomial model is extended to allow separation of prior assertions about the exposure-disease association from assertions about other parameters. The method is applied to a study of occupational risk factors for new-onset adult asthma.

Nous proposons un ajustement bayésien pour la classification erronée d'une variable d'exposition binaire dans une étude liée cas-témoins. La méthode tient compte de l'information a priori au sujet des paramètres de la classification erronée et de l'association exposition-maladie. La distribution a priori de Dirichlet, habituellement utilisée pour un modèle multinomial, est généralisée afin de permettre la séparation l'information *a priori* sur l'association exposition-maladie de l'information au sujet des autres paramètres. La méthode est appliquée à une étude sur les facteurs de risques occupationnels pour l'asthme chez les adultes.

**Session 02H**      **Lundi 24 mai • Monday, May 24, 10:25 – 11:55**      **VAN 1871**  
**Clustering, Soccer, and the Scientific Method**  
**Classification, soccer et la méthode scientifique**

[MS-27]

**Model-based Clustering via tCLUST**  
**Classification à base de modèles au moyen de tCLUST**

Jeffrey ANDREWS & Paul McNICOLAS, *University of Guelph*

Two families of mixture models derived from the multivariate t-distribution will be discussed and collectively referred to as tCLUST. Both families result from parameterizing the covariance structure, and then imposing constraints upon that structure. The first family is derived via an eigendecomposition of the covariance matrix (similar to the popular MCLUST family of models for the multivariate Gaussian distribution), while the second family results from imposing a factor analysis covariance structure. Parameter estimates are found through a variant of the expectation-maximization (EM) algorithm, and model selection is performed by the Bayesian information criterion (BIC) and integrated completed likelihood (ICL).

Deux familles de modèles de mélanges obtenues à partir de la distribution t multivariée dites tCLUST seront abordées. Les deux familles résultent d'une paramétrisation de la structure de covariance et de l'imposition de contraintes sur cette structure. La première famille est obtenue par une décomposition en valeurs propres de la matrice de covariance (similairement à la famille de modèles MCLUST pour la distribution gaussienne multivariée), alors que la seconde famille résulte de l'imposition d'une structure de covariance d'analyse factorielle. L'estimation paramétrique est effectuée grâce à un algorithme d'espérance-maximisation (EM), et la sélection de modèles utilise le critère d'information bayésien (CIB) et la vraisemblance complétée intégrée (VCI).

[MS-28]

**Bayesian High-dimensional Clustering with Variable Selection**  
**Classification bayésienne en hautes dimensions avec sélection de variables**

Vahid PARTOVI NIA, *McGill University*, Anthony DAVISON, *Ecole Polytechnique Fédérale de Lausanne*

We describe an approach to clustering and classification applicable to high-dimensional continuous data, based on Bayesian mixture models whose marginal distributions are explicitly available. The result is a fast fully automatic procedure under which hyper-parameters of the model can be estimated by empirical Bayes, variable selection can be performed, and a dendrogram can be constructed using agglomerative hierarchical clustering based on ratios of marginal posteriors. We may also introduce the basic commands of the R package, *bclust*, that implements this technique.

Nous décrivons une approche à la classification applicable à des données continues de haute dimension basée sur des modèles de mélange bayésiens dont les distributions marginales peuvent être obtenues explicitement. Ceci résulte en une procédure rapide et complètement automatisée permettant d'estimer les hyperparamètres du modèle par Bayes empirique, d'effectuer une sélection de variables, et de construire un dendrogramme au moyen de classification hiérarchique agglomérative basée sur le rapport des lois *a posteriori* marginales. Si possible, nous aborderons également les commandes de base du package R *bclust* qui permet d'exécuter cette technique.

**[MS-29]****Classification of Functional Data: Curve Clustering when Curves are Misaligned****Classification des données fonctionnelles : classification de courbes lorsque les courbes sont désalignées**

Laura SANGALLI, Piercesare SECCHI, Simone VANTINI & Valeria VITELLI, *Politecnico di Milano, Italy*

The issue of classification of functional data is considered, and the problem of curve clustering when curves are misaligned is dealt with. A novel algorithm is described, which jointly clusters and aligns curves. The proposed procedure efficiently decouples amplitude and phase variability; in particular, it is able to detect amplitude clusters while simultaneously disclosing clustering structures in the phase, pointing out features that can neither be captured by simple curve clustering nor by simple curve alignment. The procedure is illustrated via simulation studies and applications to data concerning three-dimensional cerebral vascular geometries.

Nous considérons le problème de la classification de données fonctionnelles, et traitons le problème de la classification de courbes désalignées. Un algorithme permettant à la fois de classifier et d'aligner des courbes est proposé. La procédure proposée découple efficacement la variabilité de l'amplitude et de la phase ; elle peut détecter des grappes d'amplitude tout en soulignant des structures de classification de phase, rapportant ainsi des caractéristiques ne pouvant être détectées par une simple classification de courbes, ni par un simple alignage des courbes. La procédure est illustrée par des simulations et des applications à des données liées à la géométrie vasculaire tridimensionnelle.

**[MS-30]****Constrained Clustering by a Global Optimization Method****Classification sous contraintes au moyen d'une méthode d'optimisation globale**

Yu XIA, *University of Guelph*

We consider clustering within the constrained clustering paradigm whereby both must-link and cannot-link constraints exist. Therefore, it is necessary to consider whether two data points can be put into the same group or not, and whether or not they must be in the same group. Tuy's concave cutting plane method is used to provide a global solution to the problem. We also give properties of local optimal solutions to the constrained clustering problem. Numerical examples show that this method performs favourably when compared with other constrained clustering algorithms.

Nous considérons la classification dans le cadre du paradigme de classification sous contraintes où des contraintes de type « doivent être liés » et « ne peuvent être liés » existent. Il est donc nécessaire de considérer si deux observations peuvent ou non faire partie du même groupe, et si elles doivent ou non être groupées. La méthode du plan sécant concave de Tuy est utilisée afin de fournir une solution globale au problème. Nous donnons aussi les propriétés de solutions optimales locales au problème de classification sous contraintes. Des exemples numériques montrent que cette méthode surpasse d'autres algorithmes de classification sous contraintes.

**[MS-31]****A Review of Statistical Modelling and Data Analyses for Soccer****Une revue de la modélisation statistique et de l'analyse de données pour le soccer**

David BRILLINGER, *University of California, Berkeley*

Today there exist many data sets, models, and results of model fitting for soccer. This talk will present a survey of these. The sections of the talk will be taken from this list:

1. Data Collection and Descriptive Analyses
2. Stochastic Modelling

Il existe aujourd'hui beaucoup de jeux de données, modèles et résultats d'ajustement de modèles pour le soccer. Cet exposé en fera la revue. L'exposé sera divisé en ces sections :

1. Collecte de données et analyses descriptives
2. Modélisation stochastique
3. Classement

3. Ranking
4. Tournaments and Scheduling
5. Game Theory
6. Economics And Management
7. Discussion.

4. Tournois et planification
5. Théorie des jeux
6. Économie et gestion
7. Discussion.

**[MS-32]****A Statistician's Corrective for Editor [Wired Magazine] Anderson's "Data Deluge Makes the Scientific Method Obsolete"****La rectification d'un statisticien au sujet de l'éditorial « Le Déluge de données rend la méthode scientifique obsolète », d'Anderson [dans Wired Magazine].**

G. ARTHUR MIHRAM, *Princeton, NJ*, DANIELLE MIHRAM, *University of Southern California*

Wired Magazine's Editor Anderson (2008: 16/7) asserted that current data-deluges ("Cloud") end scientific theory-making, thus [via automating our statistical regression and hypothesis-testing] scientific method's obsolete. We note his misconception of the scientific method as mathematical curve-fitting and/or statistical hypothesis-testing. Rather, Science, conducted via the Scientific Method, is a well-established model-building (yet, not a model-fitting) methodology of six stages [TEOREMA 28/2, 35–44, 2009]. Science, that human activity devoted to the search for the very explanation for (i.e., for the truth about) any particular naturally-occurring phenomenon, is not data-computation. (Statistical methodologies' naturally probabilistic foundation itself rather excludes scientific truth.)

Anderson, l'éditeur de Wired Magazine (2008 ; 16,7) a affirmé que le déluge actuel de données ("Cloud") met fin au développement des théories scientifiques, et donc [via l'automatisation de nos régressions statistiques et nos tests d'hypothèses] rendra la méthode scientifique obsolète. Nous notons qu'il a une fausse conception de la méthode scientifique comme un ajustement de courbes mathématique et/ou une procédure statistique de tests d'hypothèses. En fait, la Science, menée via la méthode scientifique, est une méthode bien établie pour la construction de modèles (et non pas l'ajustement d'une courbe) qui comprend six étapes [TEOREMA 28/2, 35–44, 2009]. La Science, cette activité humaine fixée sur la recherche de l'explication même (c'est-à-dire la vérité au sujet) d'un phénomène arbitraire se produisant naturellement, ne se limite pas au calcul de données. (La fondation intrinsèquement probabiliste des méthodologies statistiques exclut plutôt la vérité scientifique.)

**Session 02I****Lundi 24 mai • Monday, May 24, 10:25 – 11:40****VAN 1883****Mathematical Statistics****Statistique mathématique****[MS-33]****Calibrating  $P$ -values using Splines.****Calibrage des valeurs  $p$  au moyen de splines**

Mikelis BICKIS, *University of Saskatchewan*, Xiao Bo LI, *Saskatchewan Assessment Management*

Under the null hypothesis, the  $p$ -value from a continuous test statistic has a uniform distribution. Although the  $p$ -value has no direct Bayesian interpretation, there have been several attempts to connect it to a posterior probability of the null hypothesis. Under certain regularity conditions, the posterior probability of the null hypotheses can be expressed

Sous l'hypothèse nulle, la valeur  $p$  d'une statistique de test continue est uniformément distribuée. Bien que la valeur  $p$  n'ait pas d'interprétation bayésienne directe, il y a eu plusieurs tentatives visant à la lier à une probabilité *a posteriori* de l'hypothèse nulle. Sous certaines conditions de régularité, la probabilité *a posteriori* de l'hypothèse nulle peut être exprimée comme le rapport de valeurs de la densité de la va-

as a ratio of values of the density of the p-value. If one has a large number of p-values from related hypotheses, the p-values can be transformed to posterior probabilities by fitting splines to the empirical distribution of p-values.

leur p. Si nous disposons d'un grand nombre de valeurs p d'hypothèses liées, les valeurs p peuvent être transformées en probabilités a posteriori en ajustant des splines à la distribution empirique des valeurs p.

#### [MS-34]

### **Equivariance and Generalized Inference in Two-sample Location-scale Families** **Équivariance et inférence généralisée pour les lois bivariées de position-échelle**

Fuqi CHEN & Sévérien NKURUNZIZA, *University of Windsor*

We discuss statistical problems where small-sample inference presents challenges. Specifically, we consider a typical Behrens–Fisher problem in general location-scale families. We present a method of constructing a generalized pivotal quantity and generalized p-value for the difference between two location parameters. The suggested method is based on minimum risk equivariant estimators and thus is an extension of the methods based on maximum likelihood estimators and conditional inference, which have been so far applied to some specific distributions. The efficiency of the procedure is illustrated by Monte Carlo simulation. Finally, we apply the proposed method to two real data sets.

Dans cette présentation, nous abordons certains problèmes statistiques où l'inférence pour de petits échantillons pose des défis. Plus précisément, nous nous intéressons au problème de type Behrens-Fisher dans les modèles statistiques invariants par translation et changement d'échelle. Nous présentons une méthode de construction de fonction pivot et de valeur-p généralisées pour le ratio de deux paramètres d'échelle et de la différence entre deux paramètres de position. La méthode proposée est basée sur le meilleur estimateur équivariant. Par conséquent, la présente méthode constitue une extension des méthodes préexistantes qui sont basées sur l'estimateur du maximum de vraisemblance ou celles basées sur l'inférence conditionnelle, appliquées à des distributions spécifiques. De plus, nous illustrons les performances de ladite procédure grâce à une étude de simulation par Monte-Carlo. Enfin, nous appliquons la méthode à deux jeux de données réelles.

#### [MS-35]

### **Simultaneous Closeness among Order Statistics to Population Quantiles** **Proximité de l'ensemble des statistiques d'ordre aux quantiles de population**

Katherine DAVIES, *University of Manitoba*, N. BALAKRISHNAN, *McMaster University*, Jerome P. KEATING, *University of Texas at San Antonio*, Robert L. MASON, *Southwest Research Institute*

We derive expressions for the probability that an individual order statistic is closest to the target parameter among all order statistics from a complete random sample. We then apply these general results to location-scale parameter families of distributions with specific applications to estimation of percentiles. In this case, simultaneous-closeness probabilities depend upon the parameters through the value of  $p$  in the percentile and the sample size,  $n$ . The results can be applied to distributions with both bounded and unbounded supports. We illustrate the results with the estimation of percentiles for Normal and exponential distributions.

Nous obtenons des expressions pour la probabilité qu'une statistique d'ordre donnée soit la plus proche du paramètre cible parmi toutes les statistiques d'ordre d'un échantillon aléatoire complet. Nous appliquons ensuite ces résultats généraux aux familles paramétriques position-échelle de distributions et à l'estimation de quantiles. Dans ce cas, les probabilités de proximité simultanées dépendent des paramètres à travers l'ordre  $p$  du quantile et de la taille  $n$  de l'échantillon. Les résultats peuvent être appliqués à des distributions de support borné ou non borné. Nous illustrons les résultats en estimant les quantiles de distributions normales et exponentielles.

**[MS-36]****Distribution of Coefficient of Variation: a Transformation in a New Perspective****Distribution du coefficient de variation : une transformation sous une nouvelle perspective**Debaraj SEN, Yogendra P. CHAUBEY & Murari SINGH, *Concordia University*

The coefficient of variation (CV) has been used as a measure of uniformity of the observations and variability in a population. Being estimated as a ratio of the sample standard deviation to the sample mean, its distribution is computationally complex. This article examines the distribution of symmetrizing transformation along the lines of Chaubey and Mudholkar (1983) for small samples drawn from a normal distribution. For a range of values of the population CV and sample size, we examine the variance of the symmetrizing transformation, distributional symmetry of the variance stabilizing transformation, probability distributions of the above transformations, and discuss the results.

Le coefficient de variation (CV) est utilisé comme mesure d'uniformité des observations et de la variabilité d'une population. Estimé par le quotient de l'écart-type et de la moyenne expérimentaux, sa distribution est complexe. Cette communication examine la distribution d'une transformation de symétrisation, comme par exemple celle de Chaubey et Mudholkar (1983), pour de petits échantillons normalement distribués. Pour différentes tailles d'échantillon et valeurs du CV théorique de la population, nous examinons la variance de la distribution de symétrisation, la symétrie de la distribution de la transformation stabilisatrice de la variance et la distribution de ces transformations. Les résultats sont discutés.

**[MS-37]****On the Robustness of Composite Likelihood Inference****De la robustesse de l'inférence par vraisemblance composite**Ximing Xu, *University of Toronto*

Composite likelihood methods have proved useful in different complex applications, where the ordinary likelihood is too cumbersome to be evaluated. By modelling only lower dimensional marginal or conditional densities, the composite likelihood inference is widely viewed as robust against model misspecification in higher dimensions. However, the best way to study and assess this robustness systematically is still an open problem. In the present paper, a brief introduction to composite likelihood inference is provided. Emphasis is given to the definition of model misspecification and assessment of robustness which is illustrated through a simulation study on multivariate the normal model.

Les méthodes de vraisemblance composite se sont avérées utiles dans plusieurs applications complexes, où la vraisemblance usuelle est difficilement évaluable. En ne modélisant que les densités marginales ou conditionnelles de plus basses dimensions, l'inférence par vraisemblance composite est considérée comme robuste par rapport à une mauvaise spécification de modèle en haute dimension. Cependant, nous ne connaissons toujours pas la meilleure façon d'étudier et d'évaluer systématiquement cette robustesse. Dans la présente communication, une brève introduction à l'inférence à base de vraisemblance composite est donnée. L'emphase est mise sur la définition de la mauvaise spécification de modèle et sur l'évaluation de la robustesse, illustrée par une étude de simulation basée sur le modèle normal multivarié.

**Session 03A                      Lundi 24 mai • Monday, May 24, 13:15 – 14:15                      PPR**

**SSC Impact Award Address**

**Allocution du récipiendaire du prix pour impact de la SSC**

**[MS-38]**

**Preserving Wild Pacific Salmon**

**La préservation du saumon sauvage du Pacifique**

Rick ROUTLEDGE, *Simon Fraser University*

Returns of sockeye salmon to the Fraser River were so low in 2009 that the federal government called for yet another public inquiry. Could salmon farms situated along the juvenile migration route have played a role in the shortfall? According to controversial research on aquaculture impacts on other wild salmon species, the potential cannot be dismissed. The presentation will highlight some of this research with special emphasis on the reaction from stakeholders and government. There are remarkable similarities to the controversy over health impacts from cigarette smoking. Emerging evidence directly related to Fraser sockeye will also be highlighted.

Si peu de saumons rouges ont remonté le fleuve Fraser en 2009 que le gouvernement fédéral a choisi, encore une fois, de mener une enquête publique. Les élevages de saumon situés le long de la route de migration juvénile peuvent-ils avoir joué un rôle ? D'après des recherches controversées évaluant l'impact de l'aquaculture sur les autres espèces de saumon sauvage, cette hypothèse ne peut être écartée. Cette présentation soulignera certaines de ces recherches et portera une attention particulière aux réactions des parties prenantes et du gouvernement. Il y a là des similitudes remarquables avec la controverse sur les impacts du tabagisme sur la santé. Des preuves émergentes directement liées au saumon rouge de la rivière Fraser seront aussi soulignées.

**Session 04A                      Lundi 24 mai • Monday, May 24, 14:30 – 16:00                      VAN 1245A**

**Statisticians in the Fight against Cancer**

**Statisticiens dans la bataille contre le cancer**

**[MS-39]**

**Statistical Methods and Study Designs for Evaluating Cancer Risk in Families Carrying a Known Genetic Mutation**

**Méthodes statistiques et plans d'enquête pour évaluer le risque associé au cancer dans les familles porteuses d'une certaine mutation génétique**

Karen KOPCIUK, *Alberta Health Services – Cancer Care*, Yun-Hee CHOI, *University of Western Ontario*, Pat PARFREY & Jane GREEN, *Memorial University*, Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute*

Many inherited cancer-causing mutations have been identified in the human genome. Families who harbour these identified genetic alterations are often studied to provide age-specific risk estimates. However, a number of statistical and study design issues need to be considered in order to minimize bias and maximize efficiency in the resulting cancer risk estimates. In this talk, we will highlight our methodological work that has focused on identifying efficient designs and appropriate likelihood methods, estimating cancer risk (penetrance) us-

Plusieurs mutations héréditaires causant le cancer ont été identifiées dans le génome humain. Les familles portant ces altérations génétiques identifiées sont souvent étudiées afin d'obtenir des estimations du risque spécifique à l'âge. Cependant, plusieurs problèmes statistiques doivent être résolus afin de minimiser le biais et maximiser l'efficacité des estimations du risque de cancer. Dans cette présentation, nous allons partager notre travail méthodologique centré sur l'identification de plans efficaces et de méthodes appropriées basées sur la vraisemblance, l'estimation du risque de cancer (pénétrance) au moyen de méthodes de ségrégation mo-

ing modified segregation methods which adopted a generalized log-Burr hazard form, and estimating risk of successive cancers within a multi-state framework.

difiées, et l'estimation du risque de cancers successifs dans un cadre multi-états.

**[MS-40]**

**Small-area Spatial Variation in Smoking Incidence using Survey Data**

**Variation spatiale en petits domaines des taux de tabagisme pour des données d'enquête**

Patrick BROWN & Lutong ZHOU, *Cancer Care Ontario*

The Canadian Community Household Survey can be used to make inference about the variation in smoking rates at small scales, within cities, even when the data appear to be sparse. This is accomplished with generalized linear mixed models with spatially correlated random effects, which are able to predict smoking levels for the smallest census regions. This talk describes the model and the required Bayesian inference methods, as well as some extensions and ongoing research.

L'Enquête sur la santé dans les collectivités canadiennes peut être utilisée pour tirer des conclusions sur la variation des taux de tabagisme à petite échelle, dans le cadre des villes, même si les données semblent être rares. Cela s'accomplit avec les modèles linéaires généralisés mixtes à effets aléatoires spatialement corrélés, qui sont capables de prédire les taux de tabagisme pour les régions de recensement les plus petites. Cet exposé décrit le modèle et les méthodes d'inférence bayésienne requises, ainsi que des extensions possibles et des développements en cours.

**[MS-41]**

**Modern Cancer Clinical Trials: Recent Statistical Advancements and Controversies**

**Essais cliniques oncologiques modernes : percées et controverses statistiques récentes**

Gregory POND, *Ontario Clinical Oncology Group, Department of Oncology, McMaster University*

Oncology research is the fastest growing area of pharmaceutical research and development. Clinical trials in cancer have many unique methodological challenges due, in part, to the severity of the disease and associated toxicities of treatments. Cancer clinical trials can regularly involve novel aspects of imaging technologies, radiotherapy delivery, biological therapies, health economics, quality of life and more. Biostatisticians in this field must have a broad knowledge base, an ability to understand practical issues and have developed collaboration abilities. This talk will discuss the use of some controversial biostatistical techniques in cancer research, such as dynamic allocation methods and non-inferiority trials.

La recherche oncologique est le secteur de croissance la plus rapide parmi la recherche et le développement pharmaceutique. Les essais cliniques sur le cancer comportent plusieurs défis méthodologiques uniques dus, en partie, à la sévérité de la maladie et à la toxicité associée aux traitements. Les essais cliniques sur le cancer peuvent régulièrement impliquer de nouveaux aspects des technologies d'imagerie, de l'administration de radiothérapie, de thérapies biologiques, d'économie de la santé, de la qualité de vie, etc. Les biostatisticiens dans ce domaine doivent avoir une large base de connaissances, une capacité de compréhension de problèmes pratiques et des compétences de collaboration développées. Cet exposé discutera de l'utilisation de certaines techniques biostatistiques controversées en recherche oncologique, comme les méthodes d'allocation dynamique et les essais de non-infériorité.

**Session 04B      Lundi 24 mai • Monday, May 24, 14:30 – 16:00      VAN 2289A****Large (Industrial-scale) Computer Experiments for Business and Industry  
Grandes expériences par ordinateur (à échelle industrielle) pour l'industrie et la gestion****[MS-42]****Challenges in Industrial Applications of Computer Experiments  
Défis en applications industrielles d'expériences informatiques**Andrew BOOKER, *Boeing*

A wide range of applications Design and Analysis of Computer Experiments (DACE) exist at Boeing. The applications include engineering conceptual design, detailed design, multi-disciplinary design, and design under uncertainty. They include many design variables, complicated, perhaps discrete, design spaces and simulation codes (perhaps more than one) that take a very long time to run and that may fail. Thus Boeing statisticians are confronted with a number of challenges in the design of computer experiments and in approximate modelling. I will use examples to describe these challenges and directions of research that would be particularly useful in an engineering environment.

Il y a une large gamme d'applications de planification et d'analyse d'expériences informatiques chez Boeing. Les applications incluent la planification conceptuelle en génie, la planification détaillée, la planification multidisciplinaire, et la planification avec incertitudes. Elles peuvent comprendre plusieurs variables de design, complexes, peut-être discrètes, des espaces de design et des codes de simulation pouvant nécessiter beaucoup de temps d'exécution et pouvant ne pas fonctionner. Les statisticiens chez Boeing sont donc confrontés à un grand nombre de défis en planification d'expériences informatiques et en modélisation approximative. Je vais utiliser des exemples afin de décrire ces défis et les directions de recherches particulièrement utiles en génie.

**[MS-43]****Emulating a Computationally Intensive Simulation Model of Disease to Aid Decision-making on Drug Development  
Émulation d'un modèle lourd de simulation d'une maladie pour aider à la prise de décision en développement de médicaments**Anthony O'HAGAN & Jeremy OAKLEY, *University of Sheffield*

A decision-support model simulates the outcomes of a series of clinical trials for a new drug. The most important outputs of this simulation include whether a successful result is obtained in the final confirmatory trial to satisfy drug regulators, and whether those results are strong enough for the drug to be considered cost-effective by health care agencies. The cost-effectiveness module is important but highly computationally intensive, and must be run many times in order to explore different trial designs. The decision-support capability of this system was made feasible by building an emulator for this module.

Un modèle d'aide à la décision simule une série d'essais cliniques pour un nouveau médicament. Les résultats les plus importants de cette simulation sont les conclusions des dernières études confirmatoires, à savoir que le médicament satisfait la réglementation en place, et que les résultats sont suffisamment convaincants pour que le médicament soit considéré comme rentable par les agences de soins de santé? Le module traitant la rentabilité est important, mais demande beaucoup de puissance de calcul et doit être exécuté à maintes reprises afin d'explorer différents plans d'expériences. La capacité de ce système en tant qu'aide à la décision a été rendue viable par la construction d'un émulateur pour ce module.

[MS-44]

**Predictive Modelling of a Radiative Shock Physics**  
**Modélisation prédictive de la physique de chocs radiatifs**

Derek BINGHAM, *Simon Fraser University*, James HOLLOWAY & Bruce FRYXELL, *University of Michigan*

Predictive science, in some circles, has become synonymous with the use of physics modelling, often realized in complex computer codes, and field data to forecast what would be observed in a physical process. In the statistics community, this endeavour falls in the area of computer experiments. This talk deals with the design and analysis of a series of computer and field experiments to form a predictive model for a radiative shock hydrodynamics system. Specific issues addressed are (i) the role of experiment design in efficient predictive model building and (ii) uncertainty quantification. The interplay between (i) and (ii) is discussed.

Dans certains cercles, la science prédictive est devenue synonyme d'utilisation de modélisation physique, souvent à l'aide de codes informatiques complexes, et d'observations sur le terrain afin de prédire ce qui sera observé dans un processus physique. Dans la communauté statistique, ceci tombe dans le domaine des expériences informatiques. Cet exposé traite de la planification et l'analyse d'une série d'expériences informatiques et sur le terrain afin de former un modèle prédictif pour un système hydrodynamique de chocs radiatifs. Les questions spécifiques abordées sont (i) le rôle du plan d'expérience dans la construction d'un modèle efficace et (ii) la quantification des incertitudes. Nous discutons également de la relation entre (i) et (ii).

**Session 04C      Lundi 24 mai • Monday, May 24, 14:30 – 16:00      VAN 2291A**

**Optimization and Applications in MCMC**

**Optimisation et applications des algorithmes Monte-Carlo par chaînes de Markov**

[MS-45]

**Divide and Conquer: A Mixture-Based Approach to Regional Adaptation for MCMC**  
**Diviser pour régner : un algorithme régional et adaptatif pour l'échantillonnage Monte-Carlo avec des chaînes de Markov**

Radu CRAIU & Yan BAI, *University of Toronto*, Antonio DI NARZO, *University of Bologna*

Adaptive MCMC (AMCMC) algorithms allow the automatic tuning of the parameters while the simulation is in progress. A multimodal target distribution may call for regional adaptation of Metropolis–Hastings samplers so that the proposal distribution varies across regions. In the case in which the target distribution is approximated by a mixture of Gaussians, we propose an adaptation process which involves fitting the mixture using the available samples via an online EM algorithm and, based on the current mixture parameters, constructing the regional adaptive algorithm with online recursion (RAPTOR). The method is tested on simulated as well as a real data example.

L'efficacité des algorithmes de Monte-Carlo par chaînes de Markov (MCMC) dépend du choix des paramètres dirigeant la simulation. Les algorithmes adaptatifs MCMC (AMCMC) sont capables d'ajuster les paramètres de la simulation en cours de route. Dans le cas d'une loi cible multimodale, la loi instrumentale utilisée dans l'algorithme de Metropolis–Hastings peut être différente selon la région de l'espace d'échantillonnage. Dans le cas où la loi cible est bien approximée par un mélange de lois normales, nous proposons un algorithme AMCMC où les paramètres du mélange sont estimés par un algorithme EM en ligne et définissent à leur tour la loi de proposition régionale. La méthode est testée avec des simulations et une analyse génétique.

[MS-46]

### Optimising and Adapting the Metropolis Algorithm Comment optimiser et adapter l'algorithme Metropolis

Jeffrey ROSENTHAL, *University of Toronto*

The Metropolis algorithm is a very popular method of approximately sampling from complicated probability distributions, especially those arising in Bayesian inference. A wide variety of proposal distributions are available, and it can be difficult to choose among them. We will discuss optimal proposals under various circumstances. We will also consider the possibility of having the computer automatically “adapt” the algorithm while it runs, to improve and tune on the fly.

L'algorithme Metropolis est une méthode très populaire pour faire de l'échantillonnage approximatif de distributions de probabilités compliquées, en particulier celles qui apparaissent en inférence bayésienne. Plusieurs distributions instrumentales sont disponibles et choisir parmi elles peut être difficile. Nous allons discuter des distributions instrumentales optimales dans certains contextes. Nous allons aussi considérer la possibilité que l'ordinateur puisse « adapter » l'algorithme pour l'améliorer pendant l'exécution.

[MS-47]

### A Bayesian Sieves Model for Inference on Bivariate Extremes Un modèle de tamis bayésien pour l'inférence sur les lois de valeurs extrêmes bivariées

Simon GUILLOTTE, *University of Prince-Edward Island*, François PERRON, *Université de Montréal*, Johan SEGERS, *Université Catholique de Louvain*

A Bayesian sieves model for the spectral measure characterizing the tail dependence structure of a bivariate distribution function will be presented. More precisely, as an alternative to parametric modelling of the spectral measure, we propose an infinite-dimensional model that is at the same time manageable and still dense within the class of spectral measures. Inference is done in a Bayesian framework using the censored-likelihood approach. In particular, we construct a prior distribution on the class of spectral measures and develop a trans-dimensional MCMC algorithm for numerical computations. The method provides a bivariate predictive density used for predicting rare extreme outcomes.

Un modèle de tamis bayésien pour la mesure de probabilité spectrale caractérisant la structure de dépendance de la queue d'une loi bivariée sera présenté. Plus précisément, en tant qu'alternative à la modélisation paramétrique de la mesure spectrale, nous proposons un modèle qui est à la fois facilement probabilisable et dense dans l'espace des mesures spectrales. L'inférence est faite dans un cadre bayésien utilisant l'approche de la vraisemblance censurée. En particulier, nous construisons une loi a priori sur la classe des mesures spectrales et développons un algorithme MCMC à sauts réversibles pour les calculs numériques. La méthode nous donne notamment une loi prédictive bivariée servant à prévoir les événements rares.

## Session 04D      Lundi 24 mai • Monday, May 24, 14:30 – 16:00      VAN 2809A

### Panel Discussion on Opportunities and Challenges in Interdisciplinary Research and Training

#### Panel sur les opportunités et les défis de la formation et la recherche interdisciplinaire

Statistical science is inherently an interdisciplinary research field, contributing to the foundational strength of many scientific disciplines. Indeed, interdisciplinary work often provides the groundwork for new statistical breakthroughs. Many sta-

La science statistique est un domaine de recherche intrinsèquement interdisciplinaire, contribuant à la fondation de nombreuses disciplines scientifiques. Des travaux interdisciplinaires sont en effet souvent à l'origine de nouvelles avancées statistiques. Plusieurs chercheurs en statistique au Ca-

tistical researchers in Canada engage in collaborative work, walking their developments through from important new statistical methodological tools, to scientific discoveries utilizing such tools, and then to impact of the knowledge derived, including impact in the policy arena or advancements in industry and health, for example. This panel discussion focuses on this important aspect of the discipline of statistical sciences, that of collaborative research, and discusses challenges in the conduct of interdisciplinary research and how best to teach and develop interdisciplinary research skills.

Panellists discuss a broad range of issues including collaborative training environments, the increasingly prominent role of the statistician, for example in health studies and in climate change investigations where accurate measures of uncertainty for rather complicated study designs are required, and where large teams of investigators are brought to bear on critical societal problems. The discussion focuses on practical and concrete approaches for research leadership, training and mentoring. The session ends with an open discussion where engagement by the community on these issues is encouraged.

nada prennent part à des travaux en collaboration et participent à toutes les étapes, débutant avec le développement de nouveaux outils méthodologiques statistiques, passant par des découvertes scientifiques utilisant ces outils, et aboutissant à l'impact tangible des connaissances ainsi acquises, qu'il soit au niveau des politiques publiques, de l'industrie ou de la santé. L'objet de ce débat est centré sur l'aspect collaboratif des sciences statistiques, sur les défis de la recherche interdisciplinaire et les façons d'enseigner et de développer les compétences liées à ce type de recherche.

Les panélistes discutent d'une large gamme de questions incluant les environnements de formation interdisciplinaire, le rôle de plus en plus proéminent du statisticien, par exemple dans les domaines de la santé et des changements climatiques où des mesures précises d'incertitude sont nécessaires pour des plans d'enquête complexes, et où de grandes équipes de chercheurs sont amenées à travailler sur des problèmes de société. La discussion est centrée sur des approches pratiques et concrètes en direction de recherche, formation et mentorat. La séance se termine avec une discussion ouverte où la communauté est encouragée à s'exprimer au sujet de ces questions.

**Session 04E      Lundi 24 mai • Monday, May 24, 14:30 – 16:00      VAN 2811A**

**Methods and Applications in Cluster Analysis**

**Méthodes et applications de la classification automatique**

[MS-48]

**Clustering using Linear Patterns**

**Classification au moyen de patrons linéaires**

Ruben ZAMAR, *University of British Columbia*

I first describe a method called linear grouping algorithm (LGA), which can be used to detect different linear structures in a data set. LGA combines ideas from principal components, clustering methods and resampling algorithms. I show that LGA can detect several different linear relations at once, but can be affected by the presence of outliers in the data set. I will then present a robustification of LGA based on trimming and a partial likelihood extension of LGA that allows for the flexible modelling of linear clusters with different scales.

Je décris d'abord une méthode de type algorithme de groupement linéaire (AGL) pouvant être utilisée pour détecter différentes structures linéaires dans un jeu de données. L'AGL combine des idées de l'analyse en composantes principales, de méthodes de classification et d'algorithmes de rééchantillonnage. Je montre que l'AGL peut détecter simultanément différentes relations linéaires, mais qu'il peut être affecté par la présence de valeurs aberrantes. Je présente ensuite une robustification de l'AGL basée sur la troncature et une généralisation de l'AGL à la vraisemblance partielle permettant la modélisation flexible de groupements linéaires à différentes échelles.

[MS-49]

**Model Clustering and Applications****Classification à partir de modèles et applications**Rong ZHU, *McMaster University*, Abdel EL-SHAARAWI, *National Water Research Institute, Environment Canada*

Model clustering consists in grouping objects where each of them is associated with many observations. Objects within a group share common traits used for classification. In addition, ancillary variables exist. A family of models appropriate for the structure and nature of available measurements is fitted for data with the parameter vector corresponding to both interested and ancillary variables. The similarity of models is defined as the equality of interested parameters, and the p-value associated with the proposed model linking test is used as the similarity measure. This procedure is applied to an environmental study classifying *E. coli* strains.

La classification à partir modèles consiste à regrouper des objets associés à plusieurs observations. Les objets à l'intérieur d'un groupe partagent certaines caractéristiques utilisées pour la classification. Des variables auxiliaires peuvent aussi exister. Une famille appropriée de modèles pour la structure et la nature des mesures disponibles est ajustée aux données ; le vecteur des paramètres correspond alors aux variables d'intérêts et aux variables auxiliaires. La similitude entre les modèles est définie comme l'égalité des paramètres d'intérêt, et la valeur-p associée au test de jumelage de modèles proposé est utilisée comme mesure de similitude. Cette méthode est appliquée à la classification de souches *E. coli* dans une étude environnementale.

[MS-50]

**Clustering Categorical Data Based On Distance Vectors****Classification de données catégorielles à partir de vecteurs de distance**Peng ZHANG, *University of Alberta*, Xiaogang WANG, *York University*, Peter SONG, *University of Michigan*

We introduce a novel statistical procedure to cluster categorical data based on distance vectors. This algorithm sequentially extracts clusters from the given data set. Consequently, the number of clusters can be determined automatically by the algorithm. The performance of the proposed algorithm is examined by both simulated and real data sets. Comparisons with two clustering algorithms show that the proposed algorithm substantially outperforms these competitors, with the classification rate or the information gain typically improved by several orders of magnitudes. Computational complexity and runtime comparisons are also provided.

Nous présentons une nouvelle procédure statistique de classification de données catégorielles basée sur des vecteurs de distance. Cet algorithme extrait séquentiellement les grappes du jeu de données. Conséquemment, le nombre de grappes peut être déterminé automatiquement par l'algorithme. La performance de l'algorithme proposée est examinée grâce à des jeux de données réels et simulés. Des comparaisons avec deux algorithmes de classification montrent que l'algorithme proposé surpasse substantiellement ces compétiteurs, améliorant le taux de classification ou le gain d'information de plusieurs ordres de grandeur. Des comparaisons de complexité de calcul sont aussi données.

**Session 04F      Lundi 24 mai • Monday, May 24, 14:30 – 15:30      VAN 1203****Methodologies for Count Data and Ecological Data****Méthodologies pour données de dénombrement et données écologiques****[MS-51]****Analysis of Longitudinal Count Data with Measurement Error and the Application in Epidemiology****Analyse de données de dénombrement longitudinales avec erreurs de mesure et application en épidémiologie**Yunqi JI & Zhaozhi FAN, *Memorial University*

In epidemiology, the collected data are often contaminated with measurement errors because of imperfect instruments and procedures. For example, new annual cancer cases are prone to being miscounted due to the limited coverage of surveillance systems, inconsistent and incomplete case reporting, and misdiagnoses by medical facilities. We propose a new count error model and combine it with longitudinal models to analyze epidemic diseases. Approaches based on generalized estimating equations, quasi-likelihood are developed to estimate of model parameters. Analyses of the simulated data and the cancer data from USA are conducted to illustrate our models and methods.

En épidémiologie, les données recueillies sont souvent contaminées par des erreurs de mesure dues à des instruments et procédures imparfaits. Par exemple, le nombre de nouveaux cancers annuels est souvent mal compté à cause de la couverture limitée des systèmes de surveillance, de cas rapportés de façon incohérente ou incomplète, et de mauvais diagnostics. Nous proposons un nouveau modèle d'erreur de dénombrement et le combinons à des modèles longitudinaux afin d'analyser des maladies épidémiques. Des approches basées sur les équations d'estimation généralisées et sur la quasi-vraisemblance sont développées afin d'estimer les paramètres du modèle. Afin d'illustrer le modèle et les méthodes, des données simulées et des données oncologiques américaines sont analysées.

**[MS-52]****A Regression Tree Based Method for Longitudinal Count Data****Une méthode à base d'arbres de régression pour des données longitudinales de dénombrement**Xiaoqin XIONG & Joel DUBIN, *University of Waterloo*

We have interest in modeling correlated count data, for which a tree-based method will help answer an important scientific question of interest. For longitudinal count data, Lee (2005) proposed a generalized multivariate decision tree based on GEE and Pearson residual distribution. However, his approach assumes the covariates are time-invariant. We propose a subject-specific based (e.g., GLMM) regression tree method for longitudinal count data, which can deal with both time-variant and time-invariant covariates. We will discuss our proposed method, which involves incorporation of a permutation test on conditional residuals, and apply it to data from a longitudinal hemodialysis study.

Nous nous intéressons à la modélisation de données de dénombrement corrélées pour lesquelles une méthode statistique à base d'arbres aidera à répondre à un problème scientifique. Pour des données de dénombrement longitudinales, Lee (2005) propose un arbre de décision multivarié généralisé basé sur les EEG et la distribution résiduelle de Pearson. Cependant, son approche suppose que les covariables sont invariantes par rapport au temps. Nous proposons d'utiliser une méthode à base d'arbres de régression spécifiques aux sujets (p. ex., MMLG) pour des données de dénombrement longitudinales, et ce à la fois pour des covariables indépendantes ou dépendantes du temps. Nous discutons de la méthode proposée, qui implique l'incorporation d'un test de permutation sur les résidus conditionnels, et nous l'appliquons à des données d'une étude longitudinale d'hémodialyse.

**[MS-53]****Variance Function in Semiparametric Analysis of Count Data****La fonction de variance dans l'analyse semi-paramétrique de données de dénombrement**Tasneem ZAIHRA, *University of New Brunswick—Saint John*, Sudhir PAUL, *University of Windsor*

We determine an appropriate variance function which can be used in the semiparametric analysis of overdispersed count data. We use data analytic and hypothesis testing approaches through a broader class of models. The models considered are three parameter negative binomial distributions and the extended quasi-likelihood (EQL). Data analysis, tests, and simulations indicate that the three parameter negative binomial distribution does not improve in fit to count data over the simpler negative binomial distribution. Further simulations using the EQL indicate that, for data with small mean and small over-dispersion, the negative binomial variance function is preferable while  $\mu^2$  is preferable otherwise.

Nous déterminons une fonction de variance appropriée pouvant être utilisée dans l'analyse semi-paramétrique de données de dénombrement surdispersées. Nous utilisons une approche d'analyse de données et de test d'hypothèses basée sur une plus large classe de modèles. Les modèles considérés sont la distribution binomiale négative à trois paramètres et la quasi-vraisemblance étendue (QVE). Une analyse de données, des tests et des simulations indiquent que la distribution binomiale négative à trois paramètres n'améliore pas l'ajustement aux données comparativement à la distribution binomiale négative plus simple. Des simulations subséquentes utilisant la QVE indiquent que, pour des données de petite moyenne et de faible surdispersion, la fonction de variance binomiale négative est préférable, alors que  $\mu^2$  est autrement préférable.

**[MS-54]****Evaluation of Diagnostic Tests by a Latent Class Model with Three Latent Classes****Evaluation de tests diagnostiques avec le modèle de classes latentes incluant trois classes**Henrik STRYHN & Charles CARAGUEL, *University of Prince Edward Island*

A study to assess the performance of five tests for infectious salmon anemia virus used for monitoring the salmon farming industry in New Brunswick included samples from 400 fish from four populations defined to have distinct prevalences. Traditional latent class modelling failed to describe the data adequately, even after introducing multiple conditional test dependencies. Visual inspection of the data showed that testing patterns varied substantially between populations. We describe how addition of a third latent class in a Bayesian framework (similar to Dendukuri et al. (2009)) lead to better fitting models, and discuss the interpretation of the resulting model parameters.

Destinés à surveiller l'anémie infectieuse du saumon dans la salmoniculture néo-brunswickoise, cinq tests diagnostiques ont été évalués dans une étude incluant 400 échantillons de poissons provenant de quatre populations de prévalences distinctes. L'approche traditionnelle du modèle de classes latentes a échoué dans l'adaptation avec les données, même en incluant plusieurs dépendances conditionnelles entre les tests. L'inspection des données a révélé plusieurs schémas de détection qui varient entre les populations. L'addition d'une troisième classe dans un cadre bayésien (semblable à Dendukuri et coll. (2009)) conduit à une meilleure correspondance du modèle avec les données. L'interprétation des paramètres obtenus est ensuite discutée.

**Session 04G      Lundi 24 mai • Monday, May 24, 14:30 – 16:00      VAN 1211**  
**Methodology for Survival Analysis**  
**Méthodologie pour l'analyse de survie**

[MS-55]

**On Estimation of Extreme Values in Survival Analysis**  
**Estimation de valeurs extrêmes en analyse de survie**

Theresa CONROY & Mei Ling HUANG, *Brock University*

This paper proposes new weighted estimators for the tail index and extreme quantiles in survival analysis which are related to the Generalized Pareto Distribution (GPD). We derive asymptotic properties of the new estimators. Simulations results of the mean square errors of the new estimators, and comparisons to existing estimators are given. A real-world example of a clinical trial testing two different procedures on lung cancer patients is studied. Efficient estimators are essential for producing accurate results when measuring survival time. The results will show that the proposed estimators produce better model fits than the existing estimators.

Cet article propose de nouveaux estimateurs pondérés pour l'indice caudal et les quantiles extrêmes en analyse de survie qui sont liés à la distribution Pareto généralisée (DPG). Nous obtenons des propriétés asymptotiques de ces nouveaux estimateurs. Les résultats de simulation pour l'erreur quadratique moyenne des nouveaux estimateurs et des comparaisons aux estimateurs existants sont donnés. Nous étudions un exemple d'essai clinique réel où deux différentes procédures sont testées sur des patients atteints du cancer du poumon. Il est essentiel d'avoir des estimateurs efficaces afin de mesurer précisément les temps de survie. Les résultats montrent une meilleure adéquation de modèle avec les nouveaux estimateurs.

[MS-56]

**A Comparison of the Performance of Some Non-parametric Estimators of the Conditional Survival Function with Interval-censored Data**

**Une comparaison de certains estimateurs non paramétriques de la fonction de survie conditionnelle avec données censurées par intervalle**

Mohammad DEGHAN & Thierry DUCHESNE, *Université Laval*

Simple nonparametric estimates of the conditional distribution of a response variable given a continuous covariate are often useful for data exploration purposes or to help with the specification/validation of a parametric or semiparametric regression model. I give a comparison between various nonparametric estimators of the conditional survival function with censored data. Since many such estimators exist, comparing their performance may be helpful for choosing the appropriate estimator. We compare a generalization of Turnbull's estimator proposed by Dehghan & Duchesne (2010) with the Nadaraya-Watson and generalized Kaplan-Meier estimators and derive a rule-of-thumb that allows one to pick the appropriate estimator for a given dataset.

Des estimateurs non paramétriques simples de la distribution conditionnelle d'une variable réponse étant donnée une covariable continue sont souvent utiles à des fins exploratoires ou de validation/spécification de modèle de régression paramétrique ou semi-paramétrique. Je compare plusieurs estimateurs non paramétriques de la fonction de survie conditionnelle avec des données censurées. Comme il existe plusieurs tels estimateurs, une comparaison entre eux nous permettant de choisir l'estimateur approprié pour une application donnée peut s'avérer utile. On compare l'estimateur de Turnbull généralisé de Dehghan & Duchesne (2010) aux estimateurs de Nadaraya-Watson et Kaplan-Meier généralisé et proposons une « règle du pouce » pour choisir le bon estimateur. Comme je considère le cas où la variable réponse est censurée par intervalle, les estimateurs de Nadaraya-Watson et Kaplan-Meier généralisé sont utilisés en combinaison avec les méthodes d'imputation multiple et d'imputation au point milieu.

**[MS-57]****Nonparametric Confidence Intervals for the Difference of Median Failure Times****Intervalle de confiance non paramétriques pour la différence de médianes de temps de défaillance**Jianrong WU, *St Jude Children's Research Hospital*

We present six confidence intervals for the difference in medians in a general right-censored failure time framework. Simulation studies are conducted to compare the coverage probability of these intervals for small samples. The proposed intervals are applied to tumour growth delay data from an actual tumour xenograft experiment.

Nous présentons six intervalles de confiance pour la différence en médianes dans un cadre général de survie avec censure à droite. Des études de simulations sont menées afin de comparer les probabilités de couverture de ces intervalles pour petites tailles échantillonales. Les intervalles proposés sont appliqués à des données de retard de croissance de tumeurs provenant d'une expérience de xénogreffes.

**[MS-58]****G-estimation for the Comparison for Dynamic Treatment Regimens****G-estimation afin de comparer des traitements dynamiques**Cecilia COTTON, *University of Waterloo*, Patrick HEAGERTY, *University of Washington*

First introduced by Robins (1986), *g*-estimation makes an appropriate adjustment for the effect of exposure in the presence of a time-dependent confounder. Typically for the comparison of treatment regimens, a Structural Nested Failure Time Model is specified through the use of blip or shift functions to link counterfactual outcomes to observed outcomes. We discuss an application of *g*-estimation for the comparison of survival under dynamic treatment regimens in the situation where no natural baseline or placebo treatment regimen exists. Our methods are illustrated through simulation results.

Initialement présentée par Robins (1986), la *g*-estimation effectue l'ajustement approprié pour l'effet de l'exposition en présence d'une variable confondante dépendant du temps. Typiquement, pour comparer des régimes de traitement, un modèle structural emboîté de temps de défaillance est spécifié à l'aide de fonctions signal et décalage afin de lier les réponses contre-factuelles aux réponses observées. Nous discutons d'une application de la *g*-estimation à la comparaison de survies sous régimes de traitement dynamiques dans la situation où il n'existe pas de risque de base ou de régime de traitement placebo. Nos méthodes sont illustrées à l'aide de résultats de simulation.

**[MS-59]****Accuracy of Conventional and Marginal Structural Cox Model Estimators: A Simulation Study****Précision des estimateurs pour un modèle de Cox marginal structurel et conventionnel : une étude de simulation**Yongling XIAO, Michal ABRAHAMOWICZ & Erica MOODIE, *McGill University*

Marginal structural models (MSM) provide a powerful tool to control for confounding by a time-dependent confounder that is affected by prior treatment. Most applications of MSM to time-to-event data rely on pooled logistic regression rather than on models specifically for time-to-event analyses. We demonstrate how a marginal structural Cox model may be fit directly using standard software. We use simulations to illustrate the limitations of the MSM fitted via pooled logistic regression, to evaluate the inverse-probability-of-treatment weighting estimator for the marginal structural Cox

Les modèles marginaux structurels (MMS) sont un outil puissant afin de contrôler pour un confondeur dépendant du temps affecté par un traitement antérieur. La plupart des applications des MMS à des données de survie s'appuient sur une régression logistique combinée plutôt que sur des modèles spécifiques à l'analyse de survie. Nous montrons comment un modèle marginal structurel de Cox peut être ajusté directement à l'aide des logiciels habituels. Nous utilisons des simulations afin d'illustrer les limites de MMS ajustés par régression logistique combinée, d'évaluer l'estimateur à pondération inverse à la probabilité de traitement pour le MMS de Cox, et d'évaluer la performance des poids de nor-

model, and to evaluate the performance of newly-proposed normalized weights alongside the usual stabilized and unstabilized weights.

malisation nouvellement proposés par rapport aux poids habituels stabilisés et non stabilisés.

[MS-60]

### Proportional Density Mixture Cure Model

#### Modèle de mélange proportionnel de densités avec fraction non à risque

Yingwei PENG & Jianfeng XU, *Queen's University*

We consider a proportional density mixture cure model for survival data with a possible cure fraction. The Box–Cox transform is employed to increase the model flexibility. A semiparametric estimation method based on the empirical likelihood technique is proposed to estimate the unknown parameters in the model. We conduct a simulation study and it shows that the proposed model is robust to model misspecification. The model is applied to a data set from a cancer study.

Nous considérons un modèle de mélange proportionnel de densités avec fraction non à risque pour des données de survie avec une possible fraction non à risque. La transformée de Box-Cox est utilisée afin d'augmenter la flexibilité du modèle. Une méthode d'estimation semi-paramétrique basée sur la vraisemblance empirique est proposée afin d'estimer les paramètres inconnus du modèle. Nous effectuons une étude de simulation et montrons que le modèle proposé est robuste sous une mauvaise spécification de modèle. Le modèle est appliqué à un jeu de données d'une étude sur le cancer.

**Session 04H**

**Lundi 24 mai • Monday, May 24, 14:30 – 16:00**

**VAN 1871**

Computational Methods in Finance

Méthodologie computationnelle en finance

[MS-61]

### Correcting the Bias in Least-Squares Monte Carlo Estimators of American Option Values

#### Corriger le biais dans l'estimation aux moindres carrés de la valeur d'options américaines obtenus par Monte-Carlo

Kin Hung (Felix) KAN & Mark REESOR, *University of Western Ontario*

We present a method for reducing the bias inherent in a regression-based Monte Carlo estimator of American option prices. We derive an easily evaluated approximation to the bias which holds for general asset-price processes of any dimensionality and for general payoff structures. This derivation uses the large sample properties of least-squares regression estimators. A bias-corrected estimator results by subtracting the bias approximation from the uncorrected estimators at each exercise opportunity. Numerical results show that the bias-corrected estimator significantly outperforms its uncorrected counterpart across all combinations of number of exercise opportunities, option moneyness, and sample size.

Nous présentons une méthode afin de réduire le biais inhérent à un estimateur de Monte-Carlo basé sur la régression pour le prix d'options américaines. Nous obtenons une approximation facile à évaluer du biais valable pour des processus de prix d'options de n'importe quelle dimension et pour des structures de remboursement générales. L'approximation s'appuie sur les propriétés asymptotiques des estimateurs de régression par moindres carrés. Un estimateur corrigé pour le biais est obtenu en soustrayant l'approximation du biais des estimateurs non corrigés à chaque opportunité d'exercice des options. Des résultats numériques montrent que l'estimateur corrigé pour le biais surpasse significativement l'estimateur non corrigé pour toutes combinaisons du nombre d'opportunités d'exercice, du degré de parité des options, et de la taille d'échantillon.

**[MS-62]****Stochastic Volatility Models with Leverage Effect, Heavy-tailed Errors, and Jump Components.  
Modèles de volatilité stochastique avec effet de levier, erreurs à ailes relevées et sauts.**Amaan MEHRABIAN & David STEPHENS, *McGill University*

We apply MCMC and Sequential Monte-Carlo (SMC) methods for inference on Stochastic Volatility models with leverage effect, heavy-tailed errors, and jump components. The intractable likelihood of the model is estimated by particle filtering techniques and a fast non-gradient based Maximum Likelihood (ML) estimation is utilized.

We perform ML ratio test for model selection and show, through numerical simulations, that the distribution of the test statistic converges to its true distribution.

The methods are illustrated using simulated data and are applied to analyze high-frequency (five-minute level) as well as hourly and daily stock returns and foreign exchange rates.

Nous appliquons des méthodes MCMC et Monte-Carlo séquentielles (MCS) pour l'inférence avec des modèles de volatilité stochastique comportant un effet de levier, des erreurs à ailes relevées, et des sauts. La vraisemblance incalculable du modèle est estimée par des techniques de filtrage particulaire et une méthode d'estimation par maximum de vraisemblance (MV) rapide non basée sur le gradient est utilisée.

Nous utilisons le test du rapport de vraisemblance pour faire la sélection de modèle et montrons, par des simulations numériques, que la distribution de la statistique de test converge vers sa vraie distribution.

Les méthodes sont illustrées au moyen de données simulées et sont appliquées à l'analyse de rendement d'actions et de taux de change de haute fréquence (aux 5 minutes), horaires et journaliers.

**[MS-63]****Multivariate Stochastic Volatility Models: A Slice Sampler Within Gibbs Approach  
Modèles de volatilité stochastique multivariés : une approche basée sur un échantillonneur par tranches dans Gibbs**Zhongxian MEN, *University of Waterloo*

Multivariate stochastic volatility (MSV) models have been intensively studied in the past several years. For the general case of the MSV models, not many methods have been proposed. In this talk, correlations are permitted between the innovations of the asset returns and those of the volatility dynamics. We look at the MSV model in a Bayesian framework by applying an inverse Wishart distribution. The multistage slice sampler within the Gibbs algorithm is proposed to sample the persistent parameters and latent variables. Since the Metropolis–Hastings (M-H) method is avoided, our algorithm is more efficient and easier to operate.

Les modèles de volatilité stochastique multivariée (VSM) ont beaucoup été étudiés au cours des dernières années. Peu de méthodes ont été proposées pour le cas général de modèles VSM. Dans cet exposé, nous permettons des corrélations entre les innovations de rendements d'actifs et celles de la dynamique de la volatilité. Nous nous intéressons au modèle VSM dans un contexte bayésien en appliquant une distribution inverse Wishart. L'algorithme d'échantillonnage par tranches dans l'algorithme de Gibbs est proposé pour échantillonner les paramètres persistants et les variables latentes. Puisque la méthode de Metropolis-Hastings (M-H) n'est pas utilisée, notre algorithme est plus efficace et plus facile d'utilisation.

**[MS-64]****Mean-variance Portfolio Optimization in a Market with Stochastic Correlations  
Gestion moyenne-variance dans un marché avec corrélations stochastiques**François WATIER & René FERLAND, *Université du Québec à Montréal*

We consider a mean-variance investment problem in a continuous-time Brownian motion setting

Nous considérons un problème d'optimisation de type moyenne-variance dans un contexte où la dynamique des

where the stock prices' correlations are driven by a Wishart process. We construct an optimal strategy by means of solutions of BSDEs (backward stochastic differential equations). We will also give an explicit analytical expression for the optimal portfolio through solutions of matrix Riccati ODEs.

corrélations entre les prix des titres risqués est modélisée par un processus de Wishart. Nous obtenons une stratégie optimale par le biais de solutions d'équations différentielles stochastiques rétrogrades. Nous verrons de plus que le problème d'optimisation se réduit à la résolution d'équations différentielles ordinaires matricielles de type Riccati.

**[MS-65]**

**Variational Bayesian Learning for Regime-switching Model**

**Apprentissage bayésien variationnel pour modèles à changement de régime**

Hui ZHAO & Paul MARRIOTT, *University of Waterloo*

In the Bayesian analysis framework all inference problems should be answered using the posterior distribution. However, for high dimensional problems this posterior distribution is hard to work with due to the intractability of the integrals. Sampling based methods, such as MCMC, gives a computational solution but can time consuming. As an alternative, Variational Bayesian method provides an analytical approximation. This method tends to be very fast in many practical cases. An application in Regime-Switching time series modelling shows its strength, especially in computational efficiency and accuracy in state-detection or model parameter estimation.

Dans le cadre d'analyse bayésien, tous les problèmes d'inférence devraient être résolus par l'utilisation de la distribution a posteriori. Cependant, il peut être difficile de travailler avec cette distribution dans des problèmes de haute dimension à cause de l'incalculabilité des intégrales. Des méthodes basées sur l'échantillonnage, comme les MCMC, donnent une solution numérique, mais peuvent demander beaucoup de temps. Une alternative, la méthode bayésienne variationnelle, donne une approximation analytique. Cette méthode est souvent rapide en pratique. Une application à des séries chronologiques avec changement de régime montre son potentiel, particulièrement en termes de temps de calcul et de précision de la détection des états et de l'estimation des paramètres.

**[MS-66]**

**Using Count Distributions to Assess the Impact of Graduate Studies on the Birth Rate for Students**

**Utilisation des distributions de comptage pour évaluer l'impact des études supérieures sur le taux de natalité des étudiants**

Jean-Philippe BOUCHER, *Université du Québec à Montréal*

A variety of government agencies collect the population birth rate. However, no statistics regarding the parents' occupation at childbirth are gathered. It is possible to obtain an estimate for the birth rate from data gathered within a survey conducted by the *Conseil national des cycles supérieurs*. We extend a non-homogeneous Poisson model by considering models that incorporate heterogeneity, compound sums, and zero-modified distributions. We show that the birth rate for graduate students is much lower than that for the general population.

Diverses agences gouvernementales collectent le taux de fécondité de la population. Toutefois, aucune statistique n'est collectée au niveau de l'occupation des parents lors de la naissance d'un enfant. À partir d'informations obtenues dans un sondage effectué par le Conseil national des cycles supérieurs, il est possible d'obtenir un estimé du taux de natalité. Pour l'analyse, nous partons d'un modèle de Poisson non homogène que nous généralisons pour tenir compte de facteurs hétérogénéité, de sommes composées ou de distributions altérées à zéro. Nous montrons que le taux de natalité des étudiants aux cycles supérieurs est très inférieur à celui de la population en général.

**Session 04I****Lundi 24 mai • Monday, May 24, 14:30 – 15:45****VAN 1883**Distribution Theory and Copulas  
Théorie des distributions et copules**[MS-67]****Inference for a Leptokurtic Symmetric Family of Distributions Represented by the Difference of Two Gamma Variates****Inférence pour une famille de distributions symétriques et leptokurtiques représentée par la différence de deux variables aléatoires gamma**Maciej AUGUSTYNIAK & LOUIS DORAY, *Université de Montréal*

We introduce a family of leptokurtic symmetric distributions represented by the difference of two gamma variates. Properties of this family are discussed. The Laplace and normal distributions arise as special cases of this family. We propose a two-step method for fitting data to this family. First, we perform a test of symmetry, and second, we estimate the parameters by minimizing the quadratic distance between the real parts of the empirical and theoretical characteristic functions. The quadratic distance estimator obtained is consistent, robust and asymptotically normally distributed. A simulation study is provided to illustrate the theory.

Nous présentons une famille de distributions symétriques et leptokurtiques représentées par la différence de deux variables aléatoires gamma. Les propriétés de cette famille sont présentées. Les distributions Laplace et normale sont des cas particuliers de cette famille. Nous proposons une méthode à deux étapes qui permet d'ajuster un jeu de données à cette famille. Premièrement, nous effectuons un test de symétrie et, deuxièmement, nous estimons les paramètres en minimisant la distance quadratique entre les parties réelles des fonctions caractéristiques empiriques et théoriques. L'estimateur qui en découle est convergent, robuste et asymptotiquement normal. Une étude de simulation est menée afin d'analyser les propriétés des estimateurs obtenus.

**[MS-68]****On Tail Index Estimation of a Heavy Tailed Distribution****Estimation de l'indice caudal d'une distribution à ailes relevées**Holly FALSETTI & Mei Ling HUANG, *Brock University*

This paper explores techniques of estimating the tail index of the heavy tailed distribution. We propose a new estimator based on exploring second order behaviour in order to reduce the bias of the classical Hill estimator. Asymptotic, distributional properties of the new and existing estimator are derived. We assume that the threshold levels vary as the sample size increases in such a way that optimal threshold values increase as well. These theoretical results are confirmed by the results of the simulations, and the study of an example.

Cet article explore des techniques d'estimation de l'indice caudal d'une distribution à ailes relevées. Nous proposons un nouvel estimateur basé sur le comportement de second ordre afin de réduire le biais de l'estimateur de Hill classique. Les propriétés distributionnelles asymptotiques du nouvel estimateur et de l'estimateur existant sont obtenues. Nous supposons que les niveaux de seuil forment une suite de valeurs telles que la valeur du seuil optimal augmente avec la taille de l'échantillon. Ces résultats théoriques sont confirmés par des résultats de simulation et l'étude d'un exemple réel.

**[MS-69]****Nonparametric Covariate Adjustment in Conditional Copulas  
Ajustement non paramétrique des covariables pour des copules conditionnelles**Elif ACAR, Radu CRAIU & Fang YAO, *University of Toronto*

We study nonparametric inference for the copula parameter in conditional copulas. Our interest is to characterize the covariate effects on the strength of dependence when modeling the joint behavior of two or more random variables. Using local polynomial techniques we propose a nonparametric estimation procedure for the functional relationship between the copula parameter and the covariate. Our second contribution is to introduce a novel copula selection method based on cross-validated prediction errors. We demonstrate the performance of our estimation and copula selection method on both simulated and real datasets.

Nous étudions l'inférence non paramétrique pour le paramètre de copule pour des copules conditionnelles. Notre but est de caractériser les effets covariables sur la force de la dépendance lorsque le comportement conjoint d'au moins deux variables aléatoires est modélisé. À partir de techniques polynomiales locales, nous proposons une méthode d'estimation non paramétrique pour la relation fonctionnelle entre le paramètre de copule et la covariable. Notre seconde contribution est de présenter une nouvelle méthode de sélection de copules basée sur les erreurs de prévisions par validation croisée. Nous démontrons la performance de nos méthodes d'estimation et de sélection de copules sur des jeux de données réels et simulés.

**[MS-70]****Identifiability and Estimation of the Sign of a Covariate Effect in the Competing Risks Model  
Identifiabilité et estimation du signe d'un effet covariable dans un modèle à risques compétitifs**Ralf WILKE, *University of Nottingham*, Simon Lo, *Lingnan University, Hong Kong*

We show identifiability of the sign of a covariate effect on the marginal distributions of latent competing variables. In contrast to the related literature, our method does not require complete knowledge about the dependence structure (copula function) other than that it is independent of the varying covariate. We establish a general link between observable changes in subdistributions (cumulative incidence curves) and the sign of changes in marginal distributions (causal treatment effect). Our method is computationally very simple and our simulations suggest that it identifies and consistently estimates the sign of the treatment effect for large sets of duration times.

Nous démontrons l'identifiabilité du signe d'un effet covariable sur les distributions marginales de variables latentes en compétition. Contrairement à la littérature à ce sujet, notre méthode ne nécessite pas une connaissance complète de la structure de dépendance (fonction copule). Nous établissons un lien général entre les changements observables dans les sous-distributions (courbes d'incidence cumulative) et le signe des changements dans les distributions marginales (effet causal de traitement). Notre méthode nécessite peu de puissance de calculs et nos simulations suggèrent qu'elle identifie et estime de façon cohérente le signe de l'effet traitement pour de grands ensembles de temps de durée.

**[MS-71]****Comparison of Semiparametric Maximum Likelihood Estimation and Two-Stage Semiparametric Estimation in Copula Models****Comparaison de l'estimation du maximum de vraisemblance semi-paramétrique et de l'estimation semi-paramétrique en deux étapes pour des modèles de copules**Yildiz YILMAZ, *Samuel Lunenfeld Research Institute*, Jerald LAWLESS, *University of Waterloo*

We discuss the performance of two semiparametric estimation procedures for copula models based

Nous discutons de la performance de deux méthodes d'estimation semi-paramétrique pour des modèles de copules

on censored data: semiparametric maximum likelihood and two-stage semiparametric, in which copula parameters are estimated without parametric assumptions on the marginal distributions. The two-stage procedure requires less computation and the two-stage estimators of the copula parameter and marginal distributions are shown to be generally about as good as the semiparametric maximum likelihood estimators of them. When the copula model is misspecified, semiparametric maximum likelihood estimators of the marginal distributions are biased. The estimator of Kendall's tau is less biased than the two-stage estimator, however.

basés sur des données censurées : le maximum de vraisemblance semi-paramétrique, et l'estimation semi-paramétrique à deux étapes, où les paramètres de copule sont estimés sans hypothèses paramétriques sur les distributions marginales. La méthode à deux étapes nécessite moins de calculs et produit des estimateurs généralement aussi bons que ceux obtenus par maximum de vraisemblance semi-paramétrique. Lorsque le modèle de copule est mal spécifié, les estimateurs du maximum de vraisemblance semi-paramétrique des distributions marginales sont biaisés. Cependant, l'estimateur du Tau de Kendall est moins biaisé que l'estimateur à deux étapes.

**Session 04J**                      **Lundi 24 mai • Monday, May 24, 14:30 – 16:00**                      **VAN Hall**

Poster Session 1  
Séance d'affichage 1

[MS-72]

**Validation of a Micro-simulation Monte-Carlo Markov Model for Schizophrenia**

**Validation d'un modèle de Markov pour la schizophrénie avec microsimulations de Monte-Carlo**

Alice DRAGOMIR & Jean-François ANGERS, *Université de Montréal*, Jean-Eric TARRIDE, *McMaster University*, Sylvie PERREAULT, *Université de Montréal*

**Context:** To evaluate the new drug therapies, we developed a Markov model using the Monte Carlo micro-simulation in order to model the natural course of schizophrenia. **Objectives:** To determine the validity of the developed Monte-Carlo Markov model. **Methods:** The model was based on data from the RAMQ databases. Internal and external validation was performed. **Results:** The characteristics of subjects from the simulated cohort were similar to those of the study cohort. The number of simulated disease states was consistent with the number of different observed disease states. The mean of the model's predicted probabilities of transition found in the simulated cohort were consistent with those published. **Conclusion:** Our results show both internal and external validity of the model.

**Contexte :** Pour évaluer les nouveaux médicaments pour la schizophrénie, nous avons développé un modèle de Markov avec des micro-simulations de Monte-Carlo, afin de modéliser le cours naturel de la schizophrénie. **Objectifs :** Déterminer la validité du modèle de Markov développé. **Méthodes :** Le modèle a été basé sur les données de la Régie de l'assurance maladie du Québec. La validité interne et externe a été évaluée. **Résultats :** Les caractéristiques des sujets de la cohorte simulée ont été similaires à ceux de la cohorte à l'étude. Le nombre d'états simulés de la schizophrénie est compatible avec le nombre d'états différents observés. Les moyennes des probabilités de transition prédites dans la cohorte simulée ont été également conformes à ceux publiés dans la littérature. **Conclusion :** Nos résultats démontrent la validité interne et externe du modèle.

**[MS-73]****Application of the Expectation-Maximization (EM) Algorithm for Mixture Models  
Application de l'algorithme espérance-maximisation (EM) aux modèles de mélanges**Man-Kee Maggie CHU, *University of Western Ontario*

A developmental trajectory describes the course of behaviour over time, and identifying multiple trajectories within an overall developmental process would allow us to focus on subgroups of particular interest. This research introduces a SAS macro program that identifies trajectories by using the Expectation-Maximization (EM) algorithm to fit semiparametric mixtures of logistic distributions to longitudinal binary data. For performance comparison, we consider a full maximization (a SAS procedure PROC TRAJ) and standard EM, as well as two other EM-based algorithms for speeding up convergence. The EM-based methodology is illustrated with a longitudinal dataset involving adolescents smoking behaviours.

Une trajectoire développementale décrit la variation d'un comportement à travers le temps, et l'identification de multiples trajectoires à l'intérieur d'un processus développemental global nous permet de nous concentrer sur des sous-groupes d'intérêt particulier. Cette recherche présente une macro SAS identifiante des trajectoires en utilisant l'algorithme espérance-maximisation (EM) pour ajuster des mélanges semi-paramétriques de distributions logistiques à des données binaires longitudinales. Nous comparons la performance de la maximisation complète (procédure SAS PROC TRAJ), de l'algorithme EM standard, et de deux autres algorithmes basés sur le EM afin d'accélérer la convergence. La méthode basée sur le EM est illustrée à l'aide de données longitudinales de tabagisme chez les adolescents.

**[MS-74]****Increasing the Efficiency of Group Sequential Methods by Auxiliary Information  
Augmenter l'efficacité de méthodes séquentielles en groupes au moyen d'information auxiliaire**Abdulkadir HUSSEIN, *University of Windsor*, Hassen MUTTLAK & ESSAM AL-SAWI, *King Fahd University of Petroleum and Minerals*

Group sequential methods are designed to save sample and enable us to make decisions as early as possible. We propose an efficient method for incorporating auxiliary information into group sequential testing procedures. We will compare power and average sample numbers of the proposed tests to the usual group sequential tests of Pocock (1977) and O'Brien & Fleming (1979). We apply the methods to data set from HIV clinical trials.

Les méthodes séquentielles en groupes sont utilisées afin d'économiser les observations et de nous permettre de prendre une décision le plus tôt possible. Nous proposons une méthode efficace pour incorporer des informations auxiliaires à des procédures de test séquentielles. Nous allons comparer la puissance et la taille échantillonnale moyenne des tests proposés aux tests séquentiels en groupes habituels de Pocock (1977) et d'O'Brien & Fleming (1979). Nous appliquons les méthodes à des données d'essais cliniques VIH.

**[MS-75]****Mixtures of Latent Class Analyzers for Incomplete Categorical Data  
Mélanges d'analyseurs de classes latentes pour données catégorielles incomplètes**Ryan BROWNE, *University of Guelph*, Paul McNICHOLAS, *University of Guelph*

In taste testing, a product is assessed based on a consumer's perception of the senses when tasting the product. To understand consumer behaviour, preference maps are built by assessing 12 to 30 commercial products. In a complete block design, every taster gets to test every product, but several products do not facilitate this sampling plan. For

Dans des tests gustatifs, un produit est évalué à partir de la perception de consommateurs goûtant le produit. Afin de comprendre le comportement des consommateurs, des cartes préférentielles sont construites en évaluant de 12 à 30 produits commerciaux. Dans un plan à blocs complets, chaque goûteur teste chacun des produits. Malheureusement, plusieurs produits ne facilitent pas l'utilisation de ce plan

example, when tasting wines, a taster can only have three or four samples before “fatigue” sets in. The nature of the resulting data, incomplete and categorical, leads to an interesting statistical problem. Possible solutions are explored and demonstrated on real and simulated data.

d'échantillonnage. Par exemple, lors de la dégustation de vins, un goûteur ne peut goûter que trois ou quatre échantillons avant de souffrir de « fatigue ». La nature incomplète et catégorielle des données résultantes mène à un problème statistique intéressant. Des solutions possibles sont explorées et illustrées sur des données réelles et simulées.

[MS-76]

**Identifiability and Stationarity Conditions for General Multivariate GARCH Models**  
**Conditions d'identifiabilité et de stationnarité pour modèles GARCH généraux multivariés**

Weibin JIANG, Hao YU & Reg KULPERGER, *University of Western Ontario*

To prove strong consistency and asymptotic normality of quasi-maximum likelihood estimator for general multivariate GARCH models, two important intermediate results are model identifiability and strict stationarity. We construct counterexamples for some identifiability results in literature and provide a better necessary and sufficient condition. We also give some conditions under which we can prove the ergodicity of the conditional variance with an application of theorems by Meyn & Tweedie (2009). Stationarity follows assuming the model is initiated from a stationary probability measure. Under those conditions, asymptotic theory can be proved using standard compactness argument and Taylor expansion.

Afin de prouver la forte cohérence et la normalité asymptotique de l'estimateur de quasi-maximum de vraisemblance pour modèles GARCH multivariés généraux, deux résultats intermédiaires importants sont l'identifiabilité et la stricte stationnarité du modèle. Nous construisons des contre-exemples pour certains résultats d'identifiabilité de la littérature et donnons une meilleure condition nécessaire et suffisante. Nous donnons aussi quelques conditions sous lesquelles peut être prouvée l'ergodicité de la variance conditionnelle, en appliquant les théorèmes de Meyn & Tweedie (2009). Ceci implique la stationnarité si le modèle a été initialisé avec une mesure de probabilité stationnaire. Sous ces conditions, la théorie asymptotique peut être prouvée au moyen d'un argument de compacité standard et d'un développement en série de Taylor.

[MS-77]

**Simultaneous Association Test on Multiple Traits**  
**Test d'association simultané pour de multiples traits**

Michael McDONALD & Zeny FENG, *University of Guelph*

Recent biological technology allows researchers to perform genome-wide association studies on a dense panel SNPs. Methods for association studies have been focused on the association of a single trait at a time. In this paper, we will present a novel method that is based on the generalized quasi-likelihood scoring (GQLS) approach for simultaneous association test on multiple traits. The method will be applied to a data set consisting of 10K SNPs and 24 estimated breeding values (EBVs), where EBVs are considered as important economic trait in animal breeding and dairy science.

La technologie biologique actuelle permet aux chercheurs de mener des études d'association sur tout le génome sur un panel dense de SNP. Les méthodes pour études d'association sont généralement conçues pour l'association d'un trait à la fois. Dans cet article, nous allons présenter une méthode originale basée sur l'approche des scores de quasi-vraisemblance généralisée (SQVG) pour tester simultanément l'association de multiples traits. La méthode sera appliquée à un jeu de données de 10k SNP et 24 valeurs reproductives estimées (VRE), où les VRE sont considérées comme un trait économique important en zootechnie et en technologie laitière.

**[MS-78]****Kendall's Tau and Sign Test as Tests for Trend in Time Series Data****Le Tau de Kendall et le test des signes en tant que tests de tendance en analyse de séries chronologiques**Xiaolei CHEN, Paul CABILIO & Ying ZHANG, *Acadia University*

When using nonparametric tests to test time series data for trend, most require that the data in the model to be independent and identically distributed. In the case of an autoregressive process, the successive observations in the series are not stochastically independent. As a result the usual Kendall's tau test is not valid. The same considerations hold for the Sign test. This study discusses the effect of correlated error on significance levels and power of Kendall's tau and the Sign test. The modification of these tests are developed to test for trend in the AR(1) model.

Lorsque des tests non paramétriques sont utilisés pour tester la présence d'une tendance dans des données de séries chronologiques, la plupart nécessitent que les données du modèle soient indépendantes et identiquement distribuées. Dans le cas de processus autorégressifs, les observations successives dans les séries ne sont pas stochastiquement indépendantes, et le test du tau de Kendall habituel n'est donc pas valide. Le test des signes a le même problème. Cette étude discute des effets d'une erreur corrélée sur le niveau de signification et la puissance statistique des tests du tau de Kendall et des signes. Une modification à ces tests est développée afin de tester la tendance dans un modèle AR(1).

**[MS-79]****Designing Experiments to Assess the Space-time Dynamics of Plant Diseases****Planification d'expériences afin d'évaluer la dynamique spatio-temporelle de maladies des plantes**Daria MARTCHENKO, Rob DEARDON & Paul McNICHOLAS, *University of Guelph*

Infectious diseases of plants can cause havoc both economically and environmentally. In order to control such disease it is desirable to understand the dynamics, which are generally spatio-temporal in nature, of the disease spread. In some situations experiments may be carried out to assess such dynamics. However, little work has been done exploring how best to design such experiments. Here we show the results from a number of simulation studies that address the questions of spatial layout and temporal resolution over which such spatial data needs to be collected in order to understand the spatio-temporal dynamics of the disease spread.

Les maladies infectieuses des plantes peuvent être dévastatrices économiquement et environnementalement. Pour contrôler de telles maladies, il est désirable de comprendre la dynamique de leur propagation, habituellement de nature spatio-temporelle. Dans certaines situations des expériences peuvent être menées afin d'évaluer cette dynamique. Cependant, peu de travaux ont exploré comment planifier de façon optimale ces expériences. Nous montrons les résultats de plusieurs études de simulation répondant aux questions relatives à la disposition spatiale et la résolution temporelle nécessaires afin de comprendre la dynamique spatio-temporelle de la propagation de maladies.

**Session 04K****Lundi 24 mai • Monday, May 24, 14:30 – 16:00****VAN Hall****Case Study 1 “The Effects of Photo-interpreted Variables in the Estimation of Stand-level Merchantable Volumes in the Province of Québec”****Étude de cas 1 « L'effet de variables photo-interprétées sur l'estimation du volume marchand à l'échelle du peuplement dans la province du Québec »**

The Quebec forest inventory uses aerial photos of target forest management units at the first step of the sampling process. Photo interpreters then divide the territory into homogeneous polygons and, for each of those polygons, estimate the value of

L'inventaire forestier du Québec utilise des photos aériennes d'unités d'aménagement forestier comme une première étape d'un processus d'échantillonnage. Les photo-interprètes divisent alors le territoire en polygones homogènes et, pour chacun de ces polygones, estiment la valeur

specific biological and physical variables. Once the photo-interpretation is completed, polygons with similar photo-interpreted characteristics are grouped into strata. The stratification is used to direct the field inventory in which plots are established and in which forest properties are measured. The primary objective of the case study is to predict different stand-level volumes using photo-interpreted, climatic, geographical and ecological classification variables.

de variables biologiques et physiques spécifiques. Une fois la photo-interprétation terminée, les polygones avec des caractéristiques photo-interprétées similaires sont groupés en strates. La stratification est utilisée pour diriger l'inventaire du terrain dans lequel les parcelles sont établies et dans lequel les propriétés forestières sont mesurées. L'objectif primaire de cette étude de cas est de prédire les volumes à l'échelle du peuplement en utilisant les variables de classification climatiques, géographiques et écologiques photo-interprétées

**Session 04L                      Lundi 24 mai • Monday, May 24, 14:30 – 16:00                      VAN Hall**

**Case Study 2 “Metabolism of Bradykinin and Endogenous des-Arg<sup>9</sup>-Bradykinin in Human Plasma: Contribution to the Pathophysiology of Angioedema Associated with ACE Inhibitors”**

**Étude de cas 2 « Métabolisme de la bradykinine et de la des-Arg<sup>9</sup>-bradykinine endogène dans le plasma humain : contribution à la physiopathologie de l'angio-œdème associé aux inhibiteurs de l'enzyme de conversion de l'angiotensine »**

Angiotensin converting enzyme (ACE) inhibitors are an important class of drugs used over the last 30 years to treat cardiovascular diseases such as hypertension and congestive heart failure. These drugs cause potentially lethal side effects, including angioedema, in some patients. The objective of this case study is to compare kinetic profiles of two peptides between cases and controls in the search for evidence of a possible anomaly in the metabolism of these peptides for cases, those with and those without a history of angioedema, who have been treated with ACE inhibitors.

Une classe importante de médicaments utilisés depuis près de 30 ans dans le traitement des affections cardiovasculaires telles l'hypertension ou l'insuffisance cardiaque provoque chez certains patients des effets secondaires potentiellement mortels dont l'angio-œdème. L'objectif de ce projet est de comparer le profil cinétique de deux peptides chez différents groupes de sujets afin de mettre en évidence une anomalie possible des métabolismes de ces peptides affectant les patients traités par cette classe de médicament ayant une histoire d'angio-œdème

**Session 05A                      Lundi 24 mai • Monday, May 24, 16:30 – 17:30                      VAN 1245A**

**Special Invited Session of the Survey Methods Section—Cell Phone Surveys**

**Allocution de l'invité du président du groupe des méthodes d'enquête — Enquêtes à partir des cellulaires**

[MS-80]

**Dual Frame Landline and Cell Phone Surveys**

**Enquêtes à partir d'une base de sondage duale de téléphones terrestres et cellulaires**

J Michael BRICK, *Westat*

With the dramatic increase in the number of households with only cell phones, U.S. telephone surveys have begun to sample both landline and cell phone numbers. These surveys identified the potential for serious biases due to nonsampling errors. In

Avec l'augmentation spectaculaire du nombre de ménages n'utilisant que des téléphones cellulaires, les enquêtes téléphoniques conduites aux États-Unis ont commencé à échantillonner à la fois les numéros de lignes téléphoniques terrestres et les numéros de téléphones cellulaires. Ces en-

this paper, we describe the potential effects of non-sampling errors in dual frame telephone surveys; discuss sample design issues such as whether to screen for cell-only households or to conduct a survey with full overlap; and present weighting methods that may reduce the effects of the nonsampling errors on the estimates.

quêtes ont identifié les possibilités d'important biais à cause d'erreurs non dues à l'échantillonnage. Dans cet article, nous décrivons les effets potentiels d'erreurs non dues à l'échantillonnage dans les enquêtes téléphoniques à base de sondage duale; nous discutons de questions techniques de sondage telles le choix entre filtrer les ménages n'utilisant que de téléphones cellulaires ou mener une enquête avec chevauchement complet; et nous présentons des méthodes de pondération qui pourraient réduire les effets d'erreurs non dues à l'échantillonnage sur l'estimation.

**Session 05B      Lundi 24 mai • Monday, May 24, 16:30 – 18:00      VAN 2289A**

**Methods for Longitudinal and Spatial Categorical Data**

**Méthodes pour des données catégorielles longitudinales et spatiales**

[MS-81]

**Spatio-temporal Modeling of Outbreaks of Mountain Pine Beetle**

**Modélisation spatio-temporelle d'infestations de dendroctone du pin ponderosa**

Yanbing ZHENG, *University of Kentucky*, Brian AUKEMA, *Canadian Forest Service, Natural Resources Canada and University of Northern British Columbia*

The mountain pine beetle is forest insect that undergoes intermittent population eruptions. At epidemic levels, it must kill its host tree to reproduce. Currently, an outbreak in British Columbia and Alberta, Canada covers over 14 million hectares of mature pine forests. We consider two different ways of modelling mountain pine beetle outbreaks across space and over time in British Columbia, Canada. One approach is a spatial-temporal autologistic regression model in the framework of Markov random fields. The other approach is a generalized linear mixed model with spatial-temporal random effects. We devise computationally feasible algorithms for Bayesian inference in both approaches.

Le dendroctone du pin ponderosa est un insecte dont la population connaît des vagues de croissance intermittentes. Au niveau épidémique, l'insecte doit tuer son arbre hôte afin de se reproduire. Une éclosion en Colombie-Britannique et en Alberta affecte présentement plus de 14 millions d'hectares de forêts de pin adultes. Nous considérons deux façons différentes afin de modéliser les éclosions de dendroctone du pin ponderosa à travers l'espace et le temps en Colombie-Britannique. Une approche est un modèle de régression autologistique spatio-temporel dans le cadre de champs aléatoires markoviens. L'autre approche est un modèle mixte linéaire généralisé à effets aléatoires spatio-temporels. Nous obtenons des algorithmes calculables pour l'inférence bayésienne sous les deux approches.

[MS-82]

**Joint Spatial Modeling of Recurrent Infection and Growth**

**Modélisation spatiale conjointe de récurrences d'infections et de croissance**

Farouk NATHOO, *University of Victoria*

We present statistical methodology for longitudinal studies in forestry, where trees are subject to recurrent infection, and the hazard of infection depends on tree growth over time. Challenges arise for statistical analysis in this setting, with sampling

Nous présentons une méthode statistique pour des études longitudinales en foresterie, où les arbres sont sujets à une infection récurrente, et où le risque d'infection dépend de la croissance de l'arbre dans le temps. Plusieurs défis statistiques sont présents dans ce cadre, notamment des plans

schemes leading to panel data, exhibiting dynamic spatial variability, and incomplete covariate histories for hazard regression. A joint model for infection and growth is developed, where a mixed non-homogeneous Poisson process is linked with a spatially dynamic nonlinear model representing height growth. The methodology is applied for analysis in an eleven year study of recurrent weevil infestation of white spruce in British Columbia.

d'échantillonnage menant à des données de panel, une variabilité spatiale dynamique, ainsi que des covariables incomplètes. Nous développons un modèle conjoint d'infection et de croissance dans lequel un processus de Poisson mixte non homogène est lié à un modèle non linéaire spatialement dynamique représentant la croissance en hauteur. La méthode est appliquée à l'analyse d'une étude étalée sur 11 ans concernant l'infestation d'épinettes blanches par des charançons en Colombie-Britannique.

### [MS-83]

#### **An Appraisal of Methods for the Analysis of Longitudinal Categorical Data with MAR Drop-outs Une évaluation des méthodes d'analyse de données catégorielles longitudinales avec attrition aléatoire**

R. Jeanette O'HARA HINES, *University of Waterloo*, W.G.S. HINES, *University of Guelph*

Methods are compared for analyzing longitudinal ordinal categorical data with missing-at-random dropouts. Two maximum-likelihood (ML) methods employ marginal global odds ratios to model associations. Two generalized estimating equations (GEE) methods use Cholesky-decomposed standardized residuals to model the association structure, while another three extend methods for longitudinal binary data with the association structures modelled using Gaussian estimation, multivariate normal estimating equations or conditional residuals. Simulations compare biases, variances and convergence rates when the association structure is misspecified. The GEE methods Cond and ML-norm had relatively good convergence rates and mean squared errors for all sample sizes (80, 120, 300) considered.

Des méthodes d'analyse de données longitudinales ordinales catégorielles avec attrition aléatoire sont comparées. Deux méthodes du maximum de vraisemblance (MV) utilisent des rapports de cotes marginaux globaux afin de modéliser les associations. Deux méthodes d'équations d'estimation généralisées (ÉEG) utilisent des résidus standardisés décomposés par Cholesky afin de modéliser la structure d'association, alors que trois autres généralisent des méthodes pour données longitudinales binaires et modélisent les structures d'association par estimation gaussienne, équations d'estimation normales multivariées ou résidus conditionnels. Les biais, variances et taux de convergence sont comparés par simulation lorsque la structure d'association est mal spécifiée. Les méthodes ÉEG Cond et MV-norm ont généralement de bons taux de convergence et erreurs quadratiques moyennes pour les tailles échantillonnales (80, 120, 300) considérées.

## **Session 05C      Lundi 24 mai • Monday, May 24, 16:30 – 18:00      VAN 2291A**

### **Probability, Regression, and Time Series**

### **Probabilité, régression et séries chronologiques**

### [MS-84]

#### **Multivariate GARCH : Some Foundations GARCH multivarié : quelques fondements**

Reg KULPERGER, Hao YU & Weibin JIANG, *University of Western Ontario*

Univariate GARCH models are well understood in the sense of parameter identifiability and necessary and sufficient conditions for stationarity. Multivariate GARCH models do not enjoy this state of knowledge, and only relatively few papers have

Les modèles GARCH univariés sont bien compris dans le sens de l'identifiabilité des paramètres et des conditions nécessaires et suffisantes pour la stationnarité. Nos connaissances au sujet des modèles GARCH multivariés ne sont malheureusement pas à ce niveau, et relativement peu

been written on these foundational aspects.

We study these aspects for multivariate GARCH models. The natural one-dimensional tools to study stationarity are not easy to use for processes such as dvec and BEKK do not have a natural iteration for the conditional variances. Other tools have to be used. This talk will discuss some of these tools and sufficient condition results for GARCH stationarity.

[MS-85]

### Second Order Inference of Time Series

#### Inférence de second ordre pour séries chronologiques

Wei Wu, *University of Chicago*

I will present a systematic theory for asymptotics of second order statistics of time series. In particular, I will talk about asymptotic problems in the estimation of covariance, spectral densities, covariance matrices and predictions.

d'articles ont été écrits sur ces aspects fondamentaux.

Nous étudions ces aspects des modèles GARCH multivariés. Les outils unidimensionnels naturels pour étudier la stationnarité ne sont pas faciles à utiliser pour les processus comme dvec et BEKK n'a pas d'itération naturelle pour les variances conditionnelles. D'autres outils doivent être utilisés. Cet exposé discutera de certains de ces outils et de résultats sur les conditions suffisantes pour la stationnarité GARCH.

[MS-86]

### Functional Central Limit Theorems in Linear Regression Models

#### Théorèmes limites centraux fonctionnels pour modèles de régression linéaires

Yuliya MARTSYNYUK, *University of Manitoba*

Stochastic processes are introduced in  $D[0, 1]$  for the unknown slope and intercept in linear error-invariables and regression models with possibly infinite variance data. They contain the traditional estimators for these parameters. Sup-norm approximations in probability and, as a consequence, functional central limit theorems are established for Studentized versions of the introduced processes. The thus obtained data-based functional central limit theorems open up new possibilities for constructing various asymptotic confidence intervals for the slope and intercept.

Nous présentons des processus stochastiques sur  $D[0, 1]$  pour la pente et l'ordonnée à l'origine inconnues de modèles linéaires à erreur dans les variables et de modèles de régression à variance possiblement infinie. Ils contiennent les estimateurs traditionnels pour ces paramètres. Des approximations en probabilité de la norme-sup et, conséquemment, des théorèmes centraux limites fonctionnels sont obtenus pour les versions studentisées des processus présentés. Les théorèmes centraux limites fonctionnels basés sur les données ainsi obtenues ouvrent de nouvelles possibilités pour la construction de divers intervalles de confiance asymptotiques pour la pente et l'ordonnée à l'origine.

## Session 05D      Lundi 24 mai • Monday, May 24, 16:30 – 18:00      VAN 2809A

What is the Role of a Statistical Association in Insuring Rigour in the Publication of Statistical Results in Media?

Quel est le rôle d'une association de statisticiens pour assurer la rigueur dans la publication médiatique de résultats statistiques ?

Statistics are widely used in the media. Survey results are most prominent but research results from economic and demographic data, etc., are also

Les statistiques sont monnaie courante dans les médias. Les résultats de sondages sont les plus présents mais on trouve également des résultats de recherches, des données écono-

found. Sometimes, statistics are used to support claims made by a journalist, an editor or a columnist; occasionally, the statistics themselves are the news item.

How does the average reader know if the information is reliable? What roles can statistical associations play in the research of responsible use of statistics for information purposes? Three panellists are invited to debate the subject amongst themselves and to react to comments from the public.

miques, démographiques, etc. Parfois, la statistique est utilisée pour appuyer le propos d'un journaliste, d'un éditorialiste ou d'un chroniqueur; parfois, elle crée elle-même la nouvelle.

Pour un lecteur moyen, comment savoir si une donnée est fiable? Quels rôles peuvent jouer les associations de statisticiens dans la recherche d'une utilisation responsable de la statistique à des fins d'information? Trois panélistes seront invités à débattre du sujet entre eux et à réagir aux commentaires du public.

**Session 05E      Lundi 24 mai • Monday, May 24, 16:30 – 18:00      VAN 2811A**  
**Financial Modelling in Actuarial Sciences**  
**Modèles financiers en actuariat**

[MS-87]

**Pricing and Hedging Equity-indexed Annuities Using Risk Measures**

**Fixation de prix et réplcation d'annuités indexées aux capitaux propres au moyen de mesures de risques**

Patrice GAILLARDETZ, *Concordia University*

In this talk, we consider the pricing of equity-indexed annuities using risk measures and present investment strategies underlying these valuations. Since these insurance products involve mortality as well as financial risks, we combine the actuarial and financial approaches to protect insurance companies against the unhedgeable mortality risk. For a given premium, risk measures are minimized to obtain investment strategies for equity-indexed annuities. The investment strategies for dynamic risk measures are also obtained. These investment strategies are based on certain assumptions and produce discrete hedging errors. Thus, the different risk measures will be compared using their respective hedging errors.

Dans cet exposé, nous considérons la fixation de prix d'annuités indexés sur actions au moyen de mesures de risque et présentons des stratégies d'investissement sous-jacentes. Puisque les produits d'assurance impliquent des risques de mortalité en plus des risques financiers, nous combinons les approches financières et actuarielles afin de protéger les compagnies d'assurances du risque de mortalité ne pouvant être couvert. Pour une prime donnée, les mesures de risques sont minimisées afin d'obtenir des stratégies d'investissement pour annuités indexées sur actions. Les stratégies d'investissement pour mesures de risque dynamiques sont aussi obtenues. Ces stratégies d'investissement sont basées sur certaines hypothèses et produisent des erreurs de couverture discrètes. Les différentes mesures de risques sont donc comparées au moyen de leurs erreurs de couverture respectives.

[MS-88]

**A Generating Function Approach in the Calibration of Vasicek and CIR Bond Price Models**

**Une approche à base de fonctions génératrices pour calibrer les modèles de Vasicek et CIR pour le prix d'obligations**

Marianito RODRIGO, *ITAM, Mexico*, Rogemar MAMON, *University of Western Ontario*

A new method to calibrate the Vasicek and CIR models to observed market data is proposed. We introduce a generating function and develop recur-

Nous proposons une nouvelle méthode de calibration des modèles CIR et de Vasicek à partir de données de marché. Nous présentons une fonction génératrice et développons

sive formulae involving the market data and the derivatives of this generating function. The model parameters are found by solving a system of linearly independent equations arising from the recursive formulae. The proposed technique is tested numerically. For the CIR model the polylogarithm function is employed to fully evaluate the derivative of the generating function. We show that our method is accurate in recovering the model parameters.

des formules récursives impliquant les données de marché et les dérivées de cette fonction génératrice. Les paramètres du modèle sont obtenus en résolvant un système d'équations linéairement indépendantes obtenues à partir des formules récursives. La technique proposée est testée numériquement. Pour le modèle CIR la fonction polylogarithmique est utilisée pour évaluer la dérivée de la fonction génératrice. Nous montrons que notre méthode trouve les paramètres du modèle de façon précise.

[MS-89]

### **Risk Measure Based Approach to Optimal Reinsurance**

#### **Une approche basée sur les mesures de risque pour réassurance optimale**

Ken Seng TAN, *University of Waterloo*, Chengguo WENG, *Towson University*

In this presentation I will describe some recent risk measure based approaches to optimal reinsurance. I will describe an intricate role between optimal reinsurance and some commonly used risk measures, including Value-at-Risk (VaR) and Conditional Tail Expectation (CTE). I will also describe an empirical approach to optimal reinsurance designs. This new method has the advantage that it does not require any explicit assumptions on the distribution of the underlying risk. More importantly, it provides a simple and yet practical way of obtaining solution on a wide range of optimal reinsurance models.

Dans cette présentation, je décrirai quelques récentes approches de réassurance optimale basée sur des mesures de risque. Je décrirai un rôle complexe entre réassurance optimale et certaines mesures de risque fréquemment utilisées, incluant la valeur-à-risque (VaR) et l'espérance caudale conditionnelle (ECC). Je décrirai aussi une approche empirique aux plans de réassurance optimaux. Cette nouvelle méthode a l'avantage de ne pas nécessiter d'hypothèses explicites au sujet de la distribution du risque sous-jacent. De façon plus importante, elle constitue une façon simple et pratique d'obtenir la solution à une large gamme de modèles de réassurance optimaux.

**Session 05F**

**Lundi 24 mai • Monday, May 24, 16:30 – 18:00**

**VAN 1203**

Biostatistics

Biostatistique

[MS-90]

### **On the Efficiency of Propensity Score Matching**

#### **De l'efficacité de l'appariement des scores de propension**

Ashkan ERTEFAIE & David STEPHENS, *McGill University*

The propensity score is the basis of several well-established causal inference adjustment methods. Here we assume a parametric model for the propensity score and we regard those parameters as nuisance parameters. In a new result, we demonstrate that the propensity score matching estimator achieves the semiparametric efficiency bound for estimators of the average causal effect (ACE) by

Le score de propension est à la base de plusieurs méthodes bien établies d'ajustement pour inférence causale. Nous supposons ici l'existence d'un modèle paramétrique pour le score de propension et considérons ces paramètres comme des paramètres de nuisance. Nous démontrons que l'estimateur d'appariement des scores de propension atteint la borne d'efficacité semi-paramétrique pour les estimateurs d'effet causal moyen (ECM) en construisant la fonction

building the efficient influence function as a residual of the projection of the score function of the parameters of interest onto the nuisance tangent space. We demonstrate that, in terms of asymptotic variance, the propensity score estimator outperforms inverse probability weighting estimators.

d'influence efficace comme résidu d'une projection de la fonction de score des paramètres d'intérêt sur l'espace tangent des paramètres de nuisance. Nous démontrons que les estimateurs de score de propension ont une plus petite variance asymptotique que les estimateurs à pondération par probabilités inverses.

#### [MS-91]

### Testing Multiple Outcome Variables with a Priori Weights in the Presence of Non-normality and Variance Heterogeneity

### Tester de multiples variables réponses avec des poids a priori en présence de non-normalité et d'hétérogénéité de la variance

Lisa LIX & Tolulope SAJOBI, *University of Saskatchewan*

We investigate procedures for testing differences between treatment and control groups on multiple correlated outcomes when the data are non-normal, group variances are heterogeneous, and the familywise error rate (FWR) is allocated unequally amongst primary and secondary outcomes. The procedures are based on the Bonferroni inequality or Roy's (1958; *Ann. Math. Statist.*, 29: 1177–1187) step-down analysis with robust statistics, as well as resampling methods. Simulation techniques examine the effects of population distribution, group heterogeneity, correlation structure, sample size, and mean configuration. While all procedures generally controlled the FWR, all-variable power differences could be substantial, particularly for non-normal data.

Nous étudions des procédures afin de tester les différences entre les groupes traitement et contrôle pour des réponses multiples corrélées lorsque les données ne sont pas gaussiennes, que les variances de groupes sont hétérogènes, et que le taux d'erreur par famille n'est pas réparti également entre les réponses primaires et secondaires. Les procédures sont basées sur l'inégalité de Bonferroni et sur l'analyse de Roy avec statistiques robustes (1958; *Ann. Math. Statist.*, 29 : 1177-1187), ainsi que sur des méthodes de ré-échantillonnage. Des études de simulation examinent les effets de la distribution de la population, de l'hétérogénéité des groupes, de la structure de corrélation, de la taille échantillonnale et de la configuration des moyennes. Bien que toutes les procédures contrôlent généralement le taux d'erreur, les différences en puissance pour toutes les variables peuvent être substantielles, particulièrement pour des données non gaussiennes.

#### [MS-92]

### Analysis of Health Utility Data

### Analyse de données d'utilité de la santé

Eleanor PULLENAYEGUM, Jean-Eric TARRIDE, Feng XIE & Daria O'REILLY, *McMaster University*

Economic models often require estimates of health utilities. A health utility is a quality weight used to calculate a quality-adjusted life year, where a utility of 1 represents perfect health and a utility of 0 represents death. Health utilities captured using standard instruments have the property that a portion of the population achieves the upper bound of 1. In response to this, many health economists have used Tobit or CLAD models for the analysis of such data. This talk will address why this is not appropriate as well as highlighting the benefits of working collaboratively to resolve the problem.

Les modèles économiques nécessitent souvent une estimation de l'utilité de la santé. L'utilité de la santé est un poids utilisé afin de calculer une année de vie ajustée pour la qualité de vie, où une utilité de 1 représente une parfaite santé et une utilité de 0 représente la mort. Les utilités de santé recueillies habituellement ont la propriété qu'une fraction de la population atteint la borne supérieure de 1. Plusieurs économistes ont donc choisi d'utiliser des modèles Tobit ou CLAD pour l'analyse de telles données. Nous allons voir, dans cet exposé, pourquoi cette approche n'est pas appropriée ainsi que les bénéfices d'une approche collaborative pour résoudre ce problème.

**[MS-93]****Discriminant Analysis for Repeated Measures Data: Effects of Covariance Structure on Bias and Error in Discriminant Function Coefficients****Analyse discriminante pour données de mesures répétées : effets de la structure de covariance sur le biais et l'erreur dans les coefficients de la fonction discriminante**Tolulope SAJOBI, Lisa LIX, William LAVERTY & Longhai LI, *University of Saskatchewan*

Discriminant analysis (DA) is used to classify subjects and/or describe the relative importance of variables for group separation. This study focuses on procedures for repeated measures (RM) data and estimates discriminant function coefficients with parsimonious models for RM mean and covariance structures. Mean square error (MSE) and coefficient bias are studied using simulations. MSE increased with increasing numbers of RMs, but decreased as N increased. Percent bias ranged from < 1% when the covariance structure was correctly specified to > 100% when misspecified. The coefficients of the proposed procedures can be used to identify the RMs that describe group separation.

L'analyse discriminante (AD) est utilisée afin de classer des objets et/ou de décrire l'importance relative de variables pour la séparation de groupes. Cette étude est centrée sur les méthodes traitant des mesures répétées (MR) et estime les coefficients de la fonction discriminante avec des modèles parcimonieux pour les structures de moyenne et covariance des MR. L'erreur quadratique moyenne (EQM) et le biais des coefficients sont étudiés au moyen de simulations. L'EQM augmente avec le nombre de MR, mais diminue lorsque N augmente. Le biais varie de < 1 % lorsque la structure de covariance est bien spécifiée à > 100 % lorsque mal spécifiée. Les coefficients des procédures proposées peuvent être utilisés afin d'identifier les MR décrivant la séparation en différents groupes.

**[MS-94]****Carryover Effects Testing in Heterogeneous Recurrent Event Processes****Tester pour les effets résiduels en processus d'événements récurrents hétérogènes**Candemir CIGSAR & Jerald LAWLESS, *University of Waterloo*

In applications involving multiple systems or individuals, heterogeneity is often apparent. In particular, although individual processes may be adequately described by a modulated Poisson process, the process rate functions may vary across individuals. Such variation is typically due to unmeasured differences in the individuals or the environment in which the processes operate. If carryover effects tests developed for homogeneous processes are used when substantial heterogeneity is present, false indications of an effect can occur, producing an inflated Type 1 error rate. In this study, we focus on testing for carryover effects in such processes under fixed and random effects models.

L'hétérogénéité est souvent apparente dans des applications impliquant plusieurs systèmes ou individus. En particulier, bien que chaque processus puisse individuellement être décrit adéquatement par un processus de Poisson modulé, les fonctions de taux des processus peuvent varier d'un individu à l'autre. Une telle variation est souvent due à des différences non mesurées entre les individus ou entre les environnements dans lesquels opèrent les processus. Si les tests d'effets de rémanence développés pour processus homogènes sont utilisés en présence d'hétérogénéité substantielle, de fausses indications d'un effet peuvent se produire, menant à un taux d'erreur de type I gonflé. Dans cette étude, nous nous intéressons aux tests d'effet de rémanence pour de tels processus avec des modèles à effets fixes et aléatoires.

**[MS-95]****An Approach to Improving Goodness of Fit for Social Network Models****Une approche afin d'améliorer l'adéquation de modèles de réseaux sociaux**Yin LI & Keumhee Chough CARRIERE, *University of Alberta*

Exponential Random Graph Models (ERGMs) has been developed for fitting social network data on

Les modèles de graphes aléatoires exponentiels (MGAE) ont été développés afin d'ajuster des données de réseaux so-

both static and dynamic levels. However, the lack of large sample asymptotic properties makes it inadequate in assessing goodness-of-fit of these ERGMs. Simulation-based goodness-of-fit plots were proposed by Hunter et al. (2006). In this research, we propose an improved approach to assess goodness of fit of ERGMs and extend it to longitudinal network models. Our method is shown to improve sensitivity of the existing graphical techniques. We also propose a simulation based test statistic to quantify consistency of goodness of fit procedures.

ciaux à des niveaux statiques et dynamiques. Cependant, leur manque de propriétés asymptotiques les rend inefficaces pour évaluer l'adéquation du modèle. Des graphes d'adéquation basés sur des simulations furent proposés par Hunter et coll. (2006). Dans cette recherche, nous proposons une approche améliorée à l'estimation de l'adéquation de MGAE et la généralisons à des modèles de réseaux longitudinaux. Notre méthode améliore la sensibilité des techniques graphiques existantes. Nous proposons aussi une statistique de test à base de simulations afin de quantifier la cohérence de procédures d'adéquation.

**Session 05G                      Lundi 24 mai • Monday, May 24, 16:30 – 18:00                      VAN 1211**

**Regression Methodology in Biostatistics and Neuronal Data**  
**Méthodologie de régression en biostatistique et données neuronales**

**[MS-96]**

**Correlation Effects on Sample Size Calculations for Poisson and ZIP Regression Models**  
**Effet de la corrélation sur la détermination de la taille d'un échantillon dans les modèles de régression de Poisson et de Poisson gonflé à zéro**

Nabil CHANNOUF, *Groupe d'études et de recherche en analyse des décisions*, Brenda MACGIBBON, *UQAM and GERAD*, Marc FREDETTE, *HEC and GERAD*

In many biomedical applications, the Poisson distribution is often used to model the data. When this data exhibits an excess of zeros, the Zero Inflated Poisson (ZIP) model will be more appropriate. Although methods for sample size calculations for tests about the regression parameters exist, no methods are available for the ZIP regression model. Thus, we develop one here by modifying a method of Shieh (2001). Moreover, it seems intuitively obvious that the correlation structure of the explanatory variables have an effect on the sample size and it has not been studied systematically. We will address this problem here.

Dans le domaine biomédical, la régression Poisson est souvent utilisée quand la variable étudiée représente le nombre d'événements observés. Lorsque les données contiennent un excès de zéros, le modèle de Poisson gonflé à zéro devient plus approprié. Bien qu'il existe des méthodes pour la détermination de la taille d'un échantillon pour des tests sur les paramètres de régression, aucune méthode n'est disponible pour le modèle de régression de Poisson gonflé à zéro. Nous en développons une ici en modifiant une méthode de Shieh (2001).

De plus, il semble intuitivement évident que la structure de corrélation des variables explicatives peut avoir un effet sur la taille de l'échantillon, mais ceci n'a pas été étudié auparavant. Nous proposons de l'étudier ici.

**[MS-97]**

**Estimation of a Longitudinal Linear Structural Error-in-variables Model**  
**Estimation d'un modèle longitudinal linéaire structurel à erreurs dans les variables**

Laura DUMITRESCU, *University of Ottawa*

In this talk we introduce a longitudinal linear structural error-in-variables model with no subject specific random effects. We consider estimators of

Dans cet exposé, nous présentons un modèle longitudinal linéaire structurel à erreurs dans les variables sans effets aléatoires spécifiques aux sujets. Nous considérons des estima-

the regression parameters, which correspond to the modified least squares estimators of the slope and intercept, used in the univariate case. These estimators are consistent and we present a central limit theorem, which depends only on data, under the assumption that the vector of latent variables lies in the generalized domain of attraction of the normal law. Our result can be viewed as an extension of the result of Martsynyuk (2007) to include the case of longitudinal data.

teurs des paramètres de régression correspondants aux estimateurs des moindres carrés modifiés de pente et d'ordonnée à l'origine utilisés dans le cas univarié. Ces estimateurs sont cohérents et nous présentons un théorème central limite ne dépendant que des données sous l'hypothèse que le vecteur des variables latentes est dans le domaine d'attraction généralisé de la loi normale. Notre résultat peut être vu comme une généralisation du résultat de Martsynyuk (2007) afin d'inclure le cas de données longitudinales.

#### [MS-98]

##### **Power of Logistic Regression with Measurement Error in Predictor Variable and Varying Number of Observations per Subject**

##### **Puissance de la régression logistique avec des erreurs de mesure dans la variable prédictive et un nombre variable d'observations par sujet**

Olga MELNICHOUK, Salomon MINKIN, Lisa MARTIN & Norman BOYD, *Ontario Cancer Institute*

Our objective was to assess the power of a case-control study of serum biomarkers, subject to varying degree of measurement error, measured in annual blood samples and risk of breast cancer. In the linear and logistic regression analysis, measurement error leads to attenuated estimate of the regression coefficient and power loss. Our technique for determining power relies on simulation that incorporates the specific characteristics of the proposed study. To correct for measurement error when subjects' averages are based on varying number of replicates we apply the methodology proposed by Armstrong et al. (1989) and Kim & Zeleniuch-Jacquotte (1997).

Notre objectif était d'évaluer la puissance d'une étude cas-témoins de biomarqueurs du sérum sujets à divers degrés d'erreurs de mesure, mesurés à partir d'échantillons sanguins annuels et liés au risque de cancer du sein. Lors de l'analyse par régression linéaire et logistique, les erreurs de mesures mènent à des estimateurs atténués du coefficient de régression et à une perte de puissance. Notre méthode de détermination de la puissance s'appuie sur une simulation incorporant les caractéristiques spécifiques de l'étude proposée. Afin de compenser les erreurs de mesure lorsque les moyennes par sujet sont basées sur un nombre variable de réplicats, nous appliquons la méthodologie proposée par Armstrong et coll. (1989) et Kim & Zeleniuch-Jacquotte (1997).

#### [MS-99]

##### **Model Checking with Residuals for $G$ -estimation of Optimal Dynamic Treatment Regimes**

##### **Vérification de modèle avec résidus pour $g$ -estimation de régimes de traitement dynamiques optimaux**

Benjamin RICH, *McGill University*, Erica MOODIE, *McGill University*, David STEPHENS, *McGill University*

Model diagnostics for  $g$ -estimation of optimal dynamic treatment regimes are considered. Proposed by Robins,  $g$ -estimation is doubly robust to the specification of a model for treatment mechanism and a model for an expected counterfactual outcome. While consistency is guaranteed when at least one of these models is correctly specified, both must be correct for maximal efficiency. Assuming the treatment model is correct, we focus on diagnosing problems with the expected counter-

Les diagnostics de modèle en  $g$ -estimation de régimes de traitement dynamiques optimaux sont considérés. Proposée par Robins, la  $g$ -estimation est doublement robuste à la spécification d'un modèle pour le mécanisme de traitement et d'un modèle pour réponse contre-factuelle espérée. Bien que la cohérence soit garantie lorsqu'au moins un des deux modèles est correctement spécifié, les deux modèles doivent être corrects pour une efficacité maximale. En supposant que le modèle de traitement soit correct, nous nous penchons sur le diagnostic de problèmes avec le modèle contre-factuel es-

factual model using residual plots. In a simulation study, we observe that it is possible in some situations to diagnose and fix model misspecification, leading to improved efficiency.

péré au moyen de graphes des résidus. Lors d'une étude de simulation, nous observons qu'il est possible de diagnostiquer et de corriger un modèle mal spécifié dans certaines situations, menant à une efficacité accrue.

#### [MS-100]

##### **Comparison of Two Populations of Curves with an Application in Neuronal Data Analysis**

##### **Comparaison de deux populations de courbes avec une application en analyse de données neuronales**

Shojaeddin CHENOURI, *University of Waterloo*, Sam BEHSETA, *California State University, Fullerton*

Often in neurophysiological studies, scientists become interested in testing hypotheses regarding the equality of the overall intensity functions of a group of neurons when recorded under two experimental conditions. In this talk, we consider such a hypothesis testing problem. We propose two test statistics and implement these tests on smooth curves obtained via fitting Bayesian Adaptive Regression Splines to the intensity functions of neuronal Peri-Stimulus Time Histograms (PSTH). We study performance of our proposed tests via simulation. Finally, we apply our methods on a group of motor cortex neurons recorded during a reaching task.

En études neurophysiologiques, les scientifiques sont souvent intéressés à tester des hypothèses relatives à l'égalité de fonctions d'intensité globales d'un groupe de neurones enregistrés sous deux conditions expérimentales différentes. Nous proposons deux statistiques de tests et les appliquons à des courbes lisses obtenues en ajustant des splines de régression bayésiennes adaptatives aux fonctions d'intensité d'histogrammes de péri-stimulus neuronaux. Nous étudions la performance des tests proposés au moyen de simulations. Finalement, nous appliquons nos méthodes à un groupe de neurones du cortex moteur.

#### [MS-101]

##### **A Critical Review of Structural Equation Modelling in fMRI Studies of Effective Connectivity**

##### **Revue critique de la modélisation par équations structurelles pour la connectivité effective en études IRMf**

Qing GUO, Hall GEOFF & Eleanor PULLENAYEGUM, *McMaster University*

Structural equation modelling (SEM) tests a system of simultaneous equations that reflect a priori assumptions about direct and indirect effects among observed variables. With growing computational ease, SEM has been widely used in functional MRI (fMRI) studies to assess the influence of one neuronal system upon another across brain regions. Is SEM an optimal method when dealing with issues such as small sample size (10 to 16 subjects), multicollinearity among the variables and longitudinal repeated measurements, compared with classical regression modelling (e.g., mixed effects model)? These issues will be explored and illustrated through an autism study.

La modélisation par équations structurelles (MES) teste un système d'équations simultanées reflétant des hypothèses a priori sur les effets directs et indirects des variables observées. Les MES sont fréquemment utilisées en études d'IRM fonctionnelle (IRMf) afin d'évaluer l'influence d'un système neuronal sur un autre dans différentes régions du cerveau. Les MES sont-ils une méthode optimale comparativement à la modélisation par régression classique (p.e., modèles à effets mixtes) en présence d'une faible taille échantillonnale (10 à 16 sujets), de multicollinéarité dans les variables, et de mesures longitudinales répétées? Ces questions seront explorées et illustrées à l'aide d'une étude sur l'autisme.

**Session 05H      Lundi 24 mai • Monday, May 24, 16:30 – 18:00      VAN 1871**  
**Hypothesis Testing and Inference**  
**Tests d'hypothèses et inférence**

**[MS-102]**

**Estimation and Testing For Continuous Zero-modified Models**

**Estimation et tests pour modèles continus modifiés à zéro**

Félix LABRECQUE-SYNNOTT & Jean-François ANGERS, *Université de Montréal*

Zero-inflated models, both discrete and continuous, have a large variety of applications and fairly well-known properties. While some work has been done on zero-deflated and zero-modified discrete models, the usual formulation of continuous zero-inflated models—a mixture between a continuous density and a Dirac mass at zero—precludes their extension to cover the zero-deflated case as well. An alternative formulation of zero-inflated continuous models is introduced, along with a natural extension to the zero-deflated case. Bayesian and likelihood-based estimation is discussed. The model and estimation methods are illustrated with simulation results and applied to aggregate rainfall data.

Les modèles à surreprésentation de zéros, continus ou discrets, ont une variété d'applications et des propriétés connues. Les modèles discrets à sous-représentation de zéros ont été étudiés, mais la formulation usuelle des modèles continus à sur-représentation de zéros — un mélange entre une densité et une masse de Dirac à zéro — ne peut aisément être généralisée pour couvrir le cas de la sous-représentation. Nous présentons une formulation alternative pour les modèles continus à sur-représentation de zéros et une généralisation naturelle au cas de la sous-représentation. L'estimation bayésienne et basée sur la vraisemblance est abordée. Des résultats de simulation illustrent le modèle et les méthodes d'estimation, et une application à des données de précipitations agrégées est présentée.

**[MS-103]**

**Multivariate Nonparametric Two-sample Tests for Mixed Outcomes**

**Tests non paramétriques multivariés pour comparer deux échantillons avec des variables réponses de type mixtes**

Denis LAROCQUE, *HEC Montréal*, Jaakko NEVALAINEN, *University of Turku*, Hannu OJA, *University of Tampere*

Over the years, many nonparametric methods for multivariate outcomes have been proposed. These include coordinate-wise and spatial rank and sign methods and others based on different notions of depth. But these methods are aimed at the situation where all outcomes are continuous. In this talk, we will present new methods specifically designed for the case where some outcomes are categorical and others are continuous, and we will focus on the two-sample problem.

De nombreuses méthodes multivariées non paramétriques ont été développées au cours des ans. Il y a par exemple des méthodes basées sur les rangs ou les signes composantes par composantes et spatiales ainsi que d'autres basées sur différentes notions de profondeur. Ces méthodes visent la situation où toutes les variables réponses sont continues. Dans cet exposé, nous présenterons des nouvelles méthodes spécialement adaptées à la comparaison de deux échantillons lorsque certaines variables réponses sont catégorielles alors que d'autres sont continues.

**[MS-104]****Hypothesis Testing in Multivariate Mixture Models: the EM approach  
Tests d'hypothèses pour des modèles de mélanges multivariés : l'approche EM**Xiaoqing NIU & Pengfei LI, *University of Alberta*

The test of homogeneity is among the crucial problems in the application of finite mixture models. Most existing results can only be applied to one-dimensional mixture models. In this paper, we applied the EM-test to multi-dimensional mixture models and investigated its asymptotic properties. The null limiting distribution of the EM-test is shown to be the supremum of a quadratic function over a cone, which can be easily simulated in practice. Extensive simulation studies show that the finite sample distribution matches the limiting distribution of the EM-test. Two real data examples are included to illustrate the performance of the proposed test.

Le test d'homogénéité fait partie des problèmes cruciaux dans l'application de modèles de mélanges finis. La plupart des résultats existants ne peuvent être appliqués qu'à des modèles de mélanges unidimensionnels. Dans cet article, nous appliquons le test-EM à des modèles de mélanges multidimensionnels et étudions ses propriétés asymptotiques. La distribution nulle asymptotique du test-EM est le supremum d'une fonction quadratique sur un cône et peut aisément être simulée en pratique. Des études de simulation montrent que la distribution sous échantillon fini du test-EM correspond à sa distribution asymptotique. Deux exemples avec des données réelles illustrent la performance du test proposé.

**[MS-105]****The Effect of Interaction and Rounding Error in Two-way ANOVA: An Example on the Impact of Testing for Normality****Effet de l'interaction et l'erreur d'arrondi dans le modèle de l'analyse de la variance à deux facteurs : l'exemple des tests de normalité**Jesper RYDÉN & Sven Erick ALM, *Uppsala University*

A key issue in various applications of analysis of variance is testing for interaction and the interpretation of resulting analysis of variance tables. In this note, we examine by simulations whether, for a two-way ANOVA, incorporating interactions or not may have a dramatic influence when considering the usual statistical tests for normality of residuals (e.g., Shapiro–Wilk). The effect of numerical rounding is also discussed.

Un problème-clé dans plusieurs applications de l'analyse de la variance est de tester pour des interactions et d'interpréter les tables d'analyse de la variance résultantes. Dans ce travail, nous examinons par des exemples numériques l'influence de l'interaction sur des tests pour normalité des résidus (par exemple Shapiro-Wilk) en analyse de variance à deux facteurs. Nous nous intéressons aussi à l'effet possible de l'arrondissement des données.

**[MS-106]****Empirical Likelihood Inference for Two-sample Problems with Missing Data  
Inférence par vraisemblance empirique pour problèmes à deux échantillons avec données manquantes**Ying YAN & Changbao WU, *University of Waterloo*

We present empirical likelihood (EL) methods for two-sample problems with missing data, with focus on the difference of the two population means. We consider scenarios where sample data on auxiliary variables are fully observed for both samples but values of response variables are subject to missingness. Two EL methods are proposed: (1) Miss-

Nous présentons des méthodes de vraisemblance empirique (VE) pour problèmes à deux échantillons avec données manquantes, en particulier pour la différence de deux moyennes de population. Nous considérons des scénarios où des variables auxiliaires sont complètes alors que la variable réponse est manquante pour certains sujets. Deux méthodes VE sont présentées : (1) les réponses manquantes sont impu-

ing responses are imputed based on auxiliary information, and the EL method is then applied to imputed data sets; (2) A parametric conditional model is combined with the nonparametric marginal distribution, resulting in a semi-EL inference. Theoretical development, simulation results and applications will be reported.

tées à partir de l'information auxiliaire, et la méthode VE est alors appliquée aux jeux de données imputés ; (2) un modèle paramétrique conditionnel est combiné à la distribution marginale non paramétrique, résultant en une inférence semi-EL. Des développements théoriques, résultats de simulation et applications sont présentés.

[MS-107]

**Interval Hypotheses for Comparing Several Treatments with a Control**

**Hypothèses par intervalles pour comparer plusieurs traitements à un contrôle**

Jianan PENG, *Acadia University*, Hubert CHEN, *National Cheng-Kung University*, Jiancheng JIANG, *University of North Carolina at Charlotte*

In comparing treatments with a control, Dunnett's multiple comparison procedure may lead to a statistical significance even though, sometimes, the difference is not worthy of attention considering the inherent variation in the characteristic. In this talk, we use Dunnett's test statistic to test an interval hypothesis instead of a classical homogeneity hypothesis. The least favourable configuration is determined. Numerical quadrature is employed to obtain the tabulated percentage points of the test statistic. An application of this new inference method is given.

Lors d'une expérience où plusieurs traitements sont comparés à un groupe contrôle, la méthode de comparaisons multiples de Dunnett peut mener à des résultats statistiquement significatifs même si la différence ne mérite parfois pas notre attention étant donnée la variation inhérente à la caractéristique. Dans cet exposé, nous utilisons la statistique de test de Dunnett afin de tester une hypothèse par intervalles plutôt que l'hypothèse habituelle d'homogénéité. La configuration la moins favorable est déterminée. Les seuils critiques tabulés de la distribution de la statistique de test sont obtenus par quadrature numérique. Un exemple d'application de cette nouvelle méthode d'inférence est donné.

**Session 06A**

**Mardi 25 mai • Tuesday, May 25, 08:30 – 09:45**

**PPR**

**Gold Medal Award Address**

**Allocution de la récipiendaire de la médaille d'or**

[MS-108]

**Some Thoughts on the Theory of Statistics**

**Quelques idées sur la théorie de la statistique**

Nancy REID, *University of Toronto*

Statistical science is undergoing unprecedented growth in both opportunity and activity. Fields as diverse as particle physics, art history and biology find themselves turning to statistical ideas to understand new modes of reasoning and to suggest productive avenues for further research. Ideally the theory of statistics would serve to provide a foundation for developing systematic approaches to data analysis. In practice, it is more likely that the pressures of large amounts of data and limited amounts of time lead to fairly ad hoc approaches

La science statistique connaît une croissance sans précédent en termes de possibilités à exploiter et d'activité. Des domaines aussi divers que la physique des particules, l'histoire de l'art et la biologie se tournent vers les idées statistiques afin de comprendre de nouveaux modes de raisonnement et de trouver de nouvelles pistes de recherche fructueuses. Idéalement, la théorie statistique devrait servir de fondation sur laquelle bâtir des approches systématiques à l'analyse de données. En pratique, les grands volumes de données et le manque de temps mènent souvent à des approches plutôt ad hoc à l'analyse de données. Ceci est exacerbé par la

to analysis. This is exacerbated by the perception that statistical theory is really only relevant to artificially simple settings. In this talk I will discuss various aspects of the development of theoretical statistics over the past thirty years with a view to understanding how theory might indeed inform today's statistical practice.

perception que la théorie statistique n'est réellement pertinente que dans des contextes artificiellement simples. Dans cet exposé, je vais discuter de divers aspects du développement de la statistique théorique au cours des trente dernières années en tentant de montrer comme la théorie peut bel et bien contribuer à la pratique statistique d'aujourd'hui.

**Session 07A      Mardi 25 mai • Tuesday, May 25, 10:25 – 11:55      VAN 1245A**

**The Biology, Technology, and Statistical Modelling of High-throughput Genomics Data**  
**La biologie, la technologie et la modélisation statistique de données génomiques à haut débit**

[MS-109]

**The Biology, Technology and Statistical Modelling of High-throughput Genomics Data**  
**La biologie, la technologie et la modélisation statistique de données génomiques à haut débit**

Naomi ALTMAN, *Penn State University*

We will briefly review some aspects of the biology of gene expression required to understand gene expression and epigenetic studies. Some of the main microarray and sequencing methods for quantifying high throughput genomics data will be introduced. Statistical issues such as data normalization, combining data across platforms and the analysis of “small  $n$  large  $p$ ” data for differential expression and location of features such as methylation and protein binding sites will be discussed.

Nous faisons une brève revue de quelques aspects biologiques de l'expression génétique nécessaires à la compréhension d'études d'expression génétique et épigénétiques. Quelques-unes des principales méthodes utilisées pour quantifier des données génomiques à haut débit seront présentées. Des problèmes statistiques tels que la normalisation des données, la combinaison des données de plusieurs plates-formes et l'analyse de données d'expression génétique “petit  $n$  grand  $p$ ” seront abordés.

**Session 07B      Mardi 25 mai • Tuesday, May 25, 10:25 – 11:55      VAN 2289A**

**New Developments in Modelling Using Survey Data**  
**Nouveaux développements en modélisation de données d'enquêtes**

[MS-110]

**Pooling Respondents' Blood Serum for Analysis in the Canadian Health Measures Survey**  
**Regroupement du sérum sanguin de répondants pour analyse dans l'Enquête canadienne sur les mesures de la santé**

François VERRET & Suzelle GIROUX, *Statistique Canada*

The Canadian Health Measures Survey collects measured indicators of health directly from a sample of Canadians. Some tests on collected specimens are costly and require a minimum quantity of specimen to detect the components analyzed. In cycle 1, residual blood serum was pooled to measure levels of organohalogenes in 10 age-sex groups

L'Enquête canadienne sur les mesures de la santé recueille directement des indicateurs de la santé à partir d'un échantillon de Canadiens. Certains tests à effectuer sur les prélèvements recueillis sont coûteux et nécessitent une quantité minimale de spécimen pour détecter les composés à analyser. Au cycle 1, nous avons combiné le sérum sanguin résiduel pour mesurer les niveaux de composés organiques halogénés

at the national level.

This presentation describes the pooling and estimation methodologies used. Pools were created by grouping the serum of individuals having similar survey weights to minimise the bias. Dependent random groups of primary sampling units were formed across the strata to estimate the variance.

dans dix groupes d'âge-sexe au niveau national.

Lors de cette présentation, la méthodologie utilisée pour la création des bassins et pour l'estimation sera décrite. Le sérum des répondants ayant des poids d'enquête similaires a été regroupé pour minimiser le biais. Des groupes aléatoires dépendants d'unités primaires d'échantillonnage ont été formés à travers les strates pour estimer la variance.

### [MS-111]

#### **Bootstrap Variance Estimation of Model Parameters with Survey Data**

#### **Méthode bootstrap pour l'estimation de la variance de paramètres de modèle avec des données d'enquêtes**

Anne-Sophie CHAREST, *Carnegie-Mellon Statistics Department*, Jean-François BEAUMONT, *Statistique Canada*

When estimating model parameters from survey data, two sources of variability must be taken into account for inference purposes: the model assumed to have generated the finite population and the sampling design. We show how the usual estimation methods which ignore the model variability can lead to large underestimation of the variance of estimates of model parameters. We then propose a generalized bootstrap method which correctly takes into account both sources of variability, and is very simple to apply once design-based bootstrap weights are obtained. We provide simulation results and examples on a real dataset to illustrate the method.

L'inférence pour des paramètres de modèle avec des données d'enquêtes doit prendre en compte deux sources de variabilité : le modèle ayant généré la population finie et le plan d'échantillonnage. Nous montrerons comment les méthodes utilisées en pratique, qui ignorent la variabilité due au modèle, peuvent mener à une sous-estimation importante de la variance des estimateurs des paramètres du modèle. Puis nous proposerons une méthode de bootstrap généralisée qui tient compte des deux sources de variabilité, et est très simple à utiliser si des poids bootstrap valides pour le plan d'échantillonnage sont déjà fournis. Nous illustrerons notre technique avec des résultats de simulation et sur un jeu de données réel.

### [MS-112]

#### **Linear Regression Diagnostics for Survey Data**

#### **Diagnostics de régression linéaire pour données d'enquête**

Richard VALLIANT, *Universities of Michigan and Maryland*

Diagnostics for linear regression models have been developed primarily for non-survey data. The models and sampling plans used for finite populations often entail stratification, clustering, and survey weights. We adapt some diagnostics for ordinary or weighted least squares for use with survey data. Statistics considered here include leverages, DFBETAS, DFFITS, and Cook's D. Differences in the performance of ordinary least squares and survey-weighted diagnostics are compared in an empirical study where values of weights, response variables, and covariates vary substantially. We also review how the forward search method for identifying groups of influential points may be applied to survey data.

Les diagnostics pour modèles de régression linéaire ont principalement été développés pour données ne provenant pas d'enquêtes. Les modèles et plans d'échantillonnage utilisés pour des populations finies impliquent souvent stratification, classification et poids de sondage. Nous adaptons certains diagnostics pour moindres carrés ordinaires ou pondérés pour données d'enquêtes. Les statistiques considérées ici incluent les leviers, DFBETAS, DFFITS et le D de Cook. Les différences de performance entre les diagnostics pour moindres carrés ordinaires et pondérés par poids de sondage sont comparées dans une étude empirique où les valeurs des poids, des variables réponses et des covariables varient substantiellement. Nous faisons aussi la revue de la méthode de recherche avant pour identifier les groupes de points influents avec des données d'enquête.

**Session 07C      Mardi 25 mai • Tuesday, May 25, 10:25 – 11:55      VAN 2291A**  
**Quality Improvement**  
**Amélioration de la qualité**

**[MS-113]**

**New Nonparametric Control Charts**  
**Nouvelles cartes de contrôle non paramétriques**

Su-Fen YANG, *National Chengchi University, Taiwan*, Smiley W CHENG, *University of Manitoba*

One of the underlying assumptions for using control charts for variables data is 'normality'. Hence the question arises "what if the underlying distribution is NOT normal or unknown"? What can we do? What charts can we apply? In this paper, we summarize a few newly proposed "Nonparametric" charts for variables data and discuss related issues and their performances. An example is given to illustrate the proposed charts.

Une des hypothèses sous-jacentes aux cartes de contrôle pour des variables est la normalité. Que se passe-t-il si la distribution sous-jacente n'est pas normale ou si elle est inconnue? Que peut-on faire dans ce cas? Quelles cartes pouvons-nous utiliser? Dans cette communication, nous présentons quelques cartes non paramétriques pour données variables récemment proposées et discutons de leur performance et de problèmes liés à leur utilisation. Un exemple illustre les cartes proposées.

**[MS-114]**

**Some Issues in the Design and Analysis of Industrial split-plot Experiments**  
**Quelques problèmes en planification et analyse d'expériences industrielles avec subdivision des parcelles**

John BREWSTER, *University of Manitoba*

In industrial experiments, some factors are often easier to vary than others. In many cases, this leads to restrictions on randomization and a split-plot structure. In this talk, we will examine some issues that arise in both the design and analysis of such experiments. At the design stage, these issues arise in screening and optimization experiments and in designing follow-up experiments. At the analysis stage, some of the issues revolve around the use of inferential procedures that are not faithful to parameter constraints induced by the whole-plot and subplot variance components.

Lors d'expériences industrielles, il est souvent plus facile de varier certains facteurs que d'autres. Dans plusieurs cas, ceci mène à des restrictions sur la randomisation et à une structure à parcelles divisées. Dans cet exposé, nous examinerons certains problèmes survenant dans la planification et l'analyse de telles expériences. Lors de la planification, ces problèmes surviennent en expérience de dépistage et d'optimisation, ainsi que dans la planification d'expériences de suivi. Lors de l'analyse, certains des problèmes sont liés à l'utilisation de procédures inférentielles ne respectant pas les contraintes paramétriques induites par les composantes de variance des parcelles entières et des sous-parcelles.

**[MS-115]**

**Issues in Planning Experiments for Reliability Data**  
**Problèmes en planification d'expériences pour données de fiabilité**

Geoff VINING & Laura FREEMAN, *Virginia Tech*

An important characteristic of many products is the length of time that it performs its intended function. Often, people call these times life data. This paper examines the issues in planning experiments for life data. Of special concern is the proper inclusion of the true experimental error into the model and its analysis.

Une caractéristique importante de plusieurs produits est la durée pendant laquelle ils remplissent leur fonction. Souvent, ces durées sont appelées temps de survie. Cet article se penche sur la planification d'expériences pour les temps de survie. Des modèles qui tiennent compte de l'erreur expérimentale réelle sont considérés ainsi que leur analyse.

**Session 07D      Mardi 25 mai • Tuesday, May 25, 10:25 – 11:55      VAN 2809A****Applications of Statistics in Hydrology, Climatology, and Meteorology  
Statistique appliquée à l'hydrologie, la climatologie et la météorologie****[MS-116]****Mixtures of Normal, Gamma and Gumbel Distributions: An Application to Peaks and Volumes of Spring Floods****Mélanges de lois de probabilités normales, gamma et Gumbel : une application aux données de pointes et de volumes de crues printanières**LUC PERREAULT & JAMES MERLEAU, *Institut de recherche d'Hydro-Québec*, Guillaume ÉVIN, *Université Laval*

The problem of switching regimes in hydrological time series brought about serious questioning in Québec during the last decade. When we study the annual inflow time series of a number of Québec watersheds, we can identify alternating sequences of high and low inflows. We illustrate the applicability of mixtures of two distributions with normal, gamma and Gumbel components to analyze changepoints in peaks and volumes of spring floods for a group of Québec watersheds. The problem is treated from the Bayesian perspective, which enables us to use marginal densities to compare the adequacy of the different models.

Le problème de changements de régime dans les séries chronologiques hydrologiques a suscité beaucoup de questionnements au Québec ces dernières années. Si l'on examine les débits annuels de certains bassins versants québécois, on peut observer en alternance des séquences d'années de faible et de forte hydraulité. Nous présentons une application des modèles de mélanges de deux distributions normales, gamma et Gumbel pour analyser ces changements de régime dans les séries de pointes et de volumes de crues printanières pour un ensemble de bassins versants québécois. La perspective bayésienne est adoptée, ce qui permet d'utiliser les densités marginales pour comparer l'adéquation des différents modèles.

**[MS-117]****Evaluation of the Canadian Meteorological Ensemble Prediction System for Short-term Hydrological Forecasting****Evaluation du système de prévision météorologique d'ensemble canadien pour la prévision hydrologique à courte échéance**Vincent FORTIN, *Environnement Canada*, Juan Alberto VELAZQUEZ, François ANCTIL & Anne-Catherine FAVRE, *Université Laval*, Richard TURCOTTE, *Centre d'Expertise Hydrique du Québec*, Heri KUSWANTO, *Université Laval*, Nguyen Bao TRINH, *Institut National de la Recherche Scientifique*

In July 2007, the Meteorological Service of Canada improved substantially its ensemble prediction system. This improved system is evaluated for the purpose of hydrological ensemble forecasting. Results show that ensemble forecasts are more skillful than deterministic forecasts, as measured by the Continuous Ranked Probability Score (CRPS). However, the resulting hydrological ensemble forecasts are under-dispersed. We conjecture that this is due in part to the fact that uncertainty in the initial conditions of the hydrological model is not taken into account. We finally compare two statistical post-processing method for obtaining reliable probabilistic forecasts from these under dispersed ensemble forecasts.

En juillet 2007, le Service météorologique du Canada a substantiellement amélioré son système de prévision par ensembles. Ce système amélioré est évalué dans le but de prévision hydrologique par ensembles. Les résultats montrent que les prévisions par ensembles sont plus fines que les prévisions déterministes, telles que mesurées par le score de probabilité à rang continu (SPRC). Cependant, les prévisions hydrologiques par ensembles résultantes sont sous-dispersées. Nous supposons que ceci est dû en partie au fait que les incertitudes sur les conditions initiales du modèle hydrologique ne sont pas prises en considération. Finalement, nous comparons deux méthodes statistiques post-traitement afin d'obtenir des prévisions probabilistes fiables à partir de ces prévisions d'ensemble sous-dispersées.

**[MS-118]****Some Bayesian Investigations about the Bivariate Gamma Model with Applications to Environmental Data  
Quelques investigations bayésiennes autour du modèle gamma bivarié avec applications environnementales**Eric PARENT & Jacques BERNIER, *AgroParisTech*

From the study of characteristic functions of conditional Kibble's bivariate exponential model, we progressively elaborate joint asymmetric distributions for positive variables, according to a hierarchical Bayesian modelling perspective. Unlike the copula approach, we first impose to fulfill a conditional structure before sticking to the strict respect of marginal distributions. The compound Poisson distribution with exponential marks plays a surprising role in the conditional distribution for the bivariate exponential model and provides inspiring ideas to design models with gamma margins or mixtures of gamma distributions. We test their interest in the case of various environmental applications.

À partir de l'étude des fonctions caractéristiques conditionnelles du modèle exponentiel bivarié de Kibble, nous proposons une méthode de construction progressive de distributions conjointes non symétriques pour des variables positives, selon une perspective de modélisation bayésienne hiérarchique. Contrairement à l'approche par copules, nous imposons d'abord la structure conditionnelle plutôt que le strict respect des marges du modèle. Après avoir remarqué le rôle surprenant du processus de Poisson à marques exponentielles dans la loi conditionnelle du modèle exponentiel bivarié, nous construisons des modèles de marges gammas ou de mélanges de gammas et testons leur intérêt sur des cas d'applications environnementales.

**Session 07E      Mardi 25 mai • Tuesday, May 25, 10:25 – 11:55      VAN 2811A****Ruin Theory****Théorie de la ruine****[MS-119]****Vector-valued Martingales in Risk Theory  
Martingales vectorielles en théorie du risque**Andrei BADESCU, *University of Toronto*

A dependent risk process with perturbation and phase-type distributed claim sizes is analyzed. Claim arrivals are modelled by a Markovian arrival process. Using a vector-valued martingale, the Laplace transform of the time to ruin is derived algorithmically. The conditional memoryless property of the phase-type distribution yields the distribution of the deficit at ruin as a corollary.

Nous analysons un processus de risque dépendant avec perturbations dont la distribution de la taille des réclamations varie par phase. L'arrivée des réclamations est modélisée par un processus markovien. À partir d'une martingale à valeurs vectorielles, la transformée de Laplace du temps jusqu'à la ruine est obtenue de façon algorithmique. La propriété de la distribution des phases d'être conditionnellement sans mémoire nous permet d'obtenir comme corollaire la distribution du déficit au moment de la ruine.

**[MS-120]****A Threshold-based Risk Process with a Waiting Period to Pay Dividends  
Un processus de risque basé sur un seuil avec un délai pour le paiement des dividendes**Steve DREKIC, *University of Waterloo*

Using matrix analytic methods, computational procedures are constructed which enable one to calcu-

Nous utilisons des méthodes analytiques matricielles afin de construire des méthodes de calcul permettant d'obtenir des

late fundamental ruin-related probability distributions of interest for a discrete-time, delayed Sparre Andersen insurance risk model in which dividends only get paid provided that a constant threshold level is crossed and some period of “good health” is sustained above the threshold. Special cases of the general model are examined in several numerical examples.

distributions fondamentales d'intérêt liées à la ruine dans un modèle du risque d'assurance de Sparre Andersen différé à temps discret dans lequel le paiement des dividendes est conditionnel au dépassement d'un certain seuil constant et au maintien d'une période de « bonne santé » au-dessus de ce seuil. Plusieurs exemples numériques permettent d'examiner des cas particuliers du modèle.

**[MS-121]**

**On the Ruin Problem for Two New Levy Insurance Risk Processes**

**Du problème de la ruine pour deux nouveaux processus du risque d'assurance de Levy**

Manuel MORALES, *University of Montreal*

Expressions for the expected discounted penalty function (EDPF) now exist for a wide range of models, in particular for a general class of Levy insurance risk processes. However, these expressions must be computationally tractable to allow for evaluation of associated risk measures. Moreover, currently available Levy risk insurance models refer to an over-simplified reality accounting only for premium in-flow and claims out-flow. First, we introduce examples of risk insurance processes for which numerical evaluation of the EDPF can be readily implemented. Second, we discuss a new model that accounts for the possibility of two dependent lines of business.

Des expressions pour la fonction de pénalité espérée escomptée (FPEE) existent maintenant pour une large gamme de modèles, et en particulier pour une classe générale de processus de risque d'assurance de Levy. Cependant, ces expressions doivent être calculables afin de permettre l'évaluation des mesures de risque associées. En outre, les modèles du risque d'assurance de Levy actuellement disponibles réfèrent à une réalité simplifiée ne tenant compte que des entrées de primes et des sorties de réclamations. Nous présentons d'abord des exemples de processus de risque d'assurance pour lesquels l'évaluation numérique de la FPEE peut être implémentée. Ensuite, nous discutons d'un nouveau modèle tenant compte de la possibilité de deux secteurs d'activités dépendants.

**Session 07F**

**Mardi 25 mai • Tuesday, May 25, 10:25 – 11:40**

**VAN 1203**

**Methodology for Hazard Rate Estimation and Censored Data**

**Méthodologie pour l'estimation du taux de risque et les données censurées**

**[MS-122]**

**Empirical Likelihood Estimation of a Covariate-adjusted Cumulative Hazard Ratio**

**Estimation par vraisemblance empirique d'un rapport de risques cumulatifs ajusté par covariables**

Bin DONG & David E. MATTHEWS, *University of Waterloo*

In medical studies, it is often of scientific interest to evaluate the treatment effect via the ratio of cumulative hazards, especially when those hazards may be nonproportional. In this paper we employ empirical likelihood (EL) to develop a nonparametric estimator of a cumulative hazard ratio with covariate adjustment under a nonproportional hazard model involving stratification. The asymptotic properties of the empirical likelihood ratio statis-

En études médicales, il est souvent d'intérêt scientifique d'évaluer l'effet de traitement par le rapport des risques cumulatifs, surtout si ces risques peuvent ne pas être proportionnels. Dans cet article, nous utilisons la vraisemblance empirique (VE) pour développer un estimateur non paramétrique du rapport de risques cumulatifs avec ajustement pour covariables sous un modèle de risques non proportionnels avec stratification. Les propriétés asymptotiques de la statistique du rapport de vraisemblance empirique sont obtenues

tic are derived and the finite-sample properties of the EL-based estimator are assessed via simulation studies. The proposed method is illustrated using a dataset concerning the survival experience of patients with non-Hodgkin's lymphoma.

et les propriétés sous échantillon fini de l'estimateur basé sur la VE sont évaluées au moyen d'études de simulation. La méthode proposée est illustrée à l'aide d'un jeu de données de survie de patients atteints de lymphome non-Hodgkin.

**[MS-123]**

**Weakly Parametric PH Regression Model for Multivariate Survival Data**

**Modèles faiblement paramétriques de régression à risques proportionnels pour données de survie multivariées**

Na LEI & Wenqing HE, *University of Western Ontario*

Proportional hazards models are popularly used to find the covariate effects in multivariate lifetime data analysis. We apply a weakly parametric approach to model the nonlinear covariate effect on the log hazard ratio in PH model. Specifically, a spline function is employed to model the nonlinear effect in the marginal hazard functions, piecewise constants are used to model the marginal hazard baseline function. The Clayton model is employed to postulate the association in multivariate lifetime. A motivating example will be provided and simulation study is conducted to assess the performance of the proposed method.

Les modèles à risques proportionnels sont souvent utilisés afin de trouver les effets covariables en analyse multivariée de données de durée de vie. Nous appliquons une approche faiblement paramétrique à la modélisation des effets covariables non linéaires sur le log du rapport de risques du modèle à RP. Spécifiquement, une fonction spline est utilisée pour modéliser l'effet non linéaire sur les fonctions de risque marginales et une fonction constante par morceaux modélise la fonction marginale de risque de base. Le modèle de Clayton est utilisé pour modéliser l'association dans la durée de vie multivariée. Un exemple motivant ces travaux et une étude de simulation en illustrant la performance seront présentés.

**[MS-124]**

**Estimation of Treatment Effects based on Composite Endpoints**

**Estimation d'effets de traitement à partir de points terminaux composites**

Longyang WU & Richard COOK, *University of Waterloo*

Composite endpoints are used routinely in long-term clinical trials. We consider the asymptotic and empirical behaviour of estimators of treatment effect based on Cox models for composite failure time events. Proportional hazards models are assumed for each component and copula functions are used to model the association between them. The limiting value of the regression coefficient is shown to depend heavily on the treatment effects on the components, the stochastic ordering of the events, and the censoring distribution. Caution is warranted on the interpretation of results based on composite endpoints. We use a recent cardiovascular trial as an illustration.

Les bornes composites sont fréquemment utilisées en essais cliniques de longue durée. Nous considérons le comportement asymptotique et empirique d'estimateurs de l'effet traitement basés sur des modèles de Cox pour temps de défaillance composites. Nous supposons chaque composante distribuée d'après un modèle à risques proportionnels et utilisons des fonctions copules pour modéliser l'association entre elles. Nous montrons que la limite du coefficient de régression dépend fortement des effets traitement sur les composantes, de l'ordre stochastique des événements, et de la distribution de censure. Une attention particulière doit être apportée lors de l'interprétation de résultats basés sur des bornes composites. Nous utilisons un essai cardiovasculaire récent comme illustration.

**[MS-125]****Some New Smooth Density Estimators for Length Biased Data****Quelques nouveaux estimateurs lisses de densités pour données biaisées en longueur**Jun LI & Yogendra P. CHAUBEY, *Concordia University*, Pranab SEN, *University of North Carolina*

Length biased sampling occurs naturally in many statistical applications. One aspect of length biased data in which people are interested is estimating unweighted density with the observed sample. Since most length biased data is nonnegative, unweighted density has a support with a non-negative finite end point. The current proposed kernel density estimator with symmetric kernel may not be appropriate. Here we propose some new smooth density estimators and study their asymptotic properties. Moreover, through our investigations using numerical simulation and a real length biased data example, we find that the new smooth estimators perform much better than kernel estimators.

L'échantillonnage biaisé en longueur se produit naturellement dans plusieurs applications statistiques. Un aspect des données biaisées en longueur auxquels s'intéressent les gens est l'estimation non pondérée de la densité à partir de l'échantillon observé. Puisque la plupart des données biaisées en longueur sont non négatives, la densité non pondérée a un support à borne finie non négative. L'estimateur de densité à noyaux actuel, dont les noyaux sont symétriques, peut ne pas être approprié. Nous proposons ici quelques nouveaux estimateurs de densité lisses et étudions leurs propriétés asymptotiques. En outre, des simulations numériques et un exemple réel de données biaisées en longueur nous permettent de conclure que les nouveaux estimateurs lisses surpassent les estimateur à noyaux.

**[MS-126]****A Copula Model for Bivariate Observations with Below Detection Limits****Une copule pour données bidimensionnelles avec un seuil de détection**Héla ROMDHANI, M'Hamed Lajmi LAKHAL CHAIEB & Louis-Paul RIVEST, *Université Laval*

It is of interest to investigate the association between two measurements subject to truncation or fixed left censoring due to below detection limits. The truncated Kendall's tau,  $\tau_c$ , provides a measure for this association. We derive a nonparametric estimator for  $\tau_c$  and give an expression for its asymptotic variance. Under a copula model, we give an expression of  $\tau_c$  in terms of the copula parameter  $\alpha$ . Estimators for  $\alpha$  and its asymptotic variance follow. The proposal is illustrated with a real data set on HIV.

Nous nous intéressons à l'étude de l'association entre deux mesures sujettes à une troncation ou une censure à gauche fixe due à un seuil au dessous duquel les mesures ne peuvent être prises. Le tau de Kendall tronqué,  $\tau_c$ , permet de mesurer cette association. Nous développons un estimateur non paramétrique de  $\tau_c$  ainsi qu'une expression de sa variance asymptotique. Sous un modèle de copule donné, nous exprimons  $\tau_c$  en fonction du paramètre de la copule  $\alpha$ . Nous en déduisons des estimateurs de  $\alpha$  et de sa variance asymptotique. Ces résultats sont illustrés avec des données réelles sur le VIH.

**Session 07G****Mardi 25 mai • Tuesday, May 25, 10:25 – 11:40****VAN 1211****Shrinkage Estimation and Bayesian Methodology****Estimation par rétrécissement et méthodologie bayésienne****[MS-127]****Shrinkage Estimation Strategy in Non-linear Models****Stratégie d'estimation par rétrécissement pour modèles non linéaires**Saber FALLAHPUR & S. Ejaz AHMED, *University of Windsor*

In this talk we consider a shrinkage approach to the analysis of non-linear models in the presence

Dans cet exposé, nous considérons une approche par rétrécissement pour l'analyse de modèles non linéaires en

of uncertain prior information about the regression parameter vector. More specifically, we propose a shrinkage estimation strategy by using the maximum quasi-likelihood estimator (MQLE) as the benchmark estimator. Our asymptotic results clearly demonstrate the superiority of our proposed shrinkage strategy over the MQLE. Some Monte Carlo simulation results show that such a performance still holds in finite samples. We illustrate the usefulness of the suggested method by applying it to data.

[MS-128]

### Shrinkage Methods in Some Multivariate Dynamical Models

#### Méthodes par rétrécissement dans certains modèles dynamiques multidimensionnels

Sévérien NKURUNZIZA, *University of Windsor*

We consider an inference problem concerning the drift parameter matrices of  $M$  independent multivariate diffusion processes. In particular, we study cases where the  $M$ -parameters matrices are suspected to be restricted to certain candidate subspaces. Under such uncertainty, we propose shrinkage estimators which improve over the performance of the maximum likelihood estimator (MLE). Also, we establish the relative risk dominance of the proposed estimators. Further, we carry out intensive simulation studies for observation periods of small and moderate lengths of time that illustrate the performance of our method. Simulation results corroborate the theoretical finding for which shrinkage estimators outperform the MLE.

présence d'information a priori incertaine au sujet du vecteur des paramètres de régression. Plus spécifiquement, nous proposons une stratégie d'estimation par rétrécissement utilisant l'estimateur du maximum de quasi-vraisemblance (EMQV) comme repère. Nos résultats asymptotiques démontrent clairement la supériorité de la stratégie proposée par rapport à l'EMQV. Des résultats de simulation montrent que c'est toujours le cas sous échantillon fini. Nous illustrons l'utilité de la méthode suggérée par une application à des données.

Dans cette présentation, nous considérons le problème d'inférence concernant les paramètres matriciels de dérive d'un processus de diffusion multivarié  $M$ -indépendants. En particulier, nous étudions les cas où ces  $M$ -paramètres matriciels sont soupçonnés d'être restreints à certains sous-espaces. En tenant compte de ce sous-espace incertain, nous proposons des estimateurs par rétrécissement qui améliorent les performances de l'estimateur du maximum de vraisemblance (EMV). En outre, nous étudions les propriétés asymptotiques des estimateurs proposés et établissons leur efficacité asymptotique relative. De plus, nous effectuons une étude de simulation pour les courtes et moyennes périodes de temps d'observation afin d'illustrer les performances de la méthode proposée. Les résultats de simulation corroborent le résultat théorique selon lequel les estimateurs par rétrécissement dominent l'EMV.

[MS-129]

### Bayesian Confidence Intervals using Balanced Loss Functions

#### Intervalles de confiance bayésiens à l'aide de fonctions de perte pondérées

Mohamed TAMAZIGHT & ÉRIC MARCHAND, *Université de Sherbrooke*

There are number of papers which describe methods and properties of point estimation for a balanced loss functions. However, to our knowledge, no work has focused on confidence intervals, in particular with constraints, in a Bayesian framework with such loss functions. Given a balanced loss function and a target estimator, we propose a general method to build a class of Bayesian confidence intervals. We then propose a few appli-

Il existe des articles et des études qui proposent des méthodes et des propriétés d'estimation ponctuelle pour le choix d'une fonction de perte pondérée (*balanced loss function* en anglais). Mais, à notre connaissance, les intervalles de confiance, surtout pour un paramètre sous contraintes, n'ont pas fait l'objet d'une étude inférentielle bayésienne avec de telles fonctions de perte. Étant donné une fonction de perte pondérée et un estimateur cible nous proposons une méthode générale pour construire une classe d'intervalles de

cations and illustrations to study related analytical properties and coverage probabilities.

confiance bayésiens. Puis, nous proposerons quelques applications et illustrations pour étudier leurs propriétés analytiques et leurs probabilités de recouvrement.

### [MS-130]

#### Are Bayesian Inferences Weak for Wasserman’s Example?

#### Les inférences bayésiennes sont-elles inefficaces pour l’exemple de Wasserman ?

Longhai LI, *University of Saskatchewan*

An example was given in the textbook “All of Statistics” (Wasserman, 2004) for arguing that Bayesian inferences are weak in high-dimensional problems because they rely heavily on the likelihood that captures information of only a tiny fraction of the parameters. He recommended the Horwitz–Thompson estimator since it is unbiased and consistent. I compared the mean squared errors of the Horwitz–Thompson estimator with a Bayes estimator. From these comparisons, I conclude that the simple Bayes estimator works better than the Horwitz–Thompson estimator for most parameter configurations.

Dans le manuel *All of Statistics* (Wasserman, 2004), un exemple est donné afin d’illustrer que les inférences bayésiennes sont relativement inefficaces dans les problèmes à dimension élevée, car elles s’appuient fortement sur la vraisemblance qui ne donne de l’information que pour une petite fraction de l’ensemble des paramètres. Il argumente que l’estimateur Horwitz–Thompson représente une bonne option car il est sans biais et cohérent. J’ai comparé l’erreur quadratique moyenne de l’estimateur Horwitz–Thompson à celui d’un estimateur de Bayes. À partir de ces comparaisons, j’arrive à la conclusion que cet estimateur de Bayes est un meilleur choix que l’estimateur Horwitz–Thompson pour la plupart des configurations de paramètres.

### [MS-131]

#### On the Performance of Truncated Linear Estimators for Estimating a Restricted Multivariate Normal Mean Sur la performance d’estimateurs linéaires tronqués pour estimer la moyenne d’une loi multinormale sous contraintes

Othmane KORTBI & Éric MARCHAND, *Université de Sherbrooke*

We consider the problem of estimating the mean  $\theta$  of a  $N_p(\theta, I_p)$  distribution with squared error loss and under the constraint  $\|\theta\| \leq m$ , for some constant  $m > 0$ . By making use of Stein’s unbiased estimate of risk, sign change arguments, and a conditional risk analysis, we compare the risk performance of the truncated linear estimators  $\delta_a$  with that of  $\delta_{\text{mle}}$ . We obtain for a fixed  $(m, p)$  necessary and sufficient conditions on  $a$  for  $\delta_a$  to dominate  $\delta_{\text{mle}}$ . An asymptotic framework is developed, where we can show that the truncated minimax linear estimator dominates  $\delta_{\text{mle}}$ . Numerical evaluations illustrate the effectiveness of the framework, especially to approximate the risks for small values of  $p$ .

Nous considérons l’estimation sous coût quadratique de la moyenne  $\theta$  d’une loi  $N_p(\theta, I_p)$  sous la contrainte  $\|\theta\| \leq m$  pour une certaine constante  $m > 0$ . En utilisant l’identité de Stein pour obtenir un estimateur sans biais du risque, les arguments du changement du signes de Karlin et l’analyse du risque conditionnel, nous comparons la performance de risque des estimateurs linéaires tronqués  $\delta_a$  avec celui de l’estimateur du maximum de vraisemblance  $\delta_{\text{evm}}$ . Nous avons obtenu, pour  $(m, p)$  fixe, des conditions nécessaires et suffisantes sur  $a$  pour que  $\delta_a$  domine  $\delta_{\text{evm}}$ . Un cadre asymptotique est développé et ceci nous permet de montrer que l’estimateur linéaire minimax tronqué domine  $\delta_{\text{evm}}$  dans ce cadre et de mesurer avec précision le degré d’amélioration. Des évaluations numériques témoignent de l’efficacité du cadre, notamment pour approcher le comportement des estimateurs pour de petites valeurs de  $p$ .

**Session 07H      Mardi 25 mai • Tuesday, May 25, 10:25 – 11:40      VAN 1871**

**Density Estimation and Robust Regression**  
**Estimation de la densité et régression robuste**

[MS-132]

**On the Boundary Properties of Bernstein Polynomial Estimators**  
**Comportement des estimateurs de Bernstein aux frontières de l'intervalle unitaire**

Alexandre LEBLANC, *University of Manitoba*

For density and distribution functions supported on  $[0, 1]$ , Bernstein polynomial estimators are known to have optimal Mean Integrated Squared Error (MISE) properties under the usual smoothness conditions on the function to be estimated. These estimators are also known to be well-behaved in terms of bias, as they exhibit no boundary bias. In this talk, we will discuss the fact that these estimators nevertheless do experience boundary effects. However, these boundary effects are of a different nature than what is seen, for example, with usual kernel estimators.

Pour l'estimation de densités et de fonctions de répartition définies sur l'intervalle  $[0, 1]$ , il est bien connu que les estimateurs basés sur les polynômes de Bernstein atteignent des taux de convergence optimaux en Écart Quadratique Moyen Intégré (EQMI) sous les hypothèses usuelles de régularité des fonctions considérées. Il est aussi bien connu que ces estimateurs ne souffrent pas d'une augmentation de leur biais près des frontières de l'intervalle unitaire. Durant cette présentation, nous discuterons du fait que la performance de ces estimateurs est néanmoins différente près des frontières de l'intervalle. La nature des changements constatés est cependant bien différente de ce qui est typiquement observé, par exemple, avec les estimateurs à noyaux usuels.

[MS-133]

**On the Determination of the Polynomial Component of Certain Density Estimates**  
**Sur la détermination de la composante polynomiale de certains estimateurs de fonctions de densité**

Serge PROVOST, *University of Western Ontario*, Min JIANG, *Statistics Canada*

Density estimates that are expressible as the product of a base density and a linear combination of orthogonal polynomials are considered. It is shown that, in the context of density approximation, the coefficients of the linear combination can be determined either from a moment-matching technique or a least-squares approach. A kernel representation is obtained for the corresponding density estimates. Modified stopping criteria are also proposed for determining the degree of the polynomial component. By way of illustration, the density estimation methodology advocated herein is applied to two data sets.

Nous considérons des fonctions de densité qui peuvent être représentées par le produit d'une fonction de densité de base et d'une combinaison linéaire de polynômes orthogonaux. Il est démontré que deux approches équivalentes permettent de spécifier les coefficients de la combinaison linéaire lorsqu'il s'agit d'approximer une fonction de densité. D'autre part, nous exprimons les estimateurs de ce type comme des estimateurs à noyau. De plus, nous modifions certains critères connus afin de déterminer le degré de la composante polynomiale. La méthodologie ici proposée est appliquée à deux jeux de données.

**[MS-134]****Using a Linear Combination of Adjustments to Enforce Shape Constraints on a Density Estimate  
Utilisation d'une combinaison linéaire d'ajustements afin de faire respecter des contraintes de forme à un estimateur de densité**Mark WOLTERS & John BRAUN, *University of Western Ontario*

Producing accurate density estimates which satisfy given qualitative constraints is not straightforward with standard kernel density estimation. A new method for shape-constrained density estimation will be presented to address this problem. The method involves adding a linear combination of adjustment densities to an unconstrained starting estimate. The adjustment and the constraints are set up in such a way that the coefficients of the combination can be found using readily-available quadratic programming codes. The talk will discuss how to choose the adjustment densities, what types of constraints can be included, and how the new method compares to standard estimators.

L'obtention d'estimateurs de densités satisfaisant des contraintes qualitatives par la méthode des noyaux habituelle n'est pas simple. Une nouvelle méthode pour l'estimation de densités restreintes à certaines formes sera présentée afin de résoudre ce problème. Cette méthode consiste en l'ajout d'une combinaison linéaire de densités d'ajustement à un estimé de départ non restreint. L'ajustement et les contraintes sont choisis tels que les coefficients de la combinaison linéaires peuvent être obtenus à partir de méthodes bien établies. Cet exposé discutera du choix des densités d'ajustement, du type de contraintes pouvant être incluses, et de la performance de la nouvelle méthode relative aux estimateurs existants.

**[MS-135]****Shrinkage  $M$ -Estimation in Partially Linear Models  
 $M$ -estimation par rétrécissement dans des modèles partiellement linéaires**Enayetur RAHEEM & S. Ejaz AHMED, *University of Windsor*

We consider shrinkage  $M$ -estimation in the context of a partially linear model. Stein-type  $M$ -estimation combines unrestricted and restricted semiparametric estimates producing a hybrid estimator. Generally, Stein-type estimates are biased but are more efficient than their unrestricted counterparts. Asymptotic distributional bias and asymptotic distributional risk (ADR) criteria are used to evaluate the performance of the proposed estimators. We show that under certain regularity conditions, Stein-type  $M$ -estimator dominates the other estimators in terms of quadratic risk. A Monte Carlo simulation will be carried out to validate the asymptotic results.

Nous considérons la  $M$ -estimation par rétrécissement dans le contexte d'un modèle partiellement linéaire. La  $M$ -estimation de Stein combine des estimés semi-paramétriques avec et sans contraintes, produisant un estimateur hybride. Généralement, les estimateurs de Stein sont biaisés, mais plus efficaces que les estimateurs correspondants sans contraintes. Les critères du biais distributionnel asymptotique et du risque distributionnel asymptotique sont utilisés afin d'évaluer la performance des estimateurs proposés. Nous montrons que sous certaines conditions de régularité, les  $M$ -estimateurs de Stein dominent les autres estimateurs en termes de risque quadratique. Une étude de Monte-Carlo est menée afin de valider les résultats asymptotiques.

**[MS-136]****An Outlier-robust Fit for Generalized Additive Models with Applications to Outbreak Detection  
Ajustement robuste aux valeurs aberrantes de modèles additifs généralisés avec applications à la détection d'épidémies**Matias SALIBIAN-BARRERA, *University of British Columbia*, Azadeh ALIMADAD, *Simon Fraser University*

We are interested in detecting disease outbreaks using GAMs to model case counts. Unfortunately,

Nous nous intéressons à la détection d'épidémies en utilisant des MAG pour modéliser les nombres de cas. Malheureusement,

GAMs can be very sensitive to a small proportion of outliers, which may mask the outbreak. We discuss an outlier-robust fit for GAMs where we replace the MLE-based weights in the Generalized Local Scoring Algorithm with those derived from robust quasi-likelihood equations. We show that the resulting estimated mean function is resistant to outliers in the response variable. It also remains close to the GAM estimator when no outliers are present in the data. We illustrate our approach with the recent outbreak of H1N1 flu.

ment, les MAG peuvent être très sensibles à une faible proportion de valeurs aberrantes pouvant masquer l'épidémie. Nous discutons d'un ajustement robuste aux valeurs aberrantes des MAG où nous remplaçons les poids basés sur les EMV dans l'algorithme de score local généralisé par des poids obtenus d'équations de quasi-vraisemblance robuste. Nous montrons que la fonction moyenne estimée résultante est résistante aux valeurs aberrantes dans la réponse. Elle demeure également semblable à l'estimateur MAG en l'absence de valeurs aberrantes. Nous illustrons notre approche à l'aide de l'épidémie récente de grippe H1N1.

## **Session 07I      Mardi 25 mai • Tuesday, May 25, 10:25 – 11:55      VAN 1883**

### **Time Series, Spatio-temporal Data, and Statistical Education**

### **Séries chronologiques, données spatio-temporelles et l'enseignement de la statistique**

#### **[MS-137]**

#### **Forecasting ARCH/GARCH Processes using Sieve Bootstrap**

#### **Prévision de processus ARCH/GARCH via le bootstrap par tamis**

Bei CHEN & Yulia GEL, *University of Waterloo*, N. BALAKRISHNA, *Cochin University of Science and Technology*, BOVAS ABRAHAM, *University of Waterloo*

Analyzing volatilities is of great practical importance for quantifying risk and uncertainty in financial markets. The most widely applied tools for modelling volatility are ARCH/GARCH models. We propose to apply sieve bootstrap, a distribution-free re-sampling technique, to construct prediction intervals of returns and volatilities in ARCH/GARCH processes. Compared to point forecasts, prediction intervals provide extra assessment of uncertainty associated with the corresponding point forecast, which can better guide risk management decisions. In this talk, we will show the experimental results of the proposed method, both for simulated and foreign exchange rate data.

L'analyse de volatilités est d'une grande importance afin de quantifier le risque et l'incertitude de marchés financiers. Les outils les plus souvent utilisés pour modéliser des volatilités sont les modèles ARCH/GARCH. Nous proposons d'appliquer le bootstrap tamis, une méthode de ré-échantillonnage sans hypothèse distributionnelle, afin de construire des intervalles de prévision pour les rendements et volatilités avec des processus ARCH/GARCH. Les intervalles de prévision apportent une évaluation supplémentaire de l'incertitude associée à une prévision ponctuelle, permettant ainsi de mieux guider la prise de décision en gestion de risque. Dans cet exposé, nous allons montrer les résultats expérimentaux de la méthode proposée avec des données de taux de change simulées et réelles.

#### **[MS-138]**

#### **Calibrating the Block Bootstrap for Time Series**

#### **Calibration du bootstrap par blocs pour séries chronologiques**

Lengyi HAN & John BRAUN, *University of Western Ontario*

The block bootstrap for time series can give confidence intervals with very poor coverage properties. Lee & Lai (*Biometrika*, 2009) proposed an alpha-level adjustment procedure based on a dou-

Le bootstrap en blocs pour séries chronologiques peut donner des intervalles de confiance ayant de mauvaises propriétés de recouvrement. Lee & Lai (*Biometrika*, 2009) ont proposé une méthode d'ajustement du niveau alpha basée sur

ble bootstrap and showed that it behaves accurately asymptotically, but they did not provide substantial simulation results to demonstrate its performance on small samples. An alternative approach is to adjust the confidence limits directly. We compare the two procedures by simulation. The level adjustment procedure sometimes provides poor coverage in small samples. The direct procedure rarely does poorly. In terms of mean interval width, neither method outperforms the other.

un double bootstrap et ont montré qu'elle un bon comportement asymptotique, mais ils n'ont pas donné de résultats de simulation substantiels permettant d'évaluer sa performance pour de petits échantillons. Une approche alternative est d'ajuster directement les limites l'intervalle. Nous comparons les deux méthodes par des simulations. La méthode d'ajustement du niveau mène parfois à une mauvaise probabilité de recouvrement lorsque l'échantillon est petit, tandis qu'il est rare que la procédure directe ne soit pas satisfaisante à cet égard. En termes de largeur moyenne de l'intervalle, aucune des deux méthodes ne se distingue.

### [MS-139]

#### **Maximum Likelihood Estimation for Separable Variance-covariance Structures in 2D and 3D**

#### **Estimation par maximum de vraisemblance de structures de variance-covariance séparables en 2D et 3D**

Ameur MANCEUR & Pierre DUTILLEUL, *McGill University*

Bi- and multidimensional experimental data (e.g., multivariate spatial and/or temporal) with replicates can be modelled with the matrix and tensor normal distributions, under separability of the variance-covariance structure without necessarily imposing stationarity. Maximum likelihood estimates of the variance-covariance parameters are obtained with an iterative algorithm, and the mean is model-free. The empirical bias (standardized, non-standardized) and the dispersion of the estimated variance-covariance matrix of  $\text{vec}(X)$  were studied by simulation in several 2D and 3D scenarios. As expected, the bias and dispersion vanish asymptotically. Special attention is paid to situations with minimum and small numbers of replicates, particularly in 2D.

Les données expérimentales bi- et multidimensionnelles (multivariées spatiales ou temporelles) avec réplicats peuvent être modélisées avec les distributions normales matricielle et tensorielle, en supposant une structure variance-covariance séparable et non stationnaire. Les estimés par maximum de vraisemblance des paramètres de variance-covariance sont obtenus itérativement ; la moyenne est estimée sans modèle. Le biais empirique (standardisé, non standardisé) et la dispersion de la matrice variance-covariance estimée de  $\text{vec}(X)$  furent étudiés par simulation pour plusieurs scénarios 2D et 3D. Tel qu'attendu, le biais et la dispersion s'annulent asymptotiquement. Une attention plus particulière fut portée aux situations avec un nombre minimum ou petit de réplicats, spécialement en 2D.

### [MS-140]

#### **Spatio-temporal Estimates as Predictors in a Hierarchical Model**

#### **Estimateurs spatio-temporels comme prédicteurs dans un modèle hiérarchique**

Katrina ROGERS-STEWART, Patrick FARRELL & Jason NIELSEN, *Carleton University*

Spatio-temporal data has become increasingly abundant in recent years; however its use for prediction often requires estimation at coordinates with no observed values. For such data, we propose to model the spatial component with a number of two-dimensional splines and correlate these splines to reflect the temporal relationship. The resulting estimates are then used as covariates in a regression. We employ a hierarchical Bayesian approach

Les données spatio-temporelles sont devenues plus abondantes au cours des dernières années ; cependant leur utilisation en prévision nécessite souvent une estimation à des coordonnées sans observations. Pour de telles données, nous proposons de modéliser la composante spatiale avec des splines bidimensionnelles et de corréliser ces splines afin de refléter la relation temporelle. Les estimés résultants sont alors utilisés comme covariables dans un modèle de régression. Nous employons une approche hiérar-

which naturally incorporates the uncertainty of the estimated covariate. As an illustration, we use an example involving particulate matter concentration and asthma hospitalizations in Northeastern United States.

chique bayésienne incorporant naturellement l'incertitude des covariables estimées. Comme illustration, nous utilisons un exemple de concentration de matière particulaire et d'hospitalisations liées à l'asthme aux États-Unis du nord-est.

#### [MS-141]

##### **Student Attitudes to Statistics: Survey Results and Discussion**

##### **Attitudes étudiantes par rapport à la statistique : résultats d'enquête et discussion**

AnneMichele MILLAR, *Mount Saint Vincent University*, Candace SCHAU, *CS Consultants, LLC*

Students' attitudes to statistics on entering their introductory course have a strong influence on their ability to achieve their potential, their likelihood of continued study, and their use of statistics outside the classroom. We administered the "Survey of Attitudes Towards Statistics" (SATS), which assesses six attitude components, at the beginning and the end of our Fall semester introductory course. We present an analysis of the relationships between the various components, and the changes in these components over the duration of the course. We also discuss statistical issues arising from that analysis.

L'attitude des étudiants face à la statistique lors de leur premier cours aura une influence durable sur leur capacité à atteindre leur potentiel, la probabilité pour qu'ils poursuivent dans ce domaine, et leur utilisation de la statistique hors de la salle de classe. Nous avons administré le Questionnaire sur les attitudes face à la statistique (QAFS), qui évalue six composantes de l'attitude, au début et à la fin de notre cours d'introduction à la statistique du semestre d'automne. Nous présentons une analyse de la relation entre les diverses composantes, et les changements dans ces composantes au cours du semestre. Nous discutons aussi de questions statistiques soulevées par cette analyse.

#### [MS-142]

##### **Teaching Statistics Utilizing Web 2.0**

##### **Enseigner la statistique avec le Web 2.0**

Tasneem ZAIHRA, *University of New Brunswick—Saint John*, Hasan RIZVI

We discuss in detail how Web 2.0 can be incorporated in the classroom while teaching statistics courses to facilitate interactive learning. Examples of Web 2.0 discussed in this paper include, video sharing, RSS, wikis and blogs. Web 2.0 enables a teacher to eliminate performance oriented delivery of lecture and convert learning into a more interactive process. With Web 2.0 the student becomes a part of the teaching allowing richness and growth. The goal of this paper is to demonstrate how Web 2.0 teaching facilitates learning by bridging the gap between student and teacher.

Nous discutons en détail de l'incorporation du Web 2.0 aux cours de statistique afin de faciliter l'apprentissage interactif en classe. Des exemples de web 2.0 abordés dans cette communication incluent le partage de vidéos, RSS, les wikis et les blogues. Le Web 2.0 permet à un enseignant de substituer au facteur de la performance un processus d'apprentissage plus interactif. Avec le Web 2.0, l'étudiant devient part intégrante de l'enseignement, facilitant son enrichissement et sa croissance. Le but de cette communication est de montrer comment l'enseignement avec le Web 2.0 facilite l'apprentissage en comblant le vide entre étudiant et enseignant.

**Session 08A**      **Mardi 25 mai • Tuesday, May 25, 13:15 – 14:15**      **POU 1112**  
**CRM–SSC Prize in Statistics Address**  
**Allocution de la récipiendaire du prix CRM-SSC**

[MS-143]

**Statistical Analysis of “Imperfect” Data: Handling Missing Observations and Measurement Error**  
**Analyse statistique de données « imparfaites » : Tenir compte d’observations manquantes et d’erreurs de mesure**

Grace Yi, *University of Waterloo*

Standard statistical analysis is often challenged by new emerging issues and various complex features of data arising in practice. We are frequently faced with “imperfect” data that can not be analyzed directly with standard methods. Data may be “imperfect” due to the presence of missing values or imprecise measurements, which are pervasive in applications. It is well known that ignoring these features in statistical analyses could lead to severely misleading results. Although there has been rapid development in the analysis of such data in recent years, many challenging problems still remain. In this talk, I will discuss some modelling and analysis methods concerning data with missing observations or measurement error.

Les méthodes d’analyse statistique habituelles sont souvent mises à l’épreuve par de nouvelles questions émergentes et par les diverses caractéristiques complexes des données rencontrées en pratique. Nous faisons souvent face à des données « imparfaites » ne pouvant être directement analysées avec les méthodes usuelles. L’« imperfection » des données peut être due à la présence de valeurs manquantes ou de mesures imprécises, souvent rencontrées dans les applications. Il est bien connu que les analyses statistiques peuvent mener à des résultats trompeurs lorsque ces caractéristiques sont ignorées. Bien que l’analyse de telles données ait connu un développement rapide au cours des dernières années, plusieurs problèmes persistent. Dans cet exposé, je vais discuter de méthodes de modélisation et d’analyse pour données à observations manquantes ou avec erreurs de mesure.

**Session 09A**      **Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55**      **VAN 1245A**  
**Statistical Methods for Life History Analysis**  
**Méthodes statistiques pour l’analyse de l’historique de vie**

[MS-144]

**Competing Risks and Time-dependent Covariates**  
**Risques concurrents et covariables dépendant du temps**

Per Kragh ANDERSEN, *University of Copenhagen, Denmark*, Giuliana CORTESE, *Università degli studi di Padova, Italia*

In survival analysis, hazard regression models enable inclusion of time-dependent covariates. However, when including internal/endogenous covariates, the one-to-one correspondence between hazard function and cumulative survival probability is lost. This problem persists in models for competing risks. In this paper we explore some methods by which cumulative incidences in competing risks models may be estimated in the presence of an internal time-dependent covariate. One method is based on an extension of the competing risks multi-

En analyse de survie, les modèles de régression pour le risque permettent l’inclusion de covariables dépendantes du temps. Cependant, lorsque des covariables internes/endogènes sont utilisées, la correspondance biunivoque entre le taux de défaillance et la probabilité de survie cumulative est perdue. Ce problème persiste dans les modèles à risques concurrents. Dans cet article, nous explorons quelques méthodes pour estimer les incidences cumulatives dans des modèles à risques concurrents en présence d’une covariable interne dépendante du temps. Une des méthodes est basée sur une généralisation du modèle à risques concurrents multi-

state model while the other builds on the so-called “landmark” approach. Data from a bone marrow transplantation study are used to illustrate the techniques.

tats, alors que l’autre est développée à partir de l’approche des points de repère. Des données de transplantation de moelle osseuse sont utilisées afin d’illustrer les techniques.

#### [MS-145]

##### **Analysis of Time-to-pregnancy from Current Duration Data**

##### **Analyse du temps d’attente avant une grossesse à partir de durées d’attentes actuelles**

Niels KEIDING & Oluf HANSEN, *University of Copenhagen, Denmark*, Rémy SLAMA, *Institut Albert Bonniot/Inserm U823, Grenoble, France*

Time to pregnancy is the duration from a couple starts trying to become pregnant until they succeed and is considered one of the most direct methods to measure natural fecundity in humans. Classical prospective studies are hard to perform in this area, and the most commonly used retrospective (pregnancy-based) designs offer several pitfalls in the analysis.

This talk is about a third possible design, starting from a cross-sectional sample of couples currently trying to become pregnant, using current duration (backward recurrence time) as basis for the estimation. Regression analysis is then most conveniently done in the accelerated failure time model.

Le délai nécessaire pour concevoir correspond à la durée entre le moment où un couple débute une tentative de grossesse et le début d’une grossesse éventuelle ; sa mesure est considérée comme une des méthodes les plus directes pour estimer la fertilité chez l’humain. Les études prospectives classiques sont difficiles à mettre en œuvre dans ce domaine, et l’approche rétrospective (reposant sur les grossesses) qui est la plus couramment utilisée présente de nombreux écueils lors de l’analyse statistique.

Cet exposé présente une troisième approche, reposant sur un échantillon transversal de couples cherchant à obtenir une grossesse lors de l’étude, en utilisant la durée d’attente en cours (temps de récurrence rétrospectif) afin de fournir une estimation de la fertilité. Une analyse par régression peut alors être aisément réalisée à l’aide du modèle de défaillance accélérée.

#### [MS-146]

##### **The Cox Model via Generalized Estimating Equations**

##### **Le modèle de Cox par des équations d’estimation généralisées**

Daniel FAREWELL, *Cardiff University*

Survival analysis and longitudinal data analysis have traditionally been seen as distinct disciplines within statistics, but both subjects concern processes evolving in time. I will argue that there is much to be gained from cross-pollination between the two fields. In particular, understanding discrete survival times can be simplified by thinking in terms of repeated measurements, while longitudinal data analysts can learn much about the treatment of censored data from their survival analyst colleagues. To illustrate the close connection between the two, I shall derive a Cox model for failure times from a generalized estimating equation.

L’analyse de survie et l’analyse de données longitudinales sont souvent considérées comme des disciplines distinctes. Néanmoins, puisque les deux concernent des processus temporels, je discuterai des avantages à retirer de leurs nombreux aspects communs. En particulier, l’étude des temps de survie discrets peut être simplifiée en pensant en termes de mesures répétées, alors que l’analyse de données longitudinales avec censure est facilitée en adoptant une perspective de modèles de survie. Pour illustrer le lien entre les deux, je vais construire un modèle de Cox à partir d’une équation d’estimation généralisée.

**Session 09B      Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55      VAN 2289A**

Recent Advances in Modelling Long Range Dependent and Nonlinear Time Series, with Applications to Business, Econometrics, and Environmental Sciences

Développements récents dans la modélisation de la dépendance à long terme et de séries chronologiques non linéaires, avec applications en gestion, économétrie et sciences environnementales

[MS-147]

**Fast Bayesian Estimation for Sub-Gaussian Stable VARFIMA Processes via a Multivariate Preconditioned Conjugate Gradient Algorithm**

**Estimation bayésienne rapide pour processus stables VARFIMA sous-gaussiens au moyen d'un algorithme de gradient conjugué préconditionné multivarié**

Nalini RAVISHANKER, *University of Connecticut, Storrs*, Jeffrey PAI, *University of Manitoba*

We present an approach via a multivariate preconditioned conjugate gradient (MPCG) algorithm for Bayesian inference for vector ARFIMA models with sub-Gaussian stable errors. This approach involves solving a block-Toeplitz system, and treating the unobserved process history and the underlying positive stable process as unknown quantities in the joint posterior. We use a modified Gibbs sampling algorithm for inference. We present results using long time series of daily average temperatures measured at several U.S. cities, and of the absolute daily returns of several stocks.

Nous présentons une approche basée sur un algorithme de gradient conjugué préconditionné multivarié (GCPM) pour l'inférence bayésienne à partir de modèles ARFIMA vectoriels avec erreurs stables sous-gaussiennes. Cette approche nécessite la résolution d'un système Toeplitz en blocs, ainsi que le traitement de l'historique non observé des processus et des processus positifs stables sous-jacents comme des quantités inconnues dans la loi a posteriori conjointe. Nous utilisons un échantillonneur de Gibbs modifié pour l'inférence. Nous appliquons l'approche à de longues séries chronologiques de températures journalières moyennes mesurées dans plusieurs villes des États-Unis, et de rendements journaliers absolus de plusieurs actions.

[MS-148]

**On the Combined Estimating Functions Method with Applications in Finance**

**De la méthode des fonctions estimatrices combinées avec applications en finance**

Melody GHAHRAMANI, *University of Winnipeg*, Aerambamoorthy THAVANESWARAN, *University of Manitoba*

Volatility plays an important role in financial forecasting and in option pricing. Various volatility models are such that the first two conditional moments of the observed process depend on the parameter of interest. When there are two estimating functions for the same parameter, a more informative estimating function may be obtained by combining them as in Ghahramani & Thavaneswaran (2009). In this talk, some applications of the combined estimating functions method for financial time series will be discussed. Applications include hypothesis tests for ARMA models with GARCH errors, GARCH model identification and, nonlinear recursive forecasting.

La volatilité joue un rôle important en prévision financière et dans la fixation des prix d'options. Plusieurs modèles de volatilité sont tels que les deux premiers moments conditionnels du processus observé dépendent du paramètre d'intérêt. Lorsqu'il y a deux fonctions d'estimation pour le même paramètre, une fonction d'estimation plus informative peut être obtenue en les combinant, tel que décrit dans Ghahramani & Thavaneswaran (2009). Dans cet exposé, certaines applications de la méthode des fonctions d'estimation combinées seront abordées pour les séries chronologiques en finance. Les applications comprennent les tests d'hypothèses pour modèles ARMA à erreurs GARCH, l'identification de modèles GARCH et la prévision récursive non linéaire.

**[MS-149]**

**A New Quantile Function Based Model for Modeling Price Behaviours in Financial Markets**  
**Un nouveau modèle à base de fonctions quantiles pour modéliser le comportement des prix en marchés financiers**

Gemai CHEN, *University of Calgary*

This talk uses a class of quantile functions to develop a new time series model for studying financial price behaviours through the tail properties of the price instead of the volatilities (variances). The model characterizes the price behaviours using a tail order measure which helps forecast how volatile the prices will be, and a tail balance measure that helps estimate whether an investment tends to gain or tends to lose. The model parameters can be estimated using the method of maximum likelihood, and two real data sets are analyzed to show the potential usefulness of our proposed model.

Cet exposé utilise une classe de fonctions quantiles afin de développer un nouveau modèle de séries chronologiques pour l'étude du comportement de prix de produits financiers à partir des propriétés caudales du prix plutôt que des volatilités (variances). Le modèle caractérise le comportement du prix au moyen d'une mesure d'ordre de la queue aidant à prévoir à quel point les prix seront volatiles, et d'une mesure d'équilibre de la queue aidant à estimer si un investissement tend à rapporter ou non. Les paramètres du modèle peuvent être estimés par la méthode du maximum de vraisemblance, et deux jeux de données réelles sont analysés afin de démontrer l'utilité potentielle du modèle.

**Session 09C      Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55      VAN 2291A**

**Quality in Surveys**  
**Qualité dans les enquêtes**

**[MS-150]**

**The Compendium of Recommended Survey Practices of the Institut de la statistique du Québec**  
**Le Recueil de bonnes pratiques dans les enquêtes de l'Institut de la statistique du Québec**

France LAPOINTE, Sylvain VÉGIARD & Louise BOURQUE, *Institut de la statistique du Québec*

The incorporating legislation of the *Institut de la statistique du Québec*, passed in 1999, contains clear provisions concerning the quality provisions it must comply with. This is why in January 2006 the Institut adopted the General Quality Management Policy including, in particular, elements pertaining to the quality of its statistical surveys. Survey quality assurance rests on the implementation of two tools: a compendium of recommended practices and a survey quality self-assessment checklist. The presentation is on the elaboration of the Compendium of Recommended Survey Practices and its dissemination within the organization.

La loi constitutive de l'Institut de la statistique du Québec, adoptée en 1999, contient des dispositions claires quant aux obligations de qualité qu'il doit respecter. Pour cette raison, l'Institut a adopté en janvier 2006 la Politique générale en matière de gestion de la qualité comportant des éléments relatifs à la gestion de la qualité de ses enquêtes statistiques. L'assurance de la qualité des enquêtes de l'Institut repose sur l'implantation de deux outils : un recueil de bonnes pratiques et une grille d'autoévaluation de la qualité. La présentation traitera de l'élaboration du Recueil de bonnes pratiques dans les enquêtes et de sa diffusion au sein de l'organisation.

**[MS-151]****Quality in Continuing Care: How new CIHI Data and Indicators are Supporting Health Care****Qualité en soins de suivi : comment les nouvelles données et indicateurs ICIS soutiennent les soins de santé**Maureen KELLY, *Canadian Institute for Health Information*

CIHI's Continuing Care Reporting System (CCRS) is designed to collect longitudinal clinical and administrative data on residents of hospital-based and residential-based continuing care facilities. It has grown dramatically over the last three years and now contains information from around 40% of the residential care facilities in Canada. This presentation will illustrate how CIHI's data quality framework is being applied to measure and document the quality of the CCRS data. It will also describe the methodology behind the new CCRS Quality Indicators which have been developed to help health managers, planners and policy makers to improve the quality of continuing care.

Le Système d'information sur les soins de longue durée (SISLD) de l'ICIS est conçu afin de recueillir des données longitudinales cliniques et administratives sur les résidents d'établissements de soins de longue durée en hôpital et en établissements de soins pour bénéficiaires internes. Il a connu une forte croissance au cours des trois dernières années et contient maintenant de l'information sur environ 40 % des établissements de soins pour bénéficiaires internes du Canada. Cette présentation illustrera comment le cadre de qualité des données de l'ICIS est appliqué afin de mesurer et de documenter la qualité des données du SISLD. Elle va aussi décrire la méthodologie derrière les nouveaux indicateurs de qualité du SISLD développés afin d'aider les gestionnaires de la santé, les planificateurs et responsables politiques à améliorer la qualité des soins de longue durée.

**[MS-152]****Improving the Quality of Automated Coding at Statistics Canada****Améliorer la qualité du codage automatisé à Statistique Canada**Xinye (Hannah) YANG & Caroline PELLETIER, *Statistics Canada*

Coding is a data processing step that assigns responses of open-ended questions into predefined categories identified by alphanumeric or numeric codes. Coding is performed automatically and/or manually. The Automated Coding by Text Recognition (ACTR) system was developed at Statistics Canada and used by several coding applications within and outside the organization. The methodology in ACTR was recently re-evaluated in an effort to achieve a higher coding rate and improve the timeliness of data without sacrificing data quality. In this talk, we present an alternative coding method and compare it to the one implemented in ACTR.

Le codage est une étape du traitement des données qui assigne les réponses aux questions ouvertes à des catégories prédéfinies identifiées par des codes numériques ou alphanumériques. Le codage se fait de façon automatique ou manuelle. Le Système de codage automatisé par reconnaissance de texte (CART) a été développé à Statistique Canada et est utilisé par plusieurs applications de codage à l'intérieur et à l'extérieur de l'organisme. La méthodologie utilisée dans le CART a récemment été réévaluée dans l'intention d'obtenir un taux de codage plus élevé et d'améliorer l'actualité des données, sans pour autant en sacrifier la qualité. Dans cet exposé, nous présentons une méthode de codage alternative et nous la comparons à celle actuellement mise en œuvre dans le CART.

**Session 09D      Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55      VAN 2809A**

**Stochastic Dynamics in Biology  
Dynamique stochastique en biologie**

[MS-153]

**Geometric Properties of the Majority Rule Model  
Propriétés géométriques du modèle de la loi majoritaire**

Nicolas LANCHIER & Jared NEUFER, *Arizona State University*

The majority rule model is an interacting particle system with the same state space as the voter model: each vertex of the regular lattice is occupied by an individual either black or white. The population is further divided into groups, with all the individuals within the same group updating their colour according the majority within the group. Whereas the voter model exhibits a diffusive clustering in two dimensions, numerical simulations suggest that stronger spatial correlations emerge for the majority rule model. We present analytical results in support of this picture.

Le modèle de la loi majoritaire est un système infini de particules avec le même espace d'état que le modèle des votants ; chaque site du réseau est occupé par un individu noir ou blanc. La population est en outre divisée en groupes. Tous les individus au sein du même groupe mettent à jour leur couleur selon la majorité du groupe. Tandis que le modèle des votants exhibe un regroupement diffusif en dimension deux, des corrélations spatiales plus prononcées émergent des simulations numériques du modèle de la loi majoritaire. Nous présentons des résultats analytiques en accord avec cette image.

[MS-154]

**Traveling Waves of Selective Sweeps  
Vagues voyageuses de balayages sélectifs**

John MAYBERRY & Rick DURRETT, *Cornell University*

The goal of cancer genome sequencing projects is to determine the genetic alterations that cause common cancers. Many malignancies arise during the clonal expansion of a benign tumour which motivates the study of recurrent selective sweeps in an exponentially growing population. To better understand this process, we consider a Moran type model in which cells from an exponentially growing population accumulate advantageous mutations and obtain small mutation rate asymptotics for the time until the first mutant with  $k$  mutations appears on the scene.

Le but des projets de séquençage du génome pour le cancer est de déterminer les altérations génétiques causant les types de cancer les plus fréquents. Plusieurs tumeurs malignes surviennent lors de l'expansion clonale de tumeurs bénignes, ce qui motive l'étude de balayages sélectifs récurrents d'une population à croissance exponentielle. Afin de mieux comprendre de processus, nous considérons un modèle de type Moran où les cellules d'une population à croissance exponentielle accumulent des mutations avantageuses et nous obtenons le comportement asymptotique pour faible taux de mutation du temps d'apparition du premier mutant à  $k$  mutations.

[MS-155]

**A Multitype Branching Model with Local Self-regulation  
Un modèle de branchement multitype avec auto-régulation locale**

Anita WINTER, *TU München*, Andreas GREVEN, *Universität Erlangen-Nürnberg*, Peter PFAFFELHUBER, *Freiburg*, Anja STURM, *Universität Göttingen*, Iljana ZÄHLE, *Universität Erlangen-Nürnberg*

We consider a spatial multi-type branching model in which individuals migrate in  $Z^d$  according

Nous considérons un modèle de branchement spatial multi-type dans lequel les individus migrent en  $Z^d$  selon des

to random walks and reproduce according to a branching mechanism which can be sub-, super- or critically depending on carrying capacities and the local intensity of individuals of the different types. In this talk we will focus on the diffusion limit of small mass, locally many individuals and rapid reproduction in the exchangeable set-up where non of the parameters involved in the model are type dependent.

In Hutzenthaler (2007) it has been shown that there are parameter regimes allowing for survival in all dimensions. In this talk we present duality relations which allow for monotonicity statements in the parameters with regards to whether or not the different surviving types can coexist.

marches aléatoires et se reproduisent selon un mécanisme de branchement qui peut être sous-, super- ou critique selon les capacités limites et l'intensité locale des individus des différents types. Dans cet exposé, nous mettrons l'accent sur la limite de diffusion où localement plusieurs individus de petite masse se reproduisent rapidement sous des conditions d'échangeabilité où aucun des paramètres impliqués dans le modèle ne dépend du type.

Dans Hutzenthaler (2007), il a été démontré qu'il y a des régimes de paramètres qui permettent la survie dans toutes les dimensions. Dans cet exposé, nous présentons des relations duales qui permettent des énoncés de monotonie sur les paramètres concernant la coexistence ou non de plusieurs types.

## **Session 09E      Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55      VAN 2811A**

### **Privacy Legislation—Implications for Canadian Statisticians**

### **Législations sur la vie privée — Conséquences pour les statisticiens canadiens**

**[MS-156]**

**What Should you Have in Hand when you Knock on the Door of your Research Ethics Board?**

**Que devriez-vous avoir en main avant de frapper à la porte de votre comité d'éthique ?**

*DON WILLISON, Ontario Agency for Health Protection and Promotion*

Research uses of health information are becoming more sophisticated and intensive. Single time-limited studies are being replaced by research platforms, in the form of registries and biobanks, and these are being linked with administrative and clinical records. As this use intensifies, so also do concerns over protection of privacy and confidentiality. Research ethics boards (REBs) have taken a prominent role in oversight over these matters. In this presentation, the author will briefly review conceptual frameworks for managing privacy and confidentiality, identify key approaches to statistical disclosure control, and walk participants through a sample REB application involving medical record review, with a view to pointing out the privacy-related issues considered by REBs these days.

L'utilisation d'informations sur la santé à des fins de recherche devient de plus en plus sophistiquée et intensive. Les études simples limitées dans le temps sont remplacées par des plateformes de recherche, soient des registres et biobanques liés à des dossiers administratifs et cliniques. Cette intensification s'accompagne de préoccupations relatives à la protection de la vie privée et de la confidentialité. Les comités d'éthique de recherche (CÉR) ont un rôle prééminent dans la surveillance par rapport à ces questions. Dans cet exposé, l'auteur fait une brève revue des cadres conceptuels en gestion de la vie privée et de la confidentialité, identifie les approches clés en contrôle de la divulgation d'informations, et examine un exemple d'application à un CÉR impliquant l'analyse de dossiers médicaux, en soulignant les questions liées à la confidentialité prises en considérations par les CÉR.

**[MS-157]****Data, Privacy and the Law****Données, vie privée et loi**

Christian WHALEN, *Office of the Ombudsman, Government of New Brunswick*

Technology's rapid change raises new risks and ethical challenges for researchers, statisticians and anyone involved in handling large volumes of personal information, whether de-identified or not. The law is having an even harder time keeping pace with the changes in technology but it is also evolving quickly. This presentation will focus on the relationships between law and ethics, technology and market forces as drivers of change in privacy and security matters and will offer a brief update on Canadian privacy laws and on recent case law and recommendations from courts and privacy commissioners in Canada.

Le rapide progrès technologique s'accompagne de nouveaux risques et défis éthiques pour les chercheurs, statisticiens et autres professionnels traitant de grands volumes d'informations personnelles, qu'elles soient dé-identifiées ou non. Bien que la loi connaisse aussi une rapide évolution, elle a de la difficulté à suivre le rythme de la technologie. Cette présentation sera centrée sur les relations entre loi, éthique, technologie et tendances du marché comme influences sur les questions de confidentialité et de sécurité et offrira un bref retour sur les lois canadiennes sur la confidentialité et sur les recommandations des cours et des commissaires à la protection de la vie privée au Canada.

**[MS-158]****Statistics Canada's Policies and Practices Related to Privacy, Security and Confidentiality****Politiques et pratiques de Statistique Canada liées à la vie privée, à la sécurité et à la confidentialité**

Philip GILES, *Statistics Canada*

As the national statistical agency and as a federal government department, Statistics Canada is subject to legislations and policies that differ from that under which most organizations engaged in data collection and statistical research work. The presentation will describe these governing mechanisms in the context of privacy, security and confidentiality. As well, it will present the procedures and practices that Statistics Canada has implemented to respect these requirements as it meets its operational commitments.

En tant qu'agence statistique nationale et département du gouvernement fédéral, Statistique Canada est sujette à des lois et politiques qui diffèrent de celles s'appliquant à la majorité des autres organisations effectuant de la collecte de données et de la recherche statistique. Cette présentation décrira ces mécanismes dans le contexte de la vie privée, de la sécurité et de la confidentialité. Elle présentera également les procédures et pratiques mises en oeuvre par Statistique Canada afin de respecter ces exigences et de remplir ses engagements opérationnels.

**Session 09F****Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55****VAN 1203****Binary or Discrete Data****Données binaires ou discrètes****[MS-159]****Modelling of Negatively Correlated Multivariate Discrete Data****Modélisation de données discrètes multidimensionnelles corrélées négativement**

Blache Paul AKPOUE & Jean-François ANGERS, *Université de Montréal*

In many fields, especially in environmental sciences, multivariate discrete data are very com-

L'observation des données discrètes multidimensionnelles est très fréquente en sciences environnementales. Cepen-

mon. However, they are not used often because of the complex form of their probability distribution. Kano & Kawamura (1991) proposed some recurrence properties on the probability distribution function of the multivariate Poisson but it allows only positive correlations. We generalize the multivariate Poisson model by introducing Poisson and Skellam distributions as marginal distributions to allow the possibility of positively and negatively correlated discrete data. Finally, some recurrence relations to easily evaluate this probability distribution and other characteristics are presented.

dant, leurs utilisations demeurent assez restreintes à cause de la forme complexe de leurs densités. Kano & Kawamura (1991) ont proposé des propriétés de récurrence sur la densité de la loi multidimensionnelle de Poisson, mais elle ne permet que des corrélations positives. Nous généralisons le modèle multidimensionnel de Poisson en introduisant les lois de Poisson et de Skellam comme distributions marginales pour permettre la possibilité de corrélations positives et négatives entre des données discrètes. Enfin, quelques propriétés de récurrence pour le calcul de sa densité et autres caractéristiques sont présentées.

### [MS-160]

#### **Estimating Functions for Correlated Binary Data Analysis with Misclassified Responses**

#### **Fonctions estimatrices pour l'analyse de données binaires corrélées avec mauvaise classification de la réponse**

Zhijian CHEN, Grace YI & Changbao WU, *University of Waterloo*

Measurement error or misclassification has been a long standing concern in medical research and survey sampling. Much research has been directed to problems concerning error-prone covariates. There is relatively little work on measurement error or misclassification in the response variable. In this talk we focus the discussion on misclassification in clustered or longitudinal binary data. We propose a marginal analysis method to handle binary response which is subject to misclassification. Numerical studies will be presented to assess the performance of the proposed method.

Les erreurs de mesures et de mauvaise classification sont un problème depuis longtemps en recherche médicale et en échantillonnage. Beaucoup de travaux portent sur des problèmes liés à des erreurs dans les covariables ; peu ont été effectués au sujet d'erreurs de mesure ou de classification de la variable réponse. Dans cet exposé, la discussion est centrée sur la mauvaise classification en présence de données binaires en grappes ou longitudinales. Nous proposons une méthode d'analyse marginale afin de traiter la réponse binaire sujette à la mauvaise classification. Des études numériques seront présentées afin d'évaluer la performance de la méthode proposée.

### [MS-161]

#### **Estimating Time-varying Intensity Matrices in Continuous Time Multi-state Markov Models Using Local Likelihood Methods**

#### **Estimation des matrices d'intensité dépendantes du temps dans des modèles de Markov à multiétats à temps continu par méthodes de vraisemblance locale**

Gerald LEBOVIC, *University of Toronto*, Patrick BROWN, *Cancer Care Ontario*, George TOMLINSON, *University Health Network*, James STAFFORD, *University of Toronto*

Multi-State-Markov (MSM) models can be used to characterize the behaviour of categorical outcomes measured repeatedly over time. Kalbfleisch & Lawless (1985) and Gentleman et al. (1994) examine the MSM model under the assumption of time-homogeneous transition intensities. When intensities vary over time, current methods are less than ideal. We propose a local likelihood method (Tibshirani & Hastie, 1987; Loader, 1996) to esti-

Les modèles de Markov à multiétats (MMÉ) peuvent être utilisés afin de caractériser le comportement de réponses catégorielles mesurées à plusieurs moments. Kalbfleisch & Lawless (1985) et Gentleman et coll. (1994) examinent le modèle MMÉ sous l'hypothèse d'intensités de transition homogène par rapport au temps. Lorsque les intensités varient avec le temps, les méthodes actuelles ne sont pas idéales. Nous proposons une méthode à vraisemblance locale (Tibshirani & Hastie, 1987 ; Loader, 1996) afin d'estimer les

mate the transition intensities as continuous functions of time. We will discuss methods for obtaining standard errors and selecting the smoothing bandwidth and present a simulation, as well as an application to an HIV study.

intensités de transition en tant que fonctions continues du temps. Nous allons discuter de méthodes afin d'obtenir les erreurs types et de choisir la largeur de la fenêtre de lissage, et présenter une simulation et une application à une étude portant sur le VIH.

**[MS-162]**

**Tests for Proportions in the Clustered Binary Data, with Applications to Toxicological Data**

**Test de proportions pour données binaires en grappes avec application à des données de toxicologie**

Krishna SAHA, *Central Connecticut State University*

The paper develops the test procedures for testing the equality of the proportions in the analysis of clustered binary data when the dispersions across the treatment groups are unequal. We propose a simple test procedure based on a sandwich estimator of the variance of the proportion estimators using the generalized estimating equations approach. We also extend the existing test procedures of testing this hypothesis to the case of unequal dispersions. The performance of these test procedures are assessed through Monte Carlo simulation studies, in terms of size and power. The test procedures are applied to a set of real-life data.

Cet article développe des méthodes de test pour tester l'égalité de proportions en analyse de données binaires en grappes lorsque les dispersions sont inégales d'un groupe de traitement à l'autre. Nous proposons une méthode de test simple basée sur un estimateur de la variance des estimateurs des proportions de type sandwich basé sur l'approche des équations estimatrices généralisées. Nous généralisons aussi les procédures de test existantes pour cette hypothèse au cas de dispersions inégales. La performance de ces méthodes de test est évaluée, en termes de taille et de puissance, au moyen d'études de simulation. Les méthodes de test sont appliquées à un jeu de données réel.

**[MS-163]**

**Estimation of Regression Effects in Longitudinal Binary Data**

**Estimation des effets de régression pour des données binaires longitudinales**

Xuemao ZHANG & Sudhir PAUL, *University of Windsor*

This paper deals with the estimation of the regression parameters for longitudinal binary data. We consider the classical method of Gaussian estimation procedure of Whittle (1961) using a general auto-correlation matrix of the repeated observations. A Newton–Raphson algorithm is derived for estimating the regression parameters for known correlation parameters which are then estimated using moment method. A two-step iterative procedure is suggested for the joint estimation of the regression parameters and the correlation parameters. Asymptotic variances of the regression parameter estimates are also obtained. Bias and efficiency of the estimates are compared by simulations with Generalized Estimating Equations methods.

Cet article traite de l'estimation des paramètres de régression pour des données longitudinales binaires. Nous considérons la méthode classique d'estimation gaussienne de Whittle (1961) qui utilise une matrice d'autocorrélation générale des observations répétées. Un algorithme de type Newton–Raphson est obtenu pour estimer les paramètres de régression lorsque les paramètres de corrélation sont connus ; ceux-ci sont alors estimés par la méthode des moments. Une procédure itérative en deux étapes est suggérée pour l'estimation conjointe des paramètres de régression et de corrélation. Les variances asymptotiques des estimateurs des paramètres de régression sont également obtenues. Des simulations permettent de comparer le biais et l'efficacité des estimateurs à ceux obtenus par des méthodes d'équations d'estimation généralisées.

**[MS-164]****Markov Chain Monte Carlo for Markov Chain Macro Data  
Chaînes de Markov Monte-Carlo pour macro données markoviennes**David STEPHENS, *McGill University*, Martin CROWDER, *Imperial College London*

Markov chain macro data arise when data arising from Markov chains, such as the disease or economic status of individuals followed longitudinally in discrete time, are aggregated to give total counts, but where the individual trajectories are not observed directly. Previously such data have been analyzed using estimating equation approaches. We present new Bayesian approach that uses Markov chain Monte Carlo to perform likelihood-based inference.

Les macro données de chaînes de Markov surviennent lorsque des données provenant de chaînes de Markov, tel le statut économique d'individus suivis longitudinalement en temps discret, sont agrégées afin d'obtenir des totaux et que les trajectoires individuelles ne sont pas observées directement. Auparavant, de telles données étaient analysées par des équations d'estimation. Nous présentons une nouvelle approche bayésienne utilisant des chaînes de Markov Monte-Carlo afin de faire de l'inférence à base de vraisemblance.

**Session 09G****Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55****VAN 1211****Risk Theory and Probability Theory in Finance  
Théorie du risque et probabilité en finance****[MS-165]****Model Mortality Index with Markovian Drift Under the Lee-Carter Framework  
Modélisation de l'indice de mortalité avec dérive markovienne dans le cadre du modèle de Lee-Carter**Yingying GAN & Xiaoming LIU, *University of Western Ontario*

We consider a continuous version of the Lee-Carter model, where the drift term of mortality index follows a continuous time finite state Markov chain. This could capture the effect of medical advancements and health care regulations on the trend of mortality improvements. We study the uncertainty implication of the proposed model to survival probabilities and life expectancies. This model can be useful for the evaluation of contingent claims with mortality risk.

Nous considérons une version continue du modèle de Lee-Carter, où le terme de dérive de l'indice de mortalité suit une chaîne de Markov à temps continu et à nombre fini d'états. Ceci pourrait permettre de mesurer l'effet du progrès médical et de la réglementation régissant les soins de santé sur l'amélioration de la mortalité. Nous étudions les conséquences du modèle proposé sur l'incertitude relative aux probabilités de survie et aux espérances de vie. Ce modèle peut être utile pour évaluer des réclamations éventuelles liées au risque de mortalité.

**[MS-166]****The Distribution of the Compound Sums and Risk Measures  
Distribution de sommes composées et mesures de risque**Ya Fang WANG, *Concordia University*, Ghislain LÉVEILLÉ, *Université Laval*, José GARRIDO, *Concordia University*

Léveillé, Garrido & Wang (2009) derived the moment generating function (mgf) for discounted compound Poisson sums and the mgf for Erlang(2) inter-arrival times. In this talk, I show how to obtain the distribution of discounted compound PH-renewal sums by inverting the Laplace transform.

Léveillé, Garrido & Wang (2009) ont obtenu la fonction génératrice des moments (fgm) pour les sommes composées de Poisson escomptées et la fgm de temps inter-arrivées d'Erlang(2). Dans cet exposé, je montre comment obtenir la distribution de sommes composées escomptées de renouvellements-PH par inversion de la transformée de La-

A differential system is studied for the mgf of discounted compound PH-renewal sums. Series methods and transform differential equations will be discussed to approximate the solutions of differential equations, if the mean of inter-arrival times is very small. Applications such as stop-loss premiums, VaR and CTE are also investigated.

place. Un système différentiel est étudié pour la fgm de sommes composées escomptées de renouvellements-PH. On discutera de méthodes de séries et de transformées d'équations différentielles afin d'approximer les solutions d'équations différentielles si la moyenne des temps inter-arrivées est très petite. Nous étudions aussi des applications tels les primes en excédent de pertes, la VaR et CTE.

#### [MS-167]

##### **The Minimal Entropy Martingale Measure (MEMM) for a Markov-modulated Lévy Model**

##### **La mesure d'entropie martingale minimale (MEMM) dans un modèle de Lévy à modulations markoviennes**

Ben Salah ZIED & Romuald Hervé MOMEYA, *Université de Montréal*

In this paper we determine the minimal entropy martingale measure (MEMM) for a Markov-modulated Lévy model. This generalizes previous works which have treated either the geometric Lévy case (Fujiwara & Miyahara, 2003) or the exponential-additive case (Fujiwara, 2009). Many authors have studied this equivalent martingale measure in different context and provided a solution based on the MEMM for the problem of pricing contingent claims in a Lévy model. We extend these later works to a general Markov-modulated Lévy process.

Dans cette communication, nous déterminons la mesure d'entropie martingale minimale (MEMM) dans un modèle de Lévy à modulations markoviennes. Ceci généralise des travaux antérieurs portant sur le cas de Lévy géométrique (Fujiwara & Miyahara, 2003) ou le cas exponentiel additif (Fujiwara, 2009). Plusieurs auteurs ont étudié la mesure martingale dans différents contextes et ont donné une solution basée sur la MEMM pour le problème de la fixation de prix de réclamations éventuelles dans un modèle de Lévy. Nous généralisons ces derniers travaux à un processus de Lévy général à modulations markoviennes.

#### [MS-168]

##### **Derivatives of the European Carbon Market**

##### **Les produits dérivés du marché du carbone européen**

Frédéric GODIN & Manuel MORALES, *Université de Montréal*

In order to limit the amount of Greenhouse Gases emitted on its territory, the European Union initiated a cap and trade on CO<sub>2</sub> emissions known as the EU ETS. Some CO<sub>2</sub> emission certificates of the EU ETS (EUA, CER) and some derivatives related to them are openly traded among several European exchanges. The purpose of this speech is to compare different models representing the price of CO<sub>2</sub> emission certificates and also to price derivatives of this market. Some models have already been proposed in the literature (Daskalakis, 2008; Seifert, 2008), but no consensus has been reached so far.

Afin de limiter les émissions de gaz à effet de serre sur son territoire, l'Union européenne a instauré un marché d'échange d'émissions de CO<sub>2</sub> : l'EU ETS. Certains certificats d'émission de CO<sub>2</sub> sur le EU ETS (EUA, CER) ainsi que certains produits dérivés sur ceux-ci s'échangent au sein de différentes bourses européennes. Le but de la conférence est de comparer différents modèles pour représenter les prix des certificats d'émission et tarifier certains produits dérivés de ce marché. Certains modèles ont déjà été proposés dans la littérature (Daskalakis, 2008 ; Seifert, 2008), mais aucun consensus n'est pour l'instant établi.

[MS-169]

**On the Price of Risk of the Underlying Markov Chain in an Exponential Regime-Switching Lévy model**  
**La prise en compte du risque sous-jacent à une chaîne de Markov dans un modèle de Lévy avec changements de régime**

Romuald Hervé MOMEYA & Manuel MORALES, *Université de Montréal*

The classical Black–Scholes model fails to reflect the stochastic variability in the market parameters. Therefore, emerging interests in the literature have focused on the so-called Regime-switching (RS) model which better fit random market environment. In the context of derivatives pricing these models lead to incomplete markets, inducing different pricing rules. A fair amount of literature make the assumption that the risk associated with the Markov chain is not priced. Recently, Siu & Yang (2009) proposed a pricing kernel that takes into account all risk components of a RS–Black–Scholes model. In this paper, we extend their setting into a more general Lévy RS model enabling the assessment of jumps on the price of risks. We specialize this general framework to RS Jump-Diffusion and Variance-Gamma models and carry out a comparative analysis of the resulting option price formulas with existing literature such as Naik (1993) and Boyle & Draviam (2007).

Le modèle classique de Black-Scholes est inapte à refléter la variabilité stochastique dans les paramètres d'un marché. Par conséquent, un intérêt s'est développé autour des modèles à changement de régime, plus adéquats pour rendre compte des aléas d'un environnement de marché. Dans le cadre de l'évaluation de produits dérivés, ces modèles donnent lieu à des marchés incomplets et la plupart des travaux de la littérature font l'hypothèse que le risque associé à la chaîne de Markov n'est pas quantifié. Récemment, Siu & Yang (2009) ont proposé un modèle qui tient compte de toutes les composantes de risque d'un modèle de Black-Scholes à changement de régime. Dans cet article, nous étendons ce cadre à un modèle plus général de Lévy avec changement de régime, ce qui permet l'évaluation de sauts dans le prix de risques. On s'attarde sur les cas particuliers des modèles à changement de régime de types *Jump-Diffusion* et *Variance-Gamma*, et effectuons une analyse comparative des formules de prix d'options résultantes avec celles issues des travaux de Naik (1993) et de Boyle & Draviam (2007).

[MS-170]

**Risk Measures and Arbitrage Opportunities**  
**Mesures de risque et opportunités d'arbitrage**

Ramin OKHRATI, *Concordia University*, Alejandro BALBÁS, *Universidad Carlos III de Madrid, Spain*, José GARRIDO, *Concordia University*

Measuring the risk of financial derivatives is a major concern in finance. Arbitrage opportunities that arise due to miss-estimation of the underlying risk are important, closely related issues. As an alternative to the classical methods, we introduce a new approach using risk measures. We investigate a type of arbitrage defined through the properties of risk measures. Under a specific risk measure if the risk of a portfolio is negative, then the possible income of the portfolio is considered an arbitrage. A new tool to detect and measure such arbitrage opportunities in bond markets is established. Some numerical examples are provided.

Mesurer le risque de produits dérivés financiers est une question majeure en finance. Les opportunités d'arbitrage survenant à cause d'une mauvaise estimation du risque sous-jacent constituent un problème étroitement lié. Comme alternative aux méthodes classiques, nous présentons une nouvelle approche utilisant des mesures de risque. Nous étudions un type d'arbitrage défini à travers les propriétés de mesures de risque. Sous une mesure de risque donnée, si le risque d'un portefeuille est négatif, alors le revenu possible du portefeuille est considéré comme étant un arbitrage. Un nouvel outil pour détecter et mesurer de telles opportunités d'arbitrage est développé. Quelques exemples numériques sont fournis.

**Session 09H      Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55      VAN 1871****Nonresponse, Imputation, and Bootstrapping  
Non-réponse, imputation et bootstrap**

[MS-171]

**Integrating Quality Indicators in Complex Surveys  
Intégrer des indicateurs de qualité dans des enquêtes complexes**James BRENNAN & Jack LOTHIAN, *Statistics Canada*

The Quarterly Survey of Financial Statements (QFS) is Statistics Canada's principal source of current information concerning the Canadian-private-incorporated business sector. This information is used extensively by governments, banks, and other institutions. A significant portion of the QFS data is imputed from previous responses or administrative data and it is increasingly important to determine the impact of this imputed data on the quality of the QFS estimates. This paper examines various strategies for estimating the combined variance of the QFS sampling and imputation with one tool being Statistics Canada's System for Estimation of Variance due to Non-response and Imputation (SEVANI).

Le Relevé trimestriel des états financiers (EFT) est la principale source actuelle de renseignements de Statistique Canada en ce qui concerne le secteur canadien des entreprises privées constituées en société. Ces renseignements sont utilisés abondamment par les gouvernements, les institutions financières et d'autres établissements institutionnels. Une portion considérable des données du EFT est imputée à partir de réponses ou de données administratives antérieures, et il s'avère de plus en plus important de déterminer la conséquence de ces données imputées sur la qualité des estimations du EFT. Cet article examine différentes stratégies pour estimer la variance combinée de l'échantillonnage du EFT et des stratégies d'imputation au moyen d'outils clés utilisés par le Système d'estimation de la variance due à la non-réponse et à l'imputation (SEVANI) de Statistique Canada.

[MS-172]

**Calibrating Complex Survey Weights to Adjust for Refusals to Record Linkage  
Calibration de poids d'enquête complexes pour traiter le refus à l'appariement des fichiers**François VERRET, *Statistique Canada*

Statistics Canada collects information on the health of Canadians via several survey and administrative sources. Linking these sources offers great analytical potential while reducing the need for collection. Survey respondents' consent is required for record linkage, and refusing to link is a supplementary form of survey non-response.

Weight calibration is an interesting avenue for compensating for this problem. Särndal & Lundström (JOS 2008) have developed a technique that selects calibration constraints to minimize non-response bias. The method is refined to avoid multicollinearity and calibration on a small group of respondents. It is applied to Canadian Community Health Survey data.

Statistique Canada recueille des renseignements sur la santé des Canadiens au moyen de plusieurs enquêtes et sources de données administratives. Le couplage de ces sources engendre un grand pouvoir analytique tout en réduisant les besoins de collecte. Le consentement des répondants aux enquêtes est requis pour procéder au couplage d'enregistrements, sans quoi leur refus entraîne une forme de non-réponse supplémentaire.

Le calage aux marges des poids est une approche intéressante pour remédier à ce problème. Särndal & Lundström (JOS 2008) ont développé une technique de sélection des contraintes de calage minimisant le biais de non-réponse. La méthode est raffinée pour éviter la multicollinéarité et le calage sur un trop petit ensemble de répondants, et elle est appliquée aux données de l'Enquête sur la santé dans les collectivités canadiennes.

**[MS-173]****A Comparison of Reweighting and Imputation Approaches for Missing Responses in Longitudinal Surveys  
Une comparaison des approches de repondération et d'imputation pour réponses manquantes dans des enquêtes longitudinales**

Ivan CARRILLO-GARCIA, *Statistics Canada*, Milorad KOVACEVIC, *United Nations Development Programme*, Changbao WU, *University of Waterloo*

In this talk we compare the Generalized Estimating Equations (GEE), with missing responses imputed by hot-deck, with the reweighting GEE method proposed by Robins et al. (1995, JASA). In contrast to their original paper, we concentrate on complex survey situations. Using data from the National Longitudinal Survey of Children and Youth, we illustrate the differences between the two approaches and present the results of a comprehensive simulation study.

Dans cet exposé, nous comparons les équations d'estimation généralisées (ÉEG) avec réponses manquantes imputées par *hot deck* à la méthode de re-pondération des ÉEG proposée par Robins et coll. (1995, JASA). Contrairement à leur article, nous nous concentrons sur des situations d'enquêtes complexes. À l'aide de données de l'Enquête longitudinale nationale sur les enfants et les jeunes, nous illustrons les différences entre les deux approches et présentons les résultats d'une étude de simulation exhaustive.

**[MS-174]****Nonparametric Estimation of Response Probabilities in Survey Sampling: An Empirical Investigation  
Estimation non paramétrique des probabilités de réponse dans les enquêtes : une étude empirique**

Audrey BÉLIVEAU & David HAZIZA, *Université de Montréal*, Myriam MAUMY, *Université de Strasbourg*

In survey sampling, a common way of reducing the nonresponse bias is to adjust the design weight of the respondents. A popular weight adjustment procedure consists of dividing the design weight of respondents by their estimated response probability. These probabilities can be estimated using parametric methods (e.g., logistic regression model) or nonparametric methods. Nonparametric methods include kernel smoothing methods and local polynomial regression (Da Silva & Opsomer, 2006/2009) or weight adjustment procedures applied within weighting classes (which is commonly used in practice). In this talk, results of a Monte Carlo simulation study comparing these nonparametric methods will be presented.

Dans les enquêtes, il est coutume de réduire le biais de non-réponse en ajustant le poids de sondage des répondants. Une procédure d'ajustement populaire consiste à diviser le poids de sondage des répondants par leur probabilité de réponse estimée. L'estimation des probabilités de réponse peut être réalisée au moyen de méthodes paramétriques (p. ex., modèle de régression logistique) ou de méthodes non paramétriques. Parmi les méthodes non paramétriques figurent la méthode du noyau et la régression par polynômes locaux (Da Silva & Opsomer, 2006/2009) ainsi que les méthodes de repondération à l'intérieur de classes (qui sont largement utilisées en pratique). Lors de cet exposé, les résultats d'une étude par simulation de Monte-Carlo comparant ces méthodes non paramétriques seront présentés.

**[MS-175]****Mean Bootstrap Variance Estimation for Survey Data  
Estimation de la variance du bootstrap moyen pour les données d'enquêtes**

Lenka MACH, Jean-François BEAUMONT, Keven BOSA & Yves LAFORTUNE, *Statistics Canada*

Mean bootstrap, proposed by Yung (1997) for variance estimation when using Public Use Microdata Files, has been implemented for several Statistics Canada surveys. This paper will examine its properties and discuss when this variance estimator is appropriate and when it should not be used.

Le bootstrap moyen, proposé par Yung (1997) pour estimer la variance lorsque des fichiers de micro données à grande diffusion sont utilisés, a été mis en oeuvre dans plusieurs enquêtes de Statistique Canada. Cet article analysera les propriétés de cet estimateur, et une discussion aura lieu afin de déterminer quand ce dernier devrait être utilisé ou non.

**[MS-176]****Bootstrap Variance Estimation in the Presence of Imputed Data****L'estimateur bootstrap de la variance en présence de données imputées**Zeinab MASHREGHI, David HAZIZA & Christian LÉGER, *Université de Montréal*

In the context of imputation, Shao & Sitter (1996) introduced a bootstrap method where the non-respondents in the bootstrap sample are reimputed with the same method, leading to a valid variance estimator if the sampling fraction  $f$  is small. We argue that this method can be justified by the reverse framework (Shao & Steel, 1999) which suggests another valid estimator that can be obtained using complete data bootstrap methods without requiring reimputation. Finally, we obtain a third valid estimator, even for large  $f$ , by modifying the weights to account for non-response and resampling the data and response status independently.

Dans le contexte de l'imputation, Shao & Sitter (1996) ont proposé une méthode de bootstrap dans laquelle les non-répondants de l'échantillon bootstrap sont réimputés avec la même méthode que pour les données originales, conduisant à un estimateur de la variance valide lorsque la fraction de sondage  $f$  est négligeable. Nous affirmons que cela peut être justifié par l'approche inversée (Shao & Steel, 1999) qui nous amène à un autre estimateur valide obtenu par les méthodes de bootstrap pour les données complètes, sans avoir besoin de réimputation. Enfin, en modifiant les poids pour tenir compte de la non-réponse et en rééchantillonnant les données et les indicateurs de réponse indépendamment, nous obtenons un troisième estimateur de la variance valide même pour les  $f$  non négligeables.

**Session 09I****Mardi 25 mai • Tuesday, May 25, 14:25 – 15:40****VAN 1883****Quality Control and Design of Experiments****Contrôle de qualité et planification d'expériences****[MS-177]****Current Practices in Reliability Improvement and Assurance Tests in Korea****Pratiques actuelles pour l'amélioration de la fiabilité et les tests d'assurance en Corée**Jaiwook BAIK, *Korea National Open University*

During the past ten years there has been a lot of emphasis on reliability in manufacturing companies in Korea. As a result reliability assurance tests have been practiced. In this talk, a review is carried out on reliability assurance tests practiced in Korea and some enhancements will be proposed.

Au cours des dix dernières années, les manufacturiers coréens ont accordés beaucoup d'importance à la fiabilité. Conséquemment, des tests d'assurance de la fiabilité ont été développés. Dans cet exposé, nous faisons la revue des tests d'assurance de la qualité utilisés en Corée et proposons quelques améliorations.

**[MS-178]****Empirical Likelihood Based Information Criteria for Change Point Problems****Critères d'information basés sur la vraisemblance empirique pour problèmes de point de rupture**Chithran VASUDEVAN & ASOKAN VARIYATH, *Memorial University of Newfoundland*

Identification of change point in an industrial process monitoring is of great research interest. The existing methods rely on the parametric distribution of the quality characteristic and a misspecified model may leads to wrong conclusions. We are proposing an empirical likelihood based informa-

L'identification de points de rupture en surveillance de processus industriels présente un intérêt de recherche important. Les méthodes existantes s'appuient sur la distribution paramétrique de la caractéristique de qualité et un modèle mal spécifié peut mener à des conclusions erronées. Nous proposons d'identifier le point de rupture à l'aide d'un

tion criteria for identifying the change point. The main advantage of the proposed method is that we don't need to specify any distribution for the quality characteristic. Our simulation results indicate that our proposed method is equally good with the existing methods when the distribution is known and it outperforms the existing methods when the distribution is unknown or misspecified.

[MS-179]

### **Optimal Designs for Fourier Regression Models Plans optimaux pour modèles de régression de Fourier**

Xiaoli SHANG & Xiaojian XU, *Brock University*

We discuss the optimal designs for Fourier regression models. First, we have constructed D-optimal designs for estimating pairs of coefficients. Second, we investigate classical Q-optimal designs and minimax robust designs under A-, D-, and Q-optimality. We have proved the equivalency of Q-optimal and A-optimal designs. It is shown that A- and D-optimal designs are not sensitive to the misspecified order of model assumed. Last, D-optimal designs for reduced Fourier models are constructed. Both the comparison with other competitors and our simulation results show that our designs incorporating the reduced models are much more efficient.

[MS-180]

### **Optimal Implementations for Robust Designs with Densities Implémentations optimales pour plans robustes avec densités**

Wai Kong YUEN & Xiaojian XU, *Brock University*

In optimal robust design theory, many designs constructed in the literature have densities. In practice, an exact design should tell the experimenter what the support points are and how many subjects should be allocated to each of these points. We consider a practical situation in which the number of support points allowed is constrained. We propose a new implementation scheme that minimizes the Kolmogorov–Smirnov distance between an exact design and the optimal design having a density. Our simulation studies confirm the efficiency gain from our proposed scheme. Both single and multiple dimensional examples are presented.

critère d'information basé sur la vraisemblance empirique. L'avantage principal de la méthode proposée est qu'il n'est pas nécessaire de spécifier la distribution de la caractéristique de qualité. Nos résultats de simulation indiquent que notre méthode est aussi bonne que les méthodes existantes lorsque la distribution est connue et les surpasse lorsque la distribution est inconnue ou mal spécifiée.

Nous discutons de plans optimaux pour modèles de régression de Fourier. Premièrement, nous avons construit des plans D-optimaux pour l'estimation de paires de coefficients. Ensuite, nous étudions les plans Q-optimaux classiques et les plans minimax robustes sous A-, D- et Q-optimauté. Nous avons prouvé l'équivalence de plan Q-optimaux et A-optimaux. Nous montrons que les plans A- et D-optimaux ne sont pas sensibles à une mauvaise spécification de l'ordre du modèle. Finalement, nous construisons les plans D-optimaux pour modèles de Fourier réduits. La comparaison avec les autres compétiteurs et nos résultats de simulation montrent que nos plans incorporant les modèles réduits sont beaucoup plus efficaces.

En théorie de la planification optimale robuste, plusieurs plans construits dans la littérature ont des densités. En pratique, un plan exact devrait donner à l'expérimentateur les points de soutien et le nombre de sujets à allouer à chaque point. Nous considérons une situation pratique avec une contrainte sur le nombre de points de soutien. Nous proposons un nouveau schéma d'implémentation minimisant la distance de Kolmogorov–Smirnov entre un plan exact et le plan optimal ayant une densité. Nos études de simulation confirment que le schéma proposé permet un gain en efficacité. Des exemples uni- et multidimensionnels sont présentés.

[MS-181]

**The Importance of Experimental Design**  
**L'importance de la planification d'expériences**

Philip BERTRAND, *Consultant Statistician*

Experimental design enables clear calculation of effects of variables in a multivariate study to ascertain the effects of many factors on a variable of interest. This, however, is not the only benefit of good design. Not realized by experimenters in other disciplines is the time and money saved by well designed experiments. This aspect is explored in an experiment to explore the effects of 64 factors on a variable of interest in a particular sequence of experiments.

La planification d'expériences permet un calcul clair de l'effet de variables dans une étude multivariée visant à quantifier les effets de plusieurs facteurs sur une variable d'intérêt. Ceci n'est toutefois pas le seul avantage d'un bon plan. Les expérimentateurs dans d'autres disciplines ne réalisent pas toujours combien de temps et d'argent peuvent être économisés par des expériences bien planifiées. Cet aspect est exploré avec une expérience visant à étudier les effets de 64 facteurs sur une variable d'intérêt à l'aide d'une suite particulière d'expériences.

**Session 09J****Mardi 25 mai • Tuesday, May 25, 14:25 – 15:55****VAN Hall**

Poster Session 2

Séance d'affichage 2

[MS-182]

**Goodness-of-fit Measures for Individual-level Models for Infectious Disease**

**Mesures de qualité de l'ajustement pour modèles de maladies infectieuses au niveau de l'individu**

Abbie GARDNER, Rob DEARDON & Gerarda DARLINGTON, *University of Guelph*

In simple models there are a variety of ways to assess goodness-of-fit. However, in spatio-temporal epidemic models, less research has been done on how to ascertain goodness-of-fit. Often such models are fitted within a Bayesian statistical framework to allow for data uncertainty to be incorporated. The posterior predictive distribution is one method for assessing goodness-of-fit. That is, a distribution for some "test statistic" is found through simulation from the posterior distribution, and compared with that observed for the data. Here we examine different test statistics and ascertain how well they detect model misspecification via simulation study.

Il existe une variété de façon d'évaluer l'adéquation de modèles simples. Cependant, peu de recherches ont été faites au sujet de l'évaluation de l'adéquation de modèles épidémiques spatio-temporels. De tels modèles sont souvent ajustés dans un cadre statistique bayésien afin de permettre d'incorporer l'incertitude liée aux données. La distribution prédictive a posteriori peut être utilisée pour évaluer l'adéquation du modèle. La distribution d'une certaine « statistique de test » est obtenue par simulation à partir de la distribution *a posteriori* et est ensuite comparée à la valeur observée de la statistique. Nous examinons ici différentes statistiques de test et étudions leur capacité à détecter une mauvaise spécification de modèles par une étude de simulation.

**[MS-183]****Relationship Between Lung Function and Hypertension among Rural Canadians using Fractional Polynomials****Relation entre fonction pulmonaire et hypertension en milieu rural au moyen de polynômes fractionnels**

Chandima KARUNANAYAKE, Donna RENNIE & Punam PAHWA, *University of Saskatchewan*, Yue CHEN, *University of Ottawa*, James DOSMAN, *University of Saskatchewan*

The relationship between lung function and hypertension in rural populations is not well described. We analyzed cross-sectional survey data collected in 2003 from 1834 rural adults (18–79 years) living in Humboldt, Saskatchewan. Fractional polynomials methods were used with logistic regression to define the correct functional forms of the continuous lung function predictors [e.g. forced expiratory volume in first second (FEV1) and forced vital capacity (FVC)]. We observed an increased risk of hypertension with decreasing FVC after adjusting for potential risk factors (age, body mass index, parental history of blood pressure and smoking) in females but not in males.

La relation entre la fonction pulmonaire et l'hypertension dans les populations rurales n'est pas bien décrite. Nous analysons des données d'enquêtes transversales recueillies en 2003 chez 1834 adultes (18 à 79 ans) habitant Humboldt, Saskatchewan. Des méthodes à polynômes fractionnels furent utilisées conjointement à la régression logistique afin de définir les formes fonctionnelles correctes des prédicteurs continus de la fonction pulmonaire [c.-à-d. volume expiratoire maximal par seconde (FEV1) et capacité vitale forcée (FVC)]. Nous observons un risque accru d'hypertension lorsque la FVC décroît après ajustement pour facteur de risques potentiels (âge, indice de masse corporelle, antécédents parentaux de pression sanguine et tabagisme) chez les femmes seulement.

**[MS-184]****Heterogeneity Control of Capture-recapture****Contrôle de l'hétérogénéité en capture-recapture**

Lee CHEL HEE & Lisa LIX, *University of Saskatchewan*

The capture-recapture (CR) method is a popular tool for chronic disease epidemiologic research to estimate the number of missing disease cases in incomplete health databases. Homogeneity of capture probabilities is an assumption unlikely to be satisfied in epidemiologic research. In order to produce robust estimates to violations of this assumption, stratification and multinomial logistic regression methods are generally recommended. We compare these methods for ascertaining cases of osteoporosis using two health administrative datasets from Alberta. A simulation study investigates robustness of these two methods for estimating the number of missing cases when capture probabilities are heterogeneous.

La méthode de capture-recapture est un outil populaire pour la recherche épidémiologique en maladies chroniques afin d'estimer le nombre de cas de maladie manquant dans des bases de données de santé incomplètes. L'homogénéité des probabilités de capture est une hypothèse peu vraisemblablement satisfaite en recherche épidémiologique. Afin de produire des estimations robustes à un non-respect de cette hypothèse, des méthodes de stratification et de régression logistique multinomiales sont habituellement recommandées. Nous comparons ces méthodes en estimation de cas d'ostéoporose à partir de deux jeux de données administratifs de l'Alberta. Une étude de simulation permet d'évaluer la robustesse de ces deux méthodes lors de l'estimation du nombre de cas manquants lorsque les probabilités de capture sont hétérogènes.

**[MS-185]****Modelling the Spatio-temporal Dynamics of Fire Spread****Modéliser la dynamique spatio-temporelle de la propagation d'incendies**Irene VRBIK, Rob DEARDON & Feng ZENY, *University of Guelph*, Braun JOHN, *University of Western Ontario*

Individual-level models (ILMs) are a class of models originally designed to model the spread of infectious disease. They can also be considered for modelling the spatio-temporal dynamics of forest fires. Initially, we consider modelling the combustion dynamics on a piece of wax paper under controlled conditions. The models are fitted in a Bayesian framework using Markov chain Monte Carlo. The first major concern is how to model the combustion pattern itself. The second is the computational difficulty of dealing with the large data set since the data describing the spread of the fire is at a high level of definition.

Les modèles au niveau de l'individu (MNI) sont une classe de modèles initialement conçus pour modéliser la propagation de maladies infectieuses. Ils peuvent aussi servir à modéliser la dynamique spatio-temporelle de feux de forêts. Nous considérons initialement la modélisation de la dynamique de la combustion d'une feuille de papier ciré sous conditions contrôlées. Les modèles sont ajustés dans un contexte bayésien par chaînes de Markov Monte-Carlo. Le premier problème majeur est la modélisation de la structure de combustion. Le second est la difficulté informatique du traitement de grands jeux de données puisque les données décrivant la propagation du feu sont à un haut niveau de définition.

**[MS-186]****Finite Mixtures of Individual-level Models for Infectious Diseases****Mélanges finis de modèles au niveau de l'individu pour maladies infectieuses**Lorna DEETH & Rob DEARDON, *University of Guelph*

A class of models, known as individual-level models (ILMs), has been used to describe the spatiotemporal spread of infectious diseases. Currently, these ILMs do not account for hidden population heterogeneity; instead, a homogeneous population is assumed and common model parameters are used. However, if a latent grouping structure exists, the parameters driving disease spread may be significantly different amongst subgroups. Finite mixture models are a way to circumvent population heterogeneity complexities, by incorporating the group structure into the modelling procedure. A finite mixture of ILMs model, comprised of component ILMs, is proposed, and considered with both real and simulated data.

Une classe de modèles, dits modèles au niveau de l'individu (MNI), a été utilisée pour décrire la propagation spatiotemporelle de maladies infectieuses. Présentement, ces MNI ne tiennent pas compte d'hétérogénéité cachée de la population ; la population est plutôt supposée homogène et des paramètres communs sont utilisés pour les modèles. Cependant, si une structure latente de regroupements existe, les paramètres derrière la propagation de la maladie peuvent être significativement différents d'un sous-groupe à l'autre. Les modèles de mélanges finis représentent une façon de contourner les complexités liées à l'hétérogénéité de la population en incorporant la structure de groupe dans le processus de modélisation. Un mélange fini de MNI, consistant des composants ILMs, est proposé et est appliqué à des données réelles et simulées.

**[MS-187]****Fine Genetic Mapping of Polygenic Traits by the Coalescent Process with Recombination****Cartographe génétique fine de caractères polygéniques par le processus de coalescence avec recombinaison**Marie FOREST & Fabrice LARRIBE, *Université du Québec à Montréal*

The adaptation of a fine mapping method named MapArg, (Larribe et al., 2002, 2008), to the mapping of polygenic traits is proposed here. MapArg uses a continuous time stochastic process and im-

Nous proposons d'adapter une méthode de cartographie génétique fine, MapArg (Larribe et coll., 2002, 2008), à la cartographie de caractère polygénique. MapArg utilise un processus stochastique à temps continu (processus de coales-

portance sampling to construct genealogies of the sample of genetic sequences, and to evaluate the likelihood of the position of a disease mutation. Until now, MapArg assumes that a single gene influences the character and that the disease is recessive and rare. We will now assume that two genes influence the character, and we will show how to estimate the likelihood surface in this context.

cence) et l'échantillonnage pondéré pour construire des généalogies d'un échantillon de séquences de marqueurs (cas et contrôle), et évalue la vraisemblance de la position d'une mutation causant une maladie. Cette méthode suppose que le caractère est influencé par un seul gène, et que la maladie est récessive et rare. Nous supposerons ici que le caractère est influencé par deux gènes, et montrerons comment estimer la vraisemblance dans ce contexte.

**[MS-188]**

**Continuous Model in Disease Mapping with Aggregated Data**  
**Modèle continu en cartographie des maladies avec données agrégées**

Ye LI & Patrick BROWN, *University of Toronto*

Surveillance data are usually aggregated with census regions that vary greatly, being larger in rural areas. Conventional modelling results in stronger spatial dependence in less populated regions which dominate maps. To test the efficacy of two methods, we simulate continuous risk surface over GTA using actual population. The first model is standard BYM model. In the second model, we approximate the continuous risk surface by a piecewise constant Markov random field on fine lattice, estimating counts for each cell using the expectation conditioning on a constant risk surface. In both cases Bayesian inference is used with Integrated Nested Laplace Approximations.

Les données de surveillance d'une même région de recensement sont habituellement agrégées. Or, la taille de ces régions varie grandement – en particulier, les régions rurales sont plus grandes. Une modélisation conventionnelle résulte en une dépendance spatiale plus forte dans les régions moins peuplées dominant les cartes. Nous testons l'efficacité de deux méthodes en simulant une surface de risque continue sur la RMT à partir de la population réelle. Le premier modèle est le modèle de BYM standard. Pour le second modèle, nous approximations la surface de risque continue par un champ aléatoire markovien constant par morceau sur un treillis fin et estimons le total par cellule par conditionnement de l'espérance par rapport à une surface de risque constante. Dans les deux cas, l'inférence bayésienne est utilisée avec approximations de Laplace intégrées et emboîtées.

**[MS-189]**

**Bayesian Model Averaging to Estimate Causal Effects**  
**Moyennage bayésien pour l'estimation d'effets causaux**

Malorie CHABOT-BLANCHET & Geneviève LEFEBVRE, *Université du Québec à Montréal*

To estimate the causal effect of a treatment in observational studies, methods such as inverse probability weighting and doubly robust methods are commonly used. It is however not easy to correctly specify the treatment model and the estimators are particularly sensitive to the choice of the model. We investigate if the use of model averaging over a set of models could improve the estimations of such causal effects. We examine the performance of the approach when the true model is or is not part of the set of models considered.

Pour estimer l'effet causal d'un traitement dans les études d'observation, les méthodes de pondération par probabilités inversées et les méthodes doublement robustes sont couramment utilisées. Il n'est toutefois pas facile de spécifier correctement le modèle de traitement, et les estimateurs sont particulièrement sensibles au choix du modèle. Nous nous intéressons à déterminer si le fait de prendre une moyenne sur plusieurs modèles peut améliorer l'estimation des effets de type causal. Nous examinons la performance de la technique lorsque le vrai modèle fait ou ne fait pas partie des modèles considérés.

**[MS-190]****Wavelet Designs for Nonparametric Regression Models with Autocorrelated Errors****Plans à base d'ondelettes pour modèles de régression non paramétriques à erreurs autocorrélées**Selva SELVARATNAM & Alwell OYET, *Memorial University*

We consider minimax designs for estimation of nonparametric regression models using wavelet approximations of the mean response function. We assume that the error terms are autocorrelated. Since the method of estimation affects the choice of design, we argue that using OLS for estimation may lead to designs that are less efficient than designs constructed based on generalized least squares. We present some examples based on the simulated annealing algorithm.

Nous considérons des plans minimax pour l'estimation de modèles de régression non paramétriques au moyen d'approximations par ondelettes de la fonction réponse moyenne. Nous supposons que les termes d'erreur sont autocorrélés. Puisque la méthode d'estimation influence le choix du plan, nous argumentons que l'utilisation des MCO pour l'estimation peut mener à des plans moins efficaces que si les MCG étaient utilisés. Nous présentons quelques exemples basés sur l'algorithme de recuisson simulée.

**[MS-191]****Efficient Estimation for Patient-specific Rates of Disease Progression Using Nonnormal Linear Mixed Models for Multiple Characteristics****Estimation efficace de taux de progression de maladies spécifiques aux patients au moyen de modèles linéaires mixtes non normaux pour de multiples caractéristiques**Dandan LUO & Peng ZHANG, *University of Alberta*

We extend the random-effects model for multiple characteristics with normal assumption for the random effects to the case of nonnormal random effects. This new class of nonnormal linear mixed models provides efficient estimation of subject-specific disease progression in glaucoma data, in which the distribution of some random effect is negatively skewed. The log gamma distribution is assumed for some of the random effects. MLEs of the parameters for this model are estimated. To validate the adequacy of the log gamma assumption versus the usual normality assumption for the random effects, a lack-of-fit test is proposed.

Nous généralisons le modèle à effets aléatoires pour de multiples caractéristiques à effets aléatoires normaux au cas d'effets aléatoires non normaux. Cette nouvelle classe de modèles linéaires mixtes non normaux estime efficacement la progression de la maladie spécifique aux sujets lorsqu'appliquée à des données de glaucome, où la distribution d'un certain effet aléatoire est négativement asymétrique. Nous supposons que certains des effets aléatoires suivent une distribution log gamma. Les EMVs des paramètres pour ce modèle sont estimés. Un test de manque d'ajustement est proposé afin de confronter l'hypothèse de la distribution log gamma à l'hypothèse habituelle d'une distribution normale des effets aléatoires.

**[MS-192]****Genome Selection for Predicting a Quantitative Trait****Sélection de génome pour la prédiction d'un trait quantitatif**Zeny FENG, Xiaojian YANG & Paul McNICHOLAS, *University of Guelph*

In the problem of selecting genome regions for predicting a quantitative trait, SNPs are treated as genetic attributes. It is often that the available data set for the study contains tens of thousands of SNPs with a moderate sample size. This is a typical high dimensional feature selection problem. The lasso-based variables selection method (Tibshirani,

Dans le problème de la sélection de régions du génome pour prévision d'un trait quantitatif, le SNP sont traités comme des attributs génétiques. Les données disponibles contiennent souvent des dizaines de milliers de SNP pour une taille échantillonnale modeste. Ceci est un problème typique de sélection de caractéristiques en haute dimension. La méthode de sélection de variables basée sur le lasso (Tib-

1996) and the sparse partial least squares (SPLS) regression (Chun & Keles, 2009) are applicable for this problem. In this paper, we will compare the performance of the two methods via simulation studies and real data analysis.

shirani, 1996) et la régression des moindres carrés spatiaux épars (Chun & Keles, 2009) sont applicables à ce problème. Dans cette communication, nous comparons la performance des deux méthodes au moyen d'études de simulation et de l'analyse de données réelles.

#### [MS-193]

### **Implementation of a Cost Quotation Tool for Statistical Consulting in an Academic Medical Setting: Preliminary Results**

### **Implémentation d'un outil d'estimation des coûts de consultation statistique dans un cadre médico-académique : résultats préliminaires**

Rhonda ROSYCHUK, *University of Alberta*

In an academic medical setting, researchers frequently apply to grant competitions and need statistical input. Budgets for statistical work are included and are often based on a consultant's views of the scope of the work and the amount of time necessary to complete the work. We developed an estimation tool to use in a partial cost-recovery statistical consultancy. Our estimation tool is based on data available about the project (e.g., number of variables) and scope of work to be performed (e.g., number of regressions). We provide a preliminary assessment based on some completed and ongoing projects.

Dans un cadre académique médical, les chercheurs remplissent souvent des demandes de subventions et ont besoins de conseils statistiques. Une portion des budgets est allouée au travail statistique et est souvent basée sur l'opinion du consultant au sujet de l'ampleur du travail et du temps nécessaire pour effectuer le travail. Nous avons développé un outil d'estimation pour consultation statistique en récupération partielle des coûts. Notre outil d'estimation est basé sur les données disponibles au sujet du projet (p. ex., nombre de variables) et l'ampleur des travaux à effectuer (p. ex., le nombre de régressions). Nous fournissons une estimation préliminaire basée sur certains projets complétés et en cours.

#### [MS-194]

### **SAS General Parametric Step-wise Regression Macro**

### **Macro SAS de régression générale paramétrique pas-à-pas**

Ryan BROWNE & Judy-Anne CHAPMAN, *NCIC Clinical Trials Group, Queen's University*

Parametric survival analyses have recently been shown relevant for three common cancer sites (breast, lung and colon), as well as for health economics. A SAS framework is needed for FDA reporting, yet SAS only permits specified covariate fits. We introduce a general parametric macro with step-wise forward or backward regression options, and model checking.

L'analyse de survie paramétrique s'est récemment avérée pertinente pour trois types de cancers (sein, poumon et colon), ainsi qu'en économie de la santé. Malheureusement, SAS se permet en ce moment que d'ajuster certaines combinaisons spécifiques de covariables. Un macro SAS facilitera le reportage à la FDA. Nous présentons une macro paramétrique SAS générale avec options de régression antérograde, rétrograde et pas-à-pas, ainsi que pour la validation du modèle.

**[MS-195]****Mixtures of Gaussian and Uniform Distributions for Model-based Clustering****Mélanges de distributions gaussiennes et uniformes pour classification à base de modèles**Matthew SPARLING & Paul McNICHOLAS, *University of Guelph*

Mixtures of multivariate  $t$ -distributions have been propounded as a robust alternative to the multivariate Gaussian distribution for model-based clustering. We propose a mixture of Gaussian and uniform distributions as an alternative to the multivariate  $t$ -distribution for model-based clustering. Details of parameter estimation are given and the two approaches are compared on real and simulated data.

Les mélanges de distributions  $t$  multivariées ont été proposés comme alternative robuste à la distribution gaussienne multivariée pour la classification à base de modèles. Nous proposons un mélange de distributions gaussiennes et uniformes comme alternative à la distribution  $t$  multivariée dans ce contexte. Nous décrivons l'estimation paramétrique et comparons les deux approches à l'aide de données réelles et simulées.

**[MS-196]****Assessing the Fit of Epidemic Models over Time and Space****Évaluer l'adéquation de modèles épidémiques à travers le temps et l'espace**Zhang LIN & Rob DEARDON, *University of Guelph*

Model validation is a topic in infectious disease modelling. Disease spread occurs in a population in which heterogeneous mixing needs to be taken into account. For instance, consider disease spread where spatial dynamics are key or where the disease is propagated through a complex network, such as a disease in farm animals that is spread via a trading network. In such heterogeneous populations it is important to consider how model fit changes over time and space. We consider measures based upon probability scoring to test the accuracy of the model in order to predict observed infection and non-infection events.

La validation de modèles est un sujet en modélisation de maladies infectieuses. Une maladie se propage dans une population à mélange hétérogène devant être prise en considération. Par exemple, considérons la propagation d'une maladie pour laquelle la dynamique spatiale est capitale, ou encore une maladie se propageant par un réseau complexe, comme une maladie des animaux de la ferme se propageant par un réseau d'échanges. Pour de telles populations hétérogènes, il est important de considérer comment l'ajustement du modèle change à travers l'espace et le temps. Nous considérons des mesures basées sur les scores de probabilité afin de tester la précision du modèle afin de prédire les événements d'infection observés et de non-infection.

**[MS-197]****Parameter Estimation for Model-based Clustering of Longitudinal Data****Estimation paramétrique en classification à base de modèles pour données longitudinales**Sanjeena SUBEDI & Paul McNICHOLAS, *University of Guelph*

Model-based clustering techniques for longitudinal data are reviewed. Particular attention is given to a family of Gaussian mixture models, with a Cholesky-decomposed covariance structure. While parameter estimation has previously been conducted within the expectation-maximization framework, we utilize a Bayesian framework. We also deviate from the traditional Bayesian information criterion approach to model-selection. Specifically, variational approximations are used for the estima-

Nous faisons la revue des techniques de classification de données longitudinales à base de modèles. Une attention particulière est portée à une famille de modèles de mélanges gaussiens dont la structure de covariance est traitée par décomposition de Cholesky. Bien que l'estimation paramétrique dans ce cas ait déjà été traitée dans un cadre d'espérance-maximisation, nous travaillons dans un cadre bayésien. Nous utilisons aussi une autre approche à la sélection de modèles que le traditionnel critère d'information bayésien. Plus spécifiquement, des approximations va-

tion of the parameters as well as the number of components in the model, along with the use of deviance information criterion (DIC) for the validation of the model selected by the variational approximation method.

riationnelles sont utilisées afin d'estimer les paramètres et le nombre de composantes du modèle, et le critère d'information de déviance est utilisé afin de valider le modèle choisi.

**[MS-198]**

**Model-based Clustering Using Mixtures of Gamma Distributions  
Classification à base de modèles utilisant des mélanges de distributions gamma**

Brian FRANZAK & Paul McNICHOLAS, *University of Guelph*

Model-based clustering using mixtures of gamma distributions, including constrained models, is described and illustrated on real data. Maximum likelihood estimates for the parameters are calculated using an expectation-maximization algorithm. Aitken's acceleration is used to determine convergence and the Bayesian information criterion is used to select the number of mixture components. These models are used to analyze real data on the chemical and physical properties of Italian wines, and their performance is compared to other model-based clustering methods.

La classification à base de modèles utilisant des mélanges de distributions gamma, incluant des modèles avec contraintes, est décrite et illustrée avec des données réelles. Les estimateurs du maximum de vraisemblance pour les paramètres sont calculés grâce à un algorithme d'espérance-maximisation. L'accélération d'Aitken est utilisée afin de déterminer la convergence et le critère d'information bayésien est utilisé afin de choisir le nombre de composantes du mélange. Ces modèles sont utilisés afin d'analyser des données réelles de propriétés chimiques et physiques de vins italiens, et leur performance est comparée à d'autres méthodes de classification à base de modèles.

**[MS-199]**

**Periodic Autoregressive Moving Average (PARMA) Models: Analyses of Ozone Levels in Nova Scotia  
Modèles autorégressifs périodiques à moyenne mobile (MAPMM) : analyse de niveaux d'ozone en Nouvelle-Écosse**

Swarna WEERASINGHE, *Dalhousie University*, Bruce SMITH, *Dalhousie university*, Sam STEWART, *Dalhousie University*

In view of global warming, accurate estimation of extreme levels from short records of data is essential. Many environmental time series behaves as a non-stationary stochastic process with seasonally varying second order and first order properties. In this paper we will demonstrate methods of modelling such non-stationarity in the presence and absence of measurement error. Trends and patterns of the ozone data were analyzed to accurately explore the time varying trends and patterns above threshold human hazardous levels. We found a linearly increasing component, a quadratic declining annual trend and periodicity in the variance and autocorrelation.

Dans le contexte du réchauffement climatique, une estimation précise de niveaux extrêmes à partir de courtes séries de données est essentielle. Plusieurs séries chronologiques environnementales se comportent comme un processus stochastique non stationnaire avec propriétés de second ordre et de premier ordre à variance saisonnière. Dans cet article, nous présentons des méthodes de modélisation d'une telle non-stationnarité en présence et en absence d'erreurs de mesure. Des données d'ozone furent analysées afin de détecter précisément si les tendances et structures variant avec le temps dépassaient les seuils nocifs pour l'humain. Nous avons trouvé une composante augmentant linéairement, une tendance diminuant de façon quadratique annuellement et une périodicité dans la variance et l'autocorrélation.

**[MS-200]****A Bayesian Evaluation of Breast Cancer Screening****Une évaluation bayésienne du dépistage du cancer du sein**

Hedy JIANG, *Cancer Care Ontario & McMaster University*, Patrick BROWN, *Cancer Care Ontario & University of Toronto*

Estimation of sensitivity and specificity for diagnostic or screening exams usually requires independent confirmation of subjects using a “gold standard”. However, in many situations, it is infeasible to apply the gold standard or no such standard is available. Therefore, the true sensitivity and specificity are unknown and errors in measurement can occur due to the subjective clinical judgement, technical imperfections or interpretational difference etc. In this paper, the true unobserved cancer and detection status are treated as latent variables, with Bayesian inference using data augmentation to account for them. Detection rates are adjusted for both individual and group level explanatory variables.

L'estimation de la sensibilité et de la spécificité d'examen diagnostics ou de dépistage nécessite habituellement une confirmation indépendante du statut des sujets utilisant un étalon-or. Cependant, dans plusieurs situations, l'étalon-or n'est pas disponible ou est difficile à appliquer. La sensibilité et la spécificité réelles sont alors inconnues et des erreurs de mesure peuvent se produire à cause de la subjectivité du jugement clinique, d'imperfections techniques, de différences d'interprétations, etc. Dans cet article, les états réels de cancer et de détection non observés sont traités comme des variables latentes, et l'inférence bayésienne avec augmentation de données est utilisée afin d'en tenir compte. Les taux de détection sont ajustés pour variables explicatives au niveau du groupe et de l'individu.

**[MS-201]****Generalized Information Criteria for Model Selection with Penalized Maximum Likelihood Estimation****Un critère d'information généralisé pour la sélection d'un modèle avec l'estimation par maximum de vraisemblance pénalisée**

Ian McLEOD & Changjiang XU, *University of Western Ontario*

Penalized maximum likelihood estimation has been developed for high dimensional variable selection. Methods with different regularization parameters produce a set of candidate models. Previous researchers used cross-validation to select the best model. Instead we consider generalized information criteria for selecting the best model. We investigate the finite-sample and asymptotic properties of the criteria. The probability distribution of selecting a model is derived. We present the conditions of whether the criteria are consistent or not. Simulation investigation confirms our results. This approach is illustrated with an application.

L'estimation par maximum de vraisemblance pénalisée a été développée pour la sélection de variables en dimension élevée. Des méthodes ayant différents paramètres de régularisation produisent un ensemble de modèles candidats. La validation croisée est habituellement utilisée afin de choisir le meilleur modèle. Nous considérons plutôt des critères d'information généralisés afin de sélectionner le meilleur modèle. Nous étudions les propriétés asymptotiques et pour échantillons finis des critères. La distribution de probabilité de la sélection de modèles est obtenue. Nous présentons les conditions nécessaires pour la cohérence des critères. Une étude de simulation confirme nos résultats. Cette approche est illustrée à l'aide d'une application.

**[MS-202]****A Pseudo-EM Algorithm for Clustering Incomplete Longitudinal Data****Un algorithme pseudo-EM pour la classification de données longitudinales incomplètes**

Mateen SHAIKH, Paul McNICHOLAS & Anthony F. DESMOND, *University of Guelph*

A method for clustering incomplete longitudinal data, and gene expression time course data in par-

Une méthode de classification de données longitudinales incomplètes et, en particulier, de données de décours tem-

ticular, is presented. An existing method that utilizes mixtures of multivariate Gaussian distributions with modified Cholesky-decomposed covariance structure is extended to account for incomplete data. Parameter estimation is carried out using an algorithm that is similar to an expectation-maximization algorithm. We focus on the particular application of clustering incomplete gene expression time course data. The Rand and adjusted Rand indices are used to measure classification agreement, which is very good on these gene expression time course data.

porel d'expression génétique, est présentée. Une méthode existante, utilisant des mélanges de distributions gaussiennes multivariées dont la structure de covariance est une décomposition de Cholesky modifiée, est généralisée pour tenir compte des données incomplètes. L'estimation paramétrique est effectuée en utilisant un algorithme similaire à un algorithme espérance-maximisation. Nous nous concentrons sur la classification de données incomplètes de décours temporel d'expression génétique. Les indices de Rand et de Rand ajusté sont utilisés afin de mesurer l'accord de classification, qui est très bon pour les données utilisées.

**Session 10A      Mardi 25 mai • Tuesday, May 25, 16:25 – 17:25      VAN 1245A**

Isobel Loutit Invited Address of the Business and Industrial Statistics Section Special: Quality Technology in the High-tech Age

Allocution Isobel Loutit du Groupe de statique industrielle et de gestion : La technologie en qualité à l'ère de la haute technologie

[MS-203]

**Quality Technology in the High-tech Age**

**La gestion de la qualité à l'ère de la haute technologie**

C. F. Jeff Wu, *Georgia Inst of Technology*

Traditional work in quality had its main impetus from manufacturing. As the economies are increasingly dependent on high value-added products, it is natural to ask what role can quality professionals play in the high-tech age. I will use several research examples to search for hints to the future, including robust growth of nano-materials, nano-mechanics, computer-aided design of data centres, and electronics packaging. They have the following distinguished features. The responses can be sensitive to input conditions; they often require complex modelling. Finally, I will try to address the ultimate question: "will there be a paradigm shift and what to expect"?

Les travaux en gestion de la qualité ont traditionnellement été motivés par le secteur manufacturier. Les économies étant de plus en plus basées sur des produits à haute valeur ajoutée, il est naturel de s'interroger sur le rôle que peuvent jouer les professionnels de la qualité à l'ère des hautes technologies. J'utiliserai plusieurs exemples de recherche afin de chercher des pistes pour l'avenir, incluant la croissance robuste des nanomatériaux, de la nanomécanique, de la conception de centres de données assistée par ordinateur, et de conditionnement électronique. Ces exemples partagent les caractéristiques suivantes : ils nécessitent une modélisation complexe et la réponse peut être sensible aux conditions initiales. Je tenterai finalement de répondre à l'ultime question : y aura-t-il un changement de paradigme ?

**Session 10B      Mardi 25 mai • Tuesday, May 25, 16:25 – 17:55      VAN 2289A****Analysis of Next Generation Sequencing Data for Functional Genomics****Analyse de données de séquençage de la prochaine génération en génomique fonctionnelle****[MS-204]****PICS: Probabilistic Inference for ChIP-seq****PICS : Inférence probabilistique pour ChIP-Seq**Raphaël GOTTARDO, *Institut de recherches cliniques de Montréal (IRCM)*

ChIP-seq, which combines chromatin immunoprecipitation with massively parallel short-read sequencing, can profile *in vivo* genome-wide transcription factor-DNA association with higher sensitivity, specificity and spatial resolution than ChIP-chip. While it presents new opportunities for research, ChIP-seq poses new challenges for statistical analysis that derive from the complexity of the biological systems characterized and the variability and biases in its digital sequence data. We propose a method called PICS (Probabilistic Inference for ChIP-seq) for extracting information from ChIP-seq aligned-read data in order to identify regions bound by transcription factors.

ChIP-seq, qui combine l'immunoprécipitation de la chromatine à un séquençage à lecture courte massivement parallèle, peut profiler *in vivo* l'association facteur de transcription-ADN pour tout le génome et ce, avec une sensibilité, une spécificité et une résolution spatiale plus hautes que ChIP-chip. Bien qu'apportant de nouvelles opportunités de recherche, ChIP-seq pose de nouveaux défis d'analyse statistique dû à la complexité des systèmes biologiques caractérisés et à la variabilité et aux biais présents dans ses données de séquençage digital. Nous proposons une méthode appelée PICS (inférence probabiliste pour ChIP-seq) permettant d'extraire l'information de données ChIP-seq et d'identifier les régions bornées par les facteurs de transcription.

**[MS-205]****Statistical Issues in the Analysis of RNA-seq Data****Problèmes statistiques dans l'analyse de données d'ARN-seq**John MARIONI, *University of Chicago*

Next generation sequencing has revolutionized genetics by allowing the genomes, transcriptomes and epigenomes of multiple species to be studied in exquisite detail. For example, by sequencing RNA samples using RNA-seq technology, information about splice forms, expression at the exon level, and allele-specific expression can be obtained at unprecedented resolution. In this talk, I will describe some of the statistical challenges that arise when analyzing these data, and discuss how addressing them allows pertinent biological questions to be answered. Issues I will consider include:

1. Reproducibility of RNA-seq experiments;
2. Identifying genetic variants that influence gene expression and splicing.

La nouvelle génération de techniques de séquençage a révolutionné la génétique en permettant d'étudier en détail le génome, le transcriptome et l'épigénome de plusieurs espèces. Par exemple, en séquençant des échantillons d'ARN avec la technologie ARN-seq, de l'information relative à l'expression au niveau des exons et à l'expression spécifique à certains allèles peut être obtenue avec des résolutions sans précédent. Dans cet exposé, je vais décrire certains des défis statistiques rencontrés dans l'analyse de telles données. Je montrerai que surmonter ces défis permet de répondre à des questions biologiques pertinentes. Les problèmes considérés incluent :

1. la reproductibilité d'expériences ARN-seq ;
2. l'identification de variantes génétiques influençant l'expression et l'épissage génétiques.

**[MS-206]****Bias Correction and Bayesian Analysis of Digital Gene Expression Libraries  
Correction du biais et analyse bayésienne de bibliothèques digitales d'expression génétique**

Russell ZARETZKI & Michael GILCHRIST, *University of Tennessee*, William BRIGGS, *Methodist Hospital, Brooklyn N.Y.*, Artin ARMAGAN, *Duke University*

Digital Gene Expression (DGE) is used to sample the mRNA pool of an organism's transcriptome. Incomplete digestion during the tag formation process allows for multiple tags to be generated from a single transcript. The probability of forming a tag varies with its relative location leading to a biased sample of the actual transcript pool. Bayesian inference algorithms that account for tag formation effects are compared and selection of the prior is discussed. Calculations demonstrate the potential for false positive and negative findings due to variation in tag formation probabilities when testing for differential expression.

L'expression génétique digitale (EGD) est utilisée pour échantillonner l'ARNm du transcriptome d'un organisme. Une digestion incomplète lors du processus de formation d'étiquette fait en sorte que plusieurs étiquettes sont générées à partir d'une seule transcription. La probabilité de formation d'une étiquette varie selon sa position relative, menant à un échantillon biaisé de l'ensemble réel des transcriptions. Des algorithmes bayésiens d'inférence qui tiennent compte des effets de la formation des étiquettes sont comparés et la sélection d'une loi a priori est abordée. Des calculs démontrent la possibilité de faux positifs ou faux négatifs dus à la variation de probabilités de formation d'étiquettes lors de tests d'expression différentielle.

**Session 10C      Mardi 25 mai • Tuesday, May 25, 16:25 – 17:55      VAN 2291A****Stochastic Models in Population Genetics  
Modèles stochastiques en génétique des populations****[MS-207]****Sources of Randomness in Genetic Ancestry  
Sources de variation aléatoire en ascendance génétique**

John WAKELEY & Sohini RAMACHANDRAN, *Harvard University*

Models of genetic ancestry do not typically distinguish the stochasticity of Mendelian inheritance from that due to the vagaries of survival and reproduction. This can be viewed as a conceptual flaw, especially in cases where models of genetic ancestry are applied to data from different loci in the genome. Further, differences of opinion may exist as to whether the various aspects of genetic ancestry may be treated as random. Analytic and simulation results will be presented, in which the applicability of the standard model of genetic ancestry, Kingman's coalescent, is assessed with regard to the different possible sources of stochasticity.

Les modèles d'ascendance génétique ne distinguent habituellement pas la variabilité de l'hérédité mendélienne de celle due aux caprices de la survie et de la reproduction. Ceci peut être vu comme un défaut conceptuel, surtout lorsqu'un modèle d'ascendance génétique est appliqué à des données associées à différents locus dans le génome. Aussi, des différences d'opinions peuvent exister quant à la possibilité de traiter divers aspects de l'ascendance génétique comme étant aléatoires. Des résultats analytiques et de simulation seront présentés ; ils permettent d'évaluer l'applicabilité du modèle d'ascendance génétique usuel, la coalescence de Kingman, par rapport aux diverses sources de variabilité possibles.

**[MS-208]****Likelihood-based Inference of Dispersal Rates from Patterns of Genetic Diversity****Méthodes de vraisemblance pour inférer sur des taux de dispersion à partir de patrons de diversité génétique**

François ROUSSET, *CNRS & Université Montpellier 2, France*, Raphaël LEBLOIS, *Museum National d'Histoire Naturelle, Paris, France*

Population geneticists have been busy trying to infer demographic parameters (such as dispersal rates) and mutation rates from patterns of genetic diversity. Likelihood inferences have been slowly developing due to the difficulty of computing likelihoods under most genetic models of interest. Some recent advances in this field use importance sampling algorithms developed by Griffiths et al. I will describe the performance of inferences based on such algorithms. I will focus on models of isolation by distance, in which dispersal occurs preferentially among adjacent populations.

Les généticiens des populations se sont efforcés d'estimer les paramètres démographiques (tels que les taux de dispersion) et les taux de mutation à partir des patrons de diversité génétique. Les approches par vraisemblance se sont développées lentement à cause de la difficulté à calculer la vraisemblance sous la plupart des modèles d'intérêt. Certains progrès récents dans ce domaine sont fondés sur des algorithmes d'échantillonnage pondéré développés par Griffiths et coll. Je décrirai la performance d'inférences fondées sur ces algorithmes. Je m'intéresserai principalement aux modèles d'isolement par la distance, selon lesquels la migration se produit préférentiellement entre populations voisines.

**[MS-209]****Coalescent Processes without Proper Frequencies and Applications to the Poisson-Dirichlet Coalescents****Processus coalescents sans fréquences propres et applications aux coalescents Poisson-Dirichlet**

Martin MÖHLE, *University of Tübingen*

Exchangeable coalescents without proper frequencies are considered. We study the external and the total branch length and the number of mutations on the genealogical tree for large sample size  $n$ . The limits are characterized via exponential integrals of the subordinator associated with the frequency of singletons. The proofs are based on decompositions into external and internal branches. The external branches are treated via the method of moments. The internal branches do not contribute to the limit since the number of collisions for these coalescents is asymptotically negligible compared to  $n$ . The results are applied to Poisson-Dirichlet coalescents.

Nous considérons les coalescents échangeables sans fréquences propres. Nous étudions les longueurs totales et externes des branches et le nombre de mutations dans un arbre généalogique pour une grande taille d'échantillon  $n$ . Les limites sont caractérisées par des intégrales exponentielles du subordonneur associé à la fréquence de singletons. Les preuves sont basées sur des décompositions en branches externes et internes. Les branches externes sont traitées par la méthode des moments. Les branches internes ne contribuent pas à la limite puisque le nombre de collisions pour ces coalescents est asymptotiquement négligeable par rapport à  $n$ . Les résultats sont appliqués à des coalescents de Poisson-Dirichlet.

**[MS-210]****A Coalescent Dual Process in a Cannings Model with Genic Selection, and the Lambda Coalescent Limit****Un processus dual de coalescence dans un modèle de Cannings avec sélection génétique et la limite de coalescence Lambda**

Robert GRIFFITHS & Alison ETHERIDGE, *Oxford University*, Jay TAYLOR, *Arizona State University*

A coalescent dual process can be derived for a class of continuous-time Cannings models with viability selection. In these models, individuals may give birth to multiple offspring whose survival depends

Un processus dual de coalescence peut être obtenu pour une classe de modèles de Cannings en temps continu à sélection de viabilité. Dans ces modèles, les individus peuvent donner naissance à des progénitures multiples dont la survie dépend

on both the parental genotype and the brood size. In the limit of infinite population size, the non-neutral Cannings models converge to a Lambda–Fleming–Viot process. The dual is a branching-coalescing process which follows the typed ancestry of genes backwards in time with real and virtual lineages.

du génotype parental et de la taille de la progéniture. À la limite, lorsque la taille de la population tend vers l'infini, le modèle de Cannings non neutre converge vers un processus de type Lambda-Fleming-Viot. Le dual est un processus de coalescence à branchements qui trace l'ascendance de gènes en remontant le temps avec des lignées réelles et virtuelles.

**Session 10D      Mardi 25 mai • Tuesday, May 25, 16:25 – 17:55      VAN 2809A**

**Panel Discussion on Problems and Issues Faced by New Researchers**

**Panel sur les problèmes et questions auxquels font face les nouveaux chercheurs**

It is well known that junior faculty members often struggle in finding an optimal balance between research, teaching, service and personal life. Many difficulties of this career period are very common across academic institutions and generations of younger researchers. However, still many newly appointed investigators feel that they are left with their problems alone, which negatively affects their research productivity. The main focus of this roundtable session is to share experiences and discuss specific problems and difficulties that are typically met in the first years after graduation, i.e., job interviews, grant proposals, papers, consulting, collaborating, balancing life and others.

Il est bien connu qu'il est difficile pour les nouveaux professeurs de trouver un équilibre entre la recherche, l'enseignement, le service et la vie personnelle. Plusieurs difficultés rencontrées à ce stade de la carrière sont très courantes parmi diverses institutions académiques et générations de jeunes chercheurs. Cependant, beaucoup de chercheurs nouvellement engagés ont l'impression d'être seuls face à leurs problèmes, ce qui affecte négativement leur productivité en recherche. Cette table ronde servira principalement à partager des expériences et discuter de problèmes et difficultés spécifiques habituellement rencontrés lors des premières années suivant l'obtention du diplôme, p. ex., entrevues, demandes de subventions de recherche et articles, consultation et collaboration, et recherche d'équilibre.

**Session 10E      Mardi 25 mai • Tuesday, May 25, 16:25 – 17:55      VAN 2811A**

**Analysis of Multilevel Models for Complex Survey Data**

**Analyse de modèles multiniveaux pour des données d'enquêtes complexes**

[MS-211]

**Estimation of Variance Components for Multilevel Models**

**Estimation de composantes de la variance pour modèles multiniveaux**

Danny PFEFFERMANN, *Hebrew University & University of Southampton*

Multi-level models are often fitted to data obtained from complex surveys that involve unequal selection probabilities that may be related to the outcome values. Estimators of model parameters that ignore the sampling design may be severely biased in such cases. Weighting reduces the bias significantly but may not be sufficient unless the sample sizes at the various levels are large. In this presentation we propose a general approach for bias correction based on resampling procedures. The proposed approach is assessed by simulations and

Les modèles multiniveaux sont souvent ajustés à des données provenant d'enquêtes complexes impliquant des probabilités de sélection inégales pouvant être liées à la réponse. Les estimateurs des paramètres du modèle ignorant le plan d'échantillonnage peuvent alors être sévèrement biaisés. La pondération permet de réduire le biais, mais peut ne pas suffire si les tailles échantillonales aux différents niveaux ne sont pas grandes. Dans cet exposé, nous proposons une approche générale pour corriger le biais à partir de procédures de rééchantillonnage. L'approche proposée est évaluée par des simulations et réduit efficacement le biais d'estimateurs

is shown to be effective in reducing the bias of weighted and unweighted estimators under different scenarios.

pondérés et non pondérés sous différents scénarios.

[MS-212]

**Estimation of Empirical Distributions of Variance Component Estimators for Multilevel Models**  
**Estimation des distributions empiriques des estimateurs des composantes de la variance pour modèles multiniveaux**

Wilson LU, *Acadia University*, Zilin WANG, *Wilfrid Laurier University*, Mary THOMPSON, *University of Waterloo*

In complex surveys, when a multilevel model is often used to capture the hierarchical structure of the population of interest, practical issues such as unequal probability sampling schemes employed at different stages and occurrence of informative sampling complicate the empirical inference procedure. Our research interest resides in investigating the variability of variance component estimators in a multilevel model setting while accounting for the above mentioned complications.

We propose an algorithmic approach, termed as “double resampling” procedure, to obtain the empirical distributions of variance component estimators. The simulation study results are presented to show the feasibility of the procedure.

En enquêtes complexes, lorsqu’un modèle à plusieurs niveaux est souvent utilisé afin de représenter la structure hiérarchique de la population d’intérêt, des problèmes pratiques tels les plans à probabilités inégales utilisés à différentes étapes et l’échantillonnage informatif peuvent compliquer la procédure d’inférence empirique. Nous nous intéressons à l’étude de la variabilité des estimateurs des composantes de la variance dans un contexte de modèles multiniveaux tout en tenant compte de ces complications.

Nous proposons une approche algorithmique, dite procédure de « double rééchantillonnage », afin d’obtenir les distributions empiriques des estimateurs des composantes de la variance. Des résultats de simulation sont présentés afin de démontrer la faisabilité de la procédure.

[MS-213]

**A Weighted Estimating Equations Approach to Inference for Two-level Models from Survey Data**  
**Une approche d’inférence par équations estimatrices pondérées pour modèles à deux niveaux et données d’enquêtes**

J. N. K. RAO, *Carleton University*

Standard methods for two-level models can lead to erroneous inferences when applied to survey data if the corresponding two-level survey design is informative. One approach to making valid inferences is to use methods that take account of survey design features. I will first review some work based on this approach. I will then propose a weighted estimating equations approach which leads to valid inferences when the number of level 2 sample units (clusters) is large, regardless of the number of level 1 sample units within the selected clusters. This approach is particularly suited to handle generalized linearized two-level models.

Les méthodes habituelles pour les modèles à deux niveaux peuvent mener à des inférences erronées lorsqu’appliquées à des données d’enquête à plan d’échantillonnage à deux niveaux informatif. Une approche pour faire des inférences valides est d’utiliser des méthodes tenant compte des caractéristiques du plan d’enquête. Je fais d’abord la revue de certains travaux basés sur cette approche. Je propose ensuite une approche par équations d’estimation pondérées menant à des inférences valides lorsque le nombre d’unités échantillonnales de niveau 2 (grappes) est grand, indépendamment du nombre d’unités échantillonnales de niveau 1 dans les grappes sélectionnées. Cette approche est particulièrement adaptée aux modèles linéaires généralisés à deux niveaux.

**Session 10F      Mardi 25 mai • Tuesday, May 25, 16:25 – 17:40      VAN 1203****Missing or Incomplete Observations  
Observations manquantes ou incomplètes****[MS-214]****Imputation Strategies for Missing Binary Outcomes in Cluster Randomized Trials  
Stratégies d'imputation pour réponses binaires manquantes en essais cliniques randomisés par grappes**Jinhui MA, Noori AKHTAR-DANESH, Lisa DOLOVICH & Lehana THABANE, *McMaster University*

Missing data is a common problem in most trials including cluster-randomized trials (CRTs). Multiple imputation (MI) is considered the gold standard technique for imputing missing data under the assumption of missing at random. We investigated three MI strategies—predictive model, propensity score method, and Markov Chain Monte Carlo method—that take into account the intra-cluster correlation of outcomes for imputing missing data from CRTs. Using a simulation study based on the Community Hypertension Assessment Trial, we compared the performance of these strategies under different percentages of missing binary outcomes using kappa statistics and robustness of treatment effects.

Les données manquantes représentent un problème fréquemment rencontré dans la plupart des essais cliniques, incluant les essais randomisés par grappes (ERG). L'imputation multiple (IM) est considérée la méthode étalon-or pour imputer les valeurs manquantes sous l'hypothèse qu'elles sont manquantes aléatoirement. Nous avons étudié trois stratégies d'IM — modèle prédictif, scores de propension, chaînes de Markov Monte-Carlo — qui tiennent compte de la corrélation intra-grappes des réponses dans l'imputation des données manquantes d'un ERG. Au moyen d'une étude de simulation basée sur l'Étude d'évaluation de l'hypertension dans les communautés, nous avons comparé la performance de ces stratégies sous différentes proportions de réponses binaires manquantes en utilisant des statistiques kappa et la robustesse des effets de traitement.

**[MS-215]****A Unified Approach for Handling the Combined Problem of Missing Data and Measurement Error  
Une approche unifiée pour traiter le problème de données manquantes combinées à des erreurs de mesure**Michael REGIER, *McGill University*

Observational studies predicated on the secondary use of data can encounter practical problems such as covariates which suffer from the dual problem of missing information and measurement error. In this talk, we will investigate the impact of such imperfection on parameter estimation and consider the properties of a likelihood-based estimator designed to account for such imperfections. This methodology will be applied to a linked dataset composed of the BC Cancer Registry, BC Vital Statistics, and Census data as an example of how low cost data can be used to refine decisions about the collection of high cost data.

Les études observationnelles s'appuyant sur l'utilisation secondaire de données peuvent rencontrer plusieurs problèmes pratiques, comme des covariables souffrant d'information manquante et d'erreurs de mesure. Dans cet exposé, nous allons étudier l'impact de telles imperfections sur l'estimation paramétrique et considérer les propriétés d'un estimateur basé sur la vraisemblance et conçu pour traiter ces imperfections. Cette méthode sera appliquée à un jeu de données jumelées provenant de *BC Cancer Registry*, *BC Vital Statistics*, et de données de recensement comme exemple de données de faible coût informant la prise de décisions relatives à la collecte de données de coût élevé.

**[MS-216]****Analysis of Left-truncated and Right-censored Survival Times with Incomplete Covariate Data  
Analyse de données de survie tronquées à gauche et censurées à droite avec covariables manquantes**Hua SHEN & Richard COOK, *University of Waterloo*

Individuals are enrolled in clinical trials after disease onset but before an event of clinical interest. Secondary analysis often aim to examine covariate effects on time from disease onset to an event (e.g., death). This survival time is left-truncated and subject to right censoring. The left-truncation time contains information about any missing covariates which should be used. We develop an EM algorithm that leads to efficient inferences under parametric models. We consider the settings of both observational studies and clinical trials where interest lies in sub-groups defined based on an incomplete covariate.

Les individus rejoignent un essai clinique après le début de la maladie, mais avant un événement d'intérêt clinique. L'analyse secondaire vise souvent à examiner l'effet des covariables sur le temps entre le début de la maladie et un événement donné (p. ex., la mort). Ce temps de survie est tronqué à gauche et sujet à une censure à droite. Le temps de troncation à gauche contient de l'information au sujet de covariables manquantes qui devrait être utilisée. Nous développons un algorithme EM menant à des inférences efficaces sous modèles paramétriques. Nous considérons le cadre d'études observationnelles et d'essais cliniques où nous nous intéressons à des sous-groupes basés sur une covariable incomplète.

**[MS-217]****Regression Model with Covariates Missing in Non-monotonic Patterns  
Modèle de régression avec covariables manquantes de façon non-monotone**Yang ZHAO & Meng LIU, *University of Regina*

In this research we consider regression models with covariates missing in non-monotonic patterns. We consider estimating equations for regression models of interest and surrogate models for complete data and available data, respectively. We describe a conditional mean estimator for regression parameters. We also derive the asymptotic variance of the estimator. Finally, we use simulation study to examine the performance of the proposed estimator.

Dans cette recherche, nous considérons des modèles de régression à covariables manquantes de façon non monotone. Nous considérons les équations d'estimation pour des modèles de régression d'intérêt ainsi que pour des modèles substituts, et ce à la fois pour des données complètes et pour les données disponibles. Nous décrivons un estimateur de moyenne conditionnelle pour les paramètres de régression. Nous obtenons aussi la variance asymptotique de l'estimateur. Finalement, nous utilisons une étude de simulation pour examiner la performance de l'estimateur proposé.

**[MS-218]****Modelling the Lagged Effect for Incomplete Data in Tumour Xenograft Experiments  
Modélisation de l'effet de rémanence pour données incomplètes en expériences de xénogreffes**Dianliang DENG & Ting ZHANG, *University of Regina*

In the anticancer drug study, tumour xenograft experiments where mice were grafted with human cancer cells were used to evaluate the action efficacy of promising anticancer agents compound. When analyzing these data, the incompleteness due to the death of the mice and the drastic tumour shrinkage is of concern. In this presentation, we propose to develop the model which concerns the lagged effect of anticancer agents on the tumour

En études de médicaments contre le cancer, les expériences de xénogreffes de tumeurs où des souris reçoivent des greffes de cellules humaines cancéreuses furent utilisées afin d'évaluer l'efficacité de composés anticancéreux prometteurs. En analysant ces données, il faut porter attention aux données incomplètes dues à la mort de souris et à la réduction drastique de tumeurs. Dans cette présentation, nous proposons de développer un modèle de l'effet de rémanence d'agents anticancéreux sur le volume des tumeurs.

volumes. The weighted least square method to estimate the lagged effect on the response and ECM algorithm to estimate the dose-response relationship will be applied to analyze real data from a xenograft study.

La méthode des moindres carrés pondérés est utilisée afin d'estimer l'effet de rémanence sur la réponse et l'algorithme ECM, utilisé pour estimer la relation dose-réponse, est appliqué à une étude de xénogreffes réelle.

**Session 10G      Mardi 25 mai • Tuesday, May 25, 16:25 – 17:40      VAN 1211**  
**Longitudinal Data and Mixed Effects**  
**Données longitudinales et effets mixtes**

[MS-219]

**Linear Mixed Effects Models for Longitudinal Studies with Measurement Error on Covariates and Response**  
**Modèles à effets mixtes linéaires pour études longitudinales avec erreurs de mesures dans les covariables et la réponse**

Taraneh ABARIN, *Samuel Lunenfeld Research Institute*, He LI & Liquan WANG, *University of Manitoba*, Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute*, Craig PENNELL, *University of Western Australia*

We define a linear mixed effects model for longitudinal data with measurement error on covariates and the response. The model allows random effects on intercept, covariates without measurement error. Since there is no assumption concerning the functional forms of the distributions of covariates, response, and both measurement errors, the model is nonparametric. Here we show that using only the first two moment equations, we can estimate all the parameters of the model. Motivated by a large child cohort study from Australia (RAINE), we investigate the finite sample performances of the estimators, and show the effect of measurement error.

Nous définissons un modèle linéaire à effets mixtes pour données longitudinales avec erreurs de mesures des covariables et de la réponse. Le modèle permet des effets aléatoires pour l'ordonnée à l'origine et pour les covariables sans erreur de mesure. Puisque nous ne faisons pas d'hypothèse sur la forme fonctionnelle des distributions des covariables, de la réponse et des deux types d'erreur de mesure, le modèle est non paramétrique. Nous montrons ici qu'à partir des deux premiers moments, nous pouvons estimer tous les paramètres du modèle. Nous étudions la performance des estimateurs pour échantillon fini dans le contexte d'une étude cohorte pédiatrique australienne de grande envergure, et montrons l'effet de l'erreur de mesure.

[MS-220]

**Improving the Estimation of Fixed Effects in the Linear Mixed Model for Longitudinal Data**  
**Améliorer l'estimation d'effets fixes dans un modèle linéaire mixte pour données longitudinales**

Tanvir QUADIR & S. Ejaz AHMED, *University of Windsor*

This study proposes improved estimation of fixed effects in linear mixed model for longitudinal data. The relative risk of restricted estimator, preliminary test estimator, improved preliminary test estimator, James-Stein type estimators as opposed to the unrestricted maximum likelihood estimator (MLE) have been explored. Although restricted and preliminary test estimator appear to outperform the unrestricted MLE at and near the null hypothesis of the restrictions on fixed effects, improved preliminary test estimator and James-Stein type estimators

Cette étude propose une estimation améliorée des effets fixes en modèles linéaires mixtes pour données longitudinales. Nous explorons les risques relatifs de l'estimateur restreint, de l'estimateur de test préliminaire, de l'estimateur de test préliminaire amélioré, d'estimateur de type James-Stein par rapport à l'estimateur du maximum de vraisemblance non restreint (EMV). Bien que les estimateurs restreints et de test préliminaire semblent surpasser l'EMV près de l'hypothèse nulle de contraintes sur les effets fixes, les estimateurs de test préliminaire amélioré et de type James-Stein ont une meilleure performance même lorsqu'on s'éloigne plus de

demonstrate fairly an improved performance over the unrestricted MLE even when the degree of violation of restriction on fixed effects intensifies.

l'hypothèse nulle.

[MS-221]

**Generalized Linear Models with Nonparametric Random Effects for Longitudinal Data**  
**Modèles linéaires généralisés avec effets aléatoires non paramétriques pour données longitudinales**

Guohua YAN, Renjun MA & M. Tariqul HASAN, *University of New Brunswick*

We introduce a new class of generalized linear mixed models for longitudinal data based on the Tweedie distributions. Our main results depend only on the first- and second-moment assumptions of unobserved random effects, accommodating flexible temporal structures. An optimal estimating function for the regression parameters has been developed using the best linear unbiased predictor of random effects, allowing an efficient common fitting algorithm for the whole class. In addition, we obtain consistent estimators for both regression and dispersion parameters. Our approach unifies population-averaged and subject-specific inferences. We illustrate the method by analyzing both continuous and count data.

Nous présentons une nouvelle classe de modèles mixtes linéaires généralisés pour données longitudinales basée sur les distributions de Tweedie. Nos principaux résultats ne dépendent que d'hypothèses sur les premiers et seconds moments d'effets aléatoires non observés, permettant des structures temporelles flexibles. Une fonction d'estimation flexible pour les paramètres de régression a été développée à partir du meilleur prédicteur linéaire sans biais des effets aléatoires, nous donnant un algorithme efficace d'ajustement pour toute la classe. De plus, nous obtenons des estimateurs cohérents pour les paramètres de régression et de dispersion. Notre approche unifie les inférences spécifiques aux sujets et basées sur le moyennage de populations. Nous illustrons la méthode en analysant des données continues et de dénombrement.

[MS-222]

**A Composite Likelihood Approach for Longitudinal Data with Nonignorable Non-monotone Missing Observations in Both Response and Covariates**

**Une approche par vraisemblance composite pour données longitudinales avec données manquantes non ignorables et non monotones dans la réponse et les covariables**

Haocheng LI & Grace YI, *University of Waterloo*

Incomplete longitudinal data arise commonly in many fields including public health studies and survey sampling. To address incompleteness in response or covariates, likelihood-based approaches are often employed. However, such methods can become computationally infeasible when missingness arises from both responses and covariates. In this talk, we propose a composite likelihood approach under a general framework in which missing observations may occur in both response and covariates, missing data patterns can be non-monotone, and missing data mechanisms can be non-ignorable. For computationally feasible purposes, a two-stage estimation algorithm is developed. Numerical studies and an application example are reported.

On retrouve fréquemment des données longitudinales incomplètes dans plusieurs domaines, incluant les études en santé publique et les enquêtes par sondage. Des méthodes basées sur la vraisemblance sont fréquemment utilisées afin de traiter une réponse ou des covariables incomplètes. Cependant, de telles méthodes peuvent être difficilement calculables lorsque des valeurs manquantes se retrouvent à la fois dans la réponse et les covariables. Dans cet exposé, nous proposons une approche par vraisemblance composite dans un cadre général où les observations manquantes peuvent se trouver dans la réponse ou les covariables, les données peuvent être manquantes de façon non monotone et non ignorable. Un algorithme d'estimation à deux étapes est développé. Des études numériques et un exemple d'application sont présentés.

[MS-223]

**Marginal Structural Models for Competing Risks**  
**Modèles structureaux marginaux pour risques compétitifs**

Erica MOODIE & David STEPHENS, *McGill University*

Analysis of the effect of exposures on multiple-outcomes data is often complicated by the fact that the occurrence of one outcome precludes that of others and, in a longitudinal exposure setting, time-varying confounders may be mediators through which past treatment acts. We develop a novel analytic approach which accounts for time-varying confounding and mediation in a competing risks setting. We demonstrate the large-sample properties of the approach using simulated data, and apply it to a cohort of HIV-HCV co-infected patients to examine the relationship between antiretroviral therapy interruptions and death by end-stage liver disease, HIV, overdose or other causes.

L'analyse de l'effet d'expositions sur des données à réponses multiples est souvent compliquée par le fait que la réalisation d'une réponse prévient la réalisation des autres réponses et que, dans le contexte d'expositions longitudinales, un confondateur dépendant du temps peut être un médiateur par lequel agit un traitement passé. Nous développons une approche analytique originale tenant compte des confondateurs dépendants du temps et de la médiation dans un contexte de risques compétitifs. Nous montrons les propriétés asymptotiques de l'approche au moyen de données simulées et l'appliquons à une cohorte de patients co-infectés par le VIH et le VCH afin d'examiner la relation entre interruption de la thérapie antirétrovirale et décès causé par maladies du foie, VIH, surdose ou autre causes.

**Session 10H****Mardi 25 mai • Tuesday, May 25, 16:25 – 17:55****VAN 1871****Methodological Issues in Designing Surveys****Questions méthodologiques dans la planification d'enquêtes**

[MS-224]

**Methodological Issues of a Survey on Industrial Processes**  
**Enjeux méthodologiques d'une enquête portant sur les processus industriels**

Sébastien LABELLE-BLANCHET, Claude GIRARD & Martin HAMEL, *Statistique Canada*

The Survey of industrial processes (SIP) is a new type of survey at Statistics Canada. SIP aims at getting a better understanding of the industrial processes for a given industry, especially the impacts they have in the fields of environment and economy. This type of survey brings very interesting methodological issues. The presentation will give an overview of this, with the industry of the gas stations as an example. A better understanding of the industrial processes will allow us, hopefully, to estimate the emissions of benzene of which it is responsible.

L'Enquête sur les processus industriels (EPI) est une enquête d'un nouveau genre à Statistique Canada. L'EPI vise à mieux comprendre les processus industriels employés par une industrie donnée, tout particulièrement le rôle qu'ils jouent au chapitre de l'environnement et de l'économie. Ce type d'enquête pose des problèmes méthodologiques intéressantes. La présentation en donnera un exemple avec l'industrie des stations d'essence pour laquelle une meilleure compréhension des processus industriels permettra d'estimer, l'espère-t-on, les émissions polluantes de benzène dont elle est responsable.

**[MS-225]****Business Survey Data Collection Research at Statistics Canada****La recherche sur la collecte de données des enquêtes-entreprises à Statistique Canada**Jeannine CLAVEAU, Xinye (Hannah) YANG & Laurie REEDMAN, *Statistique Canada*

In an effort to collect high quality and cost-effective data, Statistics Canada regularly reviews its collection practices. Over the past few years, Statistics Canada has conducted several analytic studies using paradata. The work done until now was mainly for social surveys. Statistics Canada has decided to extend the research to business surveys as well. The objective of the research is to identify operational efficiency opportunities to improve the quality of data collected. In this presentation, we discuss some paradata investigation done at Statistics Canada for business surveys and provide some preliminary results of our analysis.

Afin de recueillir des données de grande qualité de manière rentable, Statistique Canada procède régulièrement à la révision de ses pratiques de collecte. Ces dernières années, Statistique Canada a mené plusieurs études analytiques en servant de paradonnées. Le travail fait jusqu'à présent a porté principalement sur les enquêtes sociales. Statistique Canada a décidé d'étendre la recherche aux enquêtes-entreprises. L'objectif de cette recherche est d'identifier les opportunités opérationnelles afin d'améliorer la qualité des données recueillies. Cette présentation s'intéresse aux recherches faites à Statistique Canada pour les enquêtes-entreprises et aux résultats préliminaires de notre analyse.

**[MS-226]****Redesign of Research and Development in the Canadian Industry Survey: Combining Tax and Survey Data****Remaniement de l'Enquête sur la recherche et développement dans l'industrie canadienne : Combiner des données fiscales et d'enquêtes**Javier OYARZUN & Eric PELLETIER, *Statistique Canada*

Administrative data plays an important role in Statistics Canada surveys. In 2008, the Research and Development in Canadian Industry (RDCI) Survey was reworked to enhance the use of Canada Revenue Agency administrative data. This survey aims to assess the evolution of Canadian scientific and technological activities. Approximately 2,000 companies are surveyed using a paper survey, and administrative data is used for non surveyed companies. This reworking leads to higher quality data and allows to make better use of administrative data.

Statistique Canada accorde une place importante à l'utilisation des données administratives dans ses programmes d'enquêtes. Pour 2008, un remaniement de l'Enquête sur la recherche et le développement dans l'industrie canadienne a été réalisé pour une meilleure utilisation des données administratives provenant de l'Agence du revenu du Canada. Cette enquête a pour but de suivre l'évolution des activités scientifiques et technologiques au Canada. Une enquête papier annuelle est réalisée auprès d'environ 2 000 entreprises, combinée à l'utilisation des données administratives pour les entreprises non enquêtées. Ce remaniement permet à l'enquête d'améliorer la qualité des données et d'utiliser de façon accrue les données administratives.

**[MS-227]****The Methodological Challenges of the Survey on Living with Chronic Diseases in Canada****(SLCDC)****Les défis méthodologiques de l'enquête sur les personnes ayant une maladie chronique au Canada (EPMCC)**Mamadou DIALLO, Marie-Claude DUVAL & Steven THOMAS, *Statistique Canada*

The purpose of the Survey is to provide information on the impact of chronic disease on individuals, as well as how people with chronic disease manage their health condition. In 2009, the survey covers

Le but de l'enquête est de fournir de l'information sur l'impact des maladies chroniques sur les individus et sur la façon dont ces derniers gèrent leur condition. En 2009, l'enquête s'est intéressée à l'arthrite et l'hypertension en

arthritis and hypertension, and followed-up a subsample of respondents from the Canadian Community Health Survey (CCHS) 2008. This paper will present an overview of the SLCDC 2009 survey and some of its challenges; one of them being the non-response adjustment for the higher than expected out of scope rates. It will also cover the sampling process as well as the estimation method used.

**[MS-228]**

**Multidimensional Tabular Data Suppression in the Canadian Cancer Registry**

**Suppression des données dans les tableaux multidimensionnels du Registre canadien du cancer**

Eric HORTOP, *Statistics Canada*

The Canadian Cancer Registry is a national database of cancer cases compiled from provincial and territorial cancer registry data. Among the tabulations publicly available from the registry are tables of incidence, which are cross-tabulated by geography, tumour site, sex and age. These high-dimensional tables of rare events are available via Statistics Canada's CANSIM II system and require a disclosure control strategy. This paper presents a system that counters an attacker by iteratively suppressing cells, then verifies table security using a separately defined mathematical programming problem. It also provides graphical and tabular process information to analysts.

faisant un suivi d'un sous-échantillon de répondants de l'enquête sur la santé des collectivités canadiennes (ESCC) de 2008. Le présent article discute de l'EPMCC 2009, ainsi que certains des défis méthodologiques survenus, en particulier de l'ajustement de la non-réponse pour tenir compte d'un taux plus élevé que prévu d'unités hors cibles. Le processus d'échantillonnage et la méthode d'estimation utilisés seront également présentés.

Le Registre canadien du cancer est une base de données nationale des cas de cancer compilée à partir des registres provinciaux et territoriaux sur le cancer. Parmi les tableaux du registre diffusés au grand public, on trouve des tableaux d'incidence classés selon la géographie, l'organe atteint par la tumeur, le sexe et l'âge. Ces tableaux multidimensionnels d'événements rares peuvent être consultés dans le système CANSIM II de Statistique Canada et nécessitent une stratégie de contrôle de la divulgation. Cette communication présente un système de défense contre un agresseur au moyen d'une suppression itérative des cellules. Nous vérifions ensuite la sécurité des tableaux en utilisant un problème de programmation mathématique défini séparément. En outre, nous fournissons aux analystes un résumé graphique et tabulaire du processus.

**[MS-229]**

**Evaluation of Telephone Interview Collection in First Nations Communities: 2009 Pilot Study**

**Évaluation de la collecte par entrevue téléphonique dans les communautés des Premières Nations : étude pilote de 2009**

Cindy UBARTAS, *Statistique Canada*

Given that few recent data are available on socioeconomic conditions in First Nations communities, a thematic survey is being considered for development. The use of telephone interviews that could reduce collection costs poses several methodological challenges between censuses. In 2009, a test was done to evaluate this approach and compare it to the personal interview approach generally used in these communities.

This presentation will introduce the context that brought about the production of this test. Next, it will describe the options considered, the methodology selected and the principal results.

Puisque peu de données récentes portant sur les conditions socioéconomiques des communautés des Premières Nations sont disponibles, le développement d'une enquête thématique est envisagé. L'utilisation d'entrevues téléphoniques, qui permettraient de diminuer les coûts de la collecte, suscite de nombreux enjeux méthodologiques en période intercensitaire. En 2009, un test a été effectué afin d'évaluer cette approche et de la comparer à celle d'entrevues personnelles, habituellement utilisées dans ces communautés.

Dans cette présentation, le contexte ayant mené à la production de ce test sera abordé. Les options considérées, la méthodologie choisie ainsi que les principaux résultats seront ensuite décrits.

**Session 11A Mercredi 26 mai • Wednesday, May 26, 08:30 – 09:15 POU 1112**

Pierre Robillard Award Address

Allocution de la récipiendaire du prix Pierre-Robillard

[MS-230]

**Information Matrices in Estimating Function Approach: Test for Model Misspecification and Model Selection****Matrices d'information dans l'approche des fonctions d'estimation : Test pour inadéquation du modèle et sélection de modèle**Qian ZHOU, *Harvard School of Public Health*, Peter SONG, *University of Michigan*, Mary THOMPSON, *University of Waterloo*

In this talk, we focus on the circumstances where the first moment is correctly specified, but the second moment is misspecified in quasi-likelihood inference. By comparing two forms of information matrix, we propose a powerful information ratio (IR) test for misspecification of variance/covariance structure. In addition, two model selection procedures are proposed for selecting the optimal variance/covariance structure among a collection of candidate structures. One is based on a sequence of the IR tests, and the other one is based on an “information discrepancy criterion” (IDC), which provides a measurement of discrepancy between the two forms of information matrix.

Dans cet exposé, nous étudions l'inférence par quasi-vraisemblance et mettons l'accent sur le cas où le premier moment est spécifié correctement mais le second moment ne l'est pas. En comparant deux formes de matrice d'information, nous proposons un test de rapport d'information (RI) puissant pour une erreur de spécification de la structure de variance/covariance. De plus, nous proposons deux procédures de sélection de modèle pour choisir la structure de variance/covariance optimale parmi un ensemble de structures candidates. L'une est basée sur une séquence de tests IR alors que l'autre est basée sur un « critère de désaccord d'information » (CDI) qui procure une mesure de désaccord entre deux formes de matrices d'information.

**Session 11B Mercredi 26 mai • Wednesday, May 26, 09:15 – 10:00 POU 1112***Canadian Journal of Statistics* Award AddressAllocution des récipiendaires du prix de la *Revue canadienne de statistique*

[MS-231]

**Two Sample Inference in Functional Linear Models****Inférence pour deux échantillons dans des modèles linéaires fonctionnels**Piotr KOKOSZKA, *Utah State University*, Lajos HORVÁTH, *University of Utah*, Matthew REIMHERR, *University of Chicago*

We propose a method of comparing two functional linear models in which explanatory variables are functions (curves) and responses can be either scalars or functions. In such models, the role of parameter vectors (or matrices) is played by integral operators acting on a function space. We test the null hypothesis that these operators are the same in two independent samples. The complexity of the test statistics increases as we move from scalar to functional responses and relax assumptions on the

Nous proposons une méthode pour comparer deux modèles linéaires fonctionnels, pour lesquels les variables explicatives sont des fonctions (courbes) et les réponses sont ou bien des scalaires ou bien des fonctions. Pour ces modèles, le rôle des vecteurs ou des matrices de paramètres est joué par des opérateurs d'intégration sur un espace de fonctions. Nous testons l'hypothèse nulle que ces opérateurs sont les mêmes pour les deux échantillons indépendants. La complexité des statistiques de test s'accroît quand nous passons de réponses scalaires à des réponses fonctionnelles, et relâchons les hy-

covariance structure of the regressors. They all, however, have an asymptotic chi-squared distribution with the number of degrees of freedom which depends on a specific setting. The test statistics are readily computable using the R package *fda*, and have good finite sample properties. The test is applied to egg-laying curves of Mediterranean flies and to data from terrestrial magnetic observatories.

pothèses sur la structure de covariance des régresseurs. Ces tests ont cependant une distribution asymptotique qui suit une loi du khi-deux, dont le nombre de degrés de liberté dépend d'un cadre particulier. Ces statistiques de test sont calculables avec le programme *fda* du logiciel R, et ont de bonnes propriétés en petit échantillon. La statistique de test est appliquée à des courbes de pontes de mouches méditerranéennes, et à des données d'observatoires magnétiques terrestres.

## **Session 12A Mercredi 26 mai • Wednesday, May 26, 10:30 – 12:00 VAN 1245A**

### **Data and Models on the Move: Functional Data Analysis**

#### **Données et modèles en déplacement : Analyse de données fonctionnelles**

[MS-232]

#### **Data and Models on the Move: Functional Data Analysis**

#### **Données et modèles en mouvement : l'analyse de données fonctionnelles**

Jim RAMSAY, *McGill University*

Functional data analysis involves either data or models distributed over continua, such as time, space, space/time and etc. As a rule, the data are sufficient in density and accuracy to allow estimation of derivatives, which play many roles. This talk surveys the main topics, including functional versions of regression and principal components analysis, curve registration, and dynamic models. Illustrations are drawn from data on human growth and handwriting. The talk concludes with a set of problems that seem important to consider in future research, including coordinate-free modelling, models and data on manifolds, and object-oriented data analysis.

L'analyse de données fonctionnelles implique des données ou modèles distribués sur des continus, tels le temps, l'espace, l'espace-temps, etc. De façon générale, les données sont assez denses et précises afin de permettre l'estimation de dérivées, qui jouent plusieurs rôles. Cet exposé survole la régression, l'analyse en composantes principales, les modèles dynamiques et l'enregistrement de courbes dans un cadre fonctionnel. Des données de croissance humaine et d'écriture illustrent l'exposé qui se termine par un aperçu de problèmes qui semblent importants pour la recherche future, dont la modélisation libre de coordonnées, les modèles et données sur des variétés, et l'analyse de données orientées objet.

## **Session 12B Mercredi 26 mai • Wednesday, May 26, 10:30 – 12:00 VAN 2289A**

### **Spatial Statistics and the Environment**

#### **Statistique spatiale et l'environnement**

[MS-233]

#### **Geographic Epidemiology: Finding Spatial Patterns in Regional Count Data and the Effect of Diagnostic Misclassification Bias**

#### **Épidémiologie géographique : chercher des structures spatiales dans des données régionales de dénombrement et l'effet d'un biais de mauvaise classification diagnostique**

Olaf BERKE, *University of Guelph*

Geographic epidemiology is concerned with the investigation of spatially referenced data to dis-

L'épidémiologie géographique se préoccupe de l'étude de données spatialement référencées afin de découvrir des

cover spatial patterns in the health status of populations. Such patterns may then suggest the existence of environmental risk factors and respective public health strategies. In this presentation spatial patterns of trend, clustering and cluster will be related to the epidemiological triangle of causation: environment, agent and host. Furthermore diagnostic misclassification and its effect on spatial statistics for regional count data is investigated using the concept of observed versus true disease frequency. An example is presented using data about West Nile virus infections in wild birds.

structures spatiales dans le statut de santé de populations. De telles structures peuvent alors suggérer l'existence de facteurs de risque environnementaux et des stratégies de santé publique correspondantes. Dans cet exposé, les structures spatiales de tendance, de classification et de grappes seront liées au triangle de causalité épidémiologique : environnement, agent et hôte. La mauvaise classification de diagnostics et ses effets sur les statistiques spatiales pour données régionales de dénombrement sont aussi étudiés en utilisant le concept de fréquences de maladies observées et réelles. Un exemple utilisant des données d'infection au virus du Nil occidental chez les oiseaux sauvages est présenté.

[MS-234]

### **Geostatistics, Preferential Sampling and Environmental Monitoring**

#### **Géostatistique, échantillonnage préférentiel et surveillance environnementale**

Raquel MENEZES, *University of Minho, Portugal*, Peter DIGGLE, *Lancaster University & Johns Hopkins University School of Public Health*, Ting-Li SU, *Lancaster University*

In environmental monitoring, preferential sampling happens if more monitors are placed in areas classified as high risk for pollution. Traditional geostatistical models assume sampling locations as stochastically independent of measurements, i.e. non-preferential designs. In this talk, we describe a model-based approach, which considers log-Gaussian Cox processes to model the stochastic dependence of sampling locations on spatial variable under study (Diggle, Menezes & Su, 2010). We then present an application to biomonitoring data from eastern Iberian Peninsula, where since the 90s moss has been used as biomonitor for air pollution to measure heavy metals concentration, with intensified sampling in large urban or industrial areas.

En surveillance environnementale, on fait de l'échantillonnage préférentiel si la surveillance est accrue dans des zones à haut risque de pollution. Les modèles géostatistiques traditionnels supposent des lieux d'échantillonnage qui sont indépendants des mesures, c'est-à-dire que ce sont des plans non préférentiels. Dans cet exposé, nous décrivons une approche qui considère un processus de Cox log-gaussiens pour modéliser la dépendance stochastique des lieux échantillonnés par rapport à la variable spatiale étudiée (Diggle, Menezes & Su, 2010). Nous présentons alors une application à des données de biosurveillance de l'est de la péninsule ibérique où, depuis les années 90, la mousse est utilisée afin de mesurer les concentrations de métaux lourds. Ces données proviennent d'un échantillonnage préférentiel favorisant les zones urbaines ou industrielles.

## **Session 12C Mercredi 26 mai • Wednesday, May 26, 10:30 – 12:00 VAN 2291A**

### **Statistics in Conservation**

#### **Statistique en préservation**

[MS-235]

### **A Random Effects Hurdle Model for Shark Counts**

#### **Un modèle de type « haie » avec effets aléatoires pour le comptage de requins**

Eva CANTONI, *University of Geneva, Switzerland*, Joanna MILLS FLEMMING, *Dalhousie University*, Alan WELSH, *The Australian National University, Canberra*

Sharks have been increasingly exploited in the Northwest Atlantic and conservation of their popu-

Les requins de l'Atlantique du Nord-Ouest sont de plus en plus exploités et la conservation de leurs populations est

lations is of particular concern (Baum et al., 2003). We model sharks abundance trends on data from the Myers' lab (Dalhousie, Halifax). Such count data present an excess of zeros and are structured in clusters. We define a hurdle model, where we introduce random effects on both parts to account for the clustering. We show how to estimate the model by maximum likelihood, explain how to obtain empirical best predictors for the random effects and how to obtain an estimator of their mean squared error.

donc d'un intérêt particulier (Baum et coll., 2003). Nous modélisons les tendances d'abondance des requins avec des données provenant du laboratoire de Myers (Dalhousie, Halifax). De telles données de dénombrement comportent un excès de zéros et sont structurées en grappes. Nous définissons un modèle de type « haie » où des effets aléatoires sont introduits afin de tenir compte de la structure en grappes. Nous montrons comment estimer les paramètres du modèle par maximum de vraisemblance, expliquons comment obtenir les meilleurs prédicteurs empiriques des effets aléatoires et décrivons comment obtenir un estimateur de leur erreur quadratique moyenne.

**[MS-236]**

**Estimating Abundance of Animals from Correlated Binary Data**

**Estimation de l'abondance d'animaux à partir de données binaires corrélées**

Julie HORROCKS, *University of Guelph*, David HAMILTON & Hal WHITEHEAD, *Dalhousie University*

To protect and manage wildlife populations effectively, absolute estimates of population size are required. I will introduce a new method for estimating abundance of animals when there is positive probability that some animals in the sampled area remain undetected. The only information available is whether or not any animals are detected in replicate, overlapping circles. The overlap of the circles introduces correlation, which is exploited to estimate density, as well as range and probability of detection. Our method is illustrated using real acoustic data from whale surveys.

Pour protéger et gérer les populations de la faune sauvage efficacement, des estimations de la taille absolue de ces populations sont nécessaires. Au cours de mon intervention, je présenterai une nouvelle méthode pour estimer l'abondance d'animaux dans les régions échantillonnées où il y a des probabilités positives que certains animaux demeurent non détectés. Les seules informations disponibles proviennent des animaux détectés dans des zones circulaires se chevauchant. Ces chevauchements des cercles présentent des corrélations qui sont exploitées pour estimer la densité, ainsi que la gamme et la probabilité de détection. Notre méthode est illustrée en utilisant des données acoustiques réelles provenant d'enquêtes sur les baleines.

**[MS-237]**

**Zero-inflated Spatial Models for Environmental Data: Clustered Zeros versus Spatial Outliers**

**Modèles spatiaux à surreprésentation de zéros pour données environnementales : zéros en grappe contre valeurs spatialement aberrantes**

Laurie AINSWORTH & Charmaine DEAN, *Simon Fraser University*

Many environmental applications, such as species abundance studies yield data with a preponderance of zero counts. This leads to “zero-inflation”. Such zeros provide important clues to physical characteristics associated with habitat suitability or disease resistance. Here, two zero-inflated spatial models are applied to weevil infestation data. The spatial process is modelled with conditional autoregressive or discrete random effects. A mixture model approach is used and the probability of belonging

Plusieurs études environnementales comme celles sur l'abondance des espèces produisent des données avec une prépondérance de zéros. Ceci mène à une surreprésentation de zéros. Ces zéros renferment des indices sur les caractéristiques physiques associées à la qualité de l'habitat et à la résistance aux maladies. Ici, deux modèles spatiaux à surreprésentation de zéros sont appliqués à des données d'infestation de charançons. Le processus spatial est modélisé avec des effets conditionnels autorégressifs ou des effets aléatoires discrets. Des modèles de mélanges sont utilisés et la probabilité

to the zero, “resistant”, component is estimated. The two models distinguish specific data features: clusters of resistant trees located in unsuitable habitats, or individually resistant trees located among infested trees.

d’appartenir à la composante zéro, « résistante », est estimée. Les deux modèles distinguent des caractéristiques spécifiques des données : des grappes d’arbres résistants situés dans des habitats peu adéquats, ou des arbres individuellement résistants situés parmi des arbres infestés.

## **Session 12D Mercredi 26 mai • Wednesday, May 26, 10:30 – 12:00 VAN 2809A**

### **Dependence Modelling in Risk Theory Modèles de dépendance en théorie du risque**

[MS-238]

#### **Discrete-time Risk Models Based on Time Series for Count Random Variables**

#### **Modèles de risque en temps discret basés sur les séries temporelles pour des variables aléatoires de dénombrement**

Hélène COSSETTE & Étienne MARCEAU, *Université Laval*, Véronique MAUME-DESCHAMPS, *Université Lyon 1 (France)*

We consider various specifications of the general discrete-time risk model in which a serial dependence structure is introduced between the claim numbers for each period. We consider risk models based on compound distributions assuming several examples of discrete variate time series as specific temporal dependence structures: Poisson MA(1) process, Poisson AR(1) process, Markov Bernoulli process and Markov regime-switching process. In these models, we derive expressions for a function that allows us to find the Lundberg coefficient. Specific cases for which an explicit expression can be found for the Lundberg coefficient are presented. Numerical examples are provided to illustrate different topics discussed in the paper.

Nous considérons différentes spécifications du modèle général de risque en temps discret dans lesquelles la structure de dépendance temporelle est introduite par le nombre de sinistres entre les périodes. Les modèles de risque traités sont basés sur différents exemples de séries temporelles pour variables aléatoires de dénombrement : Poisson MA(1), Poisson AR(1), processus Markov-Bernoulli et processus avec changement de régime markovien. Dans ces modèles, on identifie les expressions permettant de trouver le coefficient de Lundberg. Nous présentons des cas particuliers pour lesquels des expressions analytiques du coefficient sont obtenues. Des exemples numériques illustrant les résultats développés dans cet article sont traités.

[MS-239]

#### **TVaR-based Capital Allocation with Dependence**

#### **Allocation de capital basé sur la mesure TVaR en supposant la dépendance**

Étienne MARCEAU, Hélène COSSETTE & Mélina MAILHOT, *Université Laval*

Because of regulation projects from control organizations such as the European solvency II reform and recent economic events, insurance companies need to consolidate their capital reserve with coherent amounts allocated to the whole company and to each line of business. The present study considers an insurance portfolio consisting of several dependent risks and aims to evaluate not only the capital allocation for the overall portfolio but also the contribution of each risk over their aggregation.

L’objectif de l’article est de considérer des portefeuilles de risques dépendants et d’évaluer le capital alloué pour l’ensemble du portefeuille à l’aide de la mesure TVaR (*Tail Value at Risk*) et la contribution de chaque risque à ce capital en utilisant la règle basée sur la TVaR. Nous considérons différents modèles de risque : lois composées multivariées, modèles basés sur les copules et modèles basés sur les mélanges. Nous présentons des cas particuliers pour lesquels des expressions exactes sont obtenues. Nous suggérons aussi une méthode d’approximation. Des exemples numé-

We use the Tail Value at Risk (TVaR) as risk measure. We consider the following dependence models: multivariate compound distributions, models based on common mixtures and models based on copulas. We present special cases where exact expressions for the TVaR of the sum of the risks and for the TVaR-based allocations. We then propose numerical methods applicable for any proposed dependence models in order to approximate the TVaR of the aggregate risk for the portfolio and the contribution of each risk of the portfolio. Several numerical examples are presented in order to illustrate the topics exposed in the paper. An important aspect of the paper is to provide tools for practical applications.

riques sont aussi présentés.

**[MS-240]**

**Archimedean Dependence and Beyond  
Au-delà de la dépendance archimédienne**

Johanna NEŠLEHOVÁ, *McGill University*

Archimedean copulas are popular in many fields, including actuarial science. Recently, McNeil & Nešlehová (2009) have shown that these copulas are the survival copulas of the  $d$ -dimensional simplex distributions. In this talk, I will use this representation to show how the extremal behaviour of Archimedean copulas is related to the tail behaviour of the radial part of the corresponding simplex distribution. I will show how to use these results to build models with specific tail behaviour and I will outline implications for the aggregation properties of risk measures like VaR. Finally, I will present a non-exchangeable extension of Archimedean copulas.

Les copules archimédiennes sont populaires dans divers domaines, dont l'actuariat. Récemment, McNeil & Nešlehová (2009) ont montré que ces copules sont les copules de survie des lois simpliciales  $d$ -variées. Dans cet exposé, j'utiliserai cette représentation pour montrer comment le comportement extrémal des copules archimédiennes est lié au comportement caudal de la partie radiale de la loi simpliciale correspondante. Je montrerai comment exploiter ces résultats pour construire des modèles ayant un comportement caudal spécifique et j'en tirerai des conséquences sur les propriétés d'agrégation de mesures du risque dont la VaR. Je présenterai en outre une généralisation non échangeable des copules archimédiennes.

**Session 12E Mercredi 26 mai • Wednesday, May 26, 10:30 – 12:00 VAN 2811A**

**Genomics and Association Studies  
Génomique et études d'association**

**[MS-241]**

**Resolving Isoform Expression using RNA-seq data  
Résoudre l'expression isoforme au moyen de données d'ARN-seq**

Naomi ALTMAN, Qingyu WANG, Vishesh KARWA & Aleksandra SLAVKOVIC, *Penn State University*

The advent of ultra-high throughput sequencing technology makes it possible to directly sequence

Les technologies de séquençage à ultrahaut débit permettent de directement séquençer un grand nombre de fragments

large numbers of RNA fragments, directly measuring RNA abundance. Genes in complex organisms may encode multiple mRNAs, called isoforms, which result in different proteins. Isoform expression is often more relevant than gene expression. Restriction enzyme fragmentation reliably produces a single sequenced fragment per RNA. However, since each restriction site can belong to multiple isoforms, the counts must be partitioned among the isoforms to derive accurate isoform counts. We combine a model for fragmentation with methodology for contingency table analysis to estimate isoform abundance.

d'ARN, mesurant ainsi directement l'abondance d'ARN. Chez les organismes complexes, des gènes dits isoformes peuvent encoder de multiples ARNm résultant en des protéines différentes. L'expression isoforme est souvent plus importante que l'expression génétique. La fragmentation par enzymes de restriction produit de façon fiable un seul fragment séquencé par ARN. Cependant, puisque chaque site de restriction peut appartenir à de multiples isoformes, les totaux doivent être partitionnés entre les isoformes afin d'obtenir des totaux d'isoformes précis. Nous combinons un modèle pour fragmentation à des méthodes d'analyse de tables de contingence afin d'estimer l'abondance d'isoformes.

[MS-242]

### **Gene Selection for Microarray Data via Hilbert–Schmidt Independence Criterion**

### **Sélection de gènes pour données de micropuces au moyen du critère d'indépendance d'Hilbert-Schmidt**

Ali GHODSI & Hadi ZARKOOB, *University of Waterloo*

A novel variable selection technique for microarray gene expression data is proposed. It is based on the Hilbert-Schmidt independence criterion, and partly motivated by the Singular Value Decomposition (SVD). The algorithm selects a small set of genes such that the response variable depends mainly on this subset, at the exclusion of the rest of the genes. The algorithm is computationally very fast and scalable to large data sets, and it does not require the number of important genes as an explicit input. Experimental results of the proposed technique are presented on some synthetic and well-known microarray data sets.

Une nouvelle technique de sélection de variables pour données d'expression génétique de micropuces est proposée. Elle est basée sur le critère d'indépendance d'Hilbert-Schmidt et est partiellement motivée par la décomposition en valeurs singulières. L'algorithme sélectionne un petit ensemble de gènes dont la variable réponse dépend principalement. L'algorithme est très rapide et extensible à de grands jeux de données, et il ne nécessite pas de spécifier a priori le nombre de gènes importants. La technique proposée est appliquée à quelques jeux de données de micropuces réels bien connus et à des données synthétiques.

[MS-243]

### **Assessing the Effects of Haplotype Match for Competing Risks with Unphased Genotype Data**

### **Évaluer les effets de la compatibilité haplotype sur les risques compétitifs avec des données génotypiques déphasées**

Kyle RAYMOND & Thomas SCHEIKE, *University of Copenhagen*

A patient's risk of relapse or death following a bone marrow transplant from an unrelated donor is influenced by the degree of HLA matching between the donor and patient. Matching is typically based on unphased genotype data, consisting of their alleles at the HLA A, B and DR1 loci. However, there is interest in assessing the effect of haplotype matching between patients and donors, typically requir-

Le risque de rechute ou de décès d'un patient suite à une greffe de moelle osseuse d'un donneur non parent est influencé par le degré de compatibilité HLA entre le donneur et le patient. La compatibilité est habituellement basée sur des données génotypiques déphasées, soient leurs allèles aux loci HLA A, B et DR1. Cependant, il y a un intérêt dans l'évaluation de la compatibilité haplotype entre patients et donneurs, nécessitant des données génotypiques en phase.

ing phased genotype data. In this talk, we discuss the implementation of an EM algorithm in order to assess the effect of haplotype matching in a competing risk set-up with unphased genotype data.

Dans cet exposé, nous discutons de l'implémentation d'un algorithme EM afin d'évaluer l'effet de la compatibilité haplotype dans un contexte de risques compétitifs avec données génotypiques déphasées.

**[MS-244]**

**Predicting of Survival Time by Combining Mismeasured Gene Expression Data from Different Platforms  
Prévoir le temps de survie en combinant des données d'expression génétique mal mesurées provenant de différentes plateformes**

Juan XIONG & Wenqing HE, *University of Western Ontario*

With recent development of microarray technology, combining microarray data from different platforms is of practical importance for survival analysis of microarray data. Microarray technology is essentially a measurement tool for biological features such as gene expression, and the measured biological features are subject to measurement error. We show a prediction AFT model using data with mismeasured covariates. Two variations of the SIMEX algorithm are investigated to adjust the effect of the mismeasured covariates, and a best linear prediction is employed to estimate the true covariates from different platforms. Simulation results show that the proposed method can achieve better prediction accuracy.

Avec le développement récent de la technologie de micropuces, la combinaison de données de micropuces de différentes plateformes est d'intérêt pratique en analyse de survie de données de micropuces. Les micropuces sont essentiellement un outil de mesure pour des caractéristiques biologiques telles l'expression génétique. Ces caractéristiques sont sujettes à des erreurs de mesure. Nous montrons un modèle de prévision à temps de défaillance accéléré utilisant des données avec covariables mal mesurées. Deux variations de l'algorithme SIMEX sont étudiées afin d'ajuster l'effet des covariables mal mesurées, et la meilleure prévision linéaire est utilisée afin d'estimer les vraies covariables des différentes plateformes. Des résultats de simulation montrent que la méthode proposée peut atteindre une meilleure précision prévisionnelle.

**[MS-245]**

**Latent Class Model with Familial Dependence to Address Heterogeneity in Complex Diseases: Adapting the Approach to Family-based Association Studies  
Modèle de classes latentes avec dépendance familiale pour surmonter l'hétérogénéité des maladies complexes : adaptation de l'approche aux études d'association dans les familles**

Alexandre BUREAU, *Université Laval*, Jordie CROTEAU, *Centre de recherche Université Laval – Robert-Giffard*, Aurélie LABBE, *McGill University*, Chantal MÉRETTE, *Université Laval*

We have previously developed a latent class model allowing dependence between the latent class status of relatives within families. When applied to disease symptoms, the model succeeds at defining more genetically homogeneous disease sub-types, and genetic variants causing these sub-types can be identified using family-based association tests (FBAT). We propose to make the contribution of a subject to the FBAT statistic proportional to his posterior class membership probability. Simulations show a modest but robust power advantage compared to simply assigning each subject to his most probable class. The approach is illustrated on families from the Autism Genetics Research Exchange.

Nous avons précédemment développé un modèle de classes latentes permettant que la classe latente d'un sujet dépende de celle de ses parents. Lorsqu'appliqué à des symptômes de maladie, le modèle réussit à définir des sous-types de maladie plus homogènes, et les variantes génétiques causant ces sous-types peuvent être identifiées à l'aide de tests d'association dans les familles. Nous proposons de rendre la contribution d'un sujet à la statistique de test proportionnelle à sa probabilité *a posteriori* d'appartenance à la classe. Des simulations démontrent un gain de puissance modeste mais robuste comparé à la simple attribution de chaque sujet à sa classe la plus probable. L'approche est illustrée sur des familles du *Autism Genetics Research Exchange*.

**[MS-246]****Semiparametric Mixed Models for Multivariate Profile Traits in Genome-wide Association Studies of Complex Diseases: An Application to Lipid Density Gradient Profiles****Modèles mixtes semi-paramétriques pour l'analyse de profils multivariés dans les études d'association pangénomiques de maladies complexes : une application aux profils lipidiques**

Marie-Pierre SYLVESTRE, *Samuel Lunenfeld Research Institute*, Angelo J. CANTY, *McMaster University*, Daryl WAGGOTT, *Samuel Lunenfeld Research Institute*, John D. BRUNZELL, *University of Washington*, Andrew P. BORIGHT, Andrew D. PATERSON & Shelley BULL, *University of Toronto*

We are interested in methods for genome-wide association testing of multivariate or functional traits, specifically lipid density gradient profiles. Each individual lipid profile consists of 25 correlated values that need to be modelled simultaneously. We design semiparametric mixed models that provide a smooth estimate of the lipid profiles corresponding to each of the genotypic groups defined by the single nucleotide polymorphisms under study. We model the profiles using penalized splines taking advantage of their mixed model representation to estimate the level of smoothing required. We present inference strategies based on permutation tests to quantify the strength of the association.

Nous nous intéressons aux méthodes pour les études pangénomiques dont les traits sont multivariés ou fonctionnels, spécifiquement les profils lipidiques. Chaque profil individuel consiste en 25 valeurs corrélées qui doivent être modélisées conjointement. Nous construisons des modèles semi-paramétriques produisant des estimés lisses des profils lipidiques correspondant à chacun des groupes génotypiques définis par les polymorphismes liés à un nucléotide à l'étude. Nous modélisons les profils avec des splines pénalisées et utilisons leur représentation en modèles mixtes pour estimer le degré de lissage. Nous présentons des stratégies d'inférence basées sur des tests de permutation pour quantifier la force des associations.

**Session 12F Mercredi 26 mai • Wednesday, May 26, 10:30 – 11:45 VAN 1203****Probability****Probabilité****[MS-247]****On the Number of Runs for Bernoulli Arrays****La distribution du nombre de suites de succès de longueur deux associées à un tableau de variables aléatoires Bernoulli**

Djilali AIT AOUDIA, *Université de Montréal*, Éric MARCHAND, *Université de Sherbrooke*

We consider the study of  $(n + 1) \times r$  arrays  $T$  with Bernoulli random variables, such that the random vectors  $\underline{X}_k = (X_{k,1}, \dots, X_{k,r})'$  are independent. We study the distribution of  $S_n = \sum_{k=1}^n \underline{X}'_k \underline{X}_{k+1}$  representing the number of consecutive pairs of successes when reading the array down the columns and across the rows. The case  $r = 1$  has been studied by several authors. We obtain various distribution properties and representations of  $S_n$  for the case  $r = 2$ . Applications are also given in cases where the array  $T$  arises from a Pólya sampling scheme.

Nous considérons un tableau  $T = \{X_{k,j}\}_{1 \leq k \leq n+1, 1 \leq j \leq r}$  de variables aléatoires de Bernoulli tels que les vecteurs aléatoires  $\underline{X}_k = (X_{k,1}, \dots, X_{k,r})'$  sont indépendants. Nous nous intéressons à la distribution de  $S_n = \sum_{k=1}^n \underline{X}'_k \underline{X}_{k+1}$  qui définit le nombre de suites de succès consécutifs en descendant les colonnes et en traversant les rangées. Le cas  $r = 1$  a déjà été étudié par plusieurs auteurs. Pour le cas  $r = 2$ , une représentation de la distribution de  $S_n$  est donnée, ainsi que des applications pour des modèles de Bernoulli échangeables, dont celui du modèle de l'urne de Pólya.

**[MS-248]****Exact and Approximate Expressions of the Kolmogorov and Lévy Distances between a Dirichlet Process and Another Continuous Probability Measure with some Applications****Expressions exactes et approximatives des distances de Kolmogorov et de Lévy entre un processus de Dirichlet et une autre mesure de probabilité continue, avec quelques applications**Luai AL LABADI, *University of Ottawa*, Mahmoud ZAREPOUR, *Associate Professor*

In nonparametric Bayesian inference we require the construction of a prior on an infinite dimensional space. A well-known prior, which is a stochastic process whose sample path are probability measures, is the Dirichlet process introduced by Ferguson (1973). The Dirichlet process has different representations (exact and approximate). In this paper, we derived expressions (exact and approximate) for the Kolmogorov and Lévy distances between the Dirichlet process and another continuous probability measure. These expressions were used in a goodness-of-fit test. In particular, we examine if the concentration parameters used in the Dirichlet process are reasonable or not.

En inférence non paramétrique bayésienne, il est nécessaire de construire une loi a priori sur un espace de dimension infinie. Une loi *a priori* bien connue est le processus de Dirichlet, présenté par Ferguson (1973) ; il s'agit d'un processus stochastique dont les trajectoires sont des mesures de probabilité. Le processus de Dirichlet possède différentes représentations (exactes et approximatives). Dans cette communication, nous obtenons des expressions (exactes et approximatives) pour les distances de Kolmogorov et de Lévy entre le processus de Dirichlet et une autre mesure de probabilité continue. Ces expressions ont été utilisées pour tester l'adéquation. En particulier, nous examinons si les paramètres de concentration utilisés dans le processus de Dirichlet sont raisonnables ou non.

**[MS-249]****Fixation Probability of a Gamete under Weak Selection and the Effect of Recombination Modifiers on the Probability of Fixation****Probabilité de fixation d'un gamète sous sélection faible et effet de modificateurs du taux de recombinaison**Amir KERMANY & Sabin LESSARD, *Université de Montréal*

In this talk we discuss the probability of fixation of a gamete in a finite haploid population with emphasis on a three-locus model in which one locus controls the rate of recombination between the two other loci. In the limit of weak selection, allowing for gene interaction effects, the fixation probability is derived as the first order perturbed version of the neutral case, expressed in terms of expected coalescence times and initial gamete frequencies in the population. This is applied to gametes carrying a recombination modifier allele in order to study the evolution of recombination.

Nous étudions la probabilité de fixation d'un gamète dans une population haploïde de taille finie en mettant l'accent sur un modèle à trois loci dans lequel un locus contrôle le taux de recombinaison entre les deux autres loci qui sont sous sélection. Nous obtenons une approximation du premier ordre par rapport à l'intensité de sélection, qui est exprimée à l'aide des temps de coalescence moyens et des fréquences gamétiques initiales. On applique la procédure à l'évolution de la recombinaison.

**[MS-250]****Statistical Assessment of Models for Powder Flow****Évaluation statistique de modèles d'écoulement des poudres**Jeffrey PICKA, *University of New Brunswick*

Discrete element method (DEM) models are widely used as models for the flow of powders. These powder flows are stochastic phenomena, yet are

Les modèles à méthode à éléments discrets (MED) sont fréquemment utilisés afin de modéliser l'écoulement de poudres. Ces écoulements sont des phénomènes stochas-

generally modelled with deterministic models that are based on unverifiable assumptions. Further, the quality of the fit of these models is rarely assessed and most assessments are not capable of identifying model failings. By adding stochastic elements to the DEM models and by recognizing that the assessment process requires spatial statistical methods, a strategy for objectively assessing DEM models can be developed.

tiques, mais ils sont généralement modélisés de façon déterministe, à partir d'hypothèses invérifiables. De plus, l'adéquation de ces modèles est rarement évaluée et la plupart des évaluations ne peuvent identifier les failles de ces modèles. En ajoutant des composantes stochastiques aux modèles MED et en reconnaissant que le processus d'évaluation nécessite des méthodes de statistique spatiale, une stratégie d'évaluation objective de modèles MED peut être développée.

#### [MS-251]

##### **A Time-dependent Version of Pólya Urn**

##### **Une version dépendante du temps de l'urne de Pólya**

Djilali AIT AODIA & François PERRON, *Université de Montréal*

Permantle (1990) generalized the Pólya urn model with the single change that the number of extra balls added of the colour drawn is a function of time. A ball is drawn and replaced in the urn along with  $F(n)$  balls of the same colour with  $F: \{1, 2, \dots\} \rightarrow [0, +\infty)$  be any function. He showed that the proportion of white balls converges almost surely and the limit has no atom except possibly at 0 or 1. We consider this case of time dependent version of Pólya urn and we find various distributional properties of the proportion of white balls.

Permantle (1990) a généralisé le modèle d'urne de Pólya en faisant en sorte que le nombre de boules supplémentaires de la couleur tirée soit une fonction du temps. Une boule est tirée et remplacée dans l'urne avec  $F(n)$  boules de la même couleur, où  $F: \{1, 2, \dots\} \rightarrow [0, +\infty)$  est une fonction arbitraire. Il a montré que la proportion de boules blanches converge presque sûrement et que la limite n'a pas d'atomes, sauf peut être à 0 ou à 1. Nous considérons ce cas d'urne de Pólya dépendante du temps et nous obtenons diverses propriétés concernant la loi de la proportion de boules blanches.

## **Session 12G    Mercredi 26 mai • Wednesday, May 26, 10:30 – 11:45    VAN 1211**

### **Estimation in Survey Sampling**

### **Estimation en échantillonnage**

#### [MS-252]

##### **Calendarization Methods for Reference Period Adjustments with Applications to Monthly Surveys**

##### **Méthodes de calendarisation pour ajustements à la période de référence avec applications aux enquêtes mensuelles**

Catalin DOCHITOIU & Stuart PURSEY, *Statistics Canada*

In the context of surveys, calendarization is a step that requires assumptions for modeling micro record values. Two approaches to this modelling exercise are examined—one already implemented and one proposed. We examine the rationale behind approaches, and numerical comparisons using two monthly surveys at Statistics Canada: the Monthly Survey of Manufacturers, and the Monthly Wholesale and Retail Survey. The

Dans le cadre d'une enquête, la calendarisation est une étape de traitement des données qui nécessite certaines hypothèses afin de modéliser les valeurs des enregistrements. Deux stratégies de modélisation sont examinées — une qui est utilisée en production et l'autre qui est envisagée. Nous examinons les aspects théoriques, ainsi que des comparaisons numériques utilisant les données de deux enquêtes mensuelles à Statistique Canada : l'Enquête mensuelle sur les manufacturiers et l'Enquête mensuelle sur le commerce de gros et de

results show the impact of changing from one modelling method to the other, illustrating the conceptual differences between models. After validating the new approach, the differences found are useful in planning a strategy for phasing in the new method.

détail. Les résultats montrent l'effet du changement d'une méthode à l'autre, en illustrant les différences conceptuelles dans les modèles. Après avoir validé la nouvelle stratégie, ces différences seront utiles pour chercher une stratégie afin d'introduire de la nouvelle méthode.

**[MS-253]**

**Robust Small Area Estimation  
Estimation robuste pour petites régions**

Valery DONGMO JIONGO, David HAZIZA & Pierre DUCHESNE, *Université de Montréal*

Traditional area-specific direct estimation methods are not suitable in the small area context because of low small area specific-sample size. But indirect Empirical Best Linear Unbiased Predictor (EBLUP) presented by Rao (2003) are sensitive to the presence of outlying units. To reduce the influence of the extreme units Rao & Sinhan (2009) develop a robust version of the EBLUP estimators. We propose alternative robust estimators which have in certain conditions desirable properties in term of bias and which are competitive for the means square error. These estimators are motivated by the concept of conditional bias of a unit studied by Beaumont, Haziza & Ruiz-Gazen (2009). Results of a simulation study will be presented.

Les estimateurs directs classiques des petits domaines ne sont pas suffisamment précis en raison de taille d'échantillon faible. En présence de valeurs aberrantes, la variabilité des estimateurs linéaires sans biais optimaux (EBLUP) présentés dans Rao (2003) augmente considérablement. Des versions robustes sont développées par Rao & Sinha (2009). Nous proposons des estimateurs robustes alternatifs, possédant sous certaines conditions des caractéristiques désirables de biais et qui sont compétitifs quant à l'erreur quadratique moyenne. Ces estimateurs sont motivés à partir du concept de biais conditionnel d'une unité, Beaumont, Haziza & Ruiz-Gazen (2009). Les résultats d'une étude par simulation sont présentés.

**[MS-254]**

**Small Area Estimation under Informative Sampling  
Estimation pour petites régions sous un plan de sondage informatif**

M. HIDIROGLOU & François VERRET, *Statistics Canada*, J. N. K. RAO, *Carleton University*

Population unit level models are often used in model-based small area estimation of totals and means. These models may not hold for the sample if the sampling design is informative. As a result, standard methods, assuming the model holds for the sample, can lead to biased estimators. We propose to study alternative methods that use survey weight as additional auxiliary variable in the sampling model. We report the results of a simulation study on the bias of the proposed estimators and MSE estimators, using informative sampling schemes to generate the samples.

Les modèles au niveau des unités de population sont souvent utilisés en estimation à base de modèles de totaux et de moyennes pour petits domaines. Ces modèles peuvent ne pas être applicables à l'échantillon si le plan d'échantillonnage est informatif. Les méthodes habituelles, supposant que le modèle est approprié pour l'échantillon, peuvent donc mener à des estimateurs biaisés. Nous proposons d'étudier des méthodes alternatives utilisant les poids de sondage comme variable auxiliaire supplémentaire dans le modèle d'échantillonnage. Nous présentons les résultats d'une étude de simulation du biais des estimateurs proposés et des estimateurs EQM, lorsque les échantillons sont obtenus à partir de plans informatifs.

**[MS-255]****Penalized Likelihood Methods for Variable Selection in Analysis of Survey Data****Méthodes de sélection de variables basées sur la vraisemblance pénalisée en analyse de données d'enquête**Chen Xu, *University of British Columbia*

Variable selection plays an important role in the analysis of survey data, where one common issue of interest is to identify the influential factors that are associated with certain behaviour, social or economical indices. Traditional selection procedures, such as the best subset selection or the stepwise regression, can be computationally expensive or unstable in the selection process. Instead, the penalized likelihood methods (PLM) are now being used as computational feasible alternatives for variable selection. In this work, we extend the use of PLM in the context of surveys and discuss its corresponding asymptotic behaviours.

La sélection de variables joue un rôle important en analyse de données d'enquête, où un point d'intérêt fréquent est l'identification de facteurs influents associés à certains indices comportementaux, sociaux ou économiques. Les méthodes de sélection traditionnelles, telles la sélection du meilleur sous-ensemble ou la régression pas-à-pas, peuvent demander une grande puissance de calcul ou mener à une sélection instable. Les méthodes à vraisemblance pénalisée (MVP) sont maintenant utilisées comme alternative moins lourde pour la sélection de variables. Dans ce travail, nous généralisons l'utilisation de MVP au contexte d'enquêtes et discutons de son comportement asymptotique.

**[MS-256]****Duration Analysis of Survey Data using Inverse Probability of Censoring Weights (IPCW)****Analyse de durée pour données d'enquête au moyen de la pondération par probabilité inverse de censure**Dagmar MARIACA HAJDUCEK & Jerald F. LAWLESS, *University of Waterloo*

In duration analysis, ignoring the possibility that the event of interest may be related to the censoring mechanism may result in biased estimation. Inverse probability of censoring weights (IPCW) methods can be used and the randomness of the weights needs to be considered when performing variance estimation. IPCW estimation methods that accommodate for weights based on the sampling design are discussed. Weighted Kaplan–Meier estimates of a finite population survival distribution are examined in a simulation study, where effects of dependent censoring and of the sampling design are evaluated. The methods are illustrated with data on jobless spells.

En analyse de durée, ignorer la possibilité que l'évènement d'intérêt puisse être lié au mécanisme de censure peut biaiser l'estimation. Des méthodes de pondération par l'inverse des probabilités de censure (PIPC) peuvent être utilisées ; le caractère aléatoire des poids doit alors être considéré en estimation de variance. Les méthodes d'estimation PIPC tenant compte de poids basés sur un plan échantillonnal sont abordées. Les estimateurs de Kaplan-Meier pondérés de la distribution de survie d'une population finie sont examinés grâce à une étude de simulation, où les effets d'une censure dépendante et du plan échantillonnal sont évalués. Les méthodes sont illustrées avec des données de durée de périodes de chômage.

**Session 12H Mercredi 26 mai • Wednesday, May 26, 10:30 – 12:00 VAN 1871**

## Times Series and Climate

## Séries chronologiques et le climat

[MS-257]

**Trend Estimation by Means of the Empirical Mode Decomposition****Estimation de tendances par les moyennes de la décomposition des modes empiriques**Azadeh MOGHADERI, Patrick FLANDRIN & Pierre BORGNAT, *Ecole Normale Supérieure de Lyon, France*

We consider a non-stationary time series which is the superposition of a low-frequency trend component and a broad-band noise component. Provided the two components meet certain regularity conditions, we develop a nonparametric technique to estimate the trend component. Our technique makes fundamental use of the empirical mode decomposition. To illustrate its effectiveness, we apply it to simulated time series containing different types of additive and multiplicative trend components, as well as environmental and financial time series. Finally, we indicate how an estimate of the trend component can be used to improve estimates of Wold–Cramer evolutionary spectra.

Nous considérons une série chronologique non stationnaire composée en une superposition d'une composante évolutive de basse fréquence et d'une composante de bruit à large bande. Nous développons une technique d'estimation de tendance non paramétrique lorsque les deux composantes respectent certaines conditions de régularité. Notre technique s'appuie de façon fondamentale sur la décomposition en modes empiriques. Afin d'illustrer son efficacité, nous l'appliquons à des séries chronologiques simulées contenant différents types de composantes évolutives additives et multiplicatives, ainsi qu'à des séries chronologiques environnementales et financières. Finalement, nous indiquons comment un estimateur de la composante évolutive peut être utilisé afin d'améliorer l'estimateur du spectre évolutif de Wold-Cramer.

[MS-258]

**Nonparametric Estimation for Causality Measures****Estimation non paramétrique des mesures de causalité**Taoufik BOUEZMARNI, *McGill University*, Abderrahim TAAMOUTI, *Universidad Carlos III de Madrid*

This paper aims to propose a nonparametric approach to estimate (and make inference about) the causality measures. The estimation is model-free and allows to capture and quantify both linear and nonlinear causalities. To test and build confidence intervals around the estimated causality measures, we show that the nonparametric estimator of measure is asymptotically normally distributed. Our simulation results show that the test has good properties (size and power). We also illustrate the practical relevance of nonparametric causality measures by considering empirical applications where we quantify the strength of causality between S&P500 Index returns and many other financial variables.

Ce travail propose une approche non paramétrique pour estimer (et faire de l'inférence sur) la mesure de causalité. L'estimateur ne dépend pas d'un modèle et permet de détecter et quantifier les causalités linéaires et non linéaires. Pour tester et construire les intervalles de confiance pour les mesures de causalité, nous démontrons la normalité asymptotique de l'estimateur non paramétrique. Les résultats de simulation indiquent que le test est efficace (niveau du test et puissance). Nous illustrons l'apport pratique des mesures de causalité non paramétriques et l'application de l'estimateur par une étude des indices du S&P500 et de d'autres données financières.

**[MS-259]****Multitaper Spectral Estimation Package in R  
Package R d'estimation spectrale à fenêtres de pondération multiples**Karim RAHIM & David THOMSON, *Queen's University*

In time series analysis, the periodogram is often inappropriately used to estimate the spectral density. The multitaper spectral estimation can compensate for bias better than other spectral estimation techniques, it can provide more than two degrees of freedom thus admitting narrower confidence intervals, it can be used with jackknifing to provide an internal variance estimate, and it admits a statistical test for the presence of sinusoidal (line) components. While implemented in Matlab, the technique is not generally available in statistical software. We present a simple R package that makes simple and more advanced multitaper techniques available for use by statisticians.

En analyse de séries chronologiques, le périodogramme est souvent utilisé de façon inappropriée afin d'estimer la densité spectrale. L'estimation à fenêtres de pondération multiples peut compenser le biais plus efficacement que les autres techniques d'estimation spectrale, peut apporter plus de deux degrés de liberté, et donc des intervalles de confiance plus petits, peut être utilisée avec le jackknife afin de fournir un estimateur de variance interne, et permet un test statistique pour la présence de composantes sinusoïdales. Bien qu'implémentée dans Matlab, cette technique n'est pas généralement présente dans les logiciels statistiques. Nous présentons un package R simple permettant l'utilisation de techniques à fenêtres de pondération multiples simples et complexes.

**[MS-260]****A Graphical Test for Stationary and Uniformly Modulated Processes  
Un test graphique pour processus stationnaires et uniformément modulés**Azadeh MOGHADERI, *Ecole Normale Supérieure de Lyon*, Glen TAKAHARA & David THOMSON, *Queen's University*

Given a realization of a finite-variance nonstationary stochastic process, we propose a heuristic graphical technique to test whether or not the process is stationary. Our technique can also be used to distinguish a particular class of nonstationary processes, known as uniformly modulated processes, from other types of nonstationarity. We demonstrate the performance of our technique via simulations. We also propose a test statistic based on the two by two cofactors of an estimated time frequency matrix obtained from an estimate of the evolutionary spectrum of a Wold–Cramer process, and investigate its distribution under the null hypothesis of stationarity.

À partir d'une réalisation d'un processus stochastique non stationnaire à variance finie, nous proposons une technique graphique heuristique afin de tester la stationnarité du processus. Notre technique peut aussi être utilisée afin de détecter une classe particulière de processus non stationnaires dits processus à modulation uniforme. Nous montrons la performance de notre technique au moyen de simulations. Nous proposons aussi une statistique de test basé sur les cofacteurs deux par deux d'une matrice de fréquence estimée obtenue par estimation du spectre évolutif Wold-Cramer, et étudions sa distribution sous l'hypothèse nulle de la stationnarité.

**[MS-261]****Probabilistic Weather Forecasting using the Geostatistical Output Perturbation Method on Hierarchical Bayesian Space-time Models  
Prévisions météorologiques probabilistes utilisant la méthode de perturbation géostatistique de modèles espace-temps bayésiens hiérarchiques**Lilia RAMIREZ RAMIREZ, Ziming HUANG & Yulia GEL, *University of Waterloo*

Nowadays probabilistic weather forecasting has become increasingly popular among meteorolo-

Les prévisions météorologiques probabilistes deviennent de plus en plus populaires auprès des météorologistes et des

gists and weather users as it enables to produce predictive intervals for a future atmospheric quantity with a prespecified confidence level rather than a single point forecast, which is of particular importance for an accurate and reliable assessment of weather related uncertainties for risk management tasks. However, statistical weather scenarios typically focus on capturing either a spatial or temporal dynamic. In order to introduce more general non-stationary spatio-temporal effects, we propose to employ a hierarchical Bayesian model and we illustrate its application by constructing probabilistic scenarios of temperature in the North American Pacific Northwest.

usagers des prévisions puisqu'elles permettent de produire des intervalles de prévision pour une quantité atmosphérique future d'un niveau de confiance donné plutôt qu'une simple prévision ponctuelle. Ceci est particulièrement important pour évaluer de façon fiable et précise des incertitudes liées à la météorologie en gestion de risques. Cependant, les scénarios météorologiques statistiques habituels ne sont centrés que sur une dynamique spatiale ou sur une dynamique temporelle. Pour traiter de façon générale des effets spatio-temporels non stationnaires, nous proposons d'utiliser un modèle hiérarchique bayésien et nous illustrons son application en construisant des scénarios probabilistes de la température dans le Nord-Ouest Pacifique.

[MS-262]

**Edgeworth Expansions for Analyzing Frequency Modulation in Climate Time Series**  
**Développements d'Edgeworth afin d'analyser la modulation de fréquences en séries chronologiques climatologiques**

David THOMSON, *Queen's University*

A challenge in analyzing Holocene climate data are the intermediate strength components that are not easily represented by traditional time-series models. These are non-stationary, approximately periodic components that vary in both amplitude and frequency, and which recur, with variations, possibly as a response to solar forcing. They are not simple additive effects but, for example, change the timing of the annual cycle. We begin with a multitaper time-frequency analysis, expand the Slepian sequences in Hermite functions, then apply classical results relating the coefficients of an Edgeworth expansion and cumulants to make inferences about these embedded components.

Un défi dans l'analyse de données climatiques de l'Holocène est que les composantes de forces intermédiaires ne sont pas aisément représentées par les modèles de séries chronologiques traditionnels. Il s'agit de composantes non stationnaires et approximativement périodiques d'amplitude et de fréquence variables et récurrentes avec variations, possiblement en réponse au forçage solaire. Ce ne sont pas de simples effets additifs ; ils peuvent, par exemple, changer la cadence du cycle annuel. Nous commençons avec une analyse temps-fréquence à fenêtres de pondération multiples, généralisons les suites de Slepian dans les fonctions d'Hermite, et appliquons les résultats classiques liant les coefficients d'un développement d'Edgeworth aux cumulants afin de faire des inférences au sujet de ces composantes emboîtées.

**Session 13A Mercredi 26 mai • Wednesday, May 26, 13:20 – 14:20 POU 1112**

The Lasso

Le lasso

[MS-263]

**The Lasso: Some Novel Algorithms and Applications**  
**Le lasso : quelques algorithmes et applications originaux**

Robert TIBSHIRANI, *Stanford University*

I will discuss some procedures for modelling high-dimensional data, based on  $L_1$  (lasso)-style penal-

Je vais discuter de quelques procédures de modélisation de données de dimension élevée basées sur des pénalités de

ties. I will describe pathwise coordinate descent algorithms for the lasso, which are remarkably fast and facilitate application of the methods to very large datasets for the first time. I will then give examples of new applications of the methods to microarray classification, undirected graphical models for cell pathways, and the fused lasso for signal detection (including comparative genomic hybridization) and the matrix completion problem.

style  $L_1$  (lasso). Je vais décrire des algorithmes de descente de coordonnées chemin-par-chemin pour le lasso remarquablement rapides et facilitant la première application des méthodes à de très grands jeux de données. Je vais alors donner des exemples de nouvelles applications des méthodes à la classification de micropuces, aux modèles graphiques non orientés pour chemins de cellules, et au lasso combiné pour détection de signaux (incluant l'hybridation génomique comparative) et au problème de complétion de matrices.

## Session 14A Mercredi 26 mai • Wednesday, May 26, 14:30 – 16:00 VAN 1245A

### Response-biased Sampling and Missing Covariates

### Échantillonnage biaisé par les réponses et covariables manquantes

[MS-264]

#### A General Method to Analyze Data from Outcome Dependent Cluster Sampling Designs

#### Une méthode générale pour analyser des données de plans d'échantillonnage en grappes dépendantes de la réponse

John NEUHAUS, *University of California, San Francisco*, Alastair SCOTT & Chris WILD, *University of Auckland*, Charles McCULLOCH, *University of California, San Francisco*

Outcome dependent cluster sampling designs arise often in family studies in genetic epidemiology and in other settings with rare binary outcomes. Our previous work developed a profile/pseudo-likelihood approach that addresses the clustering and differential sampling rates that such designs feature. This talk shows that by augmenting the response of the profile/pseudo-likelihood approach, we can extend it to accommodate novel outcome dependent sampling designs including those with sampling dependent on surrogates for the outcome and longitudinal follow-up of an initial case-control sample. Data from a study of Attention Deficit Hyperactivity Disorder in children motivates the work and illustrates the findings.

Les plans d'échantillonnage par grappes dépendantes de la réponse sont souvent utilisés en épidémiologie génétique pour des études familiales, ainsi que dans d'autres contextes à réponse binaire rare. Nos travaux antérieurs ont développé une approche par pseudo-vraisemblance/vraisemblance de profil traitant les grappes et les différents taux d'échantillonnages entraînés par de tels plans. Cet exposé montre qu'en augmentant la réponse de l'approche de la pseudo-vraisemblance/vraisemblance de profil, nous pouvons la généraliser afin de traiter de nouveaux plans d'échantillonnage qui dépendent de la réponse, incluant ceux où l'échantillonnage dépend de substituts à la réponse et les suivis longitudinaux d'un échantillon cas-témoins initial. Les données d'une étude du trouble déficitaire de l'attention avec hyperactivité chez les enfants motivent et illustrent nos résultats.

[MS-265]

#### Efficient Estimation in Multiphase Case-control Studies

#### Estimation efficace en études cas-témoins à phases multiples

Chris WILD, Alan LEE & Alastair SCOTT, *University of Auckland*

We discuss the analysis of multi-phase, or multi-stage, case-control studies and present an efficient semiparametric maximum-likelihood approach that unifies and extends earlier work including Prentice & Pyke (1979), Breslow & Cain

Nous discutons de l'analyse d'études cas-témoins à phases, ou à niveaux multiples, et présentons une approche efficace par maximum de vraisemblance semi-paramétrique unifiant et généralisant des travaux antérieurs, dont ceux de Prentice & Pyke (1979), Breslow & Cain (1988), Scott

(1988), Scott & Wild (1997), and Breslow & Holubkov (1997). The theoretical derivations apply to arbitrary binary regression models but we present results for logistic regression and show that the approach can be implemented by including additional intercept terms in the logistic model and then making some simple corrections to the score and information equations used in a Newton–Raphson or Fisher-scoring maximization of the prospective log-likelihood.

& Wild (1997), et Breslow & Holubkov (1997). Les résultats théoriques s'appliquent à des modèles de régression binaires arbitraires, mais nous présentons les résultats pour la régression logistique et montrons que l'approche peut être implémentée en incluant des termes d'ordonnées à l'origine additionnels au modèle logistique et en apportant ensuite de petites corrections aux équations de score et d'information utilisées pour maximiser la log-vraisemblance.

**[MS-266]**

**Estimating Treatment Effects for Episodic Interventions Based on Response-dependent Observation**

**Estimation des effets de traitement pour interventions épisodiques à partir d'observations dépendantes de la réponse**

Richard Cook, *University of Waterloo*

We consider the problem of comparing the effects of two therapeutic interventions administered in response to episodic conditions. The data arising in such settings appears like ordinary longitudinal data and it is commonplace for analyses to be based on related methods. When there is a relation between the response to the treatment and the number of episodes requiring treatment, this ignores the dependence between the observation and response process and can lead to biased estimates. We consider some particular models to illustrate the biases that result from naive analyses and use data from a recent migraine trial for illustration.

Nous considérons la comparaison des effets de deux interventions thérapeutiques administrées en réponse à des conditions épisodiques. Les données obtenues dans de tels contextes semblent être des données longitudinales ordinaires et les analyses sont souvent basées sur des méthodes liées. Lorsqu'il y a une relation entre la réponse au traitement et le nombre d'épisodes nécessitant un traitement, cette approche ignore la dépendance entre les processus d'observation et de réponse et peut mener à des estimateurs biaisés. Nous considérons quelques modèles spécifiques afin d'illustrer les biais pouvant résulter d'analyses naïves et utilisons des données d'un essai clinique en traitement de migraines à des fins d'illustration.

**Session 14B Mercredi 26 mai • Wednesday, May 26, 14:30 – 16:00 VAN 2289A**

**Computer Experiments**

**Expérimentation numérique**

**[MS-267]**

**Simultaneous Determination of Tuning and Calibration Parameters for Computer Experiments**

**Détermination simultanée de paramètres de réglage et de calibrage pour des expériences informatiques**

Thomas Santner, *Ohio State University*

Tuning and calibration are processes for improving the representativeness of a computer simulation code to physical data. This talk introduces a statistical methodology for simultaneously determining tuning and calibration parameters where data are available from both the code and associated physical experiment. Tuning parameters are set by min-

Le réglage et la calibrage sont des procédés d'amélioration de la représentativité d'un code de simulation par ordinateur pour des données physiques. Cet exposé présente une méthodologie statistique pour déterminer simultanément les paramètres de réglage et de calibrage lorsque les données sont disponibles à la fois à partir du code et d'une expérience physique. Les paramètres de réglage sont fixés par la minimisa-

imizing a discrepancy measure; the distribution of the calibration parameters is determined using a hierarchical Bayesian model. The proposed model views the output as a realization of a Gaussian stochastic process with hyperpriors. Our methodology is compared with alternative approaches and is illustrated with a biomechanical engineering application.

#### [MS-268]

### Sequential Optimization of a Computer Model: What's so Special about Gaussian Processes Anyway?

#### Optimisation séquentielle d'un modèle informatique : pourquoi les processus gaussiens ?

Hugh CHIPMAN, Weiwei WANG & Pritam RANJAN, *Acadia University*

We consider the problem of placing numerous turbines in a tidal system to maximize total power generated. Computer models are effective at predicting power, but prohibitively slow for direct optimization. To quickly approximate the computer model, we use a statistical model. The usually ubiquitous Gaussian process models are nowhere to be seen, however. Instead, an adaptive nonparametric regression model (BART) is used to deal with nasty nonstationarities in the response surface. By providing both point estimates and uncertainty bounds for prediction, BART can be plugged into conventional sequential design criteria to find optima with few function evaluations.

tion d'une mesure d'écart ; la distribution des paramètres de calibrage est déterminée en utilisant un modèle hiérarchique bayésien. Le modèle proposé voit la sortie comme une réalisation d'un processus stochastique gaussien avec des hyperparamètres. Notre méthodologie est comparée avec des approches alternatives et est illustrée par une application en ingénierie biomécanique.

Nous considérons le problème qui consiste à placer plusieurs turbines dans un système marémoteur afin d'en maximiser la production d'énergie. Les modèles informatiques prédisent efficacement la production, mais sont trop lents pour faire de l'optimisation directe. Nous utilisons un modèle statistique afin d'approximer le modèle informatique. Les omniprésents modèles à processus gaussiens n'y figurent pas. Plutôt, un modèle de régression non paramétrique adaptatif (BART) est utilisé afin de traiter les non-stationnarités dans la surface réponse. En fournissant à la fois des estimateurs ponctuels et des bornes d'incertitudes, BART peut être emboîté dans les critères d'analyse séquentielle conventionnelle afin de trouver des optimums avec peu d'évaluations de fonction.

#### [MS-269]

### Emulating the Nonlinear Matter Power Spectrum for the Universe

#### Émuler le spectre de puissance non linéaire de la matière pour l'univers

Earl LAWRENCE, Dave HIGDON, Katrin HEITMANN & Salman HABIB, *Los Alamos National Laboratory*, Martin WHITE, *University of California, Berkeley*, Christian WAGNER, *Astrophysikalisches Institut Potsdam*

Many of the most exciting questions in astrophysics and cosmology, including most observational probes of dark energy, rely on an understanding of the nonlinear regime of structure formation. In order to fully exploit the information available from this regime and to extract cosmological constraints, accurate theoretical predictions are needed. Currently, such predictions can only be obtained from costly, numerical simulations. This work is aimed at constructing an accurate calibration of the nonlinear mass power spectrum for a wide range of currently viable cosmological models, including dark energy. We use the Coyote Universe simulation suite which comprises nearly 1,000  $N$ -body simulations.

Plusieurs des questions les plus excitantes en astrophysique et en cosmologie, incluant la plupart des sondes observationnelles de l'énergie sombre, s'appuient sur une compréhension du régime non linéaire de la formation de structures. Afin de complètement exploiter l'information disponible à partir de ce régime et d'extraire les contraintes cosmologiques, des prévisions théoriques précises sont requises. De telles prévisions ne peuvent, en ce moment, qu'être obtenues grâce à des simulations numériques dispendieuses. Ce travail vise à construire un calage précis du spectre non linéaire de puissance de la matière pour une large gamme de modèles cosmologiques présentement viables, incluant l'énergie sombre. Nous utilisons la suite de simulation *Coyote Universe*, comprenant quelque 1 000 simulations à  $N$  corps.

[MS-270]

**Reinforcing Reification with Application to a Rainfall-runoff Computer Model****Renforcement de la réification avec applications à un modèle informatique de précipitation-débit**Leanna HOUSE, *Virginia Tech*

Deterministic computer models or simulators are often imperfect representations of physical systems and introduce a layer of uncertainty into model-based inferences that is hard to quantify. To formalize the use of expert judgement in assessing the uncertainty, Goldstein & Rougier (2008) propose a method called reification that relies on a hypothetical simulator known as a “reified” model. One criticism of reification is that only expert critiques can validate the judgements used to specify a reified model. For this paper, we create an artificial case study that allows us to validate reification using both data and expert judgment.

Les modèles informatiques déterministes, ou simulateurs, sont souvent des représentations imparfaites de systèmes physiques et introduisent un niveau d’incertitude difficilement quantifiable dans les inférences à base de modèles. Afin de formaliser l’utilisation d’avis d’experts pour évaluer l’incertitude, Goldstein & Rougier (2008) proposent une méthode dite de réification s’appuyant sur un hypothétique simulateur qualifié de « réifié ». Une critique de la réification est que seules les critiques d’experts peuvent valider les jugements utilisés pour spécifier un tel modèle. Pour cet article, nous créons une étude de cas artificielle permettant de valider la réification en utilisant à la fois des données et l’avis d’experts.

**Session 14C Mercredi 26 mai • Wednesday, May 26, 14:30 – 16:00 VAN 2291A****Network Sampling****Échantillonnage par réseau**

[MS-271]

**Capture-recapture Sampling and Indirect Sampling****L’échantillonnage capture-recapture et l’échantillonnage indirect**Pierre LAVALLÉE, *Statistique Canada*, Louis-Paul RIVEST, *Université Laval*

Capture-recapture Sampling is used to estimate the number of units in a population of unknown size. An initial sample is obtained and the sampled units are marked. A second sample is obtained independently, and marked units are recorded. We estimate the population size by the ratio of the product of the two sample sizes, over the number of marked units. We can generalize Capture-recapture Sampling to the case where the sampling frames are not the same as the target population, but still related to the latter. Indirect Sampling is then used.

L’échantillonnage par capture-recapture est une méthode utilisée pour estimer le nombre d’unités dans une population de taille inconnue. Un échantillon initial est obtenu et les unités échantillonnées sont marquées. Un second échantillon est obtenu indépendamment et on note combien d’unités sont marquées. On estime la taille de la population par le rapport du produit des deux tailles d’échantillon, sur le nombre total d’unités marquées. On peut généraliser l’échantillonnage par capture-recapture au cas où les bases de sondages ne sont pas les mêmes que la population cible, mais tout de même reliées à cette dernière. On utilise alors l’échantillonnage indirect.

[MS-272]

**Estimating Mixing Between Subpopulations using Respondent Driven Sampling****Estimation d’homophilie entre sous-populations avec échantillonnage dirigé par les répondants**Art POON, *BC Centre for Excellence in HIV/AIDS*, Sergei L. KOSAKOVSKY POND, *University of California at San Diego*, Simon D. W. FROST, *University of Cambridge*

Respondent-driven sampling (RDS), a kind of chain-referral sampling, is an increasingly popular

L’échantillonnage dirigé par le répondant (EDR) est une méthode de sondage « boule de neige » qui est de plus en

approach of sampling “hidden” populations such as injection drug users and men who have sex with men. RDS relies on social networks and a dual-incentive scheme to rapidly obtain a large sample size, and as a side-effect provides information on mixing between subpopulations. We show how stochastic context-free grammars can be used to model RDS as a multitype branching process, to test for non-random mixing, independence of characteristics among recruitees of a given recruiter; and variation in recruitment behaviours among subpopulations.

plus populaire pour l'échantillonnage de populations difficiles à joindre, comme les utilisateurs de drogues injectables et les hommes ayant des relations sexuelles avec des hommes. L'EDR s'appuie sur les réseaux sociaux et un double incitatif pour obtenir rapidement une taille échantillonnale suffisante, et aussi pour fournir information sur l'homophilie entre sous-populations. Nous démontrons que les grammaires hors contexte stochastiques sont utiles pour modéliser l'homophilie non aléatoire, l'indépendance de caractéristiques entre recrues du même recruteur, et la variation en comportement de recrutement entre sous-populations.

[MS-273]

**Bayes Estimation with Adaptive Web Sampling**

**Estimation bayésienne sous échantillonnage adaptatif en toile**

Mike KWANISAI, *National Opinion Research Center*, Steven K. THOMPSON, *Simon Fraser University*

Obtaining a sample of individuals from a hidden subpopulation often requires network sampling methods, in which social links are followed from individuals in the sample to find additional members of the hidden population. Adaptive web sampling designs comprise a flexible class of such link-tracing designs. Inference from network samples can be challenging because the link-tracing procedures used to obtain the sample can lead to a sample in which the mix of characteristics differs from that of the wider population of interest. In this talk we describe computational Bayes methods which provide a practical solution to this inference problem.

L'obtention d'un échantillon d'individus provenant d'une sous-population cachée nécessite souvent des méthodes d'échantillonnage par réseaux, où les liens sociaux d'individus échantillonnés servent à trouver des membres supplémentaires de la sous-population. Les plans d'échantillonnage adaptatifs en toile sont une classe flexible de tels plans traçant des liens. L'inférence à partir d'échantillons par réseaux peut être difficile, car les méthodes de traçage de liens utilisées pour obtenir l'échantillon peuvent mener à un échantillon dont les caractéristiques diffèrent de celles de la population d'intérêt. Dans cet exposé, nous décrivons des méthodes informatiques bayésiennes apportant une solution pratique à ce problème inférentiel.

**Session 14D Mercredi 26 mai • Wednesday, May 26, 14:30 – 16:00 VAN 2809A**

What can we learn from each other about teaching statistics?

Que peut-on apprendre les uns des autres concernant l'enseignement de la statistique ?

[MS-274]

**A Radical Approach to Teaching Introductory Statistics: The CATALST Project**

**Une approche radicale à l'enseignement de la statistique élémentaire : le projet CATALST**

Robert DELMAS, Joan GARFIELD & Andrew ZIEFFLER, *University of Minnesota*

The NSF-funded CATALST (Change Agents for Teaching and Learning Statistics) project is developing a new curriculum for non-calculus based introductory statistics courses that builds on ideas

Le projet CATALST (Change Agents for Teaching and Learning Statistics), financé par la NSF, développe un nouveau programme pour une introduction à la statistique non fondée sur le calcul différentiel et s'appuyant sur des idées prove-

from statistics education, cognitive science, and mathematics education. Instead of standard statistical procedures, students learn randomization and simulation methods to carry out inferential analyses. Key to the new curriculum are Model Eliciting Activities (MEAs), which are open-ended problems that use real and meaningful data to engage students in statistical thinking. This presentation will share information on the CATALST project as well as materials developed and piloted in the first two years of the project.

nant de l'éducation statistique, des sciences cognitives, et de l'éducation mathématique. Plutôt que des procédures statistiques standards, les étudiants apprennent des méthodes de randomisation et de simulation afin de mener des analyses inférentielles. Le nouveau programme s'appuie sur les activités fondatrices de modèles (AFM), des problèmes ouverts utilisant des données réelles et significatives poussant les étudiants vers la pensée statistique. Cette présentation partagera de l'information au sujet du projet CATALST ainsi que des matériaux didactiques développés et testés au cours des deux premières années du projet.

**[MS-275]**

**CAUSE: Building Partnerships to Advance College Level Statistics Education**

**CAUSE : bâtir des partenariats pour l'évolution de l'éducation statistique universitaire**

Dennis PEARL, *Ohio State University*

The Consortium for the Advancement of Undergraduate Statistics Education (CAUSE: see [www.CAUSEweb.org](http://www.CAUSEweb.org)) is an organization that focuses on providing resources and professional development opportunities to college statistics teachers, and on advancing the fields of statistics education and statistics education research. This talk will discuss a variety of CAUSE initiatives in these focus areas including the CAUSEweb digital library, the CAUSEway and SCHEMATYC workshop programs, the CAUSE webinar series in Teaching and Learning and in Activities for Teaching, the U.S. Conference on Teaching Statistics (USCOTS), the CAUSEmos faculty community groups, and the CAUSE/ASA Statistics Education graduate programs initiative.

*Le Consortium for the Advancement of Undergraduate Statistics Education (CAUSE : voir [www.CAUSEweb.org](http://www.CAUSEweb.org)) est un organisme visant à fournir des ressources et des opportunités de développement professionnel aux enseignants de la statistique au premier cycle universitaire, et à faire progresser l'éducation statistique et la recherche en éducation statistique. Cet exposé discutera d'une variété d'initiative de CAUSE, incluant la bibliothèque digitale CAUSEweb, les programmes d'ateliers CAUSEway et SCHEMATYC, la série de webinaires CAUSE en enseignement et apprentissage et en activités en enseignement, la U.S. Conference on Teaching Statistics (USCOTS), les groupes communautaires d'enseignants CAUSEmos, et l'initiative de programmes d'études supérieures CAUSE/ASA.*

**[MS-276]**

**The Canadian Statistics Educator Network**

**Le Réseau canadien des éducateurs en statistique**

Bethany WHITE, *University of Western Ontario*

Have you ever felt isolated in your teaching endeavours and wanted to bounce teaching ideas off other statistics educators? Does the sharing of tools of the trade and learning from others' experiences in the classroom sound appealing? Through an Educational Developers Caucus grant, the Society for Teaching and Learning in Higher Education (STLHE) has supported the development of the Canadian Statistics Educator Network (CSEN),

Vous êtes-vous déjà senti isolé dans votre rôle d'enseignant, et avez-vous déjà voulu échanger des idées avec d'autres éducateurs en statistique ? Êtes-vous intéressé à partager des outils du métier et apprendre des expériences des autres dans la salle de classe ? Grâce à une subvention de l'*Educational Developers Caucus*, la Société pour l'avancement de la pédagogie dans l'enseignement supérieur (SAPES) a soutenu le développement du CSEN (*Canadian Statistics Educator Network*), une communauté virtuelle visant à encourager un

a virtual community of practice that aims to encourage more collaborative initiatives in Canadian statistics education. The purpose of this talk is to introduce you to the CSEN, demonstrate its features and solicit your feedback.

plus grand nombre de collaborations en enseignement de la statistique au Canada. Le but de cet exposé est de vous présenter le CSEN, illustrer ses caractéristiques et solliciter vos commentaires.

## **Session 14E Mercredi 26 mai • Wednesday, May 26, 14:30 – 16:00 VAN 2811A**

### **Statistical Modelling in Actuarial Sciences**

### **Modèles statistiques en actuariat**

[MS-277]

#### **Semiparametric Approaches to Model the Heterogeneity of Count Data**

#### **Approches semi-paramétriques dans la modélisation de l'hétérogénéité du nombre de réclamations**

Jean-Philippe BOUCHER, *Université du Québec à Montréal*, Montserrat GUILLÉN, *Université de Barcelone*

In count data models, overdispersion of the dependent variable can be incorporated into the model if a heterogeneity term is added into the mean parameter of the Poisson distribution. We use nonparametric estimation for the heterogeneity distribution but also introduce new semiparametric count models to actuarial sciences. Additionally, these nonparametric heterogeneity approaches allow us to give appealing interpretations of the heterogeneity.

La surdispersion observée dans les distributions de comptage peut être modélisée en ajoutant un terme d'hétérogénéité au paramètre de moyenne de la loi de Poisson. Nous utilisons des méthodes d'estimation non paramétriques de l'hétérogénéité, mais présentons aussi à la science actuarielle de nouvelles approches semi-paramétriques. Ces nouveaux modèles permettent d'interpréter différemment l'hétérogénéité dans un contexte d'assurance.

[MS-278]

#### **Goodness-of-fit Tests for Bivariate Extreme-value Copulas**

#### **Tests d'adéquation pour des copules de valeurs extrêmes bivariées**

Christian GENEST, *Université Laval*, Yvan KOJADINOVIC, *University of Auckland*, Johanna NEŠLEHOVÁ, *McGill University*, Jun YAN, *University of Connecticut at Storrs*

The maxima of actuarial and financial variables often occur simultaneously. When they do, their dependence structure can possibly be represented by an extreme-value copula. In this talk, I will first recall how Ghoudi et al. (1998) propose to test this hypothesis. I will then show how to assess the goodness-of-fit of a parametric family of extreme-value copulas using nonparametric estimators of Pickands' dependence function due to Genest & Segers (2009). I will explain how the parametric bootstrap can be used to carry out the test and I will describe its performance through a simulation study.

Les maxima de variables actuarielles et financières se produisent souvent simultanément. Le cas échéant, leur structure de dépendance peut éventuellement être représentée par une copule de type extrême. Dans cet exposé, je rappellerai d'abord comment Ghoudi et coll. (1998) proposent de tester cette hypothèse. Je montrerai ensuite comment juger de l'adéquation d'une famille paramétrique de copules de valeurs extrêmes au moyen d'estimateurs non paramétriques de la fonction de dépendance de Pickands dus à Genest & Segers (2009). J'expliquerai comment se servir du bootstrap paramétrique pour effectuer le test et j'en décrirai la performance par une étude de simulation.

[MS-279]

**Mortality Improvement; an Actuarial Perspective**  
**Amélioration de la mortalité ; une perspective actuarielle**

José GARRIDO, *Concordia University*, Ana DEBÓN, *U. Politécnica de Valencia, Spain*

We study the relation between the two basic random events associated with human mortality; birth and death. Recorded dates of birth and death provide strong evidence of a positive association between longevity and cohort. That means, people born in more recent cohorts tend to live longer. Forecasting models such as that of Lee & Carter (1992, JASA) have attempted, with limited success, to give a parametric description of this association.

We investigate a copula model for a bivariate survival function, with non-parametric marginals. The estimation and fit of extreme value copulas is presented and compared to current parametric models.

Nous étudions la relation entre les deux événements aléatoires de base en mortalité humaine ; la naissance et la mort. À partir de registres de dates de naissance et de mort, nous observons une forte évidence empirique d'une association positive entre longévité et cohorte. Cela veut dire que les populations nées de cohortes plus récentes ont tendance à vivre plus longtemps. Les modèles de prévision, tel que celui de Lee & Carter (1992, JASA) ont tenté, avec un succès mitigé, d'expliquer cette association sous forme paramétrique.

Nous explorons une fonction de survie bivariable définie à partir d'une copule, avec des marges non paramétriques. L'estimation et l'ajustement de copules à valeurs extrêmes sont présentés et comparés à celui des modèles paramétriques courants.

**Session 14F      Mercredi 26 mai • Wednesday, May 26, 14:30 – 15:45      VAN 1203**

**Mixed Effects**

**Effets mixtes**

[MS-280]

**A Semiparametric Nonlinear Mixed-effects Model with Non-ignorable Missing Data and Measurement Errors for HIV Viral Data**

**Un modèle semi-paramétrique non linéaire à effets mixtes pour données virales VIH à erreurs de mesures et données manquantes non ignorables**

Wei LIU, *York University*, Lang WU, *University of British Columbia*

Semiparametric nonlinear mixed-effects (NLME) models are very flexible in modeling long-term HIV viral dynamics. In practice, statistical analyses are often complicated due to measurement errors and missing data in covariates and non-ignorable missing data in the responses. We consider likelihood methods which simultaneously address measurement error and missing data problems. A real dataset is analyzed in detail, and a simulation study is conducted to evaluate the methods.

Les modèles non linéaires à effets mixtes (NLEM) semi-paramétriques permettent une modélisation très flexible de la dynamique virale à long terme du VIH. En pratique, les analyses statistiques sont souvent complexifiées par des erreurs de mesures, des covariables manquantes et des réponses manquantes non ignorables. Nous considérons des méthodes basées sur la vraisemblance traitant simultanément les erreurs de mesure et les données manquantes. Un jeu de données réel est analysé en détail, et une étude de simulation est menée afin d'évaluer les méthodes.

**[MS-281]****An Accurate Method for the Test of Variance Component of Random Effects in Linear Mixed Models  
Une méthode précise pour tester la composante de la variance des effets aléatoires de modèles linéaires mixtes**Yan Yan WU, Georges MONETTE & Augustine WONG, *York University*

Mixed models provide a flexible and powerful tool for the analysis of grouped data in many areas such as biology, health science etc. The existing methods based on likelihood ratio statistic and Wald statistic, calculated on ML and REML, are conservative in the test of random effects. The third-order likelihood asymptotic method provides extremely accurate approximation. These methods can be easily implemented into statistical software such as R. We will illustrate the accuracy of proposed method through simulation studies.

Les modèles mixtes constituent un outil puissant et flexible pour analyser des données groupées dans plusieurs domaines telles la biologie, les sciences de la santé, etc. Les méthodes existantes basées sur le rapport de vraisemblance et la statistique de Wald, calculée à partir du MV et du MVR, sont conservatrices lorsqu'utilisées pour tester des effets aléatoires. La méthode de la vraisemblance asymptotique de troisième ordre fournit une approximation extrêmement précise. Ces méthodes peuvent aisément être implémentées à l'aide de logiciels statistiques tels R. Nous allons illustrer la précision de la méthode proposée à l'aide d'études de simulation.

**[MS-282]****Mixed Effects Trees for Clustered Data  
Arbres à effets mixtes pour données en grappe**Ahlem HAJJEM, François BELLAVANCE & Denis LAROCQUE, *HEC Montréal*

Previous works extending tree based methods to accommodate correlated data are mainly based on the multivariate repeated measures approach. We propose a mixed effects tree method which is more flexible because the observations are viewed as nested within clusters rather than as multivariate vectors. It can handle unbalanced clusters, allows observations within a cluster to be in different nodes, and can incorporate observation-level covariates and random effects. Simulation results show that the proposed method can provide substantial improvements over standard tree algorithms.

Les travaux de recherche antérieurs qui ont adapté les méthodes d'arbre au traitement de données corrélées sont principalement basés sur l'approche multivariée des mesures répétées. Nous proposons un algorithme d'arbre plus flexible qui s'inspire des modèles à effets mixtes. Les observations sont vues comme étant nichées à l'intérieur d'une grappe. La méthode permet de traiter un nombre inégal d'observations par grappe, permet à des observations d'une même grappe de se retrouver dans différents nœuds, et peut inclure des variables explicatives observées et des effets aléatoires. Les résultats d'une étude de simulation montrent que la méthode proposée améliore substantiellement la performance des algorithmes standards d'arbres.

**[MS-283]****Bootstrap Test for Mixed Linear Models  
Test bootstrap pour modèles linéaires mixtes**J LOREDO-OSTI, *Memorial University*

Several resampling tests have been proposed for regression coefficients in a linear model; these tests may not be applicable to mixed linear models because they rely on the assumption of exchangeability. The application of the bootstrap to mixed linear models is not straightforward, although there are several possibilities depending upon the assump-

Plusieurs tests de rééchantillonnage ont été proposés pour les coefficients de régression d'un modèle linéaire ; ces tests peuvent ne pas être applicables à des modèles linéaires mixtes, car ils supposent l'échangeabilité. L'application du bootstrap à des modèles linéaires mixtes n'est pas triviale, bien qu'il existe plusieurs possibilités en fonction des hypothèses faites sur le modèle. Cette présentation discute de trois

tions made. This presentation discusses three different approaches adapted from the general framework of bootstrap techniques for regression models. These approaches are applied to regression-mapping of a major quantitative trait locus in the presence of a polygenic (infinitesimal) background and deterministic covariates.

différentes approches du cadre général du bootstrap adaptées aux modèles de régression. Ces approches sont appliquées en cartographie par régression du locus d'un trait quantitatif majeur en présence de covariables déterministes.

**[MS-284]**

**Spatial Modeling of Disease Mapping of Rates Based on Putative Sources**

**Modélisation spatiale de taux d'infection à partir de sources putatives**

Mahmoud TORABI & Rhonda ROSYCHUK, *University of Alberta*

In this talk, we examine the ability of hierarchical models to recover evidence of disease risk excess around a fixed location. This location can be a putative source of health hazard, such as an incinerator, mobile phone mast or dump site. Poisson mixed models are studied with spatially distance-based random effects. In addition to this term, the conditional autoregressive (CAR) model is also studied to capture the spatially correlated random effects among regions. We illustrate this approach using data on the number of asthma presentations made by children to Emergency Departments in the province of Alberta, Canada.

Dans cet exposé, nous examinons la capacité de modèles hiérarchiques à détecter des excès de risque de maladie autour d'une localisation fixe. Cette localisation peut être une source putative de risques pour la santé, tels un incinérateur, une station de base, ou un site d'enfouissement. Des modèles mixtes de Poisson avec effets spatiaux aléatoires basés sur la distance sont étudiés. Le modèle conditionnel autorégressif (CAR) est aussi étudié afin de traiter les effets aléatoires spatialement corrélés parmi les régions. Nous illustrons cette approche à l'aide de données de dénombrement de cas d'asthme aux urgences dans la province de l'Alberta, Canada.

**Session 14G Mercredi 26 mai • Wednesday, May 26, 14:30 – 16:00 VAN 1211**

**Data Analysis in Biostatistics**

**Analyse de données en biostatistique**

**[MS-285]**

**Underestimation of Mortality Reductions in Cancer Screening Studies: The ERSPC as a Case Study**

**Sous-estimation des réductions de la mortalité dans les études de dépistage du cancer : le ERSPC comme exemple**

James HANLEY, *McGill University*

The impact of cancer screening is typically measured as the reduction in cumulative (or average) cancer mortality. This (PH) measure includes the time before the impact of screening becomes manifest, and underestimates the impact of screening. A smooth-in-time mortality rate ratio function was fitted to the data from the European Randomized Study of Screening for Prostate Cancer (ERSPC) to measure the steady state reduction, and to identify when it reached this level. The re-analysis suggests

L'impact du dépistage du cancer est habituellement mesuré par la réduction de la mortalité cumulative (ou moyenne) due au cancer. Cette mesure inclut le temps avant que l'impact du dépistage se manifeste, et sous-estime l'impact du dépistage. Une fonction du rapport des taux de mortalité lisse par rapport au temps a été ajustée aux données de l'Étude européenne randomisée du dépistage du cancer de la prostate (ERSPC) afin de mesurer la réduction de l'état d'équilibre et de déterminer quand atteint le niveau actuel fut atteint. La ré-analyse suggère que la réduction de la mortalité due au

that the reduction in prostate cancer mortality produced by sustained PSA-based screening may be as much as 50% rather than the 20% reported.

cancer produite par le dépistage basé sur l'APS pourrait être de 50 % plutôt que le 20 % mentionné antérieurement.

#### [MS-286]

### **Stratified Cox Proportional Hazards Models to Identify Genetic Prognostic Markers in Head and Neck Cancer Based on Different Genetic Models**

#### **Ajustement de modèles de Cox stratifiés pour identifier des marqueurs génétiques pronostiques chez les patients atteints d'un cancer otorhinolaryngologique selon différents modèles génétiques**

Élodie SAMSON, *Centre de recherche en cancérologie de l'Université Laval, Québec*, Wei XU, *Biostatistics Department, Princess Margaret Hospital, Toronto*, Isabelle BAIKATI & François MEYER, *Centre de recherche en cancérologie de l'Université Laval, Québec*, Azad ABUL KALAM, *Applied Molecular Oncology, Princess Margaret Hospital, Toronto*, Geoffrey LIU, *Medical Biophysics Department, Ontario Cancer Institute/Princess Margaret Hospital, Toronto*

Recognized prognostic factors cannot accurately predict how patients with head and neck cancer (HNC) evolve following initial cancer treatment. Genetic markers were genotyped in order to assess their association with overall survival (OS). Twenty-two SNPs and two CNVs were selected based on the literature on HNC. Our data set consisted of 531 patients with HNC who participated in a randomized controlled trial. Stratified Cox proportional hazards models were applied and we explored additive, dominant, recessive and discrete genetic models. Four SNPs were associated with OS. Further exploratory analyses were applied including Gene-Gene interactions and further analyses on neighbouring genetic regions.

Les facteurs pronostiques reconnus ne prédisent pas avec précision le pronostic des patients atteints d'un cancer otorhinolaryngologique (ORL) traités par radiothérapie. Des marqueurs génétiques potentiellement pronostiques d'après la littérature (22 SNPs et deux CNVs) ont été mesurés pour évaluer s'ils étaient associés au décès de toutes causes. L'échantillon était constitué de 531 patients atteints d'un cancer ORL qui avaient participé à un essai clinique randomisé. Des modèles de Cox stratifiés ont été utilisés et plusieurs modèles génétiques (additif, dominant, récessif et discret) ont été évalués. Quatre SNPs étaient associés au décès. Des interactions gène-gène ont été testées et des analyses sur les régions génétiques avoisinantes ont été réalisées.

#### [MS-287]

### **Statistical Estimates of Absenteeism Attributable to Influenza from the Labour Force Survey**

#### **Estimations statistiques de l'absentéisme attribuable à la grippe basée sur l'Enquête sur la population active**

Dena SCHANZER, *Public Health Agency of Canada*, Jason GILMORE, *Statistics Canada*, Hui ZHENG, *Public Health Agency of Canada*

The study of influenza epidemiology is particularly challenging because influenza is poorly diagnosed and most data available to monitor the level of influenza activity are not specific to influenza. Statistical methods that have been used to estimate the mortality and hospitalization burden attributable to influenza were adapted to estimate absenteeism rates attributable to influenza using the proportion of hours lost due to illness or disability from the Labour Force Survey and proxy variables for the weekly level of influenza activity. These estimates

L'étude de l'épidémiologie de la grippe est complexe parce que l'influenza est mal diagnostiquée et la plupart des données disponibles ne sont pas spécifiques à la grippe. Les méthodes statistiques qui ont été utilisées pour estimer les taux de la mortalité et de l'hospitalisation attribuables à la grippe ont été adaptées pour estimer les taux d'absentéisme attribuable à la grippe en modélisant la proportion d'heures perdues à cause de maladie ou d'invalidité de l'Enquête sur la population active en fonction du niveau hebdomadaire de l'activité grippale. Ces estimations du taux d'absentéisme informeront la modélisation des maladies infectieuses, le far-

of the absenteeism rate will inform infectious disease modelling, the economic burden of disease, and pandemic planning.

deau économique des maladies, et la planification de pandémie.

**[MS-288]**

**The Use of Mixture Models to Examine Workplace Disability in Ontario**

**L'utilisation de modèles de mélanges pour examiner l'invalidité au travail en Ontario**

David TOLUSSO & Sheilah HOGG-JOHNSON, *Institute for Work & Health*

Ontario's Workplace Safety and Insurance Board has seen an increase in the number of days of wage replacement benefits. Other jurisdictions have found that 80% the disability days come from 20% claimants suggesting the population consist of a group of quick healers and a group who take a long time to return to work. We use discrete mixture models to examine the distribution of the time on benefits within each group. We further look at transitions between working and not working using a mover-stayer model where the underlying multi-state model is an alternating two state process.

Le *Workplace Safety and Insurance Board* de l'Ontario a constaté une augmentation du nombre de jours de bénéfices de remplacement de salaire. D'autres autorités ont trouvé que 80 % des jours d'incapacité proviennent de 20 % des prestataires, suggérant que la population consiste en un groupe de guérison rapide et un groupe nécessitant plus de temps avant un retour au travail. Nous utilisons des modèles de mélanges discrets afin d'examiner le temps d'arrêt de travail pour chaque groupe. Nous nous intéressons d'avantage à la transition entre le travail et l'arrêt de travail à l'aide d'un modèle stable-mobile où le modèle multiétats sous-jacents est un processus alternant entre deux états.

**[MS-289]**

**MDConnect Physician Perspectives and Segmentation Research Study**

**Perspectives sur MDConnect et étude de segmentation**

Delbert STEWART & David STEWART, *Brogan Inc.*

Brogan Inc. and Environics Inc. recently completed groundbreaking research study that offers vital information on how to identify different communication styles in order to effectively communicate with Canada's highest-prescribing Primary Care physicians. Given the infeasibility of surveying all General Practitioners (GPs) in Canada to obtain attitudinal and opinion information, the study's success resided on the ability to develop a probability model having the capacity to predict a GP's communication classification with acceptable accuracy and precision using prescription data. This paper discusses the statistical methodologies, procedures, and tools that were devised and successfully implemented for deriving the physician classification probability model.

Brogan Inc. et Environics Inc. ont récemment complété une étude offrant de l'information vitale relative à l'identification de différents styles de communication afin de communiquer efficacement avec les médecins de premiers recours signant le plus de prescriptions. Étant donnée l'impossibilité de sonder tous les médecins généralistes du Canada, le succès de l'étude repose sur le développement d'un modèle probabiliste pouvant prédire le style de communication de généralistes avec une précision acceptable à partir de données de prescriptions. Cet article discute des méthodologies, procédures et outils statistiques développés et implémentés afin d'obtenir le modèle de classification probabiliste des médecins généralistes.

[MS-290]

**Analysis of Incomplete and Irregularly Spaced Blood Sugar Response Data from a Longitudinal Study**  
**Analyse de données de mesures glycémiques incomplètes et irrégulièrement espacées dans une étude longitudinale**

Muhammad ISLAM, *Grant MacEwan University*

The incidence of type 2 diabetes is on the rise worldwide. There is no cure for type 2 diabetes but it can be controlled with proper diet. In this project a low glycemic index diet has been developed and its effects on blood sugar level have been compared among people with and without type 2 diabetes. The data on blood sugar levels were collected from 10 participants at irregularly spaced different time points. The GEE method was used to analyze the data. The analysis shows that the newly developed diet significantly reduced the risk of developing high blood glucose level.

L'incidence du diabète de type 2 augmente mondialement. Il n'y a pas de remède pour le diabète de type 2, mais il est possible de le contrôler par une diète appropriée. Dans ce projet, une diète à faible indice glycémique a été développée et ses effets sur le taux de glycémie ont été comparés chez des gens atteints et non atteints par le diabète de type 2. Les données de glycémie ont été recueillies auprès de 10 participants à des moments irrégulièrement espacés. Une méthode ÉEG a été utilisée afin d'analyser les données. L'analyse montre que la nouvelle diète réduit significativement le risque de développer un haut taux de glucose.

**Session 14H Mercredi 26 mai • Wednesday, May 26, 14:30 – 15:45 VAN 1871**

**Regression and Data Mining**  
**Régression et forage de données**

[MS-291]

**Prediction of Mixed Outcomes using a Weighted Combination of Univariate and Multivariate Random Forests**

**Prédiction de variables réponses multivariées mixtes avec une combinaison de forêts aléatoires univariées et multivariées**

François BELLAVANCE, Abdessamad DINE & Denis LAROCQUE, *HEC Montréal*

We consider the prediction of mixed outcomes, i.e., multiple outcomes where some are continuous and others are categorical. The predictive performance of multivariate and separate univariate trees for multivariate mixed outcomes, and of weighted combinations of both, is investigated through simulation studies. Note that multivariate tree make predictions simultaneously for all outcomes with a single tree. To achieve better prediction performances, random forests are used. Results show that the proposed weighted combinations from univariate and multivariate random forests perform well in terms of mean squared errors and misclassification rates for the continuous and categorical outcomes respectively.

Nous considérons la prédiction de variables réponses multivariées mixtes, c'est-à-dire certaines sont continues et les autres catégorielles. La performance prédictive des arbres multivariés et d'arbres univariés séparés pour des variables réponses multivariées mixtes, ainsi que des combinaisons pondérées des prédictions provenant des arbres multivariés et univariés, est investiguée à l'aide de simulations. Contrairement aux arbres univariés séparés, l'arbre multivarié donne une prédiction pour l'ensemble des variables réponses avec un seul arbre. Afin d'obtenir de meilleurs résultats pour la prédiction, les forêts aléatoires sont utilisées. Les combinaisons pondérées des prédictions des arbres multivariés et univariés donnent de bons résultats en termes d'erreurs quadratiques moyennes et de taux de mauvaise classification pour les variables réponses continues et catégorielles respectivement.

**[MS-292]****Improving the Performance of Intercept Parameter Estimation in the Multivariate Reduce-rank Regression Model****Améliorer la performance de l'estimation du paramètre d'ordonnée à l'origine dans le modèle de régression de rang réduit multivarié**Shabnam CHITSAZ & S. Ejaz AHMED, *University of Windsor*

Estimation of the intercept vector in a multivariate reduce-rank regression model is considered, when it is priori suspected that the slope parameter matrix may be restricted to a subspace. The goal is to estimate the intercept parameter in the presence of uncertain information in some optimal manner. To achieve this goal, we propose a shrinkage strategy for estimating the parameter of interest. The relative dominance picture of estimators is presented using the large sample results. Further, a testing framework is suggested allowing practitioners to test the validity of the imposed restrictions. Theoretical results and illustrative Monte Carlo results are provided.

L'estimation du vecteur d'ordonnées à l'origine dans un modèle multivarié de régression à rang réduit est considérée lorsqu'il est suspecté a priori que la matrice de paramètres de pente puisse être restreinte à un sous-espace. Le but est d'estimer le paramètre d'ordonnée à l'origine en présence d'information incertaine de façon optimale. Afin d'atteindre ce but, nous proposons une stratégie par rétrécissement afin d'estimer le paramètre d'intérêt. Les performances des estimateurs sont comparées au moyen de résultats asymptotiques. De plus, nous suggérons un cadre permettant de tester la validité des restrictions imposées. Des résultats théoriques et des résultats de Monte-Carlo illustratifs sont donnés.

**[MS-293]****Feature Selection in Mixture Regression Models with Diverging Number of Parameters****Sélection de caractéristiques en mélanges de modèles de régression à nombres de paramètres divergents**Abbas KHALILI, *McGill University*

Feature selection has become an ubiquitous statistical activity in many types of regression modeling. Khalili and Chen proposed a computationally efficient method for feature selection in finite mixture regression models when the potential number of parameters in the model is finite. However, in most feature selection problems the number of parameters of the model should be large and grow with the sample size. This talk presents some new developments on the feature selection problem in mixture regression models when the potential number of parameters in the model can grow with the sample size.

La sélection de caractéristiques est devenue une activité statistique omniprésente dans plusieurs types de modélisation de régression. Khalili et Chen ont proposé une méthode computationnellement efficace de sélection de caractéristiques pour modèles de mélanges finis de régressions lorsque le nombre potentiel de paramètres du modèle est fini. Cependant, dans la plupart des problèmes de sélection de caractéristiques, le nombre de paramètres du modèle devrait être grand et augmenter avec la taille de l'échantillon. Cet exposé présente certains nouveaux développements en sélection de caractéristiques pour mélanges de régressions lorsque le nombre potentiel de paramètres du modèle peut croître avec la taille de l'échantillon.

**[MS-294]****Direct Estimation for Density-weighted Average Derivatives when Responses are Missing at Random****Estimation directe de dérivées moyennes pondérées par les densités en présence de réponses aléatoirement manquantes**Xuewen LU, *University of Calgary*

When responses are missing at random, we propose a direct imputation estimator for the density-

Lorsque les réponses sont aléatoirement manquantes, nous proposons un estimateur d'imputation directe pour la déri-

weighted average derivative of a nonparametric multiple regression function. An imputation estimator for the normalized version of the density-weighted average derivative is constructed using instrumental variables regression. The proposed estimators are computationally simple and asymptotically normal, and provide a solution to the estimation of single-index models with responses missing at random. The developed theory generalizes the method of the density-weighted average derivatives estimation of Powell, Stock & Stoker (1989) for the non-missing data case. Monte Carlo simulation studies are conducted to study the performance of the methods.

vée moyenne pondérée par densité d'une fonction de régression multiple non-paramétrique. Un estimateur d'imputation pour la version normalisée de la dérivée moyenne pondérée par densité est construit au moyen d'une régression avec variables instrumentales. Les estimateurs proposés peuvent être calculés rapidement, sont asymptotiquement normaux, et apportent une solution à l'estimation de modèles à indice simple avec réponse aléatoirement manquante. La théorie développée généralise la méthode d'estimation des dérivées moyennes pondérées par densité de Powell, Stock & Stoker (1989). Des études de simulation de Monte-Carlo sont menées afin d'évaluer la performance des méthodes.

**[MS-295]**

**Mining Drug Data via Mixtures and Related Computational Issues**

**Forage de données de médicaments au moyen de mélanges et problèmes informatiques liés**

Xu (Sunny) WANG, *St. Francis Xavier University*, Hugh CHIPMAN, *Acadia University*

We proposed and developed constrained mixture discriminant analysis (CMDA), a model-based statistical learning method. Constraints are used to deal with the otherwise explosive growth of the number of parameters with increasing dimensionality. CMDA is designed to solve several challenges in modelling drug data sets, such as multiple mechanisms (which correspond to mixture components), the rare target problem (i.e., imbalanced classes), and the identification of relevant subspaces of descriptors (i.e., variable selection).

Nous avons proposé et développé l'analyse discriminante de mélanges sous contraintes (ADMC), une méthode d'apprentissage statistique basée sur un modèle. Les contraintes permettent de contrôler la croissance, autrement explosive, du nombre de paramètres accompagnant une augmentation du nombre de dimensions. L'ADMC est conçue afin de surmonter plusieurs défis liés à la modélisation de jeux de données de médicaments tels que les mécanismes multiples (correspondants aux composantes du mélange), le problème de la cible rare (c.-à-d., classes non équilibrées) et l'identification de sous-espaces de descripteurs pertinents (c.-à-d., sélection de variables).

The translation team did its best. Please forgive us for our misinterpretations, we are neither professional translators nor specialists in all areas of statistics.

*L'équipe de traduction a fait de son mieux. Excusez-nous pour les erreurs de traduction, nous ne sommes ni des traducteurs professionnels ni des spécialistes dans tous les domaines de la statistique.*

## Author Index • Index des auteurs

- ABARIN, Taraneh ..... 165  
 ABRAHAM, Bovas ..... 128  
 ABRAHAMOWICZ, Michal ..... 92  
 ABUL KALAM, Azad ..... 196  
 ACAR, Elif ..... 97  
 AHMED, S. Ejaz ..... 123, 127, 165, 199  
 AINSWORTH, Laurie ..... 173  
 AIT AODIA, Djilali ..... 178, 180  
 AKHTAR-DANESH, Noori ..... 163  
 AKPOUE, Blache Paul ..... 138  
 ALIMADAD, Azadeh ..... 127  
 AL LABADI, Luai ..... 179  
 ALM, Sven Erick ..... 114  
 AL-SAWI, Essam ..... 99  
 ALTMAN, Naomi ..... 116, 175  
 ANCTIL, François ..... 119  
 ANDERSEN, Per Kragh ..... 131  
 ANDREWS, Jeffrey ..... 77  
 ANGERS, Jean-François ..... 98, 113, 138  
 ARMAGAN, Artin ..... 159  
 AUGUSTYNIAK, Maciej ..... 96  
 AUKEMA, Brian ..... 103  
  
 BADESCU, Andrei ..... 120  
 BAI, Yan ..... 85  
 BAIK, Jaiwook ..... 146  
 BAIRATI, Isabelle ..... 196  
 BALAKRISHNA, N. .... 128  
 BALAKRISHNAN, N. .... 80  
 BALBÁS, Alejandro ..... 143  
 BEAUMONT, Jean-François ..... 117, 145  
 BEHSETA, Sam ..... 112  
 BÉLIVEAU, Audrey ..... 145  
 BELLAVANCE, François ..... 194, 198  
 BERKE, Olaf ..... 171  
 BERNIER, Jacques ..... 120  
 BERTRAND, Philip ..... 148  
 BICKIS, Mikelis ..... 79  
 BINGHAM, Derek ..... 85  
 BOOKER, Andrew ..... 84  
 BORGNAT, Pierre ..... 183  
 BORIGHT, Andrew P. .... 178  
  
 BOSA, Keven ..... 145  
 BOUCHER, Jean-Philippe ..... 95, 192  
 BOUEZMARNI, Taoufik ..... 183  
 BOURQUE, Louise ..... 134  
 BOYD, Norman ..... 111  
 BRAUN, John ..... 127, 128  
 BRENNAN, James ..... 144  
 BREWSTER, John ..... 118  
 BRICK, J Michael ..... 102  
 BRIGGS, William ..... 159  
 BRILLINGER, David ..... 78  
 BRIOLLAIS, Laurent ..... 82, 165  
 BROCHU, Magalie ..... 71  
 BROWN, Patrick ..... 83, 139, 151, 156  
 BROWNE, Ryan ..... 99, 153  
 BRUNZELL, John D. .... 178  
 BULL, Shelley ..... 178  
 BUREAU, Alexandre ..... 177  
 BURSTYN, Igor ..... 76  
  
 CABILIO, Paul ..... 101  
 CANTONI, Eva ..... 172  
 CANTY, Angelo J. .... 178  
 CARAGUEL, Charles ..... 90  
 CARNEVALE, Franco ..... 73  
 CARONE, Marco ..... 74  
 CARRIERE, Keumhee Chough ..... 109  
 CARRILLO-GARCIA, Ivan ..... 145  
 CAUCHON, Caroline ..... 70  
 CHABOT-BLANCHET, Malorie ..... 151  
 CHAN, Lisa ..... 73  
 CHANNOUF, Nabil ..... 110  
 CHAPMAN, Judy-Anne ..... 153  
 CHAREST, Anne-Sophie ..... 117  
 CHAUBEY, Yogendra P. .... 81, 123  
 CHEL HEE, Lee ..... 149  
 CHEN, Bei ..... 128  
 CHEN, Fuqi ..... 80  
 CHEN, Gemai ..... 134  
 CHEN, Hubert ..... 115  
 CHEN, Xiaolei ..... 101  
 CHEN, Ying-Pin ..... 76

CHEN, Yue	149	DUCHESNE, Thierry	91
CHEN, Zhijian	139	DUMITRESCU, Laura	110
CHENG, Smiley W	118	DURRETT, Rick	136
CHENOURI, Shojaeddin	112	DUTILLEUL, Pierre	129
CHERRY, Nicola	76	DUVAL, Marie-Claude	168
CHIPMAN, Hugh	188, 200	EL-SHAARAWI, Abdel	88
CHITSAZ, Shabnam	199	ENGLE, Robert	65
CHOI, Yun-Hee	82	ERTEFAIE, Ashkan	107
CHRÉTIEN, Michel	67	ETHERIDGE, Alison	160
CHU, Man-Kee Maggie	99	ÉVIN, Guillaume	119
CIGSAR, Candemir	109	FALLAHPOUR, Saber	123
CLAVEAU, Jeannine	168	FALSETTI, Holly	96
CONROY, Theresa	91	FAN, Zhaozhi	89
COOK, Richard	75, 122, 164, 187	FAREWELL, Daniel	132
COREY, Paul	68	FARRELL, Patrick	129
CORTESE, Giuliana	131	FAVRE, Anne-Catherine	119
COSSETTE, Hélène	174	FENG, Zeny	100, 152
COTTON, Cecilia	92	FERLAND, René	94
CRAIU, Radu	85, 97	FLANDRIN, Patrick	183
CROTEAU, Jordie	177	FOREST, Marie	150
CROWDER, Martin	141	FORTIN, Vincent	119
CsÖRGŐ, Miklós	66	FRANCZAK, Brian	155
DARLINGTON, Gerarda	148	FRANGAKIS, Constantine	75
DAVIES, Katherine	80	FREDETTE, Marc	110
DAVISON, Anthony	77	FREEMAN, Laura	118
DEAN, Charmaine	173	FROST, Simon D. W.	189
DEARDON, Rob	101, 148, 150, 154	FRYXELL, Bruce	85
DEBAS, Haile	67	GAILLARDETZ, Patrice	106
DEBÓN, Ana	193	GAN, Yingying	141
DEETH, Lorna	150	GARDNER, Abbie	148
DEGHAN, Mohammad	91	GARFIELD, Joan	190
DELMAS, Robert	190	GARRIDO, José	141, 143, 193
DENG, Dianliang	164	GEL, Yulia	128, 184
DESMOND, Anthony F.	156	GENEST, Christian	192
DIALLO, Mamadou	168	GEOFF, Hall	112
DIGGLE, Peter	172	GHAHRAMANI, Melody	133
DI NARZO, Antonio	85	GHODSI, Ali	176
DINE, Abdessamad	198	GILCHRIST, Michael	159
DOCHITOIU, Catalin	180	GILES, Philip	138
DOLOVICH, Lisa	163	GILMORE, Jason	196
DONG, Bin	121	GIRARD, Claude	167
DONGMO JIONGO, Valery	181	GIROUX, Suzelle	116
DORAY, Louis	96	GODIN, Frédéric	142
DOSMAN, James	149	GOSSEN, Marianne	70
DRAGOMIR, Alice	98	GOTTARDO, Raphaël	158
DREKIC, Steve	120	GREEN, Jane	82
DUBIN, Joel	89	GREVEN, Andreas	136
DUCHESNE, Pierre	181		

GRIFFITHS, Robert	160	KHALILI, Abbas	199
GUILLÉN, Montserrat	192	KOJADINOVIC, Yvan	192
GUILLOTTE, Simon	86	KOKOSZKA, Piotr	170
GUO, Qing	112	KOPCIUK, Karen	82
GUSTAFSON, Paul	76	KORTBI, Othmane	125
HABIB, Salman	188	KOSAKOVSKY POND, Sergei L.	189
HAJJEM, Ahlem	194	KOVACEVIC, Milorad	145
HAMEL, Martin	167	KULPERGER, Reg	100, 104
HAMILTON, David	173	KUSWANTO, Heri	119
HAN, Lengyi	128	KWANISAI, Mike	190
HANLEY, James	195	LABBE, Aurélie	177
HANSEN, Oluf	132	LABELLE-BLANCHET, Sébastien	167
HASAN, M. Tariqul	166	LABRECQUE-SYNNOTT, Félix	113
HAZIZA, David	145, 146, 181	LAFORTUNE, Yves	145
HE, Wenqing	73, 122, 177	LAKHAL CHAIEB, M'Hamed Lajmi	123
HEAGERTY, Patrick	92	LANCHIER, Nicolas	136
HEISZ, Andrew	70	LAPOINTE, France	134
HEITMANN, Katrin	188	LAROCQUE, Denis	113, 194, 198
HIDIROGLOU, M.	181	LARRIBE, Fabrice	150
HIGDON, Dave	188	LAVALLÉE, Pierre	189
HINES, W.G.S.	104	LAVERTY, William	109
HOGG-JOHNSON, Sheilah	197	LAWLESS, Jerald	97, 109, 182
HOLLOWAY, James	85	LAWRENCE, Earl	188
HORROCKS, Julie	173	LEBLANC, Alexandre	126
HORTOP, Eric	169	LEBLOIS, Raphaël	160
HORVÁTH, Lajos	170	LEBOVIC, Gerald	139
HOUSE, Leanna	189	LEE, Alan	186
HUANG, Mei Ling	91, 96	LEFEBVRE, Geneviève	151
HUANG, Ziming	184	LÉGER, Christian	146
HUSSEIN, Abdulkadir	99	LEI, Na	122
ISLAM, Muhammad	198	LESSARD, Sabin	179
Ji, Yunqi	89	LÉTOURNEAU, Raynald	71
JIANG, Depeng	72	LEVALLOIS, Patrick	74
JIANG, Hedy	156	LÉVEILLÉ, Ghislain	141
JIANG, Jiancheng	115	LI, Haocheng	166
JIANG, Min	126	LI, He	165
JIANG, Weibin	100, 104	LI, Jun	123
JOHN, Braun	150	LI, Longhai	109, 125
KAN, Kin Hung (Felix)	93	LI, Pengfei	114
KARUNANAYAKE, Chandima	149	LI, Xiao Bo	79
KARWA, Vishesh	175	LI, Ye	151
KAVANAGH, Peter	68	LI, Yin	109
KEATING, Jerome P.	80	LIN, Zhang	154
KEIDING, Niels	132	LIU, Geoffrey	196
KELLY, Maureen	135	LIU, Juxin	76
KERMANY, Amir	179	LIU, Meng	164
		LIU, Wei	193
		LIU, Xiaoming	141

LIX, Lisa	108, 109, 149	MÖHLE, Martin	160
LO, Simon	97	MOMEYA, Romuald Hervé	142, 143
LOEWEN, David	69	MONETTE, Georges	194
LOREDO-OSTI, J	194	MOODIE, Erica	92, 111, 167
LOTHIAN, Jack	144	MORALES, Manuel	121, 142, 143
LU, Wilson	162	MURDOCH, Duncan	73
LU, Xuewen	199	MUTTLAK, Hassen	99
LUO, Dandan	152	NADEAU, Christian	70
MA, Jinhui	163	NATHOO, Farouk	103
MA, Renjun	166	NEUFER, Jared	136
MACGIBBON, Brenda	110	NEUHAUS, John	186
MACH, Lenka	145	NEVALAINEN, Jaakko	113
MAILHOT, Méline	174	NEŠLEHOVÁ, Johanna	175, 192
MAMON, Rogemar	106	NIELSEN, Jason	129
MANCEUR, Ameer	129	NIU, Xiaoqing	114
MARCEAU, Étienne	174	NKURUNZIZA, Sévérien	80, 124
MARCHAND, Éric	124, 125, 178	OAKLEY, Jeremy	84
MARIACA HAJDUCEK, Dagmar	182	O'HAGAN, Anthony	84
MARIONI, John	158	O'HARA HINES, R. Jeanette	104
MARRIOTT, Paul	95	OJA, Hannu	113
MARTCHENKO, Daria	101	OKHRATI, Ramin	143
MARTIN, Lisa	111	O'REILLY, Daria	108
MARTSYNYUK, Yuliya	105	OYARZUN, Javier	168
MASHREGHI, Zeinab	146	OYET, Alwell	152
MASON, Robert L.	80	PAHWA, Punam	149
MATTHEWS, David E.	121	PAI, Jeffrey	133
MAUME-DESCHAMPS, Véronique	174	PARENT, Eric	120
MAUMY, Myriam	145	PARFREY, Pat	82
MAYBERRY, John	136	PARTOVI NIA, Vahid	77
McCULLOCH, Charles	186	PATERSON, Andrew D.	178
McDONALD, Mary Ellen	73	PAUL, Sudhir	90, 140
McDONALD, Michael	100	PEARL, Dennis	191
McISAAC, Michael	75	PELLETIER, Caroline	135
McLEOD, Ian	156	PELLETIER, Eric	168
McNICHOLAS, Paul	77, 99, 101, 152, 154–156	PENG, Jianan	115
MEHRABIAN, Amaan	94	PENG, Yingwei	93
MELNICHOUK, Olga	111	PENNELL, Craig	165
MEN, Zhongxian	94	PERREAULT, Luc	119
MENEZES, Raquel	172	PERREAULT, Sylvie	98
MÉRLETTE, Chantal	177	PERRON, François	86, 180
MERLEAU, James	119	PFAFFELHUBER, Peter	136
MEYER, François	196	PFEFFERMANN, Danny	161
MIHRAM, Danielle	79	PICKA, Jeffrey	179
MIHRAM, G. Arthur	79	PLATZ, Elizabeth	75
MILLAR, AnneMichele	130	POND, Gregory	83
MILLS FLEMMING, Joanna	172	POON, Art	189
MINKIN, Salomon	111	PROVOST, Serge	126
MOGHADERI, Azadeh	183, 184		

PULLENAYEGUM, Eleanor	108, 112	SEN, Debaraj	81
PURSEY, Stuart	180	SEN, Pranab	123
QUADIR, Tanvir	165	SHAHIDI, Javad	73
RAHEEM, Enayetur	127	SHAIKH, Mateen	156
RAHIM, Karim	184	SHANG, Xiaoli	147
RAMACHANDRAN, Sohini	159	SHEN, Hua	164
RAMIREZ RAMIREZ, Lilia	184	SHINOHARA, Russell	75
RAMSAY, Jim	171	SHRIER, Ian	73
RANJAN, Pritam	188	SINGH, Murari	81
RAO, J. N. K.	162, 181	SLAMA, Rémy	132
RAVISHANKER, Nalini	133	SLAVKOVIC, Aleksandra	175
RAYMOND, Kyle	176	SMITH, Bruce	155
RAYMOND, Mélanie	71	SONG, Peter	88, 170
REEDMAN, Laurie	168	SPARLING, Matthew	154
REESOR, Mark	93	STAFFORD, James	139
REGIER, Michael	163	STEELE, Russell	73
REID, Nancy	115	STEPHENS, David	94, 107, 111, 141, 167
REIMHERR, Matthew	170	STEWART, David	197
RÉMILLARD, Bruno	66	STEWART, Delbert	197
RENNIE, Donna	149	STEWART, Sam	155
RICH, Benjamin	111	STRYHN, Henrik	90
RIVEST, Louis-Paul	74, 123, 189	STURM, Anja	136
RIZVI, Hasan	130	SU, Ting-Li	172
RODRIGO, Marianito	106	SUBEDI, Sanjeena	154
ROGERS-STEWART, Katrina	129	SWISHCHUK, Anatoliy	66
ROMDHANI, Héla	123	SYLVESTRE, Marie-Pierre	178
ROSENTHAL, Jeffrey	86	TAAMOUTI, Abderrahim	183
ROSYCHUK, Rhonda	153, 195	TAKAHARA, Glen	184
ROTONDI, Michael	71	TALEBAN, Julia	72
ROUSSET, François	160	TAMAZIGHT, Mohamed	124
ROUTLEDGE, Rick	82	TAN, Ken Seng	107
RUBENTHALER, Sylvain	66	TARRIDE, Jean-Eric	98, 108
RYDÉN, Jesper	114	TAYLOR, Jay	160
SAHA, Krishna	140	THABANE, Lehana	163
SAJOBI, Tolulope	108, 109	THAVANESWARAN, Aerambamoorthy	133
SALIBIAN-BARRERA, Matias	127	THOMAS, Steven	168
SAMSON, Élodie	196	THOMPSON, Mary	162, 170
SANGALLI, Laura	78	THOMPSON, Steven K.	190
SANTNER, Thomas	187	THOMSON, David	184, 185
SAVARD, Nathalie	74	TIBSHIRANI, Robert	65, 185
SCHANZER, Dena	196	TOLUSSO, David	197
SCHAU, Candace	130	TOMLINSON, George	139
SCHEIKE, Thomas	176	TORABI, Mahmoud	195
SCOTT, Alastair	186	TOUSIGNANT, Julie	71
SECCHI, Piercesare	78	TRINH, Nguyen Bao	119
SEGRS, Johan	86	TSILIDIS, Konstantinos	75
SELVARATNAM, Selva	152	TURCOTTE, Richard	119
		UBARTAS, Cindy	169

VALLIANT, Richard	117	XU, Changjiang	156
VANTINI, Simone	78	XU, Chen	182
VARIYATH, Asokan	146	XU, Jianfeng	93
VASUDEVAN, Chithran	146	XU, Wei	196
VÉGIARD, Sylvain	134	XU, Xiaojian	147
VELAZQUEZ, Juan Alberto	119	XU, Ximing	81
VERRET, François	116, 144, 181	YAN, Guohua	166
VINING, Geoff	118	YAN, Jun	192
VITELLI, Valeria	78	YAN, Ying	114
VRBIK, Irene	150	YANG, Su-Fen	118
WAGGOTT, Daryl	178	YANG, Xiaojian	152
WAGNER, Christian	188	YANG, Xinye (Hannah)	135, 168
WAKELEY, John	159	YAO, Fang	97
WANG, Liqun	165	YI, Grace	131, 139, 166
WANG, Qingyu	175	YI, Yanqing	75
WANG, Weiwei	188	YILMAZ, Yildiz	97
WANG, Xiaogang	88	YU, Hao	100, 104
WANG, Xikui	75	YUE, Li Hua	73
WANG, Xu (Sunny)	200	YUEN, Wai Kong	147
WANG, Ya Fang	141	ZÄHLE, Iljana	136
WANG, Zilin	162	ZAIHRA, Tasneem	90, 130
WATIER, François	94	ZAMAR, Ruben	87
WEERASINGHE, Swarna	155	ZAREPOUR, Mahmoud	179
WELDON, Larry	69	ZARETZKI, Russell	159
WELSH, Alan	172	ZARKOOB, Hadi	176
WENG, Chengguo	107	ZENY, Feng	150
WHALEN, Christian	138	ZHANG, Peng	88, 152
WHITE, Bethany	191	ZHANG, Ting	164
WHITE, Martin	188	ZHANG, Xuemao	140
WHITEHEAD, Hal	173	ZHANG, Ying	101
WILD, Chris	186	ZHAO, Hui	95
WILKE, Ralf	97	ZHAO, Yang	164
WILLISON, Don	137	ZHENG, Hui	196
WINTER, Anita	136	ZHENG, Yanbing	103
WOLTERS, Mark	127	ZHOU, Lutong	83
WONG, Augustine	194	ZHOU, Qian	170
WU, C. F. Jeff	157	ZHU, Rong	88
WU, Changbao	114, 139, 145	ZIED, Ben Salah	142
WU, Jianrong	92	ZIEFFLER, Andrew	190
WU, Lang	193	ZOU, Guang Yong	72
WU, Longyang	122		
WU, Wei	105		
WU, Yan Yan	194		
XIA, Yu	78		
XIAO, Yongling	92		
XIE, Feng	108		
XIONG, Juan	177		
XIONG, Xiaoqin	89		

# Author Index (per session)

## Index des auteurs (par session)

ABARIN, Taraneh	10G	BELLAVANCE, François	00A, 14F, 14H
ABBASI, Saqib	04L	BERKE, Olaf	12B
ABDALLA, Bakri	04K	BERNIER, Jacques	07D
ABRAHAM, Bovas	01A, 07I	BERTRAND, Philip	09I
ABRAHAMOWICZ, Michal	04G	BICKIS, Mikelis	02I
ABUL KALAM, Azad	14G	BINGHAM, Derek	04B
ACAR, Elif	04I	BLACKWOOD-BECKFORD, Jessica	04K
AHMED, S. Ejaz	07G, 07H, 10G, 14H	BONNER, Ashley	04L
AINSWORTH, Laurie	04D, 12C	BOOKER, Andrew	04B
AIT AOUDIA, Djilali	12F	BORGNAT, Pierre	12H
AKHTAR-DANESH, Noori	10F	BORIGHT, Andrew P.	12E
AKPOUE, Blache Paul	09F	BOSA, Keven	09H
ALI, Herlini Salim	04K	BOUCHER, Jean-Philippe	04H, 14E
ALIMADAD, Azadeh	07H	BOUEZMARNI, Taoufik	12H
AL LABADI, Luai	12F	BOURGET, Edwin	01A
ALM, Sven Erick	05H	BOURQUE, Louise	09C
AL-SAWI, Essam	04J	BOYD, Norman	05G
ALTMAN, Naomi	07A, 12E	BRAUN, John	07H, 07I
ANCTIL, François	07D	BRENNAN, James	09H
ANDERSEN, Per Kragh	09A	BREWSTER, John	07C
ANDREWS, Jeffrey	02H	BRICK, J Michael	05A
ANGERS, Jean-François	04J, 05H, 09F	BRIGGS, William	10B
ARMAGAN, Artin	10B	BRILLINGER, David	02H
ASZTALOS, Christina	04K	BRIOLLAIS, Laurent	04A, 10G
AUGUSTYNIAK, Maciej	04I	BROCHU, Magalie	02E
AUKEMA, Brian	05B	BROWN, Patrick	04A, 09F, 09J
BADESCU, Andrei	07E	BROWNE, Ryan	04J, 09J
BAI, Yan	04C	BRUNZELL, John D.	12E
BAIK, Jaiwook	09I	BULL, Shelley	04D, 12E
BAIRATI, Isabelle	14G	BUREAU, Alexandre	12E
BALAKRISHNA, N.	07I	BURSTYN, Igor	02G
BALAKRISHNAN, N.	02I	CABILIO, Paul	04J, 10D
BALBÁS, Alejandro	09G	CANTONI, Eva	12C
BALKISSOON, Kyle	04K	CANTY, Angelo J.	12E
BARAL, Chelsia	04K	CARAGUEL, Charles	04F
BEAUMONT, Jean-François	07B, 09H	CARNEVALE, Franco	02F
BEHSETA, Sam	05G	CARONE, Marco	02G
BÉLIVEAU, Audrey	04K, 09H	CARRIERE, Keumhee Chough	05F

CARRILLO-GARCIA, Ivan	09H	DEHGHAN, Mohammad	04G
CAUCHON, Caroline	02E	DE LEON, Faye	04L
CHABOT-BLANCHET, Malorie	09J	DELMAS, Robert	14D
CHAN, Lisa	02F	DENG, Dianliang	10F
CHANNOUF, Nabil	05G	DESMOND, Anthony F.	09J
CHAPMAN, Judy-Anne	09J	DIALLO, Mamadou	10H
CHAREST, Anne-Sophie	07B	DIGGLE, Peter	12B
CHAUBEY, Yogendra P.	02I, 07F	DI NARZO, Antonio	04C
CHEL HEE, Lee	09J	DINE, Abdessamad	14H
CHEN, Bei	07I	DOCHITOIU, Catalin	12G
CHEN, Fuqi	02I	DOLOVICH, Lisa	10F
CHEN, Gemai	09B	DONG, Bin	07F
CHEN, Hubert	05H	DONGMO JIONGO, Valery	12G
CHEN, Jun (Jason)	04L	DORAY, Louis	04I
CHEN, Xiaolei	04J	DOSMAN, James	09J
CHEN, Ying-Pin	02G	DRAGOMIR, Alice	04J
CHEN, Yue	09J	DREKIC, Steve	07E
CHEN, Zhijian	09F	DUBIN, Joel	04F
CHENG, Smiley W	07C	DUCHESNE, Pierre	12G
CHENOURI, Shojaeddin	05G	DUCHESNE, Thierry	01A, 04G
CHERRY, Nicola	02G	DUMITRESCU, Laura	05G
CHIPMAN, Hugh	00A, 14B, 14H	DURRETT, Rick	09D
CHITSAZ, Shabnam	14H	DUTILLEUL, Pierre	07I
CHOI, Yun-Hee	04A	DUVAL, Marie-Claude	10H
CHRÉTIEN, Michel	02C	EL-SHAARAWI, Abdel	04E
CHU, Man-Kee Maggie	04J	ENGLE, Robert	01B
CIGSAR, Candemir	05F	ERTEFAIE, Ashkan	05F
CLAVEAU, Jeannine	10H	ETHERIDGE, Alison	10C
CONROY, Theresa	04G	ÉVIN, Guillaume	07D
COOK, Richard	02G, 04D, 07F, 10F, 14A	FALLAHPOUR, Saber	07G
COREY, Paul	02D	FALSETTI, Holly	04I
CORMIER, Eric	04L	FAN, Zhaozhi	04F
CORTESE, Giuliana	09A	FAREWELL, Daniel	09A
COSSETTE, Hélène	12D	FARRELL, Patrick	07I
COTTON, Cecilia	04G	FAVRE, Anne-Catherine	07D
CRAIU, Radu	04C, 04I	FENG, Zeny	04J, 09J
CROTEAU, Jordie	12E	FERLAND, René	04H
CROWDER, Martin	09F	FERLAND-RAYMOND, Bastien	04K
CSÖRGŐ, Miklós	02B	FLANDRIN, Patrick	12H
DARLINGTON, Gerarda	09J	FOREST, Marie	09J
DAVIES, Katherine	02I	FORTIN, Vincent	07D
DAVIS, Michael (Jack)	04L	FRANCZAK, Brian	09J
DAVISON, Anthony	02H	FRANGAKIS, Constantine	02G
DEAN, Charmaine	12C	FREDETTE, Marc	05G
DEARDON, Rob	04J, 09J	FREEMAN, Laura	07C
DEBAS, Haile	02C	FROST, Simon D. W.	14C
DEBÓN, Ana	14E	FRYXELL, Bruce	04B
DEETH, Lorna	09J	FU, Eric	04L

GAILLARDEZ, Patrice .....	05E	HOUSE, Leanna .....	14B
GAN, Yingying .....	09G	HU, Nianping .....	04L
GARDNER, Abbie .....	09J	HUANG, Mei Ling .....	04G, 04I
GARFIELD, Joan .....	14D	HUANG, Ziming .....	12H
GARRIDO, José .....	09G, 14E	HUSSEIN, Abdulkadir .....	04J
GEL, Yulia .....	07I, 12H		
GENEST, Christian .....	14E	ISLAM, Muhammad .....	14G
GEOFF, Hall .....	05G	IVANOFF, Gail .....	00A
GHAHRAMANI, Melody .....	09B		
GHODSI, Ali .....	12E	Ji, Yunqi .....	04F
GILCHRIST, Michael .....	10B	JIANG, Depeng .....	02F
GILES, Philip .....	09E	JIANG, Hedy .....	09J
GILMORE, Jason .....	14G	JIANG, Jiancheng .....	05H
GIRARD, Claude .....	10H	JIANG, Min .....	07H
GIROUX, Suzelle .....	07B	JIANG, Weibin .....	04J, 05C
GODIN, Frédéric .....	09G	JOHN, Braun .....	09J
GOSSEN, Marianne .....	02E		
GOTTARDO, Raphaël .....	10B	KAN, Kin Hung (Felix) .....	04H
GREEN, Jane .....	04A	KARUNANAYAKE, Chandima .....	09J
GREVEN, Andreas .....	09D	KARWA, Vishesh .....	12E
GRIFFITHS, Robert .....	10C	KAVANAGH, Peter .....	02C
GUILLÉN, Montserrat .....	14E	KEATING, Jerome P. ....	02I
UILLOTTE, Simon .....	04C	KEEN, Kevin .....	09E
GUO, Qing .....	05G	KEIDING, Niels .....	09A
GUSTAFSON, Paul .....	02G	KELLY, Maureen .....	09C
GUSTAFSSON, Emelie .....	04L	KERMANY, Amir .....	12F
		KHALILI, Abbas .....	14H
HABIB, Salman .....	14B	KOJADINOVIC, Yvan .....	14E
HAJJEM, Ahlem .....	14F	KOKOSZKA, Piotr .....	11B
HAMEL, Martin .....	10H	KOPCIUK, Karen .....	04A
HAMILTON, David .....	12C	KORTBI, Othmane .....	07G
HAN, Lengyi .....	07I	KOSAKOVSKY POND, Sergei L. ....	14C
HANLEY, James .....	14G	KOVACEVIC, Milorad .....	09H
HANSEN, Oluf .....	09A	KULPERGER, Reg .....	04J, 05C
HARVEY, Fanny .....	04K	KUSWANTO, Heri .....	07D
HASAN, M. Tariqul .....	10G	KWANISAI, Mike .....	14C
HAZIZA, David .....	09H, 12G		
HE, Wenqing .....	02F, 07F, 12E	LABBE, Aurélie .....	12E
HEAGERTY, Patrick .....	04G	LABELLE-BLANCHET, Sébastien .....	10H
HEISZ, Andrew .....	02E	LABRECQUE-SYNNOTT, Félix .....	05H
HEITMANN, Katrin .....	14B	LAFORTUNE, Yves .....	09H
HIDIROGLOU, M. ....	12G	LAKHAL CHAIEB, M'Hamed Lajmi .....	07F
HIGDON, Dave .....	14B	LANCHIER, Nicolas .....	09D
HINES, W.G.S. ....	05B	LAPOINTE, France .....	09C
HOGG-JOHNSON, Sheilah .....	14G	LAROCQUE, Denis .....	05H, 14F, 14H
HOLLOWAY, James .....	04B	LARRIBE, Fabrice .....	09J
HORROCKS, Julie .....	12C	LAVALLÉE, Pierre .....	14C
HORTOP, Eric .....	10H	LAVERTY, William .....	05F
HORVÁTH, Lajos .....	11B	LAWLESS, Jerald .....	04I, 05F, 12G
		LAWRENCE, Earl .....	14B

LEBLANC, Alexandre	07H	MARCEAU, Étienne	12D
LEBLOIS, Raphaël	10C	MARCHAND, Éric	07G, 12F
LEBOVIC, Gerald	09F	MARIACA HAJDUCEK, Dagmar	12G
LEE, Alan	14A	MARIONI, John	10B
LEFEBVRE, Geneviève	09J	MARRIOTT, Paul	04H
LÉGER, Christian	09H	MARTCHENKO, Daria	04J
LEI, Na	07F	MARTIN, Lisa	05G
LESPERANCE, Mary	00A	MARTSYNYUK, Yuliya	05C
LESSARD, Sabin	12F	MASHREGHI, Zeinab	04K, 09H
LÉTOURNEAU, Raynald	02E	MASON, Robert L.	02I
LEVALLOIS, Patrick	02F	MATTHEWS, David E.	07F
LÉVELLÉ, Ghislain	09G	MAUME-DESCHAMPS, Véronique	12D
LI, Haocheng	10G	MAUMY, Myriam	09H
LI, He	10G	MAYBERRY, John	09D
LI, Jun	07F	McCULLOCH, Charles	14A
LI, Longhai	05F, 07G	McDONALD, Mary Ellen	02F
LI, Pengfei	05H	McDONALD, Michael	04J
LI, Wenbin	04L	McISAAC, Michael	02G
LI, Xiao Bo	02I	McLEOD, Ian	09J
LI, Ye	09J	McNICHOLAS, Paul	02H, 04J, 09J
LI, Yin	05F	MEHRABIAN, Amaan	04H
LIANG, Yitian (Sky)	04L	MELNICHOUK, Olga	05G
LIN, Wei (Becky)	04L	MEN, Zhongxian	04H
LIN, Zhang	09J	MENEZES, Raquel	12B
LIU, Geoffrey	14G	MÉRETTE, Chantal	12E
LIU, Juxin	02G	MERLEAU, James	07D
LIU, Meng	10F	MEYER, François	14G
LIU, Wei	14F	MIHRAM, Danielle	02H
LIU, Xiaoming	09G	MIHRAM, G. Arthur	02H
LIU, Zhihui Amy	04L	MILLAR, AnneMichele	07I
LIX, Lisa	05F, 09J	MILLETTE, Lise	05D
Lo, Simon	04I	MILLS FLEMMING, Joanna	12C
LOEWEN, David	02D	MINKIN, Salomon	05G
LOREDO-OSTI, J	14F	MOGHTADERI, Azadeh	12H
LOTHIAN, Jack	09H	MÖHLE, Martin	10C
LU, Hong	04L	MOMEYA, Romuald Hervé	09G
LU, Linghong	04L	MONETTE, Georges	14F
LU, Wilson	10E	MOODIE, Erica	04G, 05G, 10G
LU, Xuewen	14H	MORALES, Manuel	07E, 09G
LUO, Dandan	09J	MURDOCH, Duncan	02F
		MUTTLAK, Hassen	04J
MA, Jinhui	10F	NADEAU, Christian	02E
MA, Renjun	10G	NATHOO, Farouk	05B
MacGIBBON, Brenda	05G	NEUFER, Jared	09D
MACH, Lenka	09H	NEUHAUS, John	14A
MAILHOT, Mélina	12D	NEVALAINEN, Jaakko	05H
MAMON, Rogemar	05E	NEŠLEHOVÁ, Johanna	12D, 14E
MANCEUR, Ameer	07I	NIELSEN, Jason	07I
MANN, Rena	04L		

NIU, Xiaoqing	05H	RANJAN, Pritam	14B
NKURUNZIZA, Séverien	.02I, 07G	RAO, J. N. K.	10E, 12G
NOREAU, Pierre-Paul	05D	RAVISHANKER, Nalini	09B
OAKLEY, Jeremy	04B	RAYMOND, Kyle	12E
O'HAGAN, Anthony	04B	RAYMOND, Mélanie	02E
O'HARA HINES, R. Jeanette	05B	REEDMAN, Laurie	10H
OJA, Hannu	05H	REESOR, Mark	04H
OKHRATI, Ramin	09G	REGIER, Michael	10F
O'REILLY, Daria	05F	REID, Nancy	06A, 10D
OYARZUN, Javier	10H	REIMHERR, Matthew	11B
OYET, Alwell	09J	RÉMILLARD, Bruno	02B
PAHWA, Punam	09J	RENNIE, Donna	09J
PAI, Jeffrey	09B	RICH, Benjamin	05G
PARENT, Eric	07D	RIDDELL, Corinne	04L
PARFREY, Pat	04A	RIVEST, Louis-Paul	00A, 02F, 07F, 14C
PARTOVI NIA, Vahid	02H	RIZVI, Hasan	07I
PATENAUDE, Valérie	04K	ROBERTS, Georgia	07B
PATERSON, Andrew D.	12E	RODRIGO, Marianito	05E
PAUL, Sudhir	04F, 09F	ROGERS-STEWART, Katrina	07I
PEARL, Dennis	14D	ROMDHANI, Héla	07F
PELLETIER, Caroline	09C	ROSENTHAL, Jeffrey	04C
PELLETIER, Eric	10H	ROSYCHUK, Rhonda	09J, 14F
PENG, Jianan	05H	ROTONDI, Michael	02F
PENG, Yingwei	04G	ROUSSET, François	10C
PENNELL, Craig	10G	ROUTLEDGE, Rick	03A
PERREAULT, Luc	07D	RUBENTHALER, Sylvain	02B
PERREAULT, Sylvie	04J	RYDÉN, Jesper	05H
PERRON, François	04C, 12F	SAHA, Krishna	09F
PETKAU, John	04D	SAJOBI, Tolulope	05F
PFAFFELHUBER, Peter	09D	SALIBIAN-BARRERA, Matias	07H
PFEFFERMANN, Danny	10E	SAMSON, Élodie	14G
PICKA, Jeffrey	12F	SANGALLI, Laura	02H
PLATZ, Elizabeth	02G	SANTNER, Thomas	14B
POND, Gregory	04A	SAULNIER, Michel	05D
POON, Art	14C	SAVARD, Nathalie	02F
PROVOST, Serge	07H	SCHANZER, Dena	14G
PULLENAYEGUM, Eleanor	05F, 05G	SCHAU, Candace	07I
PURSEY, Stuart	12G	SCHEIKE, Thomas	12E
QUADIR, Tanvir	10G	SCOTT, Alastair	14A
QURESHI, Abdul	04L	SECCHI, Piercesare	02H
RAHEEM, Enayetur	07H	SEGERS, Johan	04C
RAHIM, Karim	12H	SELVARATNAM, Selva	09J
RAIFU, Amidu	04L	SEN, Debaraj	02I
RAMACHANDRAN, Sohini	10C	SEN, Pranab	07F
RAMIREZ RAMIREZ, Lilia	12H	SHAHIDI, Javad	02F
RAMSAY, Jim	12A	SHAIKH, Mateen	09J
		SHANG, Xiaoli	09I
		SHEN, Hua	10F

SHINOHARA, Russell	02G	VARIYATH, Asokan	09I
SHRIER, Ian	02F	VASUDEVAN, Chithran	09I
SINGH, Murari	02I	VÉGIARD, Sylvain	09C
SLAMA, Rémy	09A	VELAZQUEZ, Juan Alberto	07D
SLAVKOVIC, Aleksandra	12E	VERRET, François	07B, 09H, 12G
SMITH, Bruce	09J	VINING, Geoff	07C
SONG, Peter	04E, 11A	VITELLI, Valeria	02H
SPARLING, Matthew	09J	VRBIK, Irene	09J
STAFFORD, James	09F	WAGGOTT, Daryl	12E
STEELE, Russell	02F	WAGNER, Christian	14B
STEPHENS, David	04H, 05F, 05G, 09F, 10G	WAKELEY, John	10C
STEWART, David	14G	WANG, Chunlin	04K
STEWART, Delbert	14G	WANG, Chunyi	04K
STEWART, Sam	09J	WANG, Liqun	10G
STONE, Ryan	04L	WANG, Qingyu	12E
STRYHN, Henrik	04F	WANG, Weiwei	14B
STURM, Anja	09D	WANG, Xiaogang	04E
SU, Ting-Li	12B	WANG, Xikui	02G
SUBEDI, Sanjeena	09J	WANG, Xu (Sunny)	14H
SWISHCHUK, Anatoliy	02B	WANG, Ya Fang	09G
SYLVESTRE, Marie-Pierre	12E	WANG, Zilin	10E
TAAMOUTI, Abderrahim	12H	WATIER, François	04H
TAKAHARA, Glen	12H	WEERASINGHE, Swarna	09J
TALEBAN, Julia	02F	WELDON, Larry	02D
TAMAZIGHT, Mohamed	07G	WELSH, Alan	12C
TAN, Ken Seng	05E	WENG, Chengguo	05E
TARRIDE, Jean-Eric	04J, 05F	WHALEN, Christian	09E
TAYLOR, Jay	10C	WHITE, Bethany	14D
THABANE, Lehana	10F	WHITE, Martin	14B
THAVANESWARAN, Aerambamoorthy	09B	WHITEHEAD, Hal	12C
THOMAS, Steven	10H	WILD, Chris	14A
THOMPSON, Emmanuel	04K	WILKE, Ralf	04I
THOMPSON, Mary	10E, 11A	WILLISON, Don	09E
THOMPSON, Steven K.	14C	WINTER, Anita	09D
THOMSON, David	12H	WITHANAGE, Niroshan	04K
TIBSHIRANI, Robert	02A, 13A	WOLTERS, Mark	07H
TOLUSSO, David	14G	WONG, Augustine	14F
TOMLINSON, George	09F	WU, C. F. Jeff	10A
TORABI, Mahmoud	14F	WU, Changbao	05H, 09F, 09H
TOUSIGNANT, Julie	02E	WU, Jianrong	04G
TRINH, Nguyen Bao	07D	WU, Lang	14F
TSILIDIS, Konstantinos	02G	WU, Longyang	07F
TURCOTTE, Richard	07D	WU, Wei	05C
UBARTAS, Cindy	10H	WU, Yan Yan	14F
USHEY, Kevin	04L	XIA, Bing	04K
VALLIANT, Richard	07B	XIA, Yu	02H
VANTINI, Simone	02H	XIAO, Yongling	04G
		XIE, Feng	05F

XIONG, Juan	12E
XIONG, Xiaoqin	04F
XU, Changjiang	09J
XU, Chen	12G
XU, Jianfeng	04G
XU, Wei	14G
XU, Xiaojian	09I
XU, Ximing	02I, 04L
YAN, Guohua	10G
YAN, Jun	14E
YAN, Ying	05H
YANG, Su-Fen	07C
YANG, Xiaojian	09J
YANG, Xinye (Hannah)	09C, 10H
YAO, Fang	04I
YI, Grace	08A, 09F, 10G
YI, Yanqing	02G
YILMAZ, Yildiz	04I
YU, Hao	04J, 05C
YUE, Li Hua	02F
YUEN, Wai Kong	09I
ZÄHLE, Iljana	09D
ZAIHRA, Tasneem	04F, 07I
ZAMAR, Ruben	04E
ZAREPOUR, Mahmoud	12F
ZARETZKI, Russell	10B
ZARKOOB, Hadi	12E
ZENY, Feng	09J
ZHANG, Peng	04E, 09J
ZHANG, Ting	10F
ZHANG, Xuemao	09F
ZHANG, Ying	04J
ZHAO, Hui	04H
ZHAO, Yang	10F
ZHENG, Hui	14G
ZHENG, Yanbing	05B
ZHONG, Xiaoyin	04L
ZHOU, Lutong	04A
ZHOU, Qian	11A
ZHU, Rong	04E
ZIED, Ben Salah	09G
ZIEFFLER, Andrew	14D
ZOU, Guang Yong	02F
ZOU, Kelly	10D
ZWIERS, Francis	04D