

# Table of Contents • Table des matières

<b>Table of Contents • Table des matières</b>	<b>1</b>
<b>Welcome • Bienvenue</b>	<b>2</b>
<b>Sponsors • Commanditaires</b>	<b>3</b>
<b>Organizers • Organismes</b>	<b>4</b>
<b>Exhibitors • Exposants</b>	<b>5</b>
<b>General Information • Information générale</b>	<b>5</b>
<b>Committees and Meetings • Comités et réunions</b>	<b>13</b>
<b>Social and Information Events • Événements sociaux et informatifs</b>	<b>15</b>
<b>List of Sessions • Liste des sessions</b>	<b>16</b>
<b>Program • Programme</b>	<b>19</b>
<b>Abstracts • Résumés</b>	<b>52</b>
<b>Author List • Liste des auteurs</b>	<b>208</b>

# Welcome • Bienvenue

It is with great pleasure that the members of the Department of Mathematics and Statistics at Acadia University welcome you to the 39th Annual Meeting of the Statistical Society of Canada. We are delighted to host the SSC once again, having last done so in 1993.

Acadia University has long been recognized as one of Canada's premier undergraduate institutions. The one-hundred hectare campus is situated overlooking the historic Evangeline dykelands and meadows. On campus, a botanical garden features native habitats of the Acadian Forest Region and woodland trails. Visitors can enjoy the beautiful green pathways and elegant buildings, two of which are in the Registry of Historic Places in Canada.

The campus is a two-minute walk from downtown. With about 4000 residents, Wolfville is a popular tourist destination, boasting rich history, active arts scene, beautiful scenery and world-class cuisine that often features local foods. The pedestrian-friendly downtown has many specialty shops and grand century-old homes. We hope that you will have time to experience the surrounding area, including the Grand-Pré National Historic Site, the world's highest tides, the beautiful Annapolis Valley, and many other attractions that Nova Scotia has to offer.

C'est avec grand plaisir que les membres du Département de mathématiques et de statistique vous accueillent à Acadia University pour le 39e congrès annuel de la Société statistique du Canada. Nous sommes très heureux d'accueillir à nouveau la SSC, qui s'est réunie ici pour la dernière fois en 1993.

Acadia University est reconnue de longue date comme l'un des meilleurs établissements de premier cycle au Canada. Son campus de cent hectares surplombe le lieu historique de Grand-Pré. Sur le campus, un jardin botanique met en vedette les habitats de la région de la Forêt acadienne et des sentiers boisés. Les visiteurs apprécieront les jolis sentiers de promenade et les élégants bâtiments de l'université, dont deux figurent au Répertoire canadien des lieux patrimoniaux.

Le campus est à deux minutes de marche du centre-ville. Petite ville de quelque 4 000 habitants, Wolfville est une destination touristique populaire, fière d'une riche histoire, d'un milieu artistique actif, de magnifiques paysages et d'une gastronomie de niveau international, dans laquelle figurent souvent des ingrédients locaux. Le centre-ville, bien adapté aux piétons, regorge de boutiques spécialisées et de demeures séculaires. Si vous avez le temps de profiter de la région, nous vous recommandons de visiter le lieu historique de Grand-Pré, d'observer les marées les plus hautes au monde, de parcourir la belle vallée de l'Annapolis ou encore de découvrir les nombreuses autres attractions qu'offre la Nouvelle-Écosse.

## Sponsors • Commanditaires

SSC 2011 would like to thank its numerous sponsors whose generous contribution has made this meeting possible:

SSC 2011 désire remercier ses nombreux commanditaires dont la généreuse contribution a rendu possible la tenue de ce congrès :

- Acadia University
- Atlantic Association for Research in the Mathematical Sciences
- Centre de recherches mathématiques
- Department of Mathematics and Statistics, Acadia University
- Fields Institute
- Pacific Institute for the Mathematical Sciences
- MINITAB

# Organizers • Organisateurs

## Local Arrangements Committee • Comité des arrangements locaux

Hugh Chipman (Chair • Président)

Paul Cabilio, Wilson Lu, Kenneth McRae, Jianan Peng, Pritam Ranjan, Ying Zhang

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. At Acadia, Nicole Phipps and Colleen Swail (Event Services), Becky Kennedy and Tammy Gregory (Chartwells), Korey Keeney (Cajun's Acadia Gear), Kathy French (Mathematics and Statistics) and Duane Currie (Acadia Centre for Mathematical Modelling and Computation) all provided invaluable assistance. Many members of the SSC also provided valuable help. Thanks to the organizers of previous meetings, in particular Thierry Duchesne, for 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, and to Alan Kelm and Steve La Rocque at the SSC office for electronic services related to the meeting. Finally in addition to abstract translation, Catherine Cox provided an unusually large number of translations for the conference this year, considering the "French-challenged" organizing committee.

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 translators Anne-Sophie Charest, Catherine Cox, Pascale Giroux-Denis, Caroline Gras, Nabil Kahouadji, Annie Lafrance Paiement, Géraldine Lo Siou and Caroline Pelletier, and to the Bilingualism Committee of the SSC (chaired by Cynthia Bocci) for its supervision of this work.

Il est impossible d'organiser un événement de l'envergure du congrès annuel de la SSC sans l'aide de nombreux individus et organismes. À Acadia, Nicole Phipps et Colleen Swail (coordinatrices des événements), Becky Kennedy et Tammy Gregory (Chartwells), Korey Keeney (Cajun's Acadia Gear), Kathy French (Mathématiques et statistique) et Duane Currie (centre de modélisation et calcul mathématique d'Acadia) nous ont été d'une aide inestimable. Plusieurs membres de la SSC nous ont également beaucoup aidés. Merci aux organisateurs des congrès antérieurs, en particulier à Thierry Duchesne, pour son 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, ainsi qu'à Alan Kelm et Steve La Rocque du bureau pour les services électroniques nécessaires au congrès. Enfin, en plus d'avoir participé à la traduction des résumés, Catherine Cox a contribué un nombre inhabituel de traductions pour le congrès, en raison des lacunes linguistiques du comité d'organisation.

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 Anne-Sophie Charest, Catherine Cox, Pascale Giroux-Denis, Caroline Gras, Nabil Kahouadji, Annie Lafrance Paiement, Géraldine Lo Siou et Caroline Pelletier pour ces traductions et au comité du bilinguisme de la SSC (présidé par Cynthia Bocci) pour en avoir assuré la supervision.

## Program Committee • Comité du programme

**Chair • Président:** Shelley Bull, *University of Toronto*

**Members • Membres:**

**Biostatistics Section • Groupe de biostatistique**

Lisa Lix, *University of Saskatchewan*

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

William Welch, *University of British Columbia*

**Probability Section • Groupe de probabilité**

Yiqiang Zhao, *Carleton University*

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

Pierre Lavallée, *Statistics Canada • Statistique Canada*

## Exhibitors • Exposants

Displays for examination and purchase will be available in the Festival Theatre. Confirmed exhibitors at the time this program went to press are: CRC Press, Minitab, Nelson, SAS, Pearson, W.H. Freeman & Company and Wiley.

Les kiosques des exposants seront situés dans le Festival Theatre. Au moment de mettre ce programme sous presse, les exposants suivants ont confirmé leur présence : CRC Press, Minitab, Nelson, SAS, Pearson, W.H. Freeman & Company et Wiley.

## General Information • Information générale

### Registration • Inscription

The registration desk will be located in the Festival Theatre (number 48 on campus map). The registration schedule is as follows:

- Sunday June 12: 8:00 am - 6:30 pm
- Monday June 13: 7:30 am - 5:00 pm
- Tuesday June 14: 8:00 am - 5:00 pm
- Wednesday June 15: 8:00 am - 3:00 pm

Please note that the registration desk is located in the same building as the plenary sessions and across the street from other sessions and workshops.

Le comptoir d'inscription sera situé dans le Festival Theatre (numéro 48 sur la carte). L'horaire des inscriptions sera le suivant :

- Dimanche 12 juin : 8h à 18h30

- Lundi 13 juin : 7h30 à 17h
- Mardi 14 juin : 8h à 17h
- Mercredi 15 juin : 8h à 15h

Veillez noter que le comptoir d'inscription est situé dans le même pavillon que les séances plénières et de l'autre côté de la rue des autres séances et des ateliers.

## On-campus accommodation • Logement sur le campus

Participants staying in campus residences may check in at the Festival Theater, at the Acadia Box Office (48 on campus map). Box Office hours are: Saturday 10am - 8pm; Sunday 8am - 6:30pm; Monday 7:30am - 5pm; Tuesday and Wednesday 10am - 6pm. These are slightly different from hours of the registration desk. Guest arriving outside these hours can check in at the residence life office (902-585-2152) in the Old SUB (9 on campus map).

For guests staying in residence, your accommodation includes:

- Free parking.
- Complimentary breakfast in the dining hall (on Sunday - Wednesday).
- Free access to the Acadia gymnasium.

Les participants qui logent en résidence universitaire peuvent s'enregistrer au Festival Theater, à la billetterie d'Acadia (numéro 48 sur la carte). Les horaires de la billetterie sont de 10h à 20h le samedi ; de 8h à 18h30 le dimanche ; de 7h30 à 17h le lundi ; et de 10h à 18h le mardi et mercredi. Ces horaires diffèrent légèrement de ceux du bureau des inscriptions au congrès. Les délégués qui arrivent en dehors de ces heures pourront s'enregistrer au bureau des résidences (902-585-2152) dans l'édifice Old SUB (numéro 9 sur la carte).

Pour les délégués qui logent en résidence, votre séjour inclut :

- le stationnement gratuit ;
- le petit-déjeuner gratuit à la cafétéria (du dimanche au mercredi) ;
- l'accès gratuit au gymnase d'Acadia.

## Directions • Emplacement

Campus maps are available in the conference bag and on the inside cover of this program. The campus is within 5 minutes walk of downtown Wolfville. Parking is available in a variety of locations indicated on the campus map. Parking passes will be available for sale at the registration desk. Free parking is also available at the south end of campus, in the lot near Crowell Tower (number 12 on campus map).

Almost all accommodation within Wolfville are less than 20 minutes walk from campus. The Old Orchard Inn is the closest hotel outside Wolfville, but is a long walk (4km). Participants staying at the Old Orchard Inn will need to use a cab or car to get in to town. Kings Transit (<http://www.kingstransit.ns.ca/>) has hourly buses that could be used to get in to town, although sharing a cab may be easier.

Des cartes du campus sont disponibles dans le sac du congressiste et en deuxième de couverture de ce programme. Le campus est à cinq minutes de marche du centre-ville de Wolfville. Vous pouvez stationner à divers endroits indiqués sur le plan du campus après avoir acheté une vignette de stationnement au comptoir des inscriptions. Le stationnement est gratuit à l'extrémité sud du campus, près de la tour Crowell (numéro 12 sur la carte).

Presque tous les hôtels de Wolfville sont à moins de 20 minutes de marche du campus. Le Old Orchard Inn est l'hôtel le plus proche de Wolfville, mais il est à 4 km de là. Les congressistes qui y logent devront sans doute appeler un taxi ou prendre le volant pour se rendre en ville. Les bus de Kings Transit (<http://www.kingstransit.ns.ca/>) font le trajet une fois par heure, mais il est peut-être plus simple de partager un taxi.

## Rooms • Salles

All scientific activities and annual general meetings will take place in the Beveridge Arts Centre (BAC, number 1 on campus map), except for the plenary sessions and award addresses (am and pm, Monday, Tuesday and Wednesday) that will be in the Festival Theatre (number 48 on campus map). Coffee breaks and poster sessions will also be in the Festival Theatre. Lunch is included with registration, and is served in Wheelock Dining Hall (number 11 on campus map). Lunchtime committee meetings will take place in a marked section of the same dining hall. The welcome reception and the barbecue will take place in the Fountain Commons (number 24 on campus map).

Toutes les activités scientifiques et les assemblées générales annuelles se dérouleront dans le Beveridge Arts Centre (BAC, numéro 1 sur la carte), à l'exception des conférences plénières (matin et après-midi, lundi, mardi et mercredi) qui auront lieu au Festival Theatre (numéro 48 sur la carte). Les pauses-café et les séances d'affichage auront également lieu au Festival Theatre. Les repas de midi, qui sont inclus dans les frais d'inscription, seront servis à la cafétéria Wheelock (numéro 11 sur la carte). Les réunions des différents comités auront aussi lieu dans un endroit réservé de la cafétéria. Le coquetel de bienvenue et le barbecue se dérouleront à Fountain Commons (numéro 24 sur la carte).

## Workshops • Ateliers

The four workshops organized by the four sections will take place on Sunday June 12 from 9:00 am to 4:00 pm in the Beveridge Arts Centre (BAC, number 1 on campus map). The Biostatistics Section workshop will be in room BAC 132, the Survey Methods Section workshop will be in room BAC 206, the Business and Industrial Statistics Section workshop will be in room BAC 141 and the Probability Section workshop will be in BAC 142.

Les quatre ateliers organisés par les quatre groupes auront lieu le dimanche 12 juin de 9h à 16h dans le Beveridge Arts Centre (BAC, numéro 1 sur la carte). L'atelier du groupe de biostatistique aura lieu dans la salle BAC 132, l'atelier du groupe de méthodologie d'enquête aura lieu dans la salle BAC 206, l'atelier du groupe de statistique industrielle et de gestion aura lieu dans la salle BAC 141 et l'atelier du groupe de probabilité aura lieu dans la salle BAC 142.

## Poster Session • Séance d'affichage

All poster and case studies sessions will take place in the Festival Theatre either on Monday June 13 or Tuesday June 14 between 8:30 am and 5:30 pm. Participants will be able to interact with authors during the afternoon coffee break and the second afternoon session 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 Festival Theatre soit le lundi 13 soit le mardi 14 juin, de 8h30 à 17h30. Les congressistes pourront interagir avec les auteurs pendant la pause-café et la seconde séance de l'après-midi, moment où les auteurs sont invités à être présents près de leur affiche.

## Information Session: Long-Range Planning Exercise in Mathematics and Statistics • Séance d'information : exercice de planification en mathématiques et statistique

**Sunday, June 12, 4:00 – 5:30 pm • Dimanche 12 juin, 16h – 17h30 ..... BAC 138**

The steering committee for the Long Range Planning exercise for mathematics and statistics has been meeting regularly since September 2010. Nancy Reid will give a short presentation summarizing progress to date and then open the floor for questions and feedback.

Le Comité de direction du Plan à long terme en mathématiques et statistique se réunit régulièrement depuis septembre 2010. Nancy Reid donnera une brève présentation résumant les travaux à ce jour, puis ouvrira le débat et répondra à vos questions.

## NSERC Discovery Grant Application Workshop • Atelier sur les demandes de subventions à la découverte du CRSNG

**Monday, June 13 1:30 – 3:00pm • Lundi 13 juin, 1h30 – 3h ..... Acadia Room, KC Irving Centre**

This Workshop will be presented by NSERC Research Grants staff and will cover the Form 180 process, the Discovery Grant evaluation process (principles, criteria and ratings), the Conference Model and tips for preparing a Discovery Grant application. Following the Workshop, there will be an opportunity for participants to ask questions.

Cet atelier sera présenté par le personnel des subventions de recherche du CRSNG et couvrira le processus du formulaire 180, le processus d'évaluation des subventions à la découverte (principes, critères et cotes), le modèle de conférence et des conseils sur la préparation d'une demande. À la fin de l'atelier les participants auront l'occasion de poser des questions.

## Information on 2011 NSERC Discovery Grant Competition • Information sur le concours 2011 des subventions à la découverte du CRSNG

**Tuesday June 14, 5:30 – 6:30pm • Mardi 14 juin, 17h30 – 18h30 ..... BAC 236**

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 2011 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 2011 seront fournies. Tous les chercheurs sont les bienvenus.



## Funding Opportunities from Mitacs • Possibilités de financement par Mitacs

**Tuesday June 14, 12:15 - 1:00pm • Mardi 14 juin, 12h15 - 13h.....BAC 244**

Susan Michel, Director Business Development (Nova Scotia), Mitacs, Inc. - sponsored by the Biostatistics Section  
This session will focus on current funding opportunities that link industry and academia and will highlight examples of successful collaborative projects within the Atlantic region.

Susan Michel, Directrice du développement commercial (Nouvelle-Écosse), Mitacs, Inc. - commandité par le groupe de biostatistique

Cette séance explorera les possibilités actuelles de financement à la frontière de l'industrie et de l'université et présentera des exemples de projets de collaboration menés à bien dans la région Atlantique.

## Lunch • Déjeuner

Lunch on Monday, Tuesday and Wednesday is included for all participants. You will receive a swipe card with credit for these meals. Lunch is served in the Wheelock Dining Hall (11 on campus map). Committee meetings occurring at lunch will be located in a marked area of the dining hall.

Les repas de midi sont inclus pour tous les congressistes lundi, mardi et mercredi. Vous recevrez une carte magnétique créditée pour ces repas. Le déjeuner sera servi à la cafétéria Wheelock (numéro 11 sur la carte). Les déjeuners de travail des comités auront lieu dans un endroit réservé de la cafétéria.

## Vineyard tour • Tournée des vignobles

Conference participants who have purchased tickets for this event will find them in their conference bag. Visit 3 local vineyards (Domaine de Grand Pré, L'Acadie Vineyards and Gaspereau Vineyards). Includes transportation, tastings, tour guide and complimentary wine tote bag. The tour will be divided into 3 groups of 8-10 people, with groups leaving at 1:00, 1:15 and 1:30. See your ticket for your departure time. Tours begin and end at the Festival Theatre.

Les participants au congrès qui ont acheté des billets pour cet événement les trouveront dans le sac du congressiste. Visitez 3 vignobles de la région (Domaine de Grand Pré, L'Acadie Vineyards et Gaspereau Vineyards). Inclut : transport, dégustations, guide touristique et sac fourre-tout souvenir. La visite sera divisée en trois groupes de huit à dix personnes, qui partiront à 13h, 13h15 et 13h30 respectivement. Votre billet indique l'heure de votre départ. La visite débute et finit au Festival Theatre.

## Welcome Reception • Réception de bienvenue

We cordially invite all conference participants to a welcome reception in the Fountain Commons (number 24 on campus map) on Sunday June 12th from 5:30 pm to 7:30 pm. Coupons for free drinks are available in the conference bags.

Nous invitons cordialement tous les participants à la conférence à une réception de bienvenue au Fountain Commons le dimanche 12 juin de 17h30 à 19h30. Des coupons donnant droit à des consommations gratuites se trouvent dans le sac du congressiste.

## Barbecue

A barbecue is organized in the Fountain Commons on Monday June 13th from 6:30 pm until 8:30 pm. 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 au Fountain Commons le lundi 13 juin de 18h30 à 20h30. Tous les étudiants gradués ainsi que les autres participants qui ont acheté des billets pour cet événement les trouveront dans le sac du congressiste.

## Lobster dinner • Dîner au homard

Conference participants who have purchased tickets for this event will find them in their conference bag. We will dine at the Hall's Harbour Lobster Pound & Restaurant. Fee includes a bus to Hall's Harbour, which will depart Wolfville from the Festival Theatre at 6:30pm (a second bus will leave slightly after 6:30pm)

Les participants au congrès qui ont acheté des billets pour cet événement les trouveront dans le sac du congressiste. Nous dînerons à Hall's Harbour Lobster Pound & Restaurant. Le billet inclut le transport vers Hall's Harbour en bus, qui quittera Wolfville (Festival Theatre) à 18h30 (un deuxième bus partira quelques minutes plus tard).

## Banquet

The conference banquet will take place at the Old Orchard Inn (153 Greenwich Road South) the evening of Tuesday June 14. A cocktail with cash bar will start at 18:30 and dinner service will start at 19:30. 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.

Transportation by bus from the Festival Theatre to the Old Orchard Inn will be offered to conference participants. There will be several departures between 6:15 pm and 7:30 pm. Buses will also be available to bring participants back from Old Orchard Inn to the Festival Theatre until 11:00 pm.

Le banquet du congrès se tiendra au Old Orchard Inn (153 rue Greenwich Sud) en soirée le mardi 14 juin. Un coquetel avec bar payant débutera à 18h30 et le souper sera servi à partir de 19h30. Tous les participants au congrès qui n'ont pas décliné leur participation au banquet retrouveront dans le sac du congressiste un billet de couleur qu'ils devront déposer sur leur table pour indiquer leur choix de menu.

Un transport en autobus du Festival Theatre jusqu'au Old Orchard Inn sera offert aux congressistes. Plusieurs départs auront lieu entre 18h15 et 19h30. Des autobus seront également disponibles pour transporter les congressistes depuis le Old Orchard Inn jusqu'au Festival Theatre jusqu'à 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

The SSC Committee on Women in Statistics and the Caucus for Women in Statistics have organized an informal supper at 7:00 pm, Sunday June 12, 2011 at The Ivy Deck Garden Bistro during the 2011 SSC Annual Meeting in Wolfville. Interested women and men are asked to email [jhorrock@uoguelph.ca](mailto:jhorrock@uoguelph.ca) by June 8. This event is not included in the registration fee. See the web site for more information. See you there!

The Ivy Deck Garden Bistro, 8 Elm Avenue Wolfville, 902-542 1868, <http://www.ivydeckbistro.ca/>

Le Comité sur les femmes en statistique de la SSC et le Caucus for Women in Statistics organisent un dîner informel à 19h, le dimanche 12 juin 2011 au Ivy Deck Garden Bistro, dans le cadre du congrès annuel 2011 de la SSC à Wolfville. Toutes les intéressées et tous les intéressés sont demandés d'envoyer un courriel à [jhorrock@uoguelph.ca](mailto:jhorrock@uoguelph.ca) avant le 8 juin. Cet événement n'est pas inclus dans les frais d'inscription. Veuillez voir le site Web pour plus de renseignements. En espérant vous y voir !

Ivy Deck Garden Bistro, 8 avenue Elm, Wolfville, 902-542-1868, <http://www.ivydeckbistro.ca/>

## uWaterloo Alumni Lunch • Dîner des anciens étudiants de uWaterloo

The Alumni of the University of Waterloo will host a luncheon at Acton's Grill and Café (406 Main Street, a short walk from Acadia Campus), Wolfville, Monday 12:00 - 1:30pm. Open to uWaterloo alumni, lunch is provided by the Alumni association.

Les anciens étudiants de la University of Waterloo organisent un dîner au Acton's Grill and Café (406 rue Main, à quelques minutes à pied du campus d'Acadia), Wolfville, lundi entre midi et 13h30. Les anciens étudiants de uWaterloo sont les bienvenus, le repas est offert par l'association des anciens.

## New Investigators informal get together • Rencontre informelle des nouveaux chercheurs

The New Investigators Committee is hosting an informal get together on Monday June 13. Join them in the upper hall of Paddy's Pub at 460 Main Street (a short walk from Acadia Campus). We will have dinner starting from 7:30pm. People will pay for meals and drinks individually.

Le Comité des nouveaux chercheurs organise une rencontre informelle lundi 13 juin. Rejoignez-les à l'étage de Paddy's Pub au 460 rue Main (à quelques minutes à pied du campus d'Acadia). Nous souperons à partir de 19h30. Les participants devront régler leur repas et consommations.

## Internet Access • Accès internet

Acadia has a free wireless network (called *Acadia Open - Web Only*) that provides web access only. No password is required. This network will be available in many areas of the BAC and Festival Theatre. In addition, a wired network provides full internet access.

To access the wired network, open a web browser and use the userid *statsociety* and password *conference*. Participants wanting to use the wired network are asked to bring a network cable, and are advised that there will be a 5-minute wait after you first login in using the userid and password given.

Acadia dispose d'un réseau sans fil gratuit (*Acadia Open - Web Only*) qui offre un accès limité à la navigation Web. Aucun mot de passe n'est requis. Ce réseau sera disponible dans la plupart des salles du BAC et du Festival Theatre. Par ailleurs, un réseau câblé offre un accès sans restriction à l'Internet.

Pour accéder au réseau câblé, ouvrez un navigateur Web et utilisez le nom d'utilisateur *statsociety* et le mot de passe *conference*. Nous demandons aux congressistes qui souhaitent utiliser le réseau câblé d'apporter leur propre câble réseau ; sachez que vous devrez attendre 5 minutes pour vous connecter après avoir saisi le nom d'utilisateur et le mot de passe indiqués ci-dessus.

## 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.

## Non-smoking campus • Campus non fumeur

Effective July 1, 2006 the Acadia campus became tobacco-free. Smoking is not permitted inside any building on campus and is only allowed at select locations on the edge of campus.

Le campus d'Acadia est devenu non-fumeur au 1er juillet 2006. Il est interdit de fumer dans les édifices du campus ; vous trouverez quelques coins fumeur en bordure de celui-ci.

## Committees and Meetings • Comités et réunions

### Saturday June 11

samedi 11 juin

18:00 - 22:00 Executive Committee 1 / Comité exécutif 1, Old Orchard Inn, Boardroom 303

### Sunday June 12

dimanche 12 juin

9:00 - 10:30 Finance Committee / Comité des finances, BAC 137

10:30 - 12:00 Publications Committee / Comité des publications, BAC 201

10:30 - 12:00 Department Heads Meeting / Rencontre des directeurs de départements, BAC 138

12:00 - 16:00 Board of Directors 1 / Conseil d'administration 1, KCIC Acadia Room

### Monday June 13

lundi 13 juin

11:45 - 13:00 Business and Industrial Statistics Section Officers 1 / Groupe de statistique industrielle et de gestion directeurs 1, Wheelock Dining Room

11:45 - 13:00 Statistical Education Committee / Comité d'éducation en statistique, Wheelock Dining Room

11:45 - 13:00 Committee on Women in Statistics / Comité sur les femmes en statistique, Wheelock Dining Room

11:45 - 13:00 Biostatistics Section Officers 1 / Groupe de biostatistique directeurs 1, Wheelock Dining Room

11:45 - 13:00 Probability Section Officers 1 / Groupe de probabilité directeurs 1, Wheelock Dining Room

11:45 - 13:00 Survey Methods Section Officers / Groupe de méthodes d'enquête directeurs, Wheelock Dining Room

12:00 Award for Case Studies in Data Analysis Committee 1 / Comité du prix pour les études de cas et l'analyse de données 1, Festival Theatre Room 152

17:30 - 18:30 SSC Annual General Meeting / Assemblée générale annuelle de la SSC, BAC 241

### Tuesday June 14

mardi 14 juin

10:15 Award for Case Studies in Data Analysis Committee 2 / Comité du prix pour les études de cas et l'analyse de données 2, Festival Theatre Room 152

11:45 - 13:00 CJS Editorial Board / Comité de rédaction de la RCS, Wheelock Dining Room

11:45 - 13:00 Awards Committee / Comité des prix, Wheelock Dining Room

11:45 - 13:00 Bilingualism Committee / Comité du bilinguisme, Wheelock Dining Room

11:45 - 13:00 Program Committee / Comité du programme, Wheelock Dining Room

11:45 - 13:00 Public Relations Committee / Comité des relations publiques, Wheelock Dining Room

12:15 - 13:00 Accreditation General Meeting / Assemblée générale des membres accrédités, BAC 235

17:30 - 18:30 Student Research Presentation Award / Comité du prix pour les présentations étudiants, BAC 207

17:30 - 18:30 Business and Industrial Statistics Section General Meeting / Groupe de statistique industrielle et de gestion assemblée générale, BAC 141

17:30 - 18:30 Biostatistics Section General Meeting / Groupe de biostatistique assemblée générale, BAC 132

17:30 - 18:30 Probability Section General Meeting / Groupe de probabilité assemblée générale, BAC 138

17:30 - 18:30 Survey Methods Section General Meeting / Groupe de méthodes d'enquête assemblée générale, BAC 142

### Wednesday June 15

mercredi 15 juin

11:45 - 13:00 Business and Industrial Statistics Section Officers 2 / Groupe de statistique industrielle et de gestion directeurs 2, Wheelock Dining Room

11:45 - 13:00 Accreditation Committee / Comité sur l'accréditation, Wheelock Dining Room

11:45 - 13:00 Research Committee / Comité de recherche, Wheelock Dining Room

11:45 - 13:00 Liaison, Wheelock Dining Room

BAC = Beveridge Arts Centre

KCIC = K. C. Irving Centre

11:45 - 13:00	Biostatistics Section Officers 2 / Groupe de biostatistique directeurs 2, Wheelock Dining Room
11:45 - 13:00	Probability Section Officers 2 / Groupe de probabilité directeurs 2, Wheelock Dining Room
15:00 - 16:30	Board of Directors 2 / Conseil d'administration 2, KCIC Acadia Room
16:30 - 17:30	Executive Committee 2 / Comité exécutif 2, KCIC Acadia Room

## Social and Information Events • Événements sociaux et informatifs

### Sunday June 12

**dimanche 12 juin**

13:00 - 17:30	Vineyard Tour - fee event / Tournée des vignobles - payant, Bus at / arrêt d'autobus au Festival Theatre
16:00 - 17:30	Long-Range Planning Exercise in Mathematics and Statistics / Exercice de planification à long terme en mathématiques et en statistique, BAC 138
17:30 - 19:30	Welcome Reception / Réception de bienvenue, Fountain Commons
19:00	SSC Committee on Women in Statistics and the Caucus for Women in Statistics Informal Supper - Interested women and men are asked to email jhorrock@uoguelph.ca by June 8 / Dîner informel Le Comité sur les femmes en statistique de la SSC et le Caucus for Women in Statistics - Toutes les intéressées et tous les intéressés sont demandés d'envoyer un courriel à jhorrock@uoguelph.ca avant le 8 juin, Off campus / en dehors du campus - Ivy Deck Garden Bistro

### Monday June 13

**lundi 13 juin**

9:45 - 10:15	Break / Pause-café, Festival Theatre
11:45 - 13:00	Lunch / Déjeuner, Wheelock Dining Hall
12:00 - 13:30	uWaterloo Alumni Lunch - registration is required by Wednesday, June 8 / Dîner des anciens étudiants de uWaterloo - l'inscription est obligatoire avant mercredi 8 juin au plus tard, Off campus / en dehors du campus - Acton's Bar and Grill
13:30 - 15:00	NSERC Discovery Grant Workshop / Atelier des subventions à la découverte du CRSNG, KCIC Acadia Room
14:30 - 15:00	Break / Pause-café, Festival Theatre
14:30 - 16:00	Posters from Canadian winners of the International Statistical Literacy Competition / Séance d'affichage: International Statistical Literacy Competition, Festival Theatre
18:30	Lobster Dinner - fee event / Dîner au homard - payant, Hall's Harbour - Bus at / arrêt d'autobus au Festival Theatre
18:30 - 20:30	Student BBQ - free for graduate students / Barbecue étudiants - gratuit pour les étudiants de cycle supérieur, Fountain Commons
19:30	New Investigators informal get together / Rencontre informelle des nouveaux chercheurs, Off campus / en dehors du campus - Paddy's Pub

### Tuesday June 14

**mardi 14 juin**

9:45 - 10:15	Break / Pause-café, Festival Theatre
11:45 - 13:00	Lunch / Déjeuner, Wheelock Dining Hall
12:15 - 13:00	Funding Opportunities from Mitacs Inc. / Occasions de financement par le Mitacs Inc., BAC 244
14:30 - 15:00	Break / Pause-café, Festival Theatre
14:30 - 16:00	Posters from Canadian winners of the International Statistical Literacy Competition / Séance d'affichage: International Statistical Literacy Competition, Festival Theatre
17:30 - 18:30	Information on 2011 NSERC Discovery Grant Competition / Information sur le concurs 2011 des subventions a la découverte du CRSNG, BAC 236
18:30 - 23:00	Banquet, Old Orchard Inn - Bus at / arrêt d'autobus au Festival Theatre

### Wednesday June 15

**mercredi 15 juin**

9:45 - 10:15	Break / Pause-café, Festival Theatre
11:45 - 13:00	Lunch / Déjeuner, Wheelock Dining Hall

BAC = Beveridge Arts Centre

KCIC = K. C. Irving Centre

## List of Sessions • Liste des sessions

- 1A-A1 Inaugural Session  
Séance inaugurale
- 1A-A2 SSC Presidential Invited Address  
Allocution de l'invité du Président de la SSC
- 1B-I1 Future of Functional Data Analysis  
L'avenir de l'analyse de données fonctionnelle
- 1B-I2 Imperfect Administrative Health Data: Can we do better?  
Données de santé administratives imparfaites : pouvons-nous faire mieux?
- 1B-I3 Statistical Issues in Longitudinal Data Analysis  
Problèmes statistiques en analyse de données longitudinales
- 1B-I4 Statistics in Sports  
Statistique dans le domaine des sports
- 1B-I5 Stochastic Analysis  
Analyse stochastique
- 1B-C1 Methods for Inference  
Méthodes pour l'inférence
- 1B-C2 Spatial Modelling for Infectious Disease  
Modélisation spatiale des maladies infectieuses
- 1B-C3 Survey Methods: Data Collection, Edit and Imputation  
Méthodes d'enquête : collecte de données, vérification et imputation
- 1C-I1 Advances in Fisherian Alternatives to Conventional Inference  
Progrès dans les alternatives fisheriennes à l'inférence conventionnelle
- 1C-I2 Methods for Human and Animal Surveillance  
Méthodes de surveillance des humains et des animaux
- 1C-I3 Nonparametric Function Estimation  
Estimation des fonctions non paramétriques
- 1C-I4 Random Effects Modelling for Spatial Data  
Modélisation des effets aléatoires appliquée aux données spatiales
- 1C-I5 Sampling & Biometry  
Sondage et biométrie
- 1C-C1 Estimation in Linear Models  
Estimation dans les modèles linéaires
- 1C-C2 Probability Applications in Finance and Economics  
Applications de la probabilité en finance et économie
- 1C-C3 Survival Models: Predictive and Causal Analysis  
Modèles de survie : analyse prévisionnelle et causale
- 1D-D Case Study 2: Determinants of the Presence and Volume of Brown Fat in Humans  
Étude de cas 2 : Déterminants de la présence et du volume de graisse brune chez les humains
- 1D-P Poster Session 1  
Séance d'affichage 1
- 1E-A SSC Impact Award Address  
Allocution du récipiendaire du prix pour impact de la SSC
- 1F-I1 Developing Teaching Skills in Graduate Students  
Comment former les étudiants de cycle supérieur à l'enseignement
- 1F-I2 Measuring Water Quality  
Comment mesurer la qualité des eaux
- 1F-I3 New Developments in (Disease and) Syndromic Surveillance  
Du neuf en surveillance (des maladies et) syndromique
- 1F-I4 Recent Advances in the Analysis of Correlated Data  
Récents progrès en analyse de données corrélées



- 1F-I5 Robust Designs of Statistical Experiments  
Conception robuste d'expériences statistiques
- 1F-C1 CANNeCTIN Methodology for Clinical Trials  
Méthodologie du CANNeCTIN pour les essais cliniques
- 1F-C2 Distribution Theory and Asymptotics  
Théorie de la distribution et asymptotique
- 1F-C3 Survey Methods: Precision and Analysis  
Méthodes d'enquête : précision et analyse
- 2A-A Gold Medal Address  
Allocution de la récipiendaire de la médaille d'or
- 2B-11 Recent Advances in MCMC  
Récents progrès en MCMC
- 2B-12 Recent Developments in Multiple Comparison Procedures  
Du neuf dans les procédures de comparaisons multiples
- 2B-13 Recent Developments in Shrinkage Estimation with Related Implications for Predictive Density Estimation, Loss Estimation and Model Selection  
Du neuf dans l'estimation par rétrécissement et implications quant à l'estimation de la densité prédictive, l'estimation des pertes et la sélection de modèles
- 2B-14 Sampling Theory and Practice: In Honour of Mary Thompson's Retirement  
Théorie et pratique de l'échantillonnage : À l'occasion du départ à la retraite de Mary Thompson
- 2B-15 Statistical Consulting: Communication, Interpersonal, Project Management Skills Required  
Conseil statistique : compétences requises en communications, interpersonnel et gestion de projet
- 2B-C1 High-Dimensional Data in Biology  
Données de grande dimension en biologie
- 2B-C2 Methods in Mixture Models and Spatial Statistics  
Méthodes dans les modèles de mélange et spatiale statistique
- 2B-C3 Spatial-temporal Modelling in the Natural World  
Modélisation spatio-temporelle dans le monde naturel
- 2C-11 Designs for Generalized Linear Models  
Conception de modèles linéaires généralisés
- 2C-12 New Developments in Mixture Models & Their Applications  
Du neuf dans les modèles de mélange et leurs applications
- 2C-13 On the Issues Raised by Environmental Surveys at Statistics Canada  
Au sujet des questions soulevées par les études environnementales chez Statistique Canada
- 2C-14 Panel on Online Teaching of Statistics  
Table ronde sur l'enseignement de la statistique en ligne
- 2C-15 Sequential Designs for Computer Experiments  
Conception séquentielle d'expériences informatiques
- 2C-C1 Distributions: Theory and Application  
Distributions : théorie et application
- 2C-C2 Genetic Association and Analysis of Families  
Association génétique et analyse des familles
- 2C-C3 Optimizing Products and Processes  
Optimisation des produits et processus
- 2C-C4 Survival Models: Correlated Data and Non-parametric Estimation  
Modèles de survie : données corrélées et estimation non paramétrique
- 2D-D Case Study 1: Gender Gap in Earnings among Young People  
Étude de cas 1 : Écart salarial entre les jeunes hommes et femmes
- 2D-P Poster Session 2  
Séance d'affichage 2
- 2E-A CJS Award Address  
Allocution du récipiendaire du Prix de la RCS
- 2C-A Isobel Loutit Invited Address  
Allocution Isobel Loutit

- 2F-A Why a Bayesian be?  
Conférence de vue d'ensemble : Pourquoi être Bayésien?
- 2F-I1 Financial Time Series  
Séries chronologiques financières
- 2F-I2 Gene-environment Effects in Pharmacogenetic and Related Settings  
Effets génotype – environnement en pharmacogénétique et domaines connexes
- 2F-I3 Statistics Literacy  
Littératie en statistique
- 2F-C1 Clustering and Classification  
Mise en grappes et classification
- 2F-C2 Longitudinal Models: Measurement Error, Missing Data, and Misspecification  
Modèles longitudinaux : erreurs de mesure, données manquantes et erreurs de spécification
- 2F-C3 Survey Methods: Estimation  
Méthodes d'enquête : estimation
- 3A-A Pierre Robillard Award Address  
Allocution du récipiendaire du Prix Pierre-Robillard
- 3B-A Approaches to Pathway Modeling in Genetic Epidemiology  
Approches de la modélisation des voies en épidémiologie génétique
- 3B-I1 Ecological Sampling  
Échantillonnage écologique
- 3B-I2 Order Statistics and Related Inference, Ranked Set Sampling and Censoring Methodology  
Statistiques d'ordre et inférence connexe, échantillonnage d'ensembles ordonnés et méthodologie de censure
- 3B-I3 Psychographic Surveys and Behavioural Surveys - Can information from one be used to accurately predict characteristics of the other?  
Enquêtes psychographiques et enquêtes sur le comportement – peut-on utiliser des informations provenant des unes pour prévoir avec précision les caractéristiques des autres?
- 3B-I4 Randomness in Discrete Mathematics and Biology  
Stochastocité en mathématiques discrètes et biologie
- 3B-I5 Statistical Problems Arising in the Lumber Industry  
La statistique en assurance de la qualité du bois d'œuvre
- 3B-C1 Design and Analysis of Clinical Studies and Trials  
Conception et analyse d'études et d'essais cliniques
- 3B-C2 Harmonic Analysis and Spectral Methods for Time Series  
Analyse harmonique et méthodes spectrales pour les séries chronologiques
- 3B-C3 Modelling Correlation Structure in Repeated Measures and Multivariate Data  
Modélisation de la structure de corrélation dans les mesures répétées et les données multivariées
- 3C-A CRM-SSC Prize in Statistics Address  
Allocution du récipiendaire du Prix CRM-SSC
- 3C-I1 Consultations for Statistical Consultants: Lessons Learned from the Field  
Consultations pour les consultants en statistique : leçons apprises sur le terrain (discussion en groupe)
- 3C-I2 Rank Set Sampling: New Developments and Applications  
Échantillonnage d'ensembles ordonnés : récents développements et applications
- 3C-I3 Recent Advances in Multivariate Control Charts  
Limites de contrôle multivariées

## Program • Programme

**Sunday June 12**

**dimanche 12 juin**

**9:00 - 16:00**

**BAC 132**

**Workshop of the Biostatistics Section: A Practical Introduction to Hierarchical Modelling for Spatially Referenced Data**

**Atelier du Groupe de biostatistique : Modélisation statistique pour les données géographiques**

Organizer/Responsable: Lisa Lix (University of Saskatchewan)

9:00 - 16:00

Sudipto Banerjee (School of Public Health, University of Minnesota)

**9:00 - 16:00**

**BAC 141**

**Workshop of the Business and Industrial Statistics Section: Developing Reliable and Usable Software in R**

**Atelier du Groupe de statistique industrielle et de gestion : Comment développer des logiciels fiables et utilisables en R**

Organizer/Responsable: Will Welch (University of British Columbia)

9:00 - 16:00

Duncan Murdoch (University of Western Ontario)

**9:00 - 16:00**

**BAC 142**

**Workshop of the Probability Section: Parallel Computing in R**

**Atelier du Groupe de probabilité : Traitement parallèle en R**

Organizer/Responsable: Yiqiang Zhao (Carleton University)

9:00 - 16:00

Hao Yu (University of Western Ontario)

**9:00 - 16:00**

**BAC 206**

**Workshop of the Survey Methods Section: Using STATA for Survey Data Analysis**

**Atelier du Groupe des méthodes d'enquête : Analyse de données d'enquête via STATA**

Organizer/Responsable: Pierre Lavallée (Statistique Canada)

9:00 - 16:00

Stanislav Kolenikov (University of Missouri)

**16:00 - 17:30**

**BAC 138**

**Long-Range Planning Exercise in Mathematics and Statistics**

**Exercice de planification à long terme en mathématiques et en statistique**

**Monday June 13****lundi 13 juin**

<b>8:15 - 8:30</b>	<b>Session 1A-A1</b>		<b>Festival Theatre</b>
<b>Inaugural Session</b> <b>Séance inaugurale</b>			
Special Session / Conférence spéciale Organizers/Responsables: Hugh Chipman (Acadia University) and/et Don McLeish (University of Waterloo)			
8:15 - 8:20	Ray Ivany (President and Vice-Chancellor, Acadia University)		
8:20 - 8:25	Don McLeish (SSC President)		
8:25 - 8:30	Hugh Chipman (SSC2011 Local Arrangements Chair)		
<b>8:30 - 9:45</b>	<b>Session 1A-A2</b>	(abstracts/résumés - page 53)	<b>Festival Theatre</b>
<b>SSC Presidential Invited Address</b> <b>Allocution de l'invité du Président de la SSC</b>			
Special Session / Conférence spéciale Chair/Président: Don McLeish (University of Waterloo)			
8:30 - 9:45	Ivan Fellegi (Chief Statistician of Canada Emeritus) Official Statistics, Public Confidence and Lessons from the Story of the 2011 Canadian Census / Statistiques officielles, confiance du public et leçons tirées du Recensement canadien 2011		
<b>9:45 - 10:15</b>			<b>Festival Theatre</b>
<b>Break / Pause-café</b>			
<b>10:15 - 11:45</b>	<b>Session 1B-I1</b>	(abstracts/résumés - page 54)	<b>BAC236</b>
<b>Future of Functional Data Analysis</b> <b>L'avenir de l'analyse de données fonctionnelle</b>			
Invited Paper Session / Présentations sur invitation Organizer and Chair / Responsable et président: Jiguo Cao (Simon Fraser University)			
10:15 - 10:45	Fang Yao (University of Toronto), Hans George Mueller (University of California at Davis) Additive Modeling of Functional Gradients / Modélisation additive de gradients fonctionnels		
10:45 - 11:15	Jason D Nielsen (Carleton University) A Functional Approach to the Analysis of Recurrent Event Data / Une approche fonctionnelle à l'analyse des données d'évènements récurrents		
11:15 - 11:45	Sofia Mosesova (Genentech), Hugh Chipman (Acadia University) Bayesian Hierarchical Modeling of Functional Data / Modélisation bayésienne hiérarchique pour données fonctionnelles		
<b>10:15 - 11:45</b>	<b>Session 1B-I2</b>	(abstracts/résumés - page 56)	<b>BAC 132</b>
<b>Imperfect Administrative Health Data: Can we do better?</b> <b>Données de santé administratives imparfaites : pouvons-nous faire mieux ?</b>			
Invited Paper Session / Présentations sur invitation Organizer and Chair / Responsable et président: Lisa Lix (University of Saskatchewan) Session sponsored by the Biostatistics Section Session parrainée par le Groupe de biostatistique			
10:15 - 10:45	Yulei He (Harvard University), Mary Beth Landrum & Alan Zaslavsky (Harvard Medical School) Combining Information on Health Services Variables from Multiple Sources / Comment combiner des informations de sources multiples sur les variables de services de santé		

- 10:45 - 11:15 Martin Ladouceur, Elham Rahme, Christian Pineau & Lawrence Joseph (McGill University)  
Investigating Robustness of Prevalence Estimates from Misclassified Administrative Data Using Bayesian Latent Class Models / Étude de la robustesse des estimations de la prévalence à partir de données administratives mal classées par l'utilisation de modèles de classes latentes
- 11:15 - 11:45 Khokan Sikdar & Jeffrey Dowden (Newfoundland and Labrador Centre for Health Information), Joanne Stares (Public Health Agency of Canada), Josh Squires & Kayla Collins (Newfoundland and Labrador Centre for Health Information)  
Challenges and Opportunities for Using Administrative Health Data to Improve Chronic Disease Case Ascertainment / Défis et possibilités d'utilisation des données de gestion de santé pour améliorer la détermination des cas de maladie chronique

---

**10:15 - 11:45      Session 1B-I3** (abstracts/résumés - page 58) **BAC239**

**Statistical Issues in Longitudinal Data Analysis**

**Problèmes statistiques en analyse de données longitudinales**

Invited Paper Session / Présentations sur invitation

Chair/Président: Julie Horrocks (University of Guelph)

Organizer/Responsable: Leilei Zeng (Simon Fraser University)

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

- 10:15 - 10:45 Cecilia Cotton (University of Waterloo)  
Causal Effect Estimation for Longitudinal Survival Data / Estimation d'effets causaux pour données longitudinales de survie
- 10:45 - 11:15 Juli Atherton & David Wolfson (McGill University), Edens Thaddeus  
Optimal Design for Clinical Trials with an Unknown Delay in Treatment Effect / Plan optimal pour essais cliniques avec retard inconnu de l'effet du traitement
- 11:15 - 11:45 Yang Zhao & Meng Liu (University of Regina)  
A Uniform Approach for Longitudinal Data with Data Missing in Nonmonotonic Patterns / Une approche uniforme aux données longitudinales avec données manquantes de façon non-monotone

---

**10:15 - 11:45      Session 1B-I4** (abstracts/résumés - page 60) **BAC237**

**Statistics in Sports**

**Statistique dans le domaine des sports**

Invited Paper Session / Présentations sur invitation

Chair/Président: Reg Kulperger (University of Western Ontario)

Organizer/Responsable: Paramjit Gill (University of British Columbia Okanagan)

- 10:15 - 10:45 Bill Hurley (Royal Military College of Canada), Andrey Pavlov (Queen's University)  
A Valuation Model for NFL and NHL Draft Positions / Modèle d'évaluation des rangs au repêchage de la LNF et de la LNH
- 10:45 - 11:15 Tim Swartz (Simon Fraser University)  
Desperation in Sport / Du désespoir chez les sportifs
- 11:15 - 11:45 Paramjit Gill (University of British Columbia, Okanagan), Michael Davis (UBC Okanagan)  
The Third Period in an NHL Game / La troisième période d'un match de la LNH

<b>10:15 - 11:45</b>	<b>Session 1B-15</b>	(abstracts/résumés - page 62)	<b>BAC138</b>
<b>Stochastic Analysis</b>			
<b>Analyse stochastique</b>			
Invited Paper Session / Présentations sur invitation			
Chair/Président: Rafal Kulik (University of Ottawa)			
Organizer/Responsable: Raluca Balan (University of Ottawa)			
Session sponsored by the Probability Section			
Session parrainée par le Groupe de probabilité			
10:15 - 10:45	Shui Feng (McMaster University)	Gamma-Dirichlet Structure / Structure de Gamma-Dirichlet	
10:45 - 11:15	Cody Hyndman (Concordia University)	Existence and Uniqueness for Nonlinear FBSDEs Related to Quadratic Term-Structure Models. / Existence et unicité des équations différentielles stochastiques forward-backward (FBSDE) non linéaires liées aux modèles avec structure quadratique	
11:15 - 11:45	Wei Sun (Concordia University), Zechun Hu (Nanjing University)	Hunt's Hypothesis (H) and Gettoor's Conjecture for Lévy Processes / Hypothèse de Hunt et conjecture de Gettoor pour les processus de Lévy	
<b>10:15 - 11:30</b>	<b>Session 1B-C1</b>	(abstracts/résumés - page 64)	<b>BAC141</b>
<b>Methods for Inference</b>			
<b>Méthodes pour l'inférence</b>			
Contributed Paper Session / Séance de communications libres			
Chair/Président: JC Loredano-Osti (Memorial University of Newfoundland)			
10:15 - 10:30	Michael Stephens & Richard Lockhart (Simon Fraser University), Alberto Contreras (IIMAS – UNAM)	Use of Priors in Goodness-of-Fit / Utilisation des lois a priori pour la qualité de l'ajustement	
10:30 - 10:45	Zeynep Baskurt & Michael Evans (University of Toronto)	Hypothesis Assessment via the Method of Concentration / Évaluation d'hypothèses via la méthode de concentration	
10:45 - 11:00	Osama Bataineh & Mikelis Bickis (University of Saskatchewan)	Imprecise Predictive Inference for Logistic Regression / Inférence prédictive imprécise pour la régression logistique	
11:00 - 11:15	Ximing Xu & Nancy Reid (University of Toronto)	More or Less: Two Paradoxes on the Selection of Composite Likelihood / Plus ou moins : deux paradoxes sur la sélection d'une vraisemblance composite	
11:15 - 11:30	Tharshanna Nadarajah & Asokan Mulayath Variyath (Memorial University of Newfoundland and Labrador)	Penalized Empirical Likelihood Based Variable Selection / Sélection de variables basée sur la vraisemblance empirique pénalisée	
<b>10:15 - 11:45</b>	<b>Session 1B-C2</b>	(abstracts/résumés - page 67)	<b>BAC 234</b>
<b>Spatial Modelling for Infectious Disease</b>			
<b>Modélisation spatiale des maladies infectieuses</b>			
Contributed Paper Session / Séance de communications libres			
Chair/Président: Karen Kopciuk (Alberta Health Services and University of Calgary)			
Session sponsored by the Biostatistics Section			
Session parrainée par le Groupe de biostatistique			

- 10:15 - 10:30 Daniel Gillis, Tony Desmond & John Holt (University of Guelph)  
Disease Source Classification Using Multivariate Spatial Poisson Mixture Models / Classification de sources de maladies à l'aide de modèles de mélanges spatiaux multivariés de Poisson
- 10:30 - 10:45 Ye Li (University of Toronto), Patrick Brown (Cancer Care Ontario), Havard Run (Norwegian University of Science and Technology)  
Geostatistical Model for Spatially Aggregated Disease Incidence Data / Modèles géostatistiques pour données d'incidence de maladies spatialement agrégées
- 10:45 - 11:00 Grace Pui Sze Kwong & Rob Deardon (University of Guelph)  
Computationally Efficient Forms of Spatial Infectious Disease Models / Forme computationnellement efficace pour modèles spatiaux de maladies infectieuses
- 11:00 - 11:15 Jourdan Gold, Rob Deardon & Zeny Feng (University of Guelph)  
Effects of Time-line Uncertainty in Infectious Disease Modelling / Effets de l'incertitude sur les données temporelles dans la modélisation de maladies infectieuses
- 11:15 - 11:30 Lorna Deeth & Rob Deardon (University of Guelph)  
Mixtures of Individual-level Models for Infectious Disease Modelling / Mélanges de modèles au niveau de l'individu pour la modélisation de maladies infectieuses
- 11:30 - 11:45 Rob Deardon, Abbie Gardner & Gerarda Darlington (University of Guelph)  
Posterior Predictive Approach to Goodness-of-Fit for Infectious Disease Models / Approche prédictive a posteriori pour l'adéquation de modèles de maladies infectieuses

---

**10:15 - 11:45      Session 1B-C3      (abstracts/résumés - page 70)**

**BAC142****Survey Methods: Data Collection, Edit and Imputation****Méthodes d'enquête : collecte de données, vérification et imputation**

Contributed Paper Session / Séance de communications libres

Chair/Président: Carlos León (Statistique Canada)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

- 10:15 - 10:30 Qiong Wang & Danielle Laroche (Statistics Canada)  
Identification of the Collection Units which will Receive the Letter Motivating the Response Via Internet for the 2011 Census / Identification des secteurs qui recevront une lettre destinée à promouvoir la réponse par Internet au Recensement de 2011
- 10:30 - 10:45 Kimberley Coleman & Jean-Sébastien Provençal (Statistics Canada)  
A Comparison of Dependent and Independent Interviewing for Collecting Revised Data in the Fruit and Vegetable Survey / Une comparaison de données révisées provenant d'entrevues dépendantes et indépendantes pour l'enquête sur les fruits et légumes
- 10:45 - 11:00 Naima Gouzi & Abdellatif Demnati (Statistique Canada)  
Sample Size Determination for the Integrated Business Statistics Program / Détermination de la taille de l'échantillon dans le cadre du Programme intégré de la statistique des entreprises
- 11:00 - 11:15 Brad Johnson & Mohammad Jafari Jozani (University of Manitoba)  
Randomized Minima-Maxima Nomination Sampling in Finite Populations / Échantillonnage nominatif randomisé des minima-maxima pour des populations finies
- 11:15 - 11:30 Christian Olivier Nambeu (Statistique Canada), David Haziza (Université de Montréal), Guillaume Chauvet (Laboratoire de Statistique d'Enquête, CREST/ENSAI)  
Single Imputation for Populations Containing a Large Amount of Zeroes / Imputation simple de populations contenant un nombre élevé de zéros
- 11:30 - 11:45 Carlos León (Statistique Canada)  
Optimally controlled sampling designs to guard against bad samples / Plans d'échantillonnage contrôlés pour éliminer les mauvais échantillons
-

11:45 - 13:00

Wheelock Dining Hall

Lunch / Déjeuner

13:00 - 14:30

**Session 1C-I1**

(abstracts/résumés - page 73)

**BAC236****Advances in Fisherian Alternatives to Conventional Inference****Progrès dans les alternatives fishériennes à l'inférence conventionnelle**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: David Bickel (University of Ottawa)

13:00 - 13:30

Ao Yuan (Howard University)

Bayesian Frequentist Hybrid Model with Applications to CGH Data Analysis / Modèle hybride fréquentiste bayésien avec applications à l'analyse de données HGC

13:30 - 14:00

Alan Polansky (Northern Illinois University)

Observed Confidence Levels: An Alternative to Multiple Testing Techniques / Niveaux de confiance observés : une alternative aux techniques d'épreuves multiples

14:00 - 14:30

Jeffrey Blume (Vanderbilt University School of Medicine)

Fisher's Likelihood Inference: Coming of Age / Inférence par la vraisemblance de Fisher : l'ère de la maturité

13:00 - 14:30

**Session 1C-I2**

(abstracts/résumés - page 75)

**BAC 132****Methods for Human and Animal Surveillance****Méthodes de surveillance des humains et des animaux**

Invited Paper Session / Présentations sur invitation

Chair/Président: Henrik Stryhn (Atlantic Veterinary College, PEI)

Organizers/Responsables: Lisa Lix (University of Saskatchewan) and/et Henrik Stryhn (Atlantic Veterinary College, PEI)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

13:00 - 13:30

Bernard Choi (Public Health Agency of Canada), Manisha Kulkarni Woodstock, Christine Kennedy, Maxim Trubnikov, Michael Coulthart, Susanna Ogunnaike-Cooke &amp; Purni Narayanan

Informing Human and Animal Health Surveillance: A Catalogue of Epidemiologic and Statistical Biases Revisited / À l'appui de la surveillance de la santé humaine et animale : réexamen d'un catalogue de biais épidémiologiques et statistiques

13:30 - 14:00

Dirk Pfeiffer (Royal Veterinary College, University of London, UK)

Risk Assessment Approaches and their Increasing Importance in the Context of Risk-Based Disease Surveillance Design / Approches de l'évaluation des risques et leur importance grandissante dans le contexte d'une surveillance épidémiologique conçue en fonction du risque

14:00 - 14:30

Crawford Revie, Javier Sanchez &amp; Fernanda Dorea (Atlantic Veterinary College, University of Prince Edward Island)

Syndromic Surveillance: Can Large Data Sets Really Be "Unreasonably Effective"? / Surveillance syndromique : les grands ensembles de données peuvent-ils vraiment être « excessivement efficaces » ?

13:00 - 14:30

**Session 1C-I3**

(abstracts/résumés - page 77)

**BAC237****Nonparametric Function Estimation****Estimation des fonctions non paramétriques**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Alexandre Leblanc (University of Manitoba)



- 13:00 - 13:30 Jean-Francois Angers (Université de Montreal), Brenda MacGibbon (UQAM)  
Function Estimation with Non-negative "wavelets" / Estimation de fonction utilisant des "ondelettes" non négatives
- 13:30 - 14:00 John Braun (University of Western Ontario)  
Kernel Density and Regression Function Estimation using Derivative Information / Estimation de densité et de fonctions de régression par la méthode du noyau utilisant l'information à propos des dérivées
- 14:00 - 14:30 Éric Marchand (Université de Sherbrooke), Mohammad Jafari Jozani & Alexandre Leblanc (University of Manitoba)  
On Estimating a Cumulative Distribution Function with a Balanced Loss Function / Sur l'estimation d'une fonction de répartition avec une fonction de perte pondérée

---

**13:00 - 14:30**      **Session 1C-I4**      (abstracts/résumés - page 79)      **BAC239**

**Random Effects Modelling for Spatial Data**

**Modélisation des effets aléatoires appliquée aux données spatiales**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Tariqul Hasan (University of New Brunswick)

- 13:00 - 13:30 Xin Feng & Charmaine Dean (Simon Fraser University)  
An Overview of Random Effects Models for Spatial Analyses / Vue d'ensemble des modèles d'effets aléatoires pour analyses spatiales
- 13:30 - 14:00 Renjun Ma (University of New Brunswick), Ed Hughes (Ed Hughes Consulting, Ottawa, Canada)  
Tweedie Mixed Models for Spatiotemporal Data. / Modèles mixtes de Tweedie pour les données spatio-temporelles
- 14:00 - 14:30 Gary Sneddon (Mount Saint Vincent University), M. Tariqul Hasan & Renjun Ma (University of New Brunswick)  
Analysis of Cross-Classified Spatial Data / Analyse de données spatiales recoupées

---

**13:00 - 14:30**      **Session 1C-I5**      (abstracts/résumés - page 81)      **BAC142**

**Sampling & Biometry**

**Sondage et biométrie**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Louis-Paul Rivest (Université Laval)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

- 13:00 - 13:30 Stephen Smith (Bedford Institute)  
Application of Sampling with Partial Replacement to Scallop Surveys / Application de l'échantillonnage à remplacement partiel aux enquêtes sur les pétoncles
- 13:30 - 14:00 Noel Cadigan (Fisheries and Oceans Canada)  
Accounting for Survey Variance Can Improve Fish Stock Assessment / Tenir compte de la variance expérimentale peut améliorer l'évaluation des inventaires de poissons
- 14:00 - 14:30 Simon J Bonner (University of Kentucky), Carl J Schwarz (Simon Fraser University)  
Advances in Modelling Time-Stratified Mark-Recapture Data from Salmon Outmigrations via Bayesian P-Splines / Développements dans la modélisation de données stratifiées par le temps de marquage-recapture de la migration des saumons au moyen de p-splines bayésiennes

---

**13:00 - 14:30**      **Session 1C-C1**      (abstracts/résumés - page 83)      **BAC141**

**Estimation in Linear Models**

**Estimation dans les modèles linéaires**

Contributed Paper Session / Séance de communications libres

Chair/Président: Elif Acar (McGill University)

- 13:00 - 13:15 Anne Michele Millar (Mount Saint Vincent University)  
Gain Scores and Regression to the Mean / Scores de gain et régression à la moyenne
- 13:15 - 13:30 Xiaoyu Chen & Xiaojian Xu (Brock University)  
Robust Estimation in Case of Asymmetry / Estimation robuste dans le cas asymétrique
- 13:30 - 13:45 Hensley Hubert Mariathas (Memorial University of Newfoundland), Brajendra Sutradhar & J Concepción Loredó-Osti (Memorial University of Newfoundland)  
Inferences in Linear Mixed Models with Spatially Correlated Random Effects / Inférences pour modèles linéaires mixtes avec effets aléatoires spatialement corrélés
- 13:45 - 14:00 Bingrui Sun (Memorial University), Brajendra Sutradhar (Memorial University of Newfoundland)  
GQL Estimation in Linear Dynamic Models for Panel Data / Estimation du quasi maximum de vraisemblance généralisé pour les jeux de données provenant d'un modèle linéaire dynamique
- 14:00 - 14:15 Mahmoud Torabi (University of Manitoba)  
Likelihood Inference in Generalized Linear Mixed Models with Two Components of Dispersion / Inférence de vraisemblance dans les modèles linéaires mixtes généralisés à deux composantes de dispersion
- 14:15 - 14:30 Marc Sormany (Université de Moncton), Vartan Choulakian (Université de Moncton, Dépt. de Math/Statistique)  
Picture of 2 and 3 Parameters Least Squares Solutions in Rational Regression / Régression rationnelle par moindres carrés : carte des solutions

---

**13:00 - 14:15**      **Session 1C-C2**      (abstracts/résumés - page 86)      **BAC138**

**Probability Applications in Finance and Economics**

**Applications de la probabilité en finance et économie**

Contributed Paper Session / Séance de communications libres

Chair/Président: Boxin Tang (Simon Fraser University)

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

- 13:00 - 13:15 Vickneswary Tagore (University of Saskatchewan), Brajendra C Sutradhar (Memorial University)  
Simple and Efficient Estimation Approaches for Stochastic Volatility Models / Approches simples et efficaces d'estimation des modèles de volatilité stochastique
- 13:15 - 13:30 Zhongxian Men (University of Waterloo)  
Multivariate Stochastic Volatility Models: a Probabilistic Principal Component Analysis Approach / Les modèles de volatilité stochastique multivariés : une approche probabiliste de l'analyse en composantes principales
- 13:30 - 13:45 Zhenyu Cui & Carole Bernard (University of Waterloo), William Gornall (Stanford University)  
Improved Continuity Correction for Discrete Barrier and Lookback Options / Correction de continuité améliorée pour des options discrètes de barrière et rétrospective
- 13:45 - 14:00 Francois Watier & Alexandre Scott (University of Quebec in Montreal)  
Goal Achieving Probability of a Mean-Variance Strategy Without Short-Selling / Probabilité d'atteindre le but d'une stratégie moyenne-variance sans vente à découvert
- 14:00 - 14:15 Alexandre Scott & Francois Watier (University of Quebec in Montreal)  
A Stopping Time Rule for Mean-Variance Portfolios Without Bankruptcy / Un temps d'arrêt dans le temps pour des portefeuilles moyenne-variance sans faillite

---

**13:00 - 14:30**      **Session 1C-C3**      (abstracts/résumés - page 89)      **BAC 234**

**Survival Models: Predictive and Causal Analysis**

**Modèles de survie : analyse prévisionnelle et causale**

Contributed Paper Session / Séance de communications libres

Chair/Président: Tim Ramsay (University of Ottawa)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

13:00 - 13:15	James Hanley & Zhihui Liu (McGill University) Measuring the Performance of Prognostic Models for Survival Data / Mesure de la performance des modèles pronostiques pour des données de survie
13:15 - 13:30	Amay S.M. Cheam (University of Ottawa) ROC Curve Modeling with Pearson Distributions / Modélisation de la courbe ROC à partir des distributions de Pearson
13:30 - 13:45	Olli Saarela (McGill University), Elja Arjas (University of Helsinki) On Non-parametric Bayesian Regression in Cardiovascular Disease Risk Assessment / Régression Bayésienne non-paramétrique dans l'évaluation du risque de maladies cardiovasculaires
13:45 - 14:00	Julie Horrocks & Greg Persson (University of Guelph), Anne Duffy & Sarah Doucette (Dalhousie University), Paul Grof (University of Ottawa) Statistical Models for Staging / Modèles statistiques pour la stadification
14:00 - 14:15	Meaghan Cuerden, Cecilia Cotton & Richard Cook (University of Waterloo) Statistical Models and Methods for Causal Analysis of the Effect of Storage Time of Blood in Patients Receiving Transfusions / Modèles statistiques et méthodes d'analyse causale des effets du temps de stockage du sang chez les transfusés
14:15 - 14:30	Hui Song, Yingwei Peng & Dongsheng Tu (Queen's University) A New Approach for Joint Modeling of Longitudinal Measurements and Survival Times with a Cure Fraction / Nouvelle approche pour la modélisation combinée des mesures longitudinales et des temps de survie avec une fraction guérie

**13:30 - 15:00****KCIC Acadia Room****NSERC Discovery Grant Workshop****Atelier des subventions à la découverte du CRSNG****14:30 - 15:00****Festival Theatre****Break / Pause-café****14:30 - 16:00****Session 1D-D****Festival Theatre****Case Study 2: Determinants of the Presence and Volume of Brown Fat in Humans****Étude de cas 2 : Déterminants de la présence et du volume de graisse brune chez les humains**

Case Study / Étude de case

Organizer and Chair / Responsable et président: Lajmi Lakhali-Chaieb (Université Laval)

Technological advances made in the recent years allow us today to detect and measure the volume of brown fat in adults. However, the factors enhancing the prevalence of brown fat are relatively unknown and still have to be disentangled. The main purpose of this study is to investigate the relationships between a number of factors such as Age, Body Mass Index and sex and the prevalence and volume of brown fat. The students will have to deal with the problems of zero-inflated distribution for outcomes and missing data.

Les progrès technologiques des dernières années nous permettent aujourd'hui de détecter et de mesurer le volume de graisse brune chez les adultes. Toutefois, les facteurs capables d'augmenter la prévalence de graisse brune sont relativement inconnus et n'ont pas encore été démêlés. Le principal objectif de cette étude est d'étudier les relations entre un certain nombre de facteurs tels l'âge, l'indice de masse corporelle et le sexe, et la prévalence et le volume de graisse brune. Les étudiants devront traiter les problèmes d'une distribution continue avec un poids positif à zéro et de données manquantes.

14:30 - 16:00 Julius Bankole (University of Northern British Columbia)

14:30 - 16:00 Jason Buss &amp; Lingling Fan (University of New Brunswick)

14:30 - 16:00 Jigyuan Chen, Sandra Gonzalez Maldonado, Xin Qiu &amp; Ying Wu (University of Waterloo)

14:30 - 16:00 Brian Franczak &amp; Sanjeena Subedi (University of Guelph)

14:30 - 16:00 Yang Liu, James Proudfoot &amp; Hao Luo (University of British Columbia)

14:30 - 16:00 Yaqing Liu, Dong Xun, Yuhui Huang, Xue Yao, Masud Rana &amp; Xiaojing Wu (University of Saskatchewan)

- 14:30 - 16:00 Joshua Murray, Tim Guimond, Joanne Quan & Kristen O'Brien (University of Toronto)  
 14:30 - 16:00 Lang Sirius Qin & Yumi Kondo (University of British Columbia)  
 14:30 - 16:00 Mireille Schnitzer, Zhihui Amy Liu & Esther Perez Trejo (McGill University)  
 14:30 - 16:00 Radcliffe A. Siddo & Carlos Zerpa (Lakehead University)  
 14:30 - 16:00 Melissa Song, Daniel Ramunno & Xiao Di Yang (McMaster University)

**14:30 - 16:00 Session 1D-P** (abstracts/résumés - page 92) **Festival Theatre**

**Poster Session 1**

**Séance d'affichage 1**

Contributed Poster Session / Séance d'affichage libre

- 14:30 - 16:00 Lengyi Spectrum Han & Willard John Braun (University of Western Ontario)  
 Prediction Contours for the Prometheus Fire Growth Model / Contours du modèle Prometheus de prédiction de la croissance des feux
- 14:30 - 16:00 Jonathan Lee, Hao Yu & Reg Kulperger (University of Western Ontario)  
 Parallel Computation of Spatial Lattice Models / Calculs parallèles de modèles spatiaux en treillis
- 14:30 - 16:00 Shudong Li & Zilin Wang (Wilfrid Laurier University)  
 Analysis of Generalized Linear Mixed Models for Complex Survey Data / Analyse des modèles linéaires mixtes généralisés pour des données d'enquêtes complexes
- 14:30 - 16:00 Wenkai Ma & Willard John Braun (University of Western Ontario)  
 Fitting a Stochastic Spread Model to Experimental Fire Data / L'ajustement d'un modèle de propagation stochastique à des données sur des incendies expérimentaux
- 14:30 - 16:00 Mateen Shaikh, Luiza Antonie & Paul D. McNicholas (University of Guelph)  
 Modelling with Association Rules / Modélisation avec des règles d'association
- 14:30 - 16:00 Azadeh (Fatemeh) Shohoudi (Department of Mathematics and Statistics, McGill University), Masoud Asgharian, Abbas Khalili & David Wolfson (McGill University)  
 Variable Selection in Multipath Change-point Problems / Sélection de variables dans les problèmes de point de rupture à voies multiples
- 14:30 - 16:00 Lu Xin & Mahesh Pandey (University of Waterloo)  
 An Improved Statistical Model for Predicting the Deuterium Ingress in Zirconium Alloy Pressure Tubes / Un modèle statistique amélioré pour prévoir l'introduction de deutérium dans des tubes de force en alliage de zirconium

**15:00 - 16:00 Session 1E-A** (abstracts/résumés - page 95) **Festival Theatre**

**SSC Impact Award Address**

**Allocution du récipiendaire du prix pour impact de la SSC**

Special Session / Conférence spéciale

Chair/Président: Chris Field (Dalhousie University)

- 15:00 - 16:00 Abdel El-Shaarawi (National Water Research Institute, Environment Canada and McMaster University)  
 An Overview of Environmetrics Research / La recherche en statistique appliquée aux sciences de l'environnement : une vue d'ensemble

**16:00 - 17:30 Session 1F-I1** (abstracts/résumés - page 96) **BAC 237**

**Developing Teaching Skills in Graduate Students**

**Comment former les étudiants de cycle supérieur à l'enseignement**

Invited Paper Session / Présentations sur invitation

Chair/Président: Kevin Thorpe (University of Toronto)

Organizer/Responsable: Alison Gibbs (University of Toronto)

Session sponsored by the Statistical Education Committee

Session parrainée par le Comité d'éducation en statistique

- 16:00 - 16:30 Joe Blitzstein & Xiao-Li Meng (Harvard University)  
Teaching the Teaching of Statistics / Formation relative à l'enseignement des statistiques
- 16:30 - 17:00 Bethany White (University of Western Ontario)  
Training our Graduate Students to be Statistics Educators / Formation de nos étudiants diplômés pour qu'ils deviennent des pédagogues statisticiens
- 17:00 - 17:30 Alison Gibbs, Zeynep Baskurt & Ramya Thinniyam (University of Toronto)  
Developing Graduate Students' Supervisory Skills / L'amélioration des compétences des étudiants diplômés

---

**16:00 - 17:30**      **Session 1F-I2**      (abstracts/résumés - page 98)      **BAC132**

**Measuring Water Quality**

**Comment mesurer la qualité des eaux**

Invited Paper Session / Présentations sur invitation

Chair/Président: Lisa Lix (University of Saskatchewan)

Organizer/Responsable: Abdel El-Shaarawi (National Water Research Institute, Environment Canada)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

- 16:00 - 16:30 Pierre Gosselin (Institut national de santé publique du Québec, INRS-ETE and CHUQ)  
Public Water, Public Health, Public Concerns / L'eau comme préoccupation publique, bien public, et facteur de santé publique
- 16:30 - 17:00 Sylvia Esterby (University of British Columbia, Okanagan), Abdel El-Shaarawi (National Water Research Institute and McMaster University)  
Assessing the Performance of Composite Water Quality Indices / Évaluation de la performance des indices composites de la qualité de l'eau
- 17:00 - 17:30 William Ross (Health Canada)  
Assessing Human Health Risks from the Viral Contamination of Shellfish / Évaluer les risques de la contamination virale des mollusques sur la santé humaine

---

**16:00 - 17:30**      **Session 1F-I3**      (abstracts/résumés - page 100)      **BAC234**

**New Developments in (Disease and) Syndromic Surveillance**

**Du neuf en surveillance (des maladies et) syndromique**

Invited Paper Session / Présentations sur invitation

Chair/Président: Patrick Brown (Cancer Care Ontario)

Organizer/Responsable: Gauri Datta (University of Georgia)

- 16:00 - 16:30 Andrew Lawson & Ana Corberan (Medical University of South Carolina)  
A Multivariate Bayesian Model-Based Approach for On-Line Spatio-Temporal Disease Surveillance / Une approche bayésienne multivariée pour la surveillance spatio-temporelle de maladies en ligne
- 16:30 - 17:00 Rong Wei (National Center for Health Statistics), Jiraphan Suntornchost & Melonie Heron (National Center for Health Statistics, CDC, USA), Eric Slud (University of Maryland, USA), Meena Khare (National Center for Health Statistics, CDC, USA)  
Changes in Age-Specific Mortality Rates from Six Leading Causes of Death in US Populations, 1970-2007 / Changements des taux de mortalité par âges pour six causes principales de décès aux États-Unis, 1970-2007
- 17:00 - 17:30 Gauri Datta (University of Georgia), David Banks & Matt Heaton (Duke University), Alan Karr (National Institute of Statistical Sciences), James Lynch (University of South Carolina), Francisco Vera (University of Ecuador), Frank Zou (National Institute of Statistical Sciences)  
Hierarchical Bayesian Modeling in Syndromic Surveillance / Modélisation bayésienne hiérarchique dans la surveillance syndromique

<b>16:00 - 17:30</b>	<b>Session 1F-I4</b>	(abstracts/résumés - page 102)	<b>BAC239</b>
<b>Recent Advances in the Analysis of Correlated Data</b>			
<b>Récents progrès en analyse de données corrélées</b>			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Wei Liu (York University)			
16:00 - 16:30	Richard Cook & Liqun Diao (University of Waterloo) Joint Analysis of Multiple Markov Processes under Interval Censoring / Analyse conjointe de processus de Markov multiples en présence de censure par intervalles		
16:30 - 17:00	Maria Lorenzi (British Columbia Cancer Centre), Xiaogiong Joan Hu (Simon Fraser University) Zero-inflated Longitudinal Data Analysis / Analyse de données longitudinales à surreprésentation de zéros		
17:00 - 17:30	Guohua Yan, Renjun Ma & M. Tariqul Hasan (University of New Brunswick) Analysis of Longitudinal Semi-continuous Data / Analyse de données longitudinales semi-continues		
<b>16:00 - 17:30</b>	<b>Session 1F-I5</b>	(abstracts/résumés - page 104)	<b>BAC 236</b>
<b>Robust Designs of Statistical Experiments</b>			
<b>Conception robuste d'expériences statistiques</b>			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Xiaojian Xu (Brock University)			
16:00 - 16:30	Julie Zhou & Michael Wilmut (University of Victoria) D-Optimal Minimax Design Criterion for Fractional Factorial Designs / Critère de conception minimax D-optimal pour les plans factoriels fractionnés		
16:30 - 17:00	Sanjoy Sinha (Carleton University) Robust Designs for Multivariate Logistic Regression / Plans d'expérience robustes pour la régression logistique multivariée		
17:00 - 17:30	Doug Wiens (University of Alberta), Alan Welsh (Australian National University) Robust Model-Based Sampling Designs / Plans d'échantillonnage robustes basés sur des modèles		
<b>16:00 - 17:15</b>	<b>Session 1F-C1</b>	(abstracts/résumés - page 106)	<b>BAC 141</b>
<b>CANNeCTIN Methodology for Clinical Trials</b>			
<b>Méthodologie du CANNeCTIN pour les essais cliniques</b>			
Contributed Paper Session / Séance de communications libres			
Chair/Président: Lehana Thabane (McMaster University)			
Organizer/Responsable: Robert Platt (McGill University)			
Session sponsored by the Biostatistics Section			
Session parrainée par le Groupe de biostatistique			
16:00 - 16:15	Lehana Thabane (McMaster University)		
16:15 - 16:30	Haocheng Li & Grace Yi (University of Waterloo) Variable Selection for Longitudinal Data via the Composite Likelihood Method / Sélection de variables pour données longitudinales par la méthode de vraisemblance composite		
16:30 - 16:45	Jinhui Ma, Parminder Raina, Joseph Beyene & Lehana Thabane (McMaster University) Simulation Study using a Multi-State Markov Model to Estimate the Statistical Power of Population Based Longitudinal Studies / Étude par simulation utilisant un modèle de Markov multi-états pour estimer la puissance statistique de population basée sur des études longitudinales		
16:45 - 17:00	Hoi Suen Wong, Aaron Childs & Eleanor Pullenayegum (McMaster University) Analysis of Health Utility Data / Analyse de données d'utilité sur la santé		

17:00 - 17:15 Maria Esther Perez Trejo & Robert William Platt (McGill University)  
Modeling Data from Cluster Randomized Trials in Presence of Measurement Error Induced by Double Clustering / Modélisation de données d'essais aléatoires en grappes en présence de mesures d'erreurs induites par double classement

**16:00 - 17:15**      **Session 1F-C2**      (abstracts/résumés - page 109)      **BAC138**

**Distribution Theory and Asymptotics**

**Théorie de la distribution et asymptotique**

Contributed Paper Session / Séance de communications libres

Chair/Président: Wei Sun (Concordia University)

Session sponsored by the Probability Section

Session parrainée par le Groupe de probabilité

16:00 - 16:15 Lilia Leticia Ramírez Ramírez, Mary E. Thompson & Yulia Gel (University of Waterloo)  
On the Estimation of the Degree Distribution of Random Networks / Estimation de la distribution de degrés des réseaux aléatoires

16:15 - 16:30 A. Akbar Mohsenipour & Serge Provost (The University of Western Ontario)  
Approximating the Distributions of Singular Quadratic Expressions and their Ratios / Approximer les distributions des expressions quadratiques singulières et de leurs ratios

16:30 - 16:45 John Sheriff (University of Lethbridge)  
Central Limit Theorems for Additive Functionals of the Ginzburg-Landau Process / Théorèmes limite central pour des fonctionnelles additives du processus de Ginzburg-Landau

16:45 - 17:00 Weibin Jiang, Hao Yu & Reg Kulperger (The University of Western Ontario)  
Asymptotic Theory for General Multivariate GARCH Models / Théorie asymptotique pour des modèles multivariés généraux GARCH

17:00 - 17:15 David Brillinger (University of California, Berkeley)  
Background and Applications of Particle Process Systems / Contexte et applications des systèmes de processus des particules

**16:00 - 17:30**      **Session 1F-C3**      (abstracts/résumés - page 112)      **BAC142**

**Survey Methods: Precision and Analysis**

**Méthodes d'enquête : précision et analyse**

Contributed Paper Session / Séance de communications libres

Chair/Président: David Haziza (Université de Montréal)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

16:00 - 16:15 Alastair Scott (The University of Auckland), Thomas Lumley (University of Auckland)  
Pseudo-Likelihood Methods for Survey Data / Modèles de quasi-vraisemblance pour données d'enquête

16:15 - 16:30 Anne-Sophie Charest (Carnegie Mellon University)  
Inference from Differentially-Private Synthetic Datasets / Inférence avec des jeux de données satisfaisant le critère de confidentialité différentielle

16:30 - 16:45 Martin Klein (U.S. Census Bureau), Tommy Wright  
Ranking Procedures for Several Normal Populations: An Empirical Investigation / Procédures de classement de plusieurs population normales : une étude empirique

16:45 - 17:00 Joanne Charlebois & Guylaine Dubreuil (Statistics Canada)  
Variance Estimation for the Redesignated Survey of Household Spending / Estimation de la variance pour l'Enquête sur les dépenses des ménages remaniée

17:00 - 17:15 Zeinab Mashreghi, Christian Léger & David Haziza (Université de Montréal)  
Bootstrap Variance Estimation for Imputed Data from Regression, Ratio and Hot Deck Imputation Methods / Sur l'estimateur bootstrap de la variance pour des données imputées par la régression, le ratio ou le hot-deck

17:15 - 17:30      Wei Qian & Abdellatif Demnati (Statistics Canada)  
Variance Estimation for Survey Data with Multiple Cycle Imputation / Estimation de la variance  
pour les données d'enquête ayant des cycles d'imputation multiples

---



**Tuesday June 14****mardi 14 juin****8:30 - 9:45      Session 2A-A      (abstracts/résumés - page 115)      Festival Theatre****Gold Medal Address****Allocution de la récipiendaire de la médaille d'or**

Special Session / Conférence spéciale

Chair/Président: Christian Genest (McGill University)

8:30 - 9:45      Louis-Paul Rivest (Université Laval)  
 Looking for Directions: from Directional Statistics to the Analysis of Object Data / A la recherche de directions : de la statistique directionnelle à l'analyse de données-objets

**9:45 - 10:15      Festival Theatre****Break / Pause-café****10:15 - 11:45      Session 2B-I1      (abstracts/résumés - page 116)      BAC138****Recent Advances in MCMC****Récents progrès en MCMC**

Invited Paper Session / Présentations sur invitation

Chair/Président: Jeff Rosenthal (University of Toronto)

Organizer/Responsable: Radu Craiu (University of Toronto)

Session sponsored by the Probability Section

Session parrainée par le Groupe de probabilité

10:15 - 10:45      Galin Jones (University of Minnesota), Kuo-Jung Lee (University of Rochester), Brian Caffo & Susan Spear Bassett (Johns Hopkins University)

Spatial Bayesian Variable Selection Models on fMRI Time-Series Data / Modèles bayesiens de sélection des variables spatiales sur des séries chronologiques d'imagerie par résonance magnétique fonctionnelle (IRMf)

10:45 - 11:15      David Stephens (McGill University)  
 Particle Methods in Markov Chain Monte Carlo / Méthodes à particules dans la chaîne de Markov Monte-Carlo

11:15 - 11:45      Dawn Woodard (Cornell University), Moises Goldszmidt (Microsoft Research)  
 Online Inference for Model-Based Clustering in Distributed Computing / Inférence en ligne pour classification basée sur des modèles en calcul réparti

**10:15 - 11:45      Session 2B-I2      (abstracts/résumés - page 118)      BAC 237****Recent Developments in Multiple Comparison Procedures****Du neuf dans les procédures de comparaisons multiples**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Jianan Peng (Acadia University)

10:15 - 10:45      Jason Hsu (Ohio State University), Szu-Yu Tang (The Ohio State University), Yi Liu (Millennium)  
 Statistical Principles and Methods in Quantitative Biomarker Usage / Principes et méthodes statistiques pour l'utilisation de biomarqueurs quantitatifs

10:45 - 11:15      Wei Liu (University of Southampton, UK)  
 Simultaneous Confidence Bands for Regression Analysis / Bandes de confiance simultanées pour l'analyse de régression

11:15 - 11:45      Wenge Guo (New Jersey Institute of Technology), Sanat Sarkar (Temple University)  
 Adaptive FWER and FDR Control under Block Dependence / Contrôle adapté de l'erreur de type global (ETG) et du taux de fausses découvertes (TFD) sous la dépendance en blocs

<b>10:15 - 11:45</b>	<b>Session 2B-I3</b>	(abstracts/résumés - page 120)	<b>BAC 239</b>
<b>Recent Developments in Shrinkage Estimation with Related Implications for Predictive Density Estimation, Loss Estimation and Model Selection</b>			
<b>Du neuf dans l'estimation par rétrécissement et implications quant à l'estimation de la densité prédictive, l'estimation des pertes et la sélection de modèles</b>			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Éric Marchand (Université de Sherbrooke)			
10:15 - 10:45	Bill Strawderman (Rutgers University), John Kent & Reman Abu-Shanab (University of Leeds)	Shrinkage Estimation for Matrix Loss / Estimation par rétrécissement avec perte matricielle	
10:45 - 11:15	Dominique Fourdrinier & Aurélie Boisbunon (Université de Rouen), Stéphane Canu (INSA de Rouen), William Strawderman (Rutgers University), Marty Wells (Cornell University)	Variable Selection: a Decision Theory Approach / Sélection de variable : une approche décisionnelle	
11:15 - 11:45	Andrew L. Rukhin (National Institute of Standards and Technology)	Reducing Data Nonconformity in Linear Models / Réduction de la non-conformité des données dans les modèles linéaires	
<b>10:15 - 11:45</b>	<b>Session 2B-I4</b>	(abstracts/résumés - page 122)	<b>BAC132</b>
<b>Sampling Theory and Practice: In Honour of Mary Thompson's Retirement</b>			
<b>Théorie et pratique de l'échantillonnage : À l'occasion du départ à la retraite de Mary Thompson</b>			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Changbao Wu (University of Waterloo)			
10:15 - 10:45	JNK Rao (Carleton University), Changbao Wu (University of Waterloo)	Some Applications of Estimating Equations Approach to Inference in Complex Survey Data / Quelques applications de l'approche d'inférence par équations estimatrices aux données provenant de plans de sondage complexes	
10:45 - 11:15	David Binder (Retired (Statistics Canada))	Causal Inference for Observational Data Obtained from a Complex Survey / Inférence causale pour des études observationnelles provenant d'un plan de sondage complexe	
11:15 - 11:45	Danny Pfeffermann (University of Southampton, UK)	Fitting Models to Complex Survey Data; Why is it a problem? How should we approach it? / Ajuster des modèles aux données d'enquêtes complexes; pourquoi est-ce un problème? Comment devons-nous l'aborder?	
<b>10:15 - 11:45</b>	<b>Session 2B-I5</b>	(abstracts/résumés - page 124)	<b>BAC141</b>
<b>Statistical Consulting: Communication, Interpersonal, Project Management Skills Required</b>			
<b>Conseil statistique : compétences requises en communications, interpersonnel et gestion de projet</b>			
Invited Paper Session / Présentations sur invitation			
Chair/Président: Boxin Tang (Simon Fraser University)			
Organizer/Responsable: John Amrhein (McDougall Scientific Ltd.)			
Session sponsored by the Business and Industrial Statistics Section			
Session parrainée par le Groupe de statistique industrielle et de gestion			
10:15 - 10:45	Lorne Rothman (SAS Institute Canada Inc.)	How Many Consultants Does It Take to Fit a Regression? Lessons From a Large Business Practice. / Combien de consultants faut-il pour ajuster une régression? Leçons d'une grande pratique.	
10:45 - 11:15	Fernando Camacho (Damos Inc.)	Statistical Consulting within an Engineering Environment / Consultation statistique dans un environnement d'ingénierie	

11:15 - 11:45 Rhonda Rosychuk (University of Alberta)  
Statistical Consulting in an Academic Medical Setting: If I Only Knew Then What I Know Now / Consultation statistique dans un milieu académique médical : Si j'avais pu savoir à l'époque ce que je sais maintenant

**10:15 - 11:45 Session 2B-C1** (abstracts/résumés - page 126) **BAC 234**

### High-Dimensional Data in Biology

#### Données de grande dimension en biologie

Contributed Paper Session / Séance de communications libres

Chair/Président: Yutaka Yasui (University of Alberta)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

10:15 - 10:30 Karen A Kopciuk (Alberta Health Services - Cancer Care), Yuan Wang (University of Calgary)  
Sample Size Estimation Methods for NMR Metabolomics Data / Méthodes d'estimation de la taille d'échantillon pour données métabolomiques de la résonance magnétique nucléaire (RMN)

10:30 - 10:45 Toby Kenney & Hong Gu (Dalhousie University)  
A Generalized Codon-based Model of Nucleotide Substitution for Protein-coding DNA Sequences / Modèle généralisé à base de codons pour la substitution de nucléotide dans des séquences d'ADN codant pour des protéines

10:45 - 11:00 Hong Gu, Toby Kenney & Joseph Bielawski (Dalhousie University)  
Estimation of Darwinian Positive Selection Using Generalized Codon-based Models / Estimation de la sélection positive de Darwin à l'aide de modèles généralisés à base de codons

11:00 - 11:15 Thierry Chekouo Tekougang & Alejandro Murua (Université de Montreal)  
Statistical View of Popular Biclustering Algorithms / Approche statistique des algorithmes populaires de grappes bi-dimensionnelles

11:15 - 11:30 Dena Givari, Paul McNicholas & Sanjeena Subedi (University of Guelph)  
On the Application of a Bayesian Infinite Mixture Modelling / De l'application d'un modèle de mélange bayésien infini

11:30 - 11:45 Vahid Partovi Nia & Vahid Partovi Nia (McGill University)  
Agglomerative Bayesian Biclustering / Classification double agglomérative bayésienne

**10:15 - 11:45 Session 2B-C2** (abstracts/résumés - page 129) **BAC 142**

### Methods in Mixture Models and Spatial Statistics

#### Méthodes dans les modèles de mélange et spatiale statistique

Contributed Paper Session / Séance de communications libres

Chair/Président: Eleanor Pullenayegum (McMaster University)

10:15 - 10:30 Abbas Khalili & David A. Stephens (McGill University)  
A New Order Estimation Method in Mixture of Autoregressive Models / Nouvelle méthode d'estimation de l'ordre dans les modèles de mélanges autorégressifs

10:30 - 10:45 Andrew Raim (University of Maryland), Matthias Gobbert & Nagaraj Neerchal (University of Maryland, Baltimore County), Jorge Morel (Procter and Gamble Company)  
Maximum Likelihood Estimation of the Random-Clumped Multinomial Model using High Performance Computing / Estimation du maximum de vraisemblance du modèle multinomial aléatoire groupé par le calcul de haute performance

10:45 - 11:00 Shaheena Bashir (University Health Network, Toronto), E. M. Carter (University of Guelph, Canada)  
Robust Covariate Density Defined (CDD) Mixture of Logistic Regression Models / Mélange robuste de modèles de régression logistique défini par la densité des covariables

11:00 - 11:15 Mélina Mailhot, Hélène Cossette & Étienne Marceau (Université Laval)  
TVaR-Based Capital Allocation for Bivariate Compound Distributions / Allocation de capital basée sur la mesure TVaR pour distributions composées bivariées

- 11:15 - 11:30 Lingling Fan (University of New Brunswick)  
Classification and Spatial Point Pattern Recognition / Classification et reconnaissance de motifs ponctuels spatiaux
- 11:30 - 11:45 Jeffrey Picka (University of New Brunswick)  
Spatial Statistical Analysis of Powder Flow Data / Analyse statistique spatiale de données d'écoulement des poudres

---

**10:15 - 11:45**      **Session 2B-C3**      (abstracts/résumés - page 132)      **BAC 236**

**Spatial-temporal Modelling in the Natural World**

**Modélisation spatio-temporelle dans le monde naturel**

Contributed Paper Session / Séance de communications libres

Chair/Président: Fahim Ashkar (Université de Moncton)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

- 10:15 - 10:30 Mike Dowd (Dept. of Mathematics and Statistics, Dalhousie University), Ruth Joy (Dept. of Statistics, Simon Fraser University)  
Estimating Behavioural Parameters in Stochastic Animal Movement Models / Estimation de paramètres comportementaux dans des modèles stochastiques de déplacements d'animaux
- 10:30 - 10:45 Angus Argyle (University of Victoria)  
Estimating the Number of Species using Abundance-Based Data from Multiple Sites within a Region / Estimation du nombre d'espèces à l'aide de données d'abondance provenant de différents sites d'une même région
- 10:45 - 11:00 Marylène Troupé (Université des Antilles et de la Guadeloupe), Jean Vaillant (Université des Antilles et de la Guyane), Léonide Celiny (Université Paris-Est Créteil)  
Dispersion Indexes and Thinned Point Processes / Indices de dispersion et processus ponctuels amincis
- 11:00 - 11:15 Noel Cadigan (Fisheries and Oceans Canada)  
Measurement Error Bias in Fish Stock Spawner-Recruitment Models / Biais dû aux erreurs de mesure dans les modèles de recrutement du stock de poissons reproducteur
- 11:15 - 11:30 Larissa Valmy & Jean Vaillant (Université des Antilles et de la Guyane)  
Statistical Study of Caribbean Seismic Events / Etude statistique d'événements sismiques caribéens
- 11:30 - 11:45 Khadija Bouzaachane (Université Cadi Ayyad Ecole Supérieure de Technologie d'Essaouira), Youssef Benghabrit (ENSAM)  
Spatial-temporal Model for Rainfall Data of Morocco / Modélisation spatio-temporelle des précipitations au Maroc

---

**11:45 - 13:00**

**Wheelock Dining Hall**

**Lunch / Déjeuner**

---

**12:15 - 13:00**

**BAC 244**

**Funding Opportunities from Mitacs Inc.**

**Occasions de financement par le Mitacs Inc.**

---

**13:00 - 14:29**

**Session 2C-I1**

(abstracts/résumés - page 135)

**BAC239**

**Designs for Generalized Linear Models**

**Conception de modèles linéaires généralisés**

Invited Paper Session / Présentations sur invitation

Chair/Président: Sue Lewis (University of Southampton, UK)

Organizer/Responsable: Abhyuday Mandal (University of Georgia)

- 13:00 - 13:22 Abhyuday Mandal (University of Georgia), Jie Yang & Dibyen Majumdar (University of Illinois at Chicago)  
Optimal Designs for Two-Level Factorial Experiments with Binary Response / Plans optimaux pour expériences factorielles à deux niveaux avec réponse binaire
- 13:22 - 13:44 Karabi Nandy (University of California at Los Angeles)  
Optimal Designs for Non-Monotone Dose-Response Models / Plans optimaux pour modèles non-monotones de la relation dose-effet
- 13:45 - 14:07 John Stufken (University of Georgia)  
Tools to Identify Optimal Designs for Generalized Linear Models / Outils pour identifier les plans optimaux de modèles linéaires généralisés
- 14:07 - 14:29 Dave Woods (University of Southampton, UK), Tim Waite (University of Southampton), Tim Waterhouse (Eli Lilly)  
Optimal Block Designs for Non-Normal Data / Plans optimaux en blocs pour données non-normales

---

**13:00 - 14:30**      **Session 2C-I2**      (abstracts/résumés - page 137)      **BAC132**

**New Developments in Mixture Models & Their Applications**

**Du neuf dans les modèles de mélange et leurs applications**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Abbas Khalili (McGill University)

- 13:00 - 13:30 Geoff McLachlan (University of Queensland, Australia)  
Clustering via Mixtures of Skew Normal or Skew t-Distributions / Classification via des mélanges de distributions asymétriques normales et de Student
- 13:30 - 14:00 Jiahua Chen (University of British Columbia), Pengfei Li (University of Alberta)  
Tuning the EM-test for Finite Mixture Models / Réglage du test EM pour modèles de mélanges finis
- 14:00 - 14:30 Paul McNicholas & Sanjeena Subedi (University of Guelph)  
Parameter Estimation in Model-Based Clustering / Estimation paramétrique dans l'analyse de regroupements basée sur des modèles

---

**13:00 - 14:30**      **Session 2C-I3**      (abstracts/résumés - page 139)      **BAC142**

**On the Issues Raised by Environmental Surveys at Statistics Canada**

**Au sujet des questions soulevées par les études environnementales chez Statistique Canada**

Invited Paper Session / Présentations sur invitation

Chair/Président: Pierre Lavallée (Statistique Canada)

Organizer/Responsable: Nathalie Hamel (Statistique Canada)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

- 13:00 - 13:30 John Marshall (Statistics Canada)  
The Diverse Needs and Challenges of Environmental Surveys : an Analytical Perspective / La diversité des besoins et des défis des enquêtes sur l'environnement : Une perspective analytique
- 13:30 - 14:00 Martin Hamel, Claude Girard, Nathalie Hamel & Sébastien Labelle-Blanchet (Statistique Canada)  
Application of the Methodology Commonly used in Economics Surveys to Environmental Surveys / Application de la méthodologie couramment utilisée pour les enquêtes économiques aux enquêtes environnementales
- 14:00 - 14:30 Claude Girard (Statistique Canada)  
Addressing Some of the Methodological Issues Raised by Environment Surveys / Aborder certains des enjeux méthodologiques soulevés par les enquêtes environnementales

---

**13:00 - 14:30**      **Session 2C-14**      (abstracts/résumés - page 141)      **BAC 237**

**Panel on Online Teaching of Statistics**

**Table ronde sur l'enseignement de la statistique en ligne**

Invited Paper Session / Présentations sur invitation

Chair/Président: Bethany White (University of Western Ontario)

Session sponsored by the Statistical Education Committee

Session parrainée par le Comité d'éducation en statistique

- 13:00 - 14:30      Mohamed Abdoell & Jennifer Payne (Dalhousie University)  
Teaching Statistics in Academic Medical Research Institutions / L'enseignement des méthodes de biostatistique dans les institutions académiques de recherche médicale
- 13:00 - 14:30      Joel Dubin (University of Waterloo)  
Experiences Teaching a Distance Education Biostatistics Course for Masters in Public Health Students / Expérience d'enseignement à distance d'un cours de biostatistique pour les étudiants de maîtrise en santé publique
- 13:00 - 14:30      Alison Weir (University of Toronto at Mississauga)  
The Instructor Dimension of e-Learning in Undergraduate Statistics / La dimension de l'enseignant dans la formation en ligne des étudiants de premier cycle en statistique

---

**13:00 - 14:30**      **Session 2C-15**      (abstracts/résumés - page 143)      **BAC 138**

**Sequential Designs for Computer Experiments**

**Conception séquentielle d'expériences informatiques**

Invited Paper Session / Présentations sur invitation

Chair/Président: Duncan Murdoch (University of Western Ontario)

Organizer/Responsable: Pritam Ranjan (Acadia University)

- 13:00 - 13:30      Derek Bingham & Joslin Goh (Simon Fraser University), James Holloway (University of Michigan)  
An Application of Sequential Design for Multi-fidelity Computer Models / Une application du plan séquentiel aux modèles informatiques de multi-fidélité
- 13:30 - 14:00      Jason Loepky (University of British Columbia, Okanagan), Brian Williams & Leslie Moore (Los Alamos National Laboratory)  
Gaussian Process Models for Mixture Experiments / Modèles de processus gaussiens pour des expériences de mélanges
- 14:00 - 14:30      Robert B Gramacy (University of Chicago), Herbie Lee (University of California, Santa Cruz)  
Optimization under Unknown Constraints / Optimisation dans le cadre de contraintes inconnues

---

**13:00 - 14:15**      **Session 2C-C1**      (abstracts/résumés - page 145)      **BAC 236**

**Distributions: Theory and Application**

**Distributions : théorie et application**

Contributed Paper Session / Séance de communications libres

Chair/Président: Paramjit Gill (University of British Columbia Okanagan)

- 13:00 - 13:15      Iman Mabrouk (University of Western Ontario), Serge Provost (University of Western Ontario)  
An Extended Inverse Gaussian Model / Un modèle gaussien inverse étendu
- 13:15 - 13:30      Weiyu Qiu (School of Public Health, University of Alberta), Yutaka Yasui (Department of Public Health Sciences, School of Public Health, University of Alberta)  
Elliptical Distributions with Linear Conditional Second Moments: Multivariate Pearson Type VII Family and its Applications / Distributions elliptiques avec seconds moments conditionnels linéaires : la famille de Pearson multivariée de Type VII et ces applications
- 13:30 - 13:45      Hanna Jankowski (York University), Jon Wellner (University of Washington)  
Estimation of a Discrete Monotone Distribution / Estimation d'une distribution discrète monotone

- 13:45 - 14:00 Zheng Sun (Simon Fraser University)  
K sample Cramer-von Mises Tests for Grouped Data / Test de Cramer-von Mises à k échantillons pour les données groupées
- 14:00 - 14:15 Val Andrei Fajardo (University of Waterloo), Mei Ling Huang (Brock University)  
On Confidence Regions for High Quantiles of a Heavy Tailed Distribution / À propos des régions de confiance des quantiles supérieures d'une distribution à ailes relevées

---

**13:00 - 14:15      Session 2C-C2      (abstracts/résumés - page 148)      **BAC 234****

**Genetic Association and Analysis of Families**  
**Association génétique et analyse des familles**

Contributed Paper Session / Séance de communications libres  
Chair/Président: Angelo Canty (McMaster University)  
Session sponsored by the Biostatistics Section  
Session parrainée par le Groupe de biostatistique

- 13:00 - 13:15 Karim Oualkacha (McGill University, Department of Epidemiology, Biostatistics and Occupational Health), Aurélie Labbe & Antonio Ciampi (McGill University Department of Epidemiology, Biostatistics and Occupational Health), Marc-André Roy & Michel Maziade (Centre de Recherche Université Laval Robert Giffard)  
Principal Components of Heritability for High Dimension Quantitative Traits and General Pedigrees / Des composantes principales d'héritabilité pour des traits quantitatifs de grande dimension et pour des pedigrees quelconques
- 13:15 - 13:30 Lucia Mirea (Maternal-Infant Care Research Centre (MiCare), Mount Sinai Hospital), Shelley Bull (Samuel Lunenfeld Research Institute, Mount Sinai Hospital), Lei Sun (Dalla Lana School of Public Health, University of Toronto)  
Incorporating Evidence of Population Stratification Bias in Genetic Association Analyses Combining Individual and Family Data / Incorporer la preuve d'un biais de stratification population dans les analyses d'association génétique combinant des données individuelles et familiales
- 13:30 - 13:45 Elif Fidan Acar (McGill University), Lei Sun (University of Toronto)  
A Generalized Kruskal-Wallis Test Incorporating Group Uncertainty with Application to Genome-Wide Association Studies / Test de Kruskal-Wallis généralisé avec incertitude au niveau du groupe et ses applications aux études d'association sur le génome entier
- 13:45 - 14:00 Laurent Briollais (Samuel Lunenfeld Research Institute)  
Bayesian Graphical Models (BGMs) for Genetic Association Studies / Modèles graphiques bayésiens (MGB) pour études d'association génétique
- 14:00 - 14:15 Antonio Ciampi, Lin Yang & Aurélie Labbe (McGill University), Chantal Merette (Laval University)  
PLS Regression and Hybrid Methods in Genomics Association Studies / Régression PLS et méthodes hybrides dans les études d'associations en génomiques

---

**13:00 - 14:15      Session 2C-C3      (abstracts/résumés - page 151)      **BAC141****

**Optimizing Products and Processes**  
**Optimisation des produits et processus**

Contributed Paper Session / Séance de communications libres  
Chair/Président: Vickneswary Tagore (University of Saskatchewan)  
Session sponsored by the Business and Industrial Statistics Section  
Session parrainée par le Groupe de statistique industrielle et de gestion

- 13:00 - 13:15 Shaun S. Wulff (University of Wyoming), John Brewster (University of Manitoba)  
Objective Bayesian Inference of Fixed Effects in Industrial Split-Plot Experiments / Inférence bayésienne objective des effets fixes dans des expériences par parcelles divisées
- 13:15 - 13:30 Lin Fang & Roman Viveros-Aguilera (McMaster University)  
Statistical Quality Control Charts to Monitor Processes Under Competing Risks / Les cartes statistique de contrôle de la qualité pour surveiller des processus à risques compétitifs

- 13:30 - 13:45 Lu Xin & Mahesh Pandey (University of Waterloo)  
Using Mixed-Effects Regression Model to Analyze the Deformation in Nuclear Reactor Components / L'utilisation d'un modèle de régression à effets mixtes pour analyser la déformation des composantes d'un réacteur nucléaire
- 13:45 - 14:00 Mei Tang (Sanofi Pasteur Canada)  
Statistical Power Analysis in Vaccine Development in Non-Clinical Areas / Analyse de la puissance statistique dans le développement de vaccins pour les domaines non cliniques
- 14:00 - 14:15 Andrey Fendyur & Giovanni J.C. da Silveira (University of Calgary)  
Statistical Detection of Moderation and Mediation Effects among Predictors in Industrial Productivity Research / Détection statistique des effets de modération et de médiation entre les variables explicatives dans la recherche sur la productivité industrielle

---

**13:00 - 14:30**      **Session 2C-C4**      (abstracts/résumés - page 154)      **BAC 241**

**Survival Models: Correlated Data and Non-parametric Estimation**

**Modèles de survie : données corrélées et estimation non paramétrique**

Contributed Paper Session / Séance de communications libres

Chair/Président: Julie Atherton (McGill University)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

- 13:00 - 13:15 Henrik Stryhn (Atlantic Veterinary College, University of Prince Edward Island), Adel Elghafghuf (Atlantic Veterinary College, UPEI)  
Estimating Correlated Frailties in a Cox Model / Estimer les fragilités partagées corrélées dans le modèle de Cox
- 13:15 - 13:30 Yi Niu & Yingwei Peng (Queen's University)  
A Marginal Mixture Cure Model for Clustered Survival Data / Un modèle marginal de mélange avec fraction non à risque pour données de survie en grappes
- 13:30 - 13:45 Hedy Jiang & Patrick Brown (Cancer Care Ontario and McMaster University)  
Survival Models for A Large Cohort Study in Pickering / Modèles de survie pour une vaste étude de cohorte à Pickering
- 13:45 - 14:00 David E. Matthews (University of Waterloo)  
Exact Nonparametric Likelihood Confidence Bands for the Survivor Function / Bornes de confiance par vraisemblance exacte non-paramétrique de la fonction de survie
- 14:00 - 14:15 Mohammad Dehghan & Thierry Duchesne (Laval University)  
Nonparametric Estimation of the Conditional Survival Function with Time-Varying Covariate and Interval Censoring / Estimation non-paramétrique de la fonction de survie conditionnelle avec covariable qui dépend du temps et censure par intervalle
- 14:15 - 14:30 Zhihui (Amy) Liu, Benjamin Rich & James Hanley (McGill University)  
Recovering the Raw Data behind a Kaplan-Meier or Nelson-Aalen Survival Curve / Retrouver les données brutes à partir d'une courbe de survie de Kaplan-Meier ou de Nelson-Aalen

---

**14:30 - 15:00**

**Festival Theatre**

**Break / Pause-café**

---



**14:30 - 16:00      Session 2D-D      Festival Theatre****Case Study 1: Gender Gap in Earnings among Young People****Étude de cas 1 : Écart salarial entre les jeunes hommes et femmes**

Case Study / Étude de case

Chair/Président: Lajmi Lakhal-Chaieb (Université Laval)

Organizers/Responsables: Karla Fox and/et Lenka Mach (Statistics Canada)

The average earnings for men have been higher than those of women. This difference, often referred to as the gender wage gap, has been studied by a number of researchers using different data sources collected by Statistics Canada. This study will use the cross-sectional Public-Use Microdata File (PUMF) for the Survey of Labour and Income Dynamics (SLID) which contains data on income, labour, personal and family characteristics for the population living in the ten Canadian provinces. The main objective of the study is to analyse the gender gap in employment and earnings for young people (25-34) using the design-based approach. The students should also examine the difference between the design-based and model-based approach. They will also need to deal with the problem of the missing data.

Le revenu moyen des hommes est historiquement plus élevé que celui des femmes. Cette différence, ou écart salarial entre les sexes, a fait l'objet d'études par plusieurs chercheurs utilisant diverses sources de données recueillies par Statistique Canada. Cette étude utilisera le fichier de microdonnées à grande diffusion (FMGD) de l'Enquête sur la dynamique du travail et du revenu (EDTR), qui contient des données sur le revenu, le travail et les caractéristiques personnelles et familiales des habitants des dix provinces canadiennes. Le principal objectif de l'étude est d'analyser l'écart d'emploi et de revenu entre les jeunes hommes et femmes (de 25 à 34 ans) par une approche fondée sur un plan. Les étudiants devront aussi examiner la différence entre l'approche fondée sur un plan et celle fondée sur un modèle. Ils devront aussi traiter le problème des données manquantes.

14:30 - 16:00      Min Chen, Jocelyn Tufts &amp; Lisha Yu (University of Waterloo)

14:30 - 16:00      Laura May Morrison &amp; Lucy Yao Liu (University of Victoria)

14:30 - 16:00      Lutong Zhou, Haiyan Jiang, Jingqin Zhu &amp; Anky Lai (University of Toronto)

**14:30 - 16:00      Session 2D-P      (abstracts/résumés - page 157)      Festival Theatre****Poster Session 2****Séance d'affichage 2**

Contributed Poster Session / Séance d'affichage libre

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

14:30 - 16:00      Shaheena Bashir (University Health Network), Celia Greenwood & Qazi Azad (McGill University)  
Multivariate Meta Analysis in Gene Expression Microarray Studies / Méta-analyse multivariée  
des études d'expression génique sur micro-réseaux14:30 - 16:00      Nadia Bifulchi, Rob Deardon & Zeny Feng (University of Guelph)  
The Interplay Between Spatial and Network-based Infectious Disease Models / La relation  
entre les modèles de maladies infectieuses spatiaux et de réseaux14:30 - 16:00      Sabrina Dicintio (University of Guelph), Sabrina Dicintio, Paul McNicholas & Brian Franczak  
Early Disease Detection of Veterinary Data / La détection précoce de maladies au moyen de  
données vétérinaires14:30 - 16:00      Angie Christina Dobbs & Rob Deardon (University of Guelph)  
Improving MCMC Convergence in Individual-Level Infectious Disease Models / Améliorer la  
convergence MCMC dans les modèles de maladies infectieuses à l'échelle de l'individu14:30 - 16:00      Eric Fu (BC Centre for Excellence in HIV/AIDS), John Petkau (University of British Columbia)  
Relative Efficiency of Modified Poisson Regression and Log-Binomial Regression / L'efficacité  
relative de la régression de Poisson et de la régression log-binomiale modifiées14:30 - 16:00      Rihong Hui & Hanna Jankowski (York University)  
Cox Proportional Hazards with Shape-Constraints / Risques proportionnels de Cox avec  
contraintes de forme

- 14:30 - 16:00 Xiang Ji & Hanna Jankowski (York University)  
Simultaneous Confidence Regions for Effective Dose / Régions de confiance simultanée pour dose efficace
- 14:30 - 16:00 Rajat Malik, Rob Deardon & Grace Kwong (University of Guelph), Ben Cowling (The University of Hong Kong)  
Modeling the Spread of Influenza within Households in Hong Kong / Modélisation de la propagation de l'influenza dans les maisons de Hong Kong
- 14:30 - 16:00 Rachel O'Reilly, Paul D. McNicholas, Shanna Wilson & Chevonne Carlow (University of Guelph)  
Supervised Gene Selection in Gene Expression Microarray Data / Sélection de gènes supervisée dans les données d'expression génique sur micro-réseaux
- 14:30 - 16:00 Irene Urbik & Paul McNicholas (University of Guelph)  
Model-based Classification of Time Course Data Using the Skew t Distribution / Classification basée sur des modèles de données de décours temporel au moyen de la distribution t asymétrique
- 14:30 - 16:00 Monica Wong, Paul D. McNicholas & David M. Mutch (University of Guelph)  
Application of Clustering in Nutrigenomics to Identify Diet Intervention Sub-populations / Application de classification en nutriginomique pour déterminer l'intervention alimentaire chez les sous-populations
- 14:30 - 16:00 Yuan Xu (University of Victoria)  
Use of Closed Population Models to Estimate the Number of Injection Drug Users in Victoria / L'utilisation de modèles de population fermée pour estimer le nombre d'utilisateurs de drogues par injection à Victoria
- 14:30 - 16:00 Lin Zhang (University of Guelph), Rob Deardon (Department of Mathematics and Statistics, University of Guelph), Dongmei Chen (Department of Geography, Queen's University)  
Modeling 2009 H1N1 Epidemic in Southern Ontario over Time and Space / Modélisation temporelle et spatiale de l'épidémie H1N1 de 2009 dans le sud de l'Ontario

---

**15:00 - 16:00**      **Session 2E-A**      (abstracts/résumés - page 163)      **Festival Theatre**

**CJS Award Address**

**Allocution du récipiendaire du Prix de la RCS**

Special Session / Conférence spéciale

Chair/Président: Duncan Murdoch (University of Western Ontario)

Organizer/Responsable: Mary Lesperance (University of Victoria)

- 15:00 - 16:00 Isabel Molina (Carlos III University-Madrid), J. N. K. Rao (Carleton University)  
Small Area Estimation of Poverty Indicators / Estimation des mesures de la pauvreté pour petits domaines

---

**15:00 - 16:00**      **Session 2C-A**      (abstracts/résumés - page 164)      **BAC 241**

**Isobel Loutit Invited Address**

**Allocution Isobel Loutit**

Special Session / Conférence spéciale

Organizer and Chair / Responsable et président: Will Welch (University of British Columbia)

Session sponsored by the Business and Industrial Statistics Section

Session parrainée par le Groupe de statistique industrielle et de gestion

- 15:00 - 16:00 Dennis KJ Lin (Pennsylvania State University)  
BIG Statistics / Statistiques BIG
-

<b>16:00 - 17:30</b>	<b>Session 2F-A</b>	(abstracts/résumés - page 165)	<b>BAC241</b>
<b>Why a Bayesian be?</b>			
<b>Conférence de vue d'ensemble : Pourquoi être Bayésien ?</b>			
Introductory Overview Lecture / Conférence de vue d'ensemble			
Chair/Président: John Petkau (University of British Columbia)			
Organizer/Responsable: Shelley Bull (University of Toronto)			
16:00 - 17:30	Andrew Willan (SickKids Research Institute and University of Toronto)	Why a Bayesian Be? / Pourquoi être Bayésien ?	
<b>16:00 - 17:30</b>	<b>Session 2F-I1</b>	(abstracts/résumés - page 166)	<b>BAC138</b>
<b>Financial Time Series</b>			
<b>Séries chronologiques financières</b>			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Rafal Kulic (University of Ottawa)			
Session sponsored by the Probability Section			
Session parrainée par le Groupe de probabilité			
16:00 - 16:30	Philippe Soulier (Université Paris X), Rafal Kulic (Université d'Ottawa), Hurvich Clifford (New York University)	Extremal Properties of Stochastic Volatility Models / Propriétés extrémales de processus à volatilité stochastique	
16:30 - 17:00	Clifford Hurvich (New York University.), Alexander Aue (University of California, Davis), Lajos Horvath (University of Utah), Philippe Soulier (Université Paris X)	Limit Laws in Transaction-Level Asset Price Models / Lois limites dans les modèles de niveaux de transaction des prix des actifs	
17:00 - 17:30	Zhou Zhou (University of Toronto)	Generalized Likelihood Ratio Test for Nonstationary Time Series / Test du rapport de vraisemblance généralisé pour les séries chronologiques non stationnaires	
<b>16:00 - 17:30</b>	<b>Session 2F-I2</b>	(abstracts/résumés - page 168)	<b>BAC 234</b>
<b>Gene-environment Effects in Pharmacogenetic and Related Settings</b>			
<b>Effets génotype – environnement en pharmacogénétique et domaines connexes</b>			
Invited Paper Session / Présentations sur invitation			
Organizer and Chair / Responsable et président: Judy-Anne Chapman (Queen's University)			
16:00 - 16:30	Brooke Fridley (Mayo Clinic, Minnesota)	Analysis of Pharmacogenomics Studies: Past, Present, Future / Analyse d'études pharmacogénétiques : passé, présent, futur	
16:30 - 17:00	Josée Dupuis (Boston University), Alisa K. Manning & Michael LaValley (Boston University School of Public Health), James B. Meigs (Massachusetts General Hospital and Harvard Medical School), L. Adrienne Cupples (Boston University School of Public Health)	Meta-Analysis of Genome-Wide Association Results Allowing for Gene-by-Environment Interactions / Méta-analyse d'études d'associations pangénomiques permettant l'incorporation d'un facteur modifiant environnemental	
17:00 - 17:30	Fei Zou (University of North Carolina at Chapel Hill)	Bayesian Semiparametric QTL Mapping for Gene-Environment Interactions / Cartographie bayésienne semi-paramétrique de loci à caractère quantitatif dans les interactions gènes-environnement	

---

**16:00 - 17:30**      **Session 2F-I3**      (abstracts/résumés - page 170)      **BAC132**

**Statistics Literacy**

**Littératie en statistique**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Bovas Abraham (University of Waterloo)

- 16:00 - 16:30      Neville Davies (Royal Statistical Society Centre for Statistical Education)  
The Royal Statistical Society getstats Campaign: Ten years to Statistical Literacy? / La campagne « getstats » de la Royal Statistical Society : littératie statistique d'ici dix ans ?
- 16:30 - 17:00      Robert Rodriguez (President-Elect American Statistical Association)  
Making the Case for Statistical Literacy: Recent ASA Initiatives / Le bien-fondé de l'instruction statistique : les récentes initiatives de l'ASA
- 17:00 - 17:30      Mary Townsend (Statistics Canada)  
Statistical Literacy in Youth / Littératie statistique chez les jeunes
- 

**16:00 - 17:15**      **Session 2F-C1**      (abstracts/résumés - page 172)      **BAC 141**

**Clustering and Classification**

**Mise en grappes et classification**

Contributed Paper Session / Séance de communications libres

Chair/Président: Wayne Olford (University of Waterloo)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

- 16:00 - 16:15      Adrian Waddell & R. Wayne Oldford (University of Waterloo)  
Interactive Visual Clustering of High Dimensional Data using Navigation Graphs / Classement interactif visuel des données de haute dimension par l'utilisation de graphiques de navigation
- 16:15 - 16:30      Jeffrey L. Andrews & Paul D. McNicholas (University of Guelph)  
Semi-Supervised Variable Selection / Sélection de variables semi-supervisée
- 16:30 - 16:45      Brian Franczak, Paul McNicholas & Ryan P. Browne (University of Guelph)  
Model-based Clustering Using Mixtures of Asymmetric Laplace Distributions / Classification à base de modèles utilisant des mélanges de distributions de Laplace asymétriques
- 16:45 - 17:00      Antonio Ciampi, Luc Villandré & Benjamin Rich (McGill University)  
Classification and Regression Trees with Soft Nodes / Arbres de classification et de régression à noeuds tendres
- 17:00 - 17:15      Jessy Donelle & Xu Wang (St. Francis Xavier University), Barbara Caffery (Yorkville Eye Clinic)  
Differentiating Between Sjogren's Syndrome and Dry Eye Disease: An Analysis using Random Forests / Différenciation du syndrome de Sjögren et du syndrome des yeux secs : Analyse au moyen de forêts aléatoires
- 

**16:00 - 17:30**      **Session 2F-C2**      (abstracts/résumés - page 175)      **BAC 239**

**Longitudinal Models: Measurement Error, Missing Data, and Misspecification**

**Modèles longitudinaux : erreurs de mesure, données manquantes et erreurs de spécification**

Contributed Paper Session / Séance de communications libres

Chair/Président: Joel Dubin (University of Waterloo)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

- 16:00 - 16:15      Wei Liu (York University, Canada), Grace Y. Yi (University of Waterloo, Canada), Lang Wu (University of British Columbia, Canada)  
Simultaneous Inference and Bias Analysis for Longitudinal Data with Covariate Measurement Error and Missing Responses / Inférence simultanée et analyse de biais pour données longitudinales avec erreur de mesure des covariables et réponses manquantes

- 16:15 - 16:30 Yunqi Ji & Fan Zhaozhi (Memorial University of Newfoundland)  
Analysis of Longitudinal Categorical Data Subject to Misclassification and Informative "Unsure" Responses / Analyse de données catégorielles longitudinales sujettes à la mauvaise classification et aux réponses informatives « Incertain »
- 16:45 - 17:00 Eleanor Pullenayegum (McMaster University)  
Longitudinal Data Subject to Intermittent Measurement / Données longitudinales exposées à des mesures intermittentes
- 17:00 - 17:15 Mireille Schnitzer & Erica Moodie (McGill University), Robert Platt (McGill University, Montreal Children's Hospital Research Institute)  
Targeted Maximum Likelihood for Longitudinal Marginal Structural Models Under Density Misspecification / Les méthodes de maximum de vraisemblance ciblé pour modèles structuraux marginaux où la densité est mal déterminée
- 17:15 - 17:30 Abderazzak Mouiha (Centre de Recherche Université Laval Robert Giffard), Simon Duchesne (Radiology Department, and Centre de Recherche Robert-Giffard, Université Laval, Canada)  
Alzheimer's Disease Biomarkers Dynamics: Best Fit Models / Dynamique des biomarqueurs de la maladie d'Alzheimer : les modèles les mieux ajustés

---

**16:00 - 17:30**      **Session 2F-C3**      (abstracts/résumés - page 178)      **BAC142**

**Survey Methods: Estimation**

**Méthodes d'enquête : estimation**

Contributed Paper Session / Séance de communications libres

Chair/Président: Geneviève Demers (Statistique Canada)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

- 16:00 - 16:15 Pierre Lavallée & Sébastien Labelle-Blanchet (Statistique Canada)  
Indirect Sampling Applied to Skewed Populations / Le sondage indirect appliqué aux populations asymétriques
- 16:15 - 16:30 Steve Matthews & Tyler Kirkland (Statistics Canada)  
Statistics Canada's Survey on Commercial and Institutional Energy Use: An Application of Indirect Sampling / L'Enquête sur la consommation d'énergie du secteur commercial et institutionnel de Statistique Canada : Une application du sondage indirect
- 16:30 - 16:45 Kyle Vincent (Simon Fraser University)  
On the Use of Adaptive Sampling for Estimating Population Sizes / Sur l'utilisation de l'échantillonnage adaptatif pour estimer les tailles de population
- 16:45 - 17:00 Abdellatif Demnati (Statistique Canada)  
On a Calibrated Estimator for Two-phase Sampling Design in the Presence of Nonresponse / Estimateur calé pour plan de sondage à deux phases en présence de non-réponse
- 17:00 - 17:15 Geneviève Demers (Statistique Canada)  
Ever-alive Units in the Unified Enterprise Survey: Is a Special Treatment Worth it? / Les unités "ever-alive" dans l'Enquête unifiée auprès des entreprises : Un traitement particulier vaut-il la peine ?
- 17:15 - 17:30 Kate Wilder (Statistics Canada), Steven Thomas  
Combining Data from Multiple Surveys / Combiner les données de plusieurs enquêtes

---

**17:30 - 18:30**      **BAC 236**

**Information on 2011 NSERC Discovery Grant Competition**

**Information sur le concurs 2011 des subventions à la découverte du CRSNG**

**Wednesday June 15****mercredi 15 juin****9:00 - 9:45      Session 3A-A      (abstracts/résumés - page 181)      Festival Theatre****Pierre Robillard Award Address****Allocution du récipiendaire du Prix Pierre-Robillard**

Special Session / Conférence spéciale

Chair/Président: Keumhee Carriere Chough (University of Alberta)

9:00 - 9:45      Gun Ho Jang (University of Pennsylvania)  
 Invariant Procedures for Model Checking, Checking for Prior-Data Conflict and Bayesian Inference / Procédures invariantes pour la vérification d'un modèle, la détection de conflits données-*a priori* et l'inférence bayésienne

**9:45 - 10:15      Festival Theatre****Break / Pause-café****10:15 - 11:45      Session 3B-A      (abstracts/résumés - page 182)      BAC241****Approaches to Pathway Modeling in Genetic Epidemiology****Approches de la modélisation des voies en épidémiologie génétique**

Introductory Overview Lecture / Conférence de vue d'ensemble

Organizer and Chair / Responsable et président: Shelley B Bull (University of Toronto)

10:15 - 11:45      Duncan Thomas (University of Southern California)  
 Approaches to Pathway Modeling in Genetic Epidemiology / Approches de modélisation de processus réactionnels en épidémiologie génétique

**10:15 - 11:45      Session 3B-11      (abstracts/résumés - page 183)      BAC 236****Ecological Sampling****Échantillonnage écologique**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Chris Field (Dalhousie University)

10:15 - 10:45      Alan Welsh (Australian National University)  
 A Sample Survey View of Occupancy/Detection Surveys / Une perspective d'échantillonnage d'enquête pour les enquêtes d'occupation et de détection

10:45 - 11:15      Joanna Mills Flemming (Dalhousie University)  
 The Ocean Tracking Network: A Data Perspective / Le « Ocean Tracking Network » : les données en vedette

11:15 - 11:45      Rick Routledge, Discussant (Simon Fraser University)

**10:15 - 11:45      Session 3B-12      (abstracts/résumés - page 185)      BAC 234****Order Statistics and Related Inference, Ranked Set Sampling and Censoring Methodology****Statistiques d'ordre et inférence connexe, échantillonnage d'ensembles ordonnés et méthodologie de censure**

Invited Paper Session / Présentations sur invitation

Chair/Président: Alexandre Leblanc (University of Manitoba)

Organizer/Responsable: Katherine Davies (University of Manitoba)

10:15 - 10:45      Mohammad Jafari Jozani (University of Manitoba), Sayed Jamal Mirkamali (Allameh Tabatabaie University)  
 Maxima Nomination Sampling with Application in Constructing Quality Control Charts for Attributes / Échantillonnage par nomination de maxima avec application dans la construction de cartes de contrôle pour attributs

- 10:45 - 11:15 Katherine Davies (University of Manitoba), N. Balakrishnan (McMaster University), Jerome P. Keating (University of Texas at San Antonio), Robert L. Mason (Southwest Research Institute)  
Optimal Plotting Points Based on Pitman Closeness / Choix optimal du tracé de points selon la proximité de Pitman
- 11:15 - 11:45 Barry Arnold (University of California at Riverside)  
Cantor Order Statistics Without Applications / Statistiques d'ordre de Cantor sans applications

---

**10:15 - 11:45      Session 3B-I3      (abstracts/résumés - page 187)      BAC142**

**Psychographic Surveys and Behavioural Surveys - Can information from one be used to accurately predict characteristics of the other?**

**Enquêtes psychographiques et enquêtes sur le comportement – peut-on utiliser des informations provenant des unes pour prévoir avec précision les caractéristiques des autres ?**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Delbert Stewart (IMS Brogan)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

- 10:15 - 10:35 John S. Lyons (University of Ottawa)  
Measurement as Communication: Communimetric Theory and Survey Design / Mesures comme communication : Théorie du communimetric et plan d'enquête
- 10:35 - 10:55 Jane E. Ledingham, Jessica Kurta, Tiffany Pursoo, Leah Puddester & Laurie Clark (University of Ottawa)  
Children's Choices of Where to Play on the Playground as a Function of Emotional State / Les choix des enfants d'où jouer sur le terrain de jeu en fonction de leur état émotif
- 10:55 - 11:15 David W. Jamieson (Environics Research Group), Delbert E. Stewart & David Stewart (IMS Brogan)  
The Use of Demographic and Behavioural Characteristics to Predict Psychographic Profiles / L'emploi de caractéristiques démographiques et comportementales pour prédire les profils psychographiques
- 11:15 - 11:45 Discussion

---

**10:15 - 11:45      Session 3B-I4      (abstracts/résumés - page 189)      BAC138**

**Randomness in Discrete Mathematics and Biology**

**Stochastocité en mathématiques discrètes et biologie**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Ed Perkins (University of British Columbia)

Session sponsored by the Probability Section

Session parrainée par le Groupe de probabilité

- 10:15 - 10:45 Richard Liang (University of British Columbia), Steve Evans (University of California, Berkeley)  
Stepping-Stone Models with Continuous Geography and Delayed Coalescence / Modèles du « stepping-stone » avec zone géographique continue et coalescence retardée
- 10:45 - 11:15 Ori Gurel-Gurevich (University of British Columbia), Itai Benjamini (Weizmann Institute of Science), Russell Lyons (Indiana University), Oded Schramm (Microsoft Research)  
Recurrence of the Simple Random Walk Path / Récurrence de la trajectoire de la marche aléatoire simple
- 11:15 - 11:45 Xiaowen Zhou (Concordia University)  
A Generalized Fleming-Viot Process / Un processus de Fleming-Viot généralisé
-

<b>10:15 - 11:45</b>	<b>Session 3B-15</b>	(abstracts/résumés - page 191)	<b>BAC141</b>
<b>Statistical Problems Arising in the Lumber Industry</b>			
<b>La statistique en assurance de la qualité du bois d'œuvre</b>			
Invited Paper Session / Présentations sur invitation			
Chair/Président: Will Welch (University of British Columbia)			
Organizer/Responsable: Jim Zidek (University of British Columbia)			
Session sponsored by the Business and Industrial Statistics Section			
Session parrainée par le Groupe de statistique industrielle et de gestion			
10:15 - 10:45	Conroy Lum (FP Innovations, Vancouver), James Zidek & Carolyn Taylor (University of British Columbia) An Overview of the use of Statistics and Stochastic Modeling in Forest Products Research / Un aperçu de l'utilisation de la modélisation statistique et stochastique dans la recherche sur les produits de la forêt		
10:45 - 11:15	Jun Zhu (University of Wisconsin, Madison), Yanbing Zheng (University of Kentucky), Brian Aukema (University of Minnesota), Allan Carroll (University of British Columbia), Kishan Sambaraju (University of Northern British Columbia) Spatial-Temporal Statistics for Studying Mountain Pine Beetle Infestation of Pine Forests in Western Canada / Des statistiques spatio-temporelles pour étudier l'infestation de dendroctone du pin ponderosa dans les forêts de pin de l'ouest canadien		
11:15 - 11:45	Jiguo Cao & Jing Cai (Simon Fraser University), Jim Zidek (University of British Columbia) Comparing and Selecting Damage Accumulation Models / Comparer et choisir des modèles d'accumulation des dommages		
<b>10:15 - 11:30</b>	<b>Session 3B-C1</b>	(abstracts/résumés - page 193)	<b>BAC 132</b>
<b>Design and Analysis of Clinical Studies and Trials</b>			
<b>Conception et analyse d'études et d'essais cliniques</b>			
Contributed Paper Session / Séance de communications libres			
Chair/Président: James Hanley (McGill University)			
Session sponsored by the Biostatistics Section			
Session parrainée par le Groupe de biostatistique			
10:15 - 10:30	Stephen D Walter (McMaster University) Optimal Treatment Allocation in Two-Stage Randomised Clinical Trials / Optimisation de l'allocation aux traitements dans les essais cliniques aléatoires à deux phases		
10:30 - 10:45	Qing Guo, Geoff Hall, Margaret McKinnon, Lehana Thabane & Eleanor Pullenayegum (McMaster University) Setting Sample Size Using Cost-efficiency in fMRI Studies / Établissement de la taille d'échantillon selon le ratio coût-efficacité en études d'IRMf		
10:45 - 11:00	Xuan Li & Xikui Wang (University of Manitoba) Response Adaptive Designs with Misclassified Responses / Plans de réponse adaptative avec réponses mal classées		
11:00 - 11:15	Yanqing Yi (Memorial University of Newfoundland) Small Sample Inference for Response Adaptive Designs / Inférence sur des échantillons de petite taille pour les plans à réponses adaptées		
11:15 - 11:30	Erica Moodie (McGill University), Bibhas Chakraborty (Columbia University) Estimation of Optimal Dynamic Treatment Rules with Shared Parameters and Non-Regularity / Estimation pour les régimes de traitement dynamique optimaux avec des paramètres partagés et de la non-régularité		



---

**10:15 - 11:45**      **Session 3B-C2**      (abstracts/résumés - page 196)      **BAC237**

**Harmonic Analysis and Spectral Methods for Time Series**

**Analyse harmonique et méthodes spectrales pour les séries chronologiques**

Contributed Paper Session / Séance de communications libres

Chair/Président: Mary Thompson (University of Waterloo)

Organizer/Responsable: Wesley Burr (Queen's University)

- 10:15 - 10:30      David J. Thomson (Queen's University)  
Crossing Problems for Spectrum Estimates / Problèmes croisés pour les estimateurs du spectre
- 10:30 - 10:45      Wesley Burr (Queen's University Dept. of Mathematics and Statistics)  
Cross-Spectral Method for Gapfilling Time Series / Méthode interspectrale de complétion des séries chronologiques
- 10:45 - 11:00      Azadeh Moghtaderi & Glen Takahara (Queen's University)  
Tests for Stationarity Based on Minors and Singular Values of the Nonstationary Spectrum / Tests pour la stationnarité basés sur les mineurs et valeurs singulières du spectre non stationnaire
- 11:00 - 11:15      Charlotte Lisa Haley (Queen's University)  
Detection of Changes in Activation in ECG Time Series Modeling / Détection des changements d'activation dans la modélisation de séries chronologiques d'électrocardiogrammes (ECG)
- 11:15 - 11:30      Karim J. Rahim & David J. Thomson (Queen's University)  
Spectral Analysis of Burgundy Pinot Noir Grape Harvest Dates / Analyse spectral des dates de récoltes des raisins Pinot noirs de Bourgogne
- 11:30 - 11:45      Wesley Burr (Queen's University Dept. of Mathematics and Statistics), Azadeh Moghtaderi & Glen Takahara (Queen's University), Hwashin H. Shin (Health Canada)  
Estimation and Extraction of Harmonic Features from Pollution and Mortality Time Series / Estimation et extraction des caractéristiques harmoniques de séries chronologiques sur la pollution et la mortalité

---

**10:15 - 11:30**      **Session 3B-C3**      (abstracts/résumés - page 199)      **BAC 239**

**Modelling Correlation Structure in Repeated Measures and Multivariate Data**

**Modélisation de la structure de corrélation dans les mesures répétées et les données multivariées**

Contributed Paper Session / Séance de communications libres

Chair/Président: Sohee Kang (Trent University)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

- 10:15 - 10:30      Tolulope T. Sajobi & Lisa M. Lix (University of Saskatchewan), Bolanle M. Dansu (University of Agriculture, Abeokuta, Nigeria)  
Robust Descriptive Discriminant Analysis for Repeated Measures Data / Analyse discriminante descriptive robuste pour les mesures répétées
- 10:30 - 10:45      Vineetha K.V. Warriyar & Sutradhar Brajendra (Memorial University of Newfoundland)  
Generalized Least Square Estimation in Semiparametric Linear Models with Correlated Errors / Estimation des moindres carrés généralisée en modèles linéaires semi-paramétriques avec erreurs corrélées
- 10:45 - 11:00      Alina Dyachenko (McGill University), Antonio Ciampi, Martin Cole & Jane McCusker (McGill University)  
Latent Classes and Hidden Markov Chains to Describe State and Course of a Disease: Delirium as an Example / Classes latentes et chaînes de Markov cachées pour décrire l'état et le cours d'une maladie : l'exemple du Délirium

- 11:00 - 11:15 Héla Romdhani, Louis-Paul Rivest & Lajmi Lakhal Chaieb (Département de mathématiques et de statistique, Université Laval)  
A Kendall's  $\tau$  Estimator for Exchangeable Data with an Application to One Random Factor Analysis of Variance. / Un estimateur  $\tau$  de Kendall pour les données permutable et une application à l'analyse de variance a un facteur aléatoire
- 11:15 - 11:30 Swarna Weerasinghe (Dalhousie University)  
Multilevel Time Series Decomposition Methods to Explore the Association between Asthma Hospital Admissions and Ozone Levels / Méthodes de décomposition à plusieurs niveaux de séries chronologiques pour explorer l'association entre les admissions d'asthmatiques dans les hôpitaux et les niveaux d'ozone

11:45 - 13:00

Wheelock Dining Hall

Lunch / Déjeuner

13:00 - 14:00 **Session 3C-A** (abstracts/résumés - page 202) **Festival Theatre****CRM-SSC Prize in Statistics Address****Allocution du récipiendaire du Prix CRM-SSC**

Special Session / Conférence spéciale

Chair/Président: David Stephens (McGill University)

- 13:00 - 14:00 Edward Susko (Dalhousie University)  
Statistical Issues in Molecular Evolution / Problèmes statistiques dans le domaine de l'évolution moléculaire

13:00 - 14:30 **Session 3C-I1** (abstracts/résumés - page 203) **BAC 132****Consultations for Statistical Consultants: Lessons Learned from the Field****Consultations pour les consultants en statistique : leçons apprises sur le terrain (discussion en groupe)**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Depeng Jiang (University of Manitoba)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

- 13:00 - 13:05 Depeng Jiang (University of Manitoba)  
Round-Table Discussion - Consultations for Statistical Consultants: Lessons Learned from the Field / Table ronde - consultations pour les conseillers en statistique : leçons apprises sur le terrain

- 13:00 - 14:30 John Amrhein (McDougall Scientific Ltd)  
13:00 - 14:30 Tim Ramsay (Ottawa Hospital Research Institute and University of Ottawa)  
13:00 - 14:30 Lehana Thabane (McMaster University)

13:00 - 14:00 **Session 3C-I2** (abstracts/résumés - page 204) **BAC 138****Rank Set Sampling: New Developments and Applications****Échantillonnage d'ensembles ordonnés : récents développements et applications**

Invited Paper Session / Présentations sur invitation

Chair/Président: Sylvia Esterby (University of British Columbia)

Organizer/Responsable: Abdulkadir Hussein (University of Windsor)

- 13:00 - 13:30 S. Ejaz Ahmed (University of Windsor)  
Meta Analysis and Shrinkage Estimation Strategies in Ranked Set Sampling / Méta-analyse et stratégies d'estimation par rétrécissement dans l'échantillonnage d'ensembles ordonnés
- 13:30 - 14:00 Abdulkadir Hussein (University of Windsor), Hassen Muttlak & Essam Al-Sawi (King Fahd University of Petroleum and Minerals, Saudi Arabia)  
Group Sequential Methods Based on Ranked Set Samples / Méthodes séquentielles groupées fondées sur des échantillons d'ensembles ordonnés

---

<b>13:00 - 14:30</b>	<b>Session 3C-I3</b>	(abstracts/résumés - page 206)	<b>BAC141</b>
<b>Recent Advances in Multivariate Control Charts</b>			
<b>Limites de contrôle multivariées</b>			
Invited Paper Session / Présentations sur invitation			
Chair/Président: David Matthews (University of Waterloo)			
Organizer/Responsable: Shoja'eddin Chenouri (University of Waterloo)			
Session sponsored by the Business and Industrial Statistics Section			
Session parrainée par le Groupe de statistique industrielle et de gestion			
13:00 - 13:30	Shoja'eddin Chenouri (University of Waterloo), Bitā Roushanaei Depth Based Multivariate Nonparametric Control Charts / Cartes de contrôle non paramétriques multivariées basées sur la profondeur		
13:30 - 14:00	Asokan Variyath (Memorial University of Newfoundland), Shojaeddin Chenouri (University of Waterloo) Robust Multivariate Control Charts for Monitoring Process Mean and Covariance Matrix / Cartes de contrôle multivariées pour surveiller la moyenne d'un processus et la matrice de covariance		
14:00 - 14:30	Roman Viveros-Aguilera (McMaster University), Jock MacKay & Stefan Steiner (University of Waterloo) Monitoring Profiles / Profils de surveillance		

---

---

Abstracts • Résumés

**1A-A2: SSC Presidential Invited Address  
Allocution de l'invité du Président de la SSC**

---

Chair/Président: Don McLeish (University of Waterloo)

**Room/Salle: Festival Theatre**

**Abstract/Résumé**

---

**IVAN FELLEGI** (Chief Statistician of Canada Emeritus / Statisticien en chef émérite du Canada)

[Monday June 13 / lundi 13 juin, 8:30]

*Official Statistics, Public Confidence and Lessons from the Story of the 2011 Canadian Census*  
*Statistiques officielles, confiance du public et leçons tirées du Recensement canadien 2011*

The paper is inspired by events concerning the 2011 Census long form being designated by the Canadian government as a voluntary survey. It outlines first of all why official statistics are important for Canada: in terms of democratic accountability, evidence-based decision making in all walks of life, research, and finally in terms of what I call "honest information brokering". It then considers why public confidence is crucial for these roles to be successfully fulfilled and the characteristics of statistical agencies that lead to public trust, including freedom from inappropriate political interference. It proceeds to describe and provide examples of various kinds of inappropriate political interference, ranging from the blatant to the more subtle. Next, the paper examines the events surrounding the 2011 Census long form and concludes that they represent inappropriate, though subtle, political interference. It ends by analysing some possible remedies for the future and recommends next steps.

Cette présentation est inspirée par les événements ayant suivi la décision du gouvernement canadien de faire du formulaire détaillé du Recensement 2011 une enquête à participation volontaire. Je commence par exposer pourquoi les statistiques officielles sont importantes pour le Canada : pour la responsabilité démocratique, la prise de décisions fondées sur les faits à tous les niveaux, pour la recherche et enfin pour ce que j'appelle le "courtage en information honnête". J'explore ensuite pourquoi il est essentiel d'avoir la confiance du public pour que ces rôles soient bien remplis et quelles caractéristiques des organismes statistiques suscitent cette confiance, notamment l'absence de toute interférence politique inappropriée. Je décris et donne des exemples d'interférences politiques inappropriées, certaines flagrantes, d'autres plus subtiles. Puis, j'examine les événements entourant le formulaire détaillé du Recensement 2011 et conclus qu'ils représentent une interférence politique inappropriée, bien que subtile. Pour finir, j'analyse d'éventuelles solutions pour l'avenir et je recommande certaines mesures à prendre.

## 1B-I1: Future of Functional Data Analysis L'avenir de l'analyse de données fonctionnelle

---

Organizer and Chair / Responsable et président: Jiguo Cao (Simon Fraser University)

**Room/Salle: BAC236**

**Abstracts/Résumés**

---

**FANG YAO** (University of Toronto), **HANS GEORGE MUELLER** (University of California at Davis)

[Monday June 13 / lundi 13 juin, 10:15]

*Additive Modeling of Functional Gradients*

*Modélisation additive de gradients fonctionnels*

In the framework of a functional regression with scalar response, we propose the concept of functional gradients that are defined by directional derivatives and indicate how changes in the predictor function in a specified direction are associated with the scalar response. We develop estimation procedure within an additive regression framework, which requires nothing more than estimating derivatives of one-dimensional nonparametric regressions, and thus is computationally very straightforward to implement. We demonstrate the estimation and interpretation of the resulting functional derivatives and functional gradient fields in a study of the dependence of lifetime fertility of flies on early life reproductive trajectories.

Dans le cadre d'une régression fonctionnelle avec réponse scalaire, nous proposons le concept des gradients fonctionnels qui sont définis par les dérivées directionnelles et qui indiquent comment les changements de la fonction prédictive dans une direction spécifiée sont associés à la réponse scalaire. Nous développons une procédure d'estimation dans le cadre d'une régression additive qui n'exige rien de plus que d'estimer les dérivés des régressions unidimensionnelles non-paramétriques, et qui est donc très direct à implémenter au niveau calculatoire. Nous montrons l'estimation et l'interprétation des dérivées fonctionnelles résultantes et des champs gradients fonctionnels dans une étude de la dépendance de la durée de vie de la fertilité des mouches sur les trajectoires reproductrices en début de vie.

---

**JASON D NIELSEN** (Carleton University)

[Monday June 13 / lundi 13 juin, 10:45]

*A Functional Approach to the Analysis of Recurrent Event Data*

*Une approche fonctionnelle à l'analyse des données d'évènements récurrents*

I will discuss an adaptive semi-parametric model for the analysis of recurrent event data. The basic model assumes that the data arise from a mixed nonhomogenous Poisson process where frailties account for heterogeneity. The intensity of the process is modeled as a family of smooth functions generated via a differential operator. This allows for flexible estimates of the intensity function with an underlying structure governed by the solution of a differential equation. Estimation and inference are outlined and the finite sample properties of the proposed methodology investigated. Finally, the method will be illustrated on some health outcome data.

Je vais parler d'un modèle adaptatif semi-paramétrique pour l'analyse des données d'évènements récurrents. Le modèle fondamental suppose que les données proviennent d'un processus de Poisson mixte non-homogène où des faiblesses expliquent l'hétérogénéité. L'intensité du processus est modélisée comme une famille de fonctions lisses engendrée par un opérateur différentiel. Ceci permet une estimation flexible de la fonction d'intensité avec une structure sous-jacente dirigée par la solution d'une équation différentielle. L'estimation et l'inférence sont mises en évidence, et les propriétés d'échantillon fini de la méthodologie proposée sont examinées. Enfin, cette méthode sera illustrée sur quelques données de résultats cliniques.

## 1B-I1: Future of Functional Data Analysis L'avenir de l'analyse de données fonctionnelle

---

**SOFIA MOSESOVA** (Genentech), **HUGH CHIPMAN** (Acadia University)

[Monday June 13 / lundi 13 juin, 11:15]

*Bayesian Hierarchical Modeling of Functional Data*

*Modélisation bayésienne hiérarchique pour données fonctionnelles*

The goal of this talk is to introduce an unsupervised learning technique for analysis of functional data. A Bayesian hierarchical model is used to cluster low-dimensional summaries of the functional curves into meaningful groups. The belief is that the clusters correspond to distinct types of processes (e.g. “good” or “faulty” assembly in manufacturing). New observations can be assigned to one of these by calculating the probabilities of belonging to each cluster. Mahalanobis distances are used to identify new observations not belonging to any of the existing clusters. Synthetic and real data are used to validate the results.

Le but de cet exposé est d'introduire une technique d'apprentissage non-supervisée pour l'analyse de données fonctionnelles. Un modèle bayésien hiérarchique est utilisé pour grouper les résumés basse dimension des courbes fonctionnelles de manière sensée. On espère que les grappes correspondent à des types de processus distincts (par ex. l'assemblage « bon » ou « défectueux » dans l'industrie). De nouvelles observations peuvent être assignées à l'une d'entre elles en calculant les probabilités d'appartenir à chaque grappe. La distance de Mahalanobis est utilisée afin d'identifier les nouvelles observations qui n'appartiennent pas aux grappes existantes. Des données synthétiques et réelles sont utilisées pour valider les résultats.

## 1B-I2: Imperfect Administrative Health Data: Can we do better? Données de santé administratives imparfaites : pouvons-nous faire mieux?

---

Organizer and Chair / Responsable et président: Lisa Lix (University of Saskatchewan)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 132**

**Abstracts/Résumés**

---

**YULEI HE** (Harvard University), **MARY BETH LANDRUM & ALAN ZASLAVSKY** (Harvard Medical School)

[Monday June 13 / lundi 13 juin, 10:15]

*Combining Information on Health Services Variables from Multiple Sources*

*Comment combiner des informations de sources multiples sur les variables de services de santé*

Measures of certain services variables (e.g., hospice use) are important quality indicators of cancer care. The Cancer Care Outcomes Research and Surveillance (CanCORS) study collected these variables from several sources including patient survey, medical records abstraction, cancer registry databases, and Medicare claims data. The sample coverages of these sources are different, and the reporting might be subject to error. We first estimate the patterns of misreporting assuming one source (e.g., claims) is a gold standard. The misreported outcome in other sources can therefore be corrected/imputed using a statistical model. Valid analysis can then be based on multiply imputed data.

Les mesures de certaines variables de services (p. ex., recours aux centres de soins palliatifs) sont d'importants indicateurs de la qualité des soins aux personnes atteintes de cancer. L'étude « Cancer Care Outcomes Research and Surveillance » (CanCORS) a recueilli ces variables via diverses sources (enquêtes auprès des patients, résumés de dossiers médicaux, registres sur le cancer et données tirées des demandes d'assurance-maladie). La représentativité des échantillons varie selon les sources et la communication des données peut être entachée d'erreurs. Nous commençons par estimer les types d'erreurs de communication en prenant une source (p. ex., demandes) comme gold standard. Les résultats entachés d'erreurs provenant d'autres sources peuvent ensuite être corrigés/imputés à l'aide d'un modèle statistique. Une analyse valable peut alors être effectuée sur la base de données en provenance de sources multiples.

---

**MARTIN LADOUCEUR, ELHAM RAHME, CHRISTIAN PINEAU & LAWRENCE JOSEPH** (McGill University)

[Monday June 13 / lundi 13 juin, 10:45]

*Investigating Robustness of Prevalence Estimates from Misclassified Administrative Data Using Bayesian Latent Class Models*

*Étude de la robustesse des estimations de la prévalence à partir de données administratives mal classées par l'utilisation de modèles de classes latentes*

Information collected in medical administrative databases is typically collected for reasons other than research, and many such databases have been shown to contain substantial proportions of misclassification errors. For example, errors in ascertaining which individuals have a given disease may bias prevalence estimates. We attempt to es-

On ne recueille habituellement pas de l'information provenant de bases de données administratives médicales pour la recherche. Un grand nombre de ces données contient des proportions substantielles d'erreurs de classement. Par exemple, des erreurs dans l'attribution d'une certaine maladie à certains individus peuvent biaiser les estimations de la prévalence. Nous tenterons d'estimer la prévalence de l'os-



## 1B-I2: Imperfect Administrative Health Data: Can we do better? Données de santé administratives imparfaites : pouvons-nous faire mieux?

---

estimate the prevalence of osteoarthritis (OA) among elderly Quebecers using a government administrative database. We compare a naive estimate relying solely on the physician diagnoses of OA listed in the database to estimates from several different Bayesian latent class models which adjust for misclassified physician diagnostic codes via use of other available diagnostic clues.

téoarthrite chez les personnes âgées québécoises par l'utilisation d'une base de données administrative du gouvernement. Nous comparerons une estimation naïve qui repose seulement sur les diagnostics d'ostéoarthrite des médecins généralistes. Ces diagnostics sont énumérés dans la base de données avec des estimations provenant de divers modèles bayésiens de classes latentes, qui ajustent les codes de diagnostic mal classés au moyen des indices diagnostiques disponibles.

---

**KHOKAN SIKDAR & JEFFREY DOWDEN** (Newfoundland and Labrador Centre for Health Information), **JOANNE STARES** (Public Health Agency of Canada), **JOSH SQUIRES & KAYLA COLLINS** (Newfoundland and Labrador Centre for Health Information)

[Monday June 13 / lundi 13 juin, 11:15]

*Challenges and Opportunities for Using Administrative Health Data to Improve Chronic Disease Case Ascertainment*  
*Défis et possibilités d'utilisation des données de gestion de santé pour améliorer la détermination des cas de maladie chronique*

Administrative data plays an important role in population health research and disease surveillance, especially in case ascertainment. These data are often readily available, cost-effective and typically encompass large populations. However, gaps in clinical information including physician billing practices and coding can compromise the ability to ascertain valid chronic disease cases; this can affect accuracy and completeness of data. This presentation uses various methods including ratio estimation and capture-recapture to adjust for data incompleteness in estimating diabetes incidence and prevalence in Newfoundland and Labrador. This research will show how these methods can aid in improving case ascertainment from incomplete data.

Les données de gestion jouent un rôle important dans la recherche sur la santé des populations et la surveillance des maladies, particulièrement en matière de détermination des cas. Ces données sont souvent facilement accessibles, économiques et englobent généralement de vastes populations. Toutefois, des lacunes dans l'information clinique, notamment en ce qui concerne les pratiques de facturation des médecins et le codage, peuvent affecter notre capacité à déterminer les cas de maladie chronique valides ; cela peut affecter l'exactitude et la complétude des données. Cette présentation utilise diverses méthodes, dont l'estimation du ratio et la méthode capture-recapture, pour tenir compte de l'incomplétude des données et estimer l'incidence et la prévalence du diabète en Terre-Neuve et Labrador. Nos recherches montreront comment ces méthodes peuvent aider à améliorer la détermination des cas en la présence de données incomplètes.

## 1B-I3: Statistical Issues in Longitudinal Data Analysis Problèmes statistiques en analyse de données longitudinales

---

Chair/Président: Julie Horrocks (University of Guelph)  
Organizer/Responsable: Leilei Zeng (Simon Fraser University)

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

**Room/Salle: BAC239**

### Abstracts/Résumés

---

**CECILIA COTTON** (University of Waterloo)

[Monday June 13 / lundi 13 juin, 10:15]

*Causal Effect Estimation for Longitudinal Survival Data*

*Estimation d'effets causaux pour données longitudinales de survie*

The problem of estimating the causal effect of some treatment on a survival outcome in the presence of an evolving history of time-varying covariates has received considerable attention of late. I will provide a brief introduction to the statistical framework for causal inference and then discuss the competing but related methods of inverse probability weighting and g-estimation. The required assumptions for causal inference as well as issues of missing data and censoring will also be discussed. Finally an application of the methods to a study from transfusion medicine will be presented.

Le problème de l'estimation de l'effet causal d'un traitement sur un résultat de survie en présence d'une histoire évoluant avec des covariables variant dans le temps a récemment suscité un intérêt considérable. Je fournirai une brève introduction au cadre statistique pour l'inférence causale, et discuterai ensuite des méthodes concurrentes mais reliées de pondération par probabilités inverses et de g-estimation. Les hypothèses requises pour l'inférence causale ainsi que les problèmes de données manquantes et de censure seront traités. Enfin, l'utilisation des méthodes pour une étude de médecine transfusionnelle sera présentée.

---

**JULI ATHERTON & DAVID WOLFSON** (McGill University), **EDENS THADDEUS**

[Monday June 13 / lundi 13 juin, 10:45]

*Optimal Design for Clinical Trials with an Unknown Delay in Treatment Effect*

*Plan optimal pour essais cliniques avec retard inconnu de l'effet du traitement*

In clinical trials, if there is an unknown delay before a treatment takes effect or if the treatment does not take effect, the subject specific times-to-effect can be viewed as the unknown change-points in a multi-path change-point setting. Once the data are collected, i) they may be used to estimate the size of each treatment effect and ii) to estimate the proportion of subjects that will experience a treatment effect. By introducing a design measure, we consider, optimal design for multi-path change-point problems, focusing on problems motivated by i) and ii) above.

Lors d'essais cliniques, s'il y a un retard inconnu avant que le traitement produise un effet, ou si le traitement ne produit pas d'effet, le retard d'effet propre au sujet peut être considéré comme un point de rupture inconnu dans le cadre de trajets multiples avec points de rupture. Une fois les données collectées, elles peuvent être utilisées pour estimer i) l'étendue de l'effet de chaque traitement et ii) la proportion des sujets qui vont connaître un effet du traitement. En implémentant une mesure du plan, nous examinons le plan optimal pour les problèmes de points de trajets multiples avec points de rupture, en se concentrant sur les problèmes découlant de i) et ii).

## 1B-I3: Statistical Issues in Longitudinal Data Analysis Problèmes statistiques en analyse de données longitudinales

---

**YANG ZHAO & MENG LIU** (University of Regina)

[Monday June 13 / lundi 13 juin, 11:15]

*A Uniform Approach for Longitudinal Data with Data Missing in Nonmonotonic Patterns*

*Une approche uniforme aux données longitudinales avec données manquantes de façon non-monotone*

Statistical analysis for longitudinal data with data missing in nonmonotonic patterns is very difficult. In this research we explain a uniform procedure to deal with longitudinal data with data missing in arbitrary nonmonotonic patterns, which including sorting data according to missing patterns, estimating regression parameters using complete observations and available observations based on the regression model of interest and parametric working models, respectively. Simulation studies will be provided to illustrate the proposed method.

L'analyse statistique de données longitudinales avec des données manquantes en motif non-monotone est très difficile. Dans cette recherche, nous expliquons une procédure uniforme pour traiter les données longitudinales avec données manquantes en motifs non-monotones et arbitraires, y compris le triage des données selon le motif des valeurs manquantes et l'estimation des paramètres de régression par des observations complètes et disponibles basées sur le modèle de régression d'intérêt et sur les modèles de travail paramétriques, respectivement. Plusieurs études de simulation seront fournies pour illustrer la méthode proposée.

## 1B-14: Statistics in Sports Statistique dans le domaine des sports

---

Chair/Président: Reg Kulperger (University of Western Ontario)  
Organizer/Responsable: Paramjit Gill (University of British Columbia Okanagan)

**Room/Salle: BAC237**

### Abstracts/Résumés

---

**BILL HURLEY** (Royal Military College of Canada), **ANDREY PAVLOV** (Queen's University)

[Monday June 13 / lundi 13 juin, 10:15]

*A Valuation Model for NFL and NHL Draft Positions*

*Modèle d'évaluation des rangs au repêchage de la LNF et de la LNH*

This paper describes an approach to the valuation of draft positions for the annual NFL and NHL entry drafts. We then use this valuation to evaluate and construct exchanges of draft positions.

Cet article décrit une approche d'évaluation des rangs au repêchage annuel dans la LNF et la LNH. Nous utilisons ensuite cette évaluation pour analyser et créer des échanges de rangs.

---

**TIM SWARTZ** (Simon Fraser University)

[Monday June 13 / lundi 13 juin, 10:45]

*Desperation in Sport*

*Du désespoir chez les sportifs*

One of the common assumptions in sport is that repeated matches between two teams are characterized as independent and identical trials. This study explores the psychological impact on teams when being ahead and when falling behind in playoff series. A novel aspect of this study is that the data consists of point spreads between teams rather than the more variable match outcomes. Regression models are considered for playoff series in both the National Basketball Association and the National Hockey League. We attempt to quantify the effect of desperation in both of these leagues.

En sport, les matchs successifs entre deux équipes sont souvent considérés comme des essais indépendants et identiques. Cette étude explore l'impact psychologique subi par les équipes prenant l'avance ou tirant de l'arrière lors de séries éliminatoires. Une innovation dans cette étude consiste à utiliser les écarts de points entre les équipes, des données moins variables que les résultats des matchs. Nous explorons des modèles de régression pour les séries éliminatoires de la NBA et de la Ligue nationale de hockey. Nous tentons de quantifier l'effet du désespoir dans ces deux ligues.

---

**PARAMJIT GILL** (University of British Columbia, Okanagan), **MICHAEL DAVIS** (UBC Okanagan)

[Monday June 13 / lundi 13 juin, 11:15]

*The Third Period in an NHL Game*

*La troisième période d'un match de la LNH*

With the adoption of the shootout rule in the NHL, ties are eliminated from the standings. A team is awarded two points for a win, zero for a regulation loss and one point for a game lost in overtime/shootout. So, having a game continue into

Avec l'adoption du règlement des tirs de barrage dans la LNH, les matchs nuls sont éliminés du classement. Une équipe reçoit deux points pour une victoire, zéro point pour une défaite en temps réglementaire et un point pour un match perdu en prolongation ou en tirs de barrage. Les pro-

## 1B-I4: Statistics in Sports Statistique dans le domaine des sports

---

overtime is no longer zero-sum. If game is tied during the third period, does it affect the scoring intensity to keep the game tied? We examine the scoring intensity under different game situations during the course of game. A time-truncated exponential model is used to estimate the goal intensity for rolling 5-minute intervals.

longations ne sont donc plus des jeux à somme nulle. Si un match est à égalité en troisième période, l'intensité des tirs au but s'en trouve-t-elle affectée afin que le match reste à égalité ? Nous examinons l'intensité des tirs au but dans diverses situations de jeu pendant tout un match. Un modèle exponentiel tronqué dans le temps est utilisé pour estimer l'intensité des buts marqués sur des intervalles mobiles de cinq minutes.

## 1B-15: Stochastic Analysis Analyse stochastique

---

Chair/Président: Rafal Kulik (University of Ottawa)  
Organizer/Responsable: Raluca Balan (University of Ottawa)

Session sponsored by the Probability Section  
Session parrainée par le Groupe de probabilité

**Room/Salle: BAC138**

### Abstracts/Résumés

---

**SHUI FENG** (McMaster University)

[Monday June 13 / lundi 13 juin, 10:15]

*Gamma-Dirichlet Structure*  
*Structure de Gamma-Dirichlet*

This talk focuses on the connections between the gamma process and the Dirichlet process. These include Lukacs's characterization of the gamma distribution, the formal Hamiltonians, quasi-invariance, stochastic dynamics and large deviations.

Cette présentation met l'emphase sur les liens entre les processus de gamma et de Dirichlet. Parmi ceux-ci, il y a la caractérisation de Lukacs de la distribution gamma, les hamiltoniens formels, la quasi-invariance, les dynamiques stochastiques et les grandes déviations.

---

**CODY HYNDMAN** (Concordia University)

[Monday June 13 / lundi 13 juin, 10:45]

*Existence and Uniqueness for Nonlinear FBSDEs Related to Quadratic Term-Structure Models.*  
*Existence et unicité des équations différentielles stochastiques forward-backward (FBSDE) non linéaires liées aux modèles avec structure quadratique*

We consider existence and uniqueness results for some nonlinear forward-backward stochastic differential equations (FBSDEs) related to quadratic term-structure models of interest rates. The forward component of the FBSDE is a Gaussian diffusion or an affine diffusion. We also investigate the case of FBSDEs related to affine term-structure models where the forward component of the FBSDE is a Wishart process.

Nous considérons les résultats d'existence et d'unicité pour des équations différentielles stochastiques forward-backward (FBSDE) non linéaires liées à des modèles quadratiques des taux d'intérêt. La composante forward des FBSDE est une diffusion gaussienne ou affine. Nous examinons également le cas de FBSDE liées à des modèles affines où la composante forward de la FBSDE est un processus de Wishart.

---

**WEI SUN** (Concordia University), **ZECHUN HU** (Nanjing University)

[Monday June 13 / lundi 13 juin, 11:15]

*Hunt's Hypothesis (H) and Gettoor's Conjecture for Lévy Processes*  
*Hypothèse de Hunt et conjecture de Gettoor pour les processus de Lévy*

Let  $X$  be a nice Markov process. Hunt's hypothesis (H) says that "every semipolar set  $A$  of  $X$  is polar". Intuitively speaking, (H) means that if  $A$  cannot be

Soit  $X$  un processus de Markov possédant de belles propriétés. L'hypothèse de Hunt précise que « chaque ensemble semi-polaire  $A$  de  $X$  est polaire ». De façon intuitive, l'hypo-

## 1B-I5: Stochastic Analysis Analyse stochastique

---

immediately hit by  $X$  with arbitrary starting point, then  $A$  will not be hit by  $X$  forever. (H) plays a crucial role in the potential theory of (dual) Markov processes. Some forty years ago, Gettoor conjectured that essentially all Lévy processes satisfy (H). In this talk, we will introduce our recent progress on Gettoor's conjecture.

thèse de Hunt signifie que si  $A$  ne peut pas être visité immédiatement par  $X$  avec un point de départ arbitraire, alors  $A$  ne sera jamais visité par  $X$ . Cette hypothèse joue un rôle fondamental dans la théorie du potentiel des processus (duals) de Markov. Il y a environ quarante ans, Gettoor a avancé la conjecture que tous les processus de Lévy satisfont l'hypothèse de Hunt. Dans cet exposé, nous présentons nos progrès récents au sujet de la conjecture de Gettoor.

## 1B-C1: Methods for Inference Méthodes pour l'inférence

---

Chair/Président: JC Loredano-Osti (Memorial University of Newfoundland)

Room/Salle: BAC141

Abstracts/Résumés

---

**MICHAEL STEPHENS & RICHARD LOCKHART** (Simon Fraser University), **ALBERTO CONTRERAS** (IIMAS – UNAM)

[Monday June 13 / lundi 13 juin, 10:15]

*Use of Priors in Goodness-of-Fit*

*Utilisation des lois a priori pour la qualité de l'ajustement*

The Dirichlet process is used to generate random alternatives to a discrete distribution or a parametric family of discrete distributions. Statistics which maximize the expected power are derived and evaluated. The idea is applied to the Poisson and discrete uniform distributions. For the Poisson distribution, conditional tests are implemented.

Le processus de Dirichlet est utilisé pour générer des alternatives aléatoires à une distribution discrète ou une famille paramétrique de distributions discrètes. Des statistiques qui maximisent la puissance espérée sont dérivées et évaluées. L'idée est appliquée aux distributions de Poisson et uniformes discrètes. Pour la distribution de Poisson, des tests conditionnels sont mis en place.

---

**ZEYNEP BASKURT & MICHAEL EVANS** (University of Toronto)

[Monday June 13 / lundi 13 juin, 10:30]

*Hypothesis Assessment via the Method of Concentration*

*Évaluation d'hypothèses via la méthode de concentration*

Suppose we assess a hypothesis  $H_0 \subset \Theta$  where  $H_0 = \Psi^{-1}(\psi_0)$  for some  $\Psi(\theta) = \psi$  by calculating a relative belief ratio or Bayes factor and calibrate this by a posterior P-value. When  $\Pi(H_0) = 0$  ( $\Pi$  is a prior on  $\Theta$ ), a common approach to obtain a Bayes factor is to introduce a prior probability for  $H_0$  together with a conditional prior  $\Pi_0$  on  $H_0$  and use the mixture prior. I discuss an inconsistency about the prior assignments on  $H_0$  and an approach to avoiding it by choosing a  $\Psi$  function from a class of constant volume distortion generators of  $H_0$ .

Supposons que nous évaluons une hypothèse  $H_0 \subset \Theta$  où  $H_0 = \Psi^{-1}(\psi_0)$  pour  $\Psi(\theta) = \psi$  en calculant un ratio de confiance relative ou le facteur de Bayes calibrés à l'aide d'un seuil observé a posteriori. Lorsque  $\Pi(H_0) = 0$  ( $\Pi$  est la loi a priori de  $\Theta$ ), une approche assez commune pour obtenir un facteur de Bayes consiste à introduire une probabilité a priori pour  $H_0$  ainsi qu'une loi a priori conditionnelle  $\Pi_0$  sur  $H_0$  et d'utiliser un mélange a priori. Je discute d'une incohérence des lois a priori sur  $H_0$  et d'une façon de l'éviter en choisissant une fonction  $\Psi$  parmi une classe de générateurs de distorsion à volume constant de  $H_0$ .

---

**OSAMA BATAINEH & MIKELIS BICKIS** (University of Saskatchewan)

[Monday June 13 / lundi 13 juin, 10:45]

*Imprecise Predictive Inference for Logistic Regression*

*Inférence prédictive imprécise pour la régression logistique*

Walley (1991) proposed an inferential paradigm that applies Bayes theorem to a family of prior distributions, yielding interval posterior probabilities.

Walley (1991) propose un paradigme d'inférence qui applique le théorème de Bayes à une famille de distributions a priori, produisant des probabilités a posteriori par inter-



## 1B-C1: Methods for Inference Méthodes pour l'inférence

---

This approach is applied to logistic regression by assuming a Dirichlet prior on the increments of response probabilities at selected values of the explanatory variable, thereby providing for prior dependence. Generalizing Walley's (1996) imprecise Dirichlet model to this situation gives an interval of estimates corresponding to a family of priors. The so-called imprecise estimates of model parameters and predictions are computed using Metropolis sampling. The sensitivity of these estimates to prior correlation and design effects is studied by simulation.

valles. Nous appliquons cette approche à la régression logistique en supposant une loi a priori de Dirichlet sur les incréments des probabilités de réponse pour des valeurs sélectionnées de la variable explicative, tenant ainsi compte de dépendance des lois a priori. En généralisant le modèle de Dirichlet imprécis de Walley (1996) à cette situation, nous obtenons un intervalle d'estimés correspondant à une famille de lois a priori. Nous calculons les estimés des paramètres dits imprécis et les prévisions en utilisant l'algorithme de Metropolis. À l'aide de simulations, nous étudions la sensibilité de ces estimés à la corrélation des lois a priori et aux effets de plan.

---

**XIMING XU & NANCY REID** (University of Toronto)

[Monday June 13 / lundi 13 juin, 11:00]

*More or Less: Two Paradoxes on the Selection of Composite Likelihood*

*Plus ou moins : deux paradoxes sur la sélection d'une vraisemblance composite*

Composite likelihood methods have been receiving growing interest in different complex applications where the ordinary likelihood is too cumbersome to be evaluated. Intuitively a more efficient estimator can be obtained by adding more log likelihoods or using higher dimensional component likelihoods. In this paper we will show that such strategies do not always work. In addition, in most statistical models, the efficiency of an estimator of a parameter of interest usually increases when the nuisance parameters are known. However the opposite phenomenon may occur for composite likelihood inference, which will be illustrated through some examples.

Les méthodes de vraisemblance composite sont de plus en plus utilisées dans diverses applications complexes où il est trop difficile d'évaluer la vraisemblance ordinaire. Intuitivement, on devrait obtenir un estimateur plus efficace en ajoutant des log-vraisemblances ou en utilisant la vraisemblance de composantes de dimension supérieure. Dans cet article, nous montrerons que ces stratégies ne réussissent pas toujours. De plus, dans la plupart des modèles statistiques, l'efficacité d'un estimateur d'un paramètre d'intérêt augmente généralement lorsque les paramètres nuisibles sont connus. Toutefois, le phénomène contraire peut se produire pour l'inférence reposant sur la vraisemblance composite, ce que nous illustrerons par plusieurs exemples.

---

**THARSHANNA NADARAJAH & ASOKAN MULAYATH VARIYATH** (Memorial University of Newfoundland and Labrador)

[Monday June 13 / lundi 13 juin, 11:15]

*Penalized Empirical Likelihood Based Variable Selection*

*Sélection de variables basée sur la vraisemblance empirique pénalisée*

A new class of variable selection methods based on penalized likelihood is lauded for its computational efficiency and stability. Variable selection and estimation of coefficients are simultaneously carried out in this approach. A well defined parametric likelihood is a crucial part. To overcome this problem, we developed a penalized empirical likelihood variable selection procedure based on a set

Grâce à son efficacité et stabilité computationnelle, une nouvelle classe de méthode de sélection de variables basée sur la vraisemblance pénalisée est favorisée. Cette approche effectue simultanément la sélection de variables et l'estimation des coefficients. Il est crucial d'avoir une vraisemblance paramétrique bien définie. Pour surmonter cette difficulté, nous avons développé une procédure de sélection de variables par vraisemblance empirique pénalisée basée sur un

## 1B-C1: Methods for Inference Méthodes pour l'inférence

---

of estimating equations. The asymptotic properties of the proposed method are investigated. Our simulation studies show that the proposed method has comparable performance when a parametric model is plausible where as it outperforms when the parametric model is mis-specified.

jeu d'équations d'estimation. Les propriétés asymptotiques de la méthode proposée sont étudiées. Nos études par simulation montrent que la méthode proposée atteint une performance comparable lorsque le modèle paramétrique est plausible et une performance supérieure lorsque le modèle paramétrique est mal spécifié.

## 1B-C2: Spatial Modelling for Infectious Disease Modélisation spatiale des maladies infectieuses

---

Chair/Président: Karen Kopciuk (Alberta Health Services and University of Calgary)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 234**

**Abstracts/Résumés**

---

**DANIEL GILLIS, TONY DESMOND & JOHN HOLT** (University of Guelph)

[Monday June 13 / lundi 13 juin, 10:15]

*Disease Source Classification Using Multivariate Spatial Poisson Mixture Models*

*Classification de sources de maladies à l'aide de modèles de mélanges spatiaux multivariés de Poisson*

Enteric disease data obtained from the Canadian Institute for Health Information provides motivation to extend mixture literature to label disease based on infection source. Mixtures are used to classify data as foodborne or waterborne. Two spatial models are presented and compared to the standard CAR spatial model described by Besag et al., 1991. The models account for spatially indexed disease by applying independent or dependent conditionally autoregressive spatial priors in the log linear term of each of the mixture components. All models are compared via simulation, with application to Alberta Gastrointestinal disease data (1992-1998).

Des données sur les maladies entériques obtenues de l'Institut canadien d'information sur la santé apportent une motivation pour étendre la littérature sur les modèles de mélanges (MM) afin d'identifier les maladies selon la source d'infection. Les MM sont utilisés pour classer les données comme étant d'origine alimentaire ou hydrique. Nous présentons deux modèles spatiaux que nous comparons au modèle autorégressif conditionnel (ARC) spatial décrit par Besag et al. (1991). Les modèles tiennent compte de maladies spatialement indexées en appliquant des lois a priori autorégressives conditionnelles spatiales indépendantes ou dépendantes au terme log linéaire de chaque composante du mélange. Tous les modèles sont comparés au moyen de simulations, avec application à des données d'Alberta sur les maladies gastro-intestinales (1992-1998).

---

**YE LI** (University of Toronto), **PATRICK BROWN** (Cancer Care Ontario), **HAVARD RUN** (Norwegian University of Science and Technology)

[Monday June 13 / lundi 13 juin, 10:30]

*Geostatistical Model for Spatially Aggregated Disease Incidence Data*

*Modèles géostatistiques pour données d'incidence de maladies spatialement agrégées*

Clinical data on the locations of syphilis cases in North Carolina are modelled with the aim of finding areas of abnormally high risk. Inference is complicated by uncertainty of case locations due to the aggregation in the population reporting census regions. A model is introduced consisting of a continuous random spatial surface with aggregated responses and fixed covariate effects. The process is modelled on a fine lattice and Bayesian inference is performed using Markov Chain Monte

Carlo. Les données cliniques sur les emplacements de cas de syphilis en Caroline du Nord sont modélisées dans le but d'y trouver des régions à risque anormalement élevé. L'inférence est compliquée par l'incertitude des emplacements de cas causée par l'agrégation des données aux régions de recensement. Un modèle de surface spatiale aléatoire continue avec des réponses agrégées et des effets de covariables fixes est présenté. Le processus est modélisé sur un treillis fin et l'inférence bayésienne est effectuée en utilisant une chaîne de Markov Monte-Carlo avec augmentation de

## 1B-C2: Spatial Modelling for Infectious Disease Modélisation spatiale des maladies infectieuses

---

Carlo with data augmentation. Simulations studies are carried out to compare with the Besag-York-Mollié model and a model assuming the exact locations are known.

données. Des études par simulation sont utilisées pour comparer avec le modèle de Besag-York-Mollié et un modèle qui suppose que l'exactitude des emplacements soit connu.

---

**GRACE PUI SZE KWONG & ROB DEARDON** (University of Guelph)

[Monday June 13 / lundi 13 juin, 10:45]

*Computationally Efficient Forms of Spatial Infectious Disease Models*

*Forme computationnellement efficace pour modèles spatiaux de maladies infectieuses*

Individual-level models (ILMs) for infectious diseases, fitted in a Bayesian MCMC framework, are an intuitive and flexible class of models that can take into account population heterogeneity via various individual-level covariates. ILMs containing a geometric distance kernel to account for geographic heterogeneity provide a natural way to model the spatial spread of many diseases. However, in even only moderately large populations, the likelihood calculations required can be prohibitively time consuming. It is possible to speed up the computation via a technique which makes use linearized distance kernels. We examine some methods of carrying out this approximation and compare their performances.

Les modèles au niveau de l'individu (MNI) pour maladies infectieuses, calculés dans un cadre bayésien de chaînes de Markov Monte-Carlo, sont une classe de modèles intuitive et flexible dont l'hétérogénéité de la population est tenue compte par divers covariables au niveau de l'individu. Les MNI ayant une distance géométrique à noyaux pour tenir compte de l'hétérogénéité géographique modélisent de façon naturelle la propagation spatiale de plusieurs maladies. Cependant, même dans le cas de populations à taille modérée, les calculs de la vraisemblance peuvent demander beaucoup de temps. Il est possible d'accélérer les calculs au moyen d'une technique qui utilise la distance linéarisée à noyaux. Nous examinons quelques méthodes utilisant cette approximation et comparons leurs performances.

---

**JOURDAN GOLD, ROB DEARDON & ZENY FENG** (University of Guelph)

[Monday June 13 / lundi 13 juin, 11:00]

*Effects of Time-line Uncertainty in Infectious Disease Modelling*

*Effets de l'incertitude sur les données temporelles dans la modélisation de maladies infectieuses*

Because of the complexity of infectious disease systems, one often makes simplifying assumptions in their modeling process. These assumptions, while computationally convenient, could lead to a poorly fitted model. We can account for data uncertainty explicitly but this may cause computational problems. A simulation study was performed to ascertain the effects of ignoring timeline uncertainty. Results will be presented that quantify the trade-off between model inferential quality and computational-time, using a family of discrete-time heterogeneous infectious disease models known as individual-level models. Modeling approaches will vary from those under 'fixed data' assumptions to those under a 'full data augmentation approach'.

En raison de la complexité des systèmes de maladies infectieuses, des hypothèses simplificatrices sont souvent utilisées dans les processus de modélisation. Bien que pratiques d'un point de vue calculatoire, ces hypothèses peuvent mener à un modèle mal ajusté. Il est possible de tenir compte de l'incertitude des données de façon explicite, mais en risquant de causer des problèmes calculatoires. Nous avons effectué une étude de simulation pour évaluer les effets lorsque l'incertitude sur les données temporelles est ignorée. Les résultats présentés quantifient le compromis entre la qualité de l'inférence du modèle et le temps de calcul à l'aide d'une famille de modèles de maladies infectieuses hétérogènes à temps discret connue sous le nom de modèles au niveau de l'individu. Les approches de modélisation proposées varient, allant des hypothèses « à données fixes » à celles d'une « approche d'augmentation

## 1B-C2: Spatial Modelling for Infectious Disease Modélisation spatiale des maladies infectieuses

---

de données ».

---

**LORNA DEETH & ROB DEARDON** (University of Guelph)

[Monday June 13 / lundi 13 juin, 11:15]

*Mixtures of Individual-level Models for Infectious Disease Modelling*

*Mélanges de modèles au niveau de l'individu pour la modélisation de maladies infectieuses*

Individual-level models (ILMs) are a class of models that have been used to describe the spatiotemporal spread of infectious diseases. However, current ILMs do not account for (hidden) population heterogeneity, and instead assume a homogeneous population. A mixture ILM (MILM) is proposed that is an adjustment to a simple, spatial ILM, and allows for population heterogeneity with respect to certain model parameters. A simulation study was conducted in which both the simple ILM and MILM were fit to simulated epidemic data, and the respective posterior predictive abilities were assessed.

Les modèles au niveau de l'individu (MNI) sont une classe de modèles utilisés pour décrire la propagation spatiotemporelle de maladies infectieuses. Cependant, les MNI actuels ne tiennent pas compte d'une hétérogénéité (cachée) dans la population et supposent une population homogène. Nous proposons un mélange de MNI (MMNI) qui constitue l'ajustement d'un MNI spatial simple et qui permet une hétérogénéité de la population par rapport à certains paramètres du modèle. Nous avons procédé à une étude de simulation dans laquelle le MNI simple et le MMNI ont été ajustés pour simuler des données épidémiques, et nous avons étudié les capacités prédictives a posteriori respectives.

---

**ROB DEARDON, ABBIE GARDNER & GERARDA DARLINGTON** (University of Guelph)

[Monday June 13 / lundi 13 juin, 11:30]

*Posterior Predictive Approach to Goodness-of-Fit for Infectious Disease Models*

*Approche prédictive a posteriori pour l'adéquation de modèles de maladies infectieuses*

In complex non-linear models, such as spatio-temporal infectious disease models, it is often unclear how best to ascertain goodness-of-fit. Often such models are fitted within a Bayesian statistical framework, since such a framework is ideally placed to account for the many areas of data uncertainty. Within a Bayesian context, a major tool in assessing goodness-of-fit is the posterior predictive distribution. Here, we examine different test statistics and ascertain how well they can detect model misspecification via a simulation study.

Dans les modèles non linéaires complexes tels que les modèles de maladies infectieuses spatio-temporels, il n'est pas toujours évident de déterminer l'adéquation. De tels modèles sont souvent ajustés dans un cadre statistique bayésien, puisqu'un tel contexte est idéal pour tenir compte des diverses formes d'incertitude dans les données. Dans un contexte bayésien, la distribution prédictive a posteriori constitue un outil remarquable pour évaluer l'adéquation. Ici, nous examinons diverses variables de test et déterminons, dans quelle mesure elles permettent de détecter les erreurs de spécification du modèle à l'aide d'une étude de simulation.

## 1B-C3: Survey Methods: Data Collection, Edit and Imputation Méthodes d'enquête : collecte de données, vérification et imputation

---

Chair/Président: Carlos León (Statistique Canada)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

**Room/Salle: BAC142**

**Abstracts/Résumés**

---

**QIONG WANG & DANIELLE LAROCHE** (Statistics Canada)

[Monday June 13 / lundi 13 juin, 10:15]

*Identification of the Collection Units which will Receive the Letter Motivating the Response Via Internet for the 2011 Census*

*Identification des secteurs qui recevront une lettre destinée à promouvoir la réponse par Internet au Recensement de 2011*

To motivate the response via the Internet for the 2011 Census, it has been decided to send letters instead of the traditional paper questionnaire to 60 percent of Canadian dwellings. It is very important to target the enumeration areas in which households are more likely to respond via the Internet, while at the same time avoiding the households less likely to respond to the Census. This document describes a model developed to identify these enumeration areas and explains the method used to integrate the 2011 geography into the 2006 Census data. It also presents the key results.

Au Recensement de la population de 2011, il a été décidé de motiver la réponse par Internet par l'envoi de lettres à 60 pourcent des logements au lieu de l'envoi traditionnel du questionnaire en papier. Il est très important de bien cibler les secteurs où les ménages sont les plus susceptibles de répondre par Internet, tout en évitant les ménages les moins enclins à répondre au recensement. Ce document décrit un modèle proposé pour identifier ces secteurs. Il présente également la méthode qui intègre la géographie de 2011 aux données du Recensement de 2006 et en expose les principaux résultats.

---

**KIMBERLEY COLEMAN & JEAN-SÉBASTIEN PROVENÇAL** (Statistics Canada)

[Monday June 13 / lundi 13 juin, 10:30]

*A Comparison of Dependent and Independent Interviewing for Collecting Revised Data in the Fruit and Vegetable Survey*

*Une comparaison de données révisées provenant d'entrevues dépendantes et indépendantes pour l'enquête sur les fruits et légumes*

The fall occasion of Statistics Canada's Fruit and Vegetable Survey gathers information regarding crops grown in the current year, including acreage, quantity sold and value received. For the fall 2009 survey, a section was added to collect revised data from the previous year. For 2010, a split sample was used to assess the impact of dependent interviewing (where the respondent is provided last year's response in the question) versus independent interviewing. This paper describes the mo-

L'enquête d'automne sur les fruits et légumes de Statistique Canada recueille des données sur la superficie, les quantités produites et les ventes de différentes cultures pour l'année courante. Pour l'année 2009, une section a été ajoutée afin de recueillir des données révisées pour l'année précédente. En 2010, un échantillon fractionné a été sélectionné afin d'évaluer l'impact d'utiliser des entrevues dépendantes (où les questions réfèrent aux réponses rapportées l'année précédente) versus indépendantes. Cet article présente les motifs entourant l'ajout de cette section, les avantages et

## 1B-C3: Survey Methods: Data Collection, Edit and Imputation Méthodes d'enquête : collecte de données, vérification et imputation

---

tivation for the additional section along with the benefits and challenges associated with each interviewing method, and presents results from a series of evaluations.

défis associés aux deux formes d'entrevues, ainsi que des résultats liés à l'évaluation de leur performance.

---

**NAIMA GOUZI & ABDELLATIF DEMNATI** (Statistique Canada)

[Monday June 13 / lundi 13 juin, 10:45]

*Sample Size Determination for the Integrated Business Statistics Program*

*Détermination de la taille de l'échantillon dans le cadre du Programme intégré de la statistique des entreprises*

Most business surveys are multi-objective. In order to reach these objectives, we aggregate the population under each objective. So, one business can belong to more than one group. We assume that each group has a frame and we use the multiple frames theory. One of the problems met with this approach is the sample size determination. We explore another approach based on a single frame. We evaluate the efficiency with a sample size determination which minimizes the cost under constraints over the variance of (calibration) total estimation of each objective.

La majorité des enquêtes-entreprises ont des objectifs multiples. Pour atteindre ces objectifs, nous regroupons la population selon chaque objectif. Une entreprise peut donc appartenir à plusieurs groupes. Nous considérons que chaque groupe est une base de sondage et nous utilisons la théorie des bases multiples. L'une des difficultés éprouvées avec cette approche est la détermination de la taille de l'échantillon. Nous explorons une alternative centrée sur une base unique. Nous évaluons son efficacité en déterminant la taille de l'échantillon qui minimise le coût sujet à des contraintes sur la variance de l'estimation (calibrée) du total de chaque objectif.

---

**BRAD JOHNSON & MOHAMMAD JAFARI JOZANI** (University of Manitoba)

[Monday June 13 / lundi 13 juin, 11:00]

*Randomized Minima-Maxima Nomination Sampling in Finite Populations*

*Échantillonnage nominatif randomisé des minima-maxima pour des populations finies*

We propose a randomized minima-maxima nomination sampling design for the use in finite populations which generalizes maxima nomination sampling. The design is simple to implement and results in easy to calculate inclusion probabilities for population elements, estimators and variance formulae. We show, through numerical studies, that the proposed design can result in significant improvements in efficiency compared to simple random sampling without replacement designs for a wide choice of population types.

Nous proposons un plan de sondage nominatif randomisé des minima-maxima pour les populations finies qui généralise l'échantillonnage nominatif des maxima. Ce plan est facile à mettre en oeuvre et permet de calculer facilement les probabilités d'inclusion des éléments de la population, des estimateurs et des formules de variance. Nous montrons à l'aide d'exemples numériques que le plan proposé peut entraîner des améliorations importantes de l'efficacité par rapport à des plans de sondage simples sans remise pour un vaste choix de types de population.

---

**CHRISTIAN OLIVIER NAMBEU** (Statistique Canada), **DAVID HAZIZA** (Université de Montréal), **GUILLAUME CHAUVET** (Laboratoire de Statistique d'Enquête, CREST/ENSAI)

[Monday June 13 / lundi 13 juin, 11:15]

*Single Imputation for Populations Containing a Large Amount of Zeroes*

*Imputation simple de populations contenant un nombre élevé de zéros*

## 1B-C3: Survey Methods: Data Collection, Edit and Imputation Méthodes d'enquête : collecte de données, vérification et imputation

---

Simple imputation is often used in surveys to compensate for item nonresponse. In some cases, the variable requiring imputation contains a large amount of zeroes. In this paper, we study the properties of two imputation procedures frequently used in practice and show that they generally lead to biased estimators. Motivated by a mixture regression model, we then propose three imputation procedures and we study their properties in terms of bias and we consider a jackknife variance estimator consistent for the true variance, provided the overall sampling fraction is negligible. A simulation study illustrates the performance of point and variance estimators.

L'imputation simple est souvent utilisée dans les enquêtes pour compenser la non-réponse partielle. La variable devant être imputée contient parfois un nombre élevé de zéros. Dans cet article, nous étudions les propriétés de deux procédures courantes d'imputation et montrons qu'elles produisent des estimateurs biaisés. S'appuyant sur un modèle de régression mixte, nous proposons ensuite trois procédures d'imputation et nous étudions leurs propriétés en terme de biais et nous considérons un estimateur jackknife de la variance cohérent avec la vraie variance, en supposant que la fraction de sondage globale est négligeable. Une étude par simulation illustre la performance des estimateurs ponctuels et de la variance.

---

**CARLOS LEÓN** (Statistique Canada)

[Monday June 13 / lundi 13 juin, 11:30]

*Optimally controlled sampling designs to guard against bad samples*

*Plans d'échantillonnage contrôlés pour éliminer les mauvais échantillons*

We propose a method for constructing sampling plans weakly equivalent to simple random sampling without replacement that achieve the exclusion of certain bad samples. We show how these plans can be obtained as solutions to a constrained optimization problem and we present some numerical examples.

Nous proposons une méthode pour construire des plans d'échantillonnage faiblement équivalents à l'échantillonnage aléatoire simple sans remise mais qui arrivent à exclure certains échantillons jugés indésirables. Nous expliquons comment ces plans s'obtiennent comme solutions d'un problème d'optimisation sous contraintes et nous donnons quelques exemples numériques.



## 1C-11: Advances in Fisherian Alternatives to Conventional Inference Progrès dans les alternatives fisheriennes à l'inférence conventionnelle

---

Organizer and Chair / Responsable et président: David Bickel (University of Ottawa)

Room/Salle: BAC236

Abstracts/Résumés

---

**AO YUAN** (Howard University)

[Monday June 13 / lundi 13 juin, 13:00]

*Bayesian Frequentist Hybrid Model with Applications to CGH Data Analysis*

*Modèle hybride fréquentiste bayésien avec applications à l'analyse de données HGC*

Comparative genomic hybridization (CGH) is a common technique to screen the GCN changes in mutant cells genome-wide. Existing statistical methods for analyzing such data are either frequentist, or full Bayesian. The former has advantage of objectivity, while the latter can incorporate useful prior information. In an attempt to take full advantages of both approaches, we develop a Bayesian-frequentist hybrid procedure, in which part of the parameters are inferred in a Bayesian fashion, while the remaining parameters are inferred in frequentist way. This is especially useful when sound prior information is available only on part of the parameters, and the sample size is relatively small.

L'hybridation génomique comparative (HGC) est une technique communément utilisée pour dépister des changements au nombre de copies d'un gène dans les cellules mutées, à l'échelle du génome. Les méthodes statistiques existantes permettant l'analyse de ces données sont soit fréquentistes, soit entièrement bayésiennes. Alors que les premières ont l'avantage d'être objectives, les secondes peuvent inclure des informations a priori utiles. Pour tenter d'exploiter les avantages des deux approches, nous avons mis au point une procédure hybride bayésienne – fréquentiste dans laquelle une partie des paramètres est inférée par une approche bayésienne, et les autres de façon fréquentiste. Cette procédure s'avère particulièrement utile lorsque de solides informations a priori ne sont disponibles que pour certains paramètres et que la taille de l'échantillon est relativement limitée.

---

**ALAN POLANSKY** (Northern Illinois University)

[Monday June 13 / lundi 13 juin, 13:30]

*Observed Confidence Levels: An Alternative to Multiple Testing Techniques*

*Niveaux de confiance observés : une alternative aux techniques d'épreuves multiples*

Many applications in statistics do not fall within the framework of traditional hypothesis testing. These problems typically require a sequence of hypothesis tests. The sequence and choice of hypotheses in such an application is important, as different choices can often lead to different overall conclusions. Observed confidence levels provide a new approach to these problems by computing a simultaneous measure of the relative truth of each hypothesis. This approach does not require the specification of a sequence of hypothesis tests. A general theory for the application of this method is developed and the method is demonstrated in several examples.

En statistique, de nombreuses applications dépassent le cadre du test d'hypothèses classique. Ces problèmes exigent généralement une séquence de tests d'hypothèses. La séquence et le choix des hypothèses sont alors importants, car des choix différents peuvent mener à des conclusions globales différentes. Les niveaux de confiance observés offrent une nouvelle approche à ces problèmes en permettant le calcul simultané d'une mesure de véracité relative de chaque hypothèse. Cette approche évite de spécifier une séquence de tests d'hypothèses. Nous développons une théorie générale pour l'application de cette méthode et nous l'illustrons par plusieurs exemples.

## 1C-I1: Advances in Fisherian Alternatives to Conventional Inference Progrès dans les alternatives fisheriennes à l'inférence conventionnelle

---

**JEFFREY BLUME** (Vanderbilt University School of Medicine)

[Monday June 13 / lundi 13 juin, 14:00]

*Fisher's Likelihood Inference: Coming of Age*

*Inférence par la vraisemblance de Fisher : l'ère de la maturité*

Likelihood methods for measuring statistical evidence have evolved slowly since R.A. Fisher first introduced them in the early 1920's. Nearly a century later the likelihood paradigm has matured enough to warrant careful consideration. Likelihood methods are the natural compromise between Bayesian and frequentist approaches; they retain the desirable properties of both paradigms (irrelevance of sample space, good performance probabilities) while shedding undesirable ones (dependence on prior distributions, ad-hoc adjustments to control error probabilities). In this talk, I will introduce the modern likelihood paradigm, show how this evidential framework resolves multiplicity paradoxes, and discuss recent advances.

Les méthodes fondées sur la vraisemblance ont évolué lentement depuis leur introduction par R.A. Fisher au début des années 1920. Près d'un siècle plus tard, le paradigme de la vraisemblance a assez mûri pour être sérieusement étudié. Les méthodes fondées sur la vraisemblance constituent un compromis naturel entre les approches bayésienne et fréquentiste : elles conservent les propriétés souhaitables des deux paradigmes (affranchissement de l'ensemble fondamental, bonnes probabilités de performance) tout en éliminant certaines propriétés moins souhaitables (dépendance vis-à-vis des distributions a priori, ajustements ad hoc pour contrôler les probabilités d'erreur). Dans cette présentation, j'introduirai le paradigme moderne de la vraisemblance, je montrerai comment ce cadre résout les paradoxes de la multiplicité, et je discuterai des récents progrès dans le domaine.

## 1C-I2: Methods for Human and Animal Surveillance Méthodes de surveillance des humains et des animaux

---

Chair/Président: Henrik Stryhn (Atlantic Veterinary College, PEI)  
Organizers/Responsables: Lisa Lix (University of Saskatchewan) and/et Henrik Stryhn (Atlantic Veterinary College, PEI)

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

**Room/Salle: BAC 132**

**Abstracts/Résumés**

---

**BERNARD CHOI** (Public Health Agency of Canada), **MANISHA KULKARNI WOODSTOCK**, **CHRISTINE KENNEDY**, **MAXIM TRUBNIKOV**, **MICHAEL COULTHART**, **SUSANNA OGUNNAIKE-COOKE** & **PURNI NARAYANAN**

[Monday June 13 / lundi 13 juin, 13:00]

*Informing Human and Animal Health Surveillance: A Catalogue of Epidemiologic and Statistical Biases Revisited*  
*À l'appui de la surveillance de la santé humaine et animale : réexamen d'un catalogue de biais épidémiologiques et statistiques*

The One Health model seeks to promote human, animal and ecosystem health through multidisciplinary collaboration. Methods for human and animal health surveillance can inform each other, leading to integrated approaches. This paper describes biases in human and animal health surveillance, building on a catalogue of 109 biases previously published in Encyclopedia of Biostatistics. Seven categories of biases are presented: literature review; study design; study execution; data collection; analysis; interpretation of results; and publication. The updated catalogue of biases serves as a resource for human and animal health practitioners, to help identify potential problems when collecting, analyzing and interpreting surveillance data.

Le modèle « One Health » cherche à promouvoir la santé des humains, des animaux et des écosystèmes via une collaboration pluridisciplinaire. Les méthodes de surveillance de la santé humaine et animale peuvent s'alimenter mutuellement, donnant lieu à des approches intégrées. Cet article décrit les biais en matière de surveillance de la santé humaine et animale, sur la base d'un catalogue de 109 biais publiés dans la « Encyclopedia of Biostatistics ». Nous présentons sept catégories de biais : l'analyse documentaire, la planification d'études, l'exécution d'études, la collecte de données, l'analyse, l'interprétation des résultats, et la publication. Ce catalogue des biais mis à jour servira de ressource aux praticiens de la santé humaine et animale et les aidera à identifier des problèmes potentiels lors de la collecte, de l'analyse et de l'interprétation des données de surveillance.

---

**DIRK PFEIFFER** (Royal Veterinary College, University of London, UK)

[Monday June 13 / lundi 13 juin, 13:30]

*Risk Assessment Approaches and their Increasing Importance in the Context of Risk-Based Disease Surveillance Design*

*Approches de l'évaluation des risques et leur importance grandissante dans le contexte d'une surveillance épidémiologique conçue en fonction du risque*

Animal health surveillance is aimed at providing information about the status of wild and domestic animal populations in relation to zoonotic and animal disease risks as well as animal welfare. The re-

La surveillance de la santé animale vise à fournir des informations sur l'état des populations d'animaux sauvages et domestiques, notamment en ce qui concerne les risques de zoonoses et de maladies animales, ainsi que le bien-

## 1C-I2: Methods for Human and Animal Surveillance Méthodes de surveillance des humains et des animaux

---

sulting information informs risk assessments which in turn will lead to targeted management of these risks which may then include tailored risk-based surveillance activities. Risk assessment models are used to determine the importance of different risk pathways for the occurrence of disease risks, based on the best scientific evidence in relation to the underlying mechanisms. This can be achieved most effectively by combining biological knowledge with quantitative methods.

être des animaux. Les informations qui résultent de la surveillance servent lors de l'évaluation des risques et la gestion ciblée de ces derniers, voire des activités de surveillance adaptées selon les risques. On utilise des modèles d'évaluation des risques pour déterminer l'importance de différentes voies à risque pour l'occurrence de certains risques de maladie, sur la base des meilleures observations scientifiques par rapport aux mécanismes sous-jacents. On peut y arriver de façon efficace en combinant la connaissance biologique et les méthodes quantitatives.

---

**CRAWFORD REVIE, JAVIER SANCHEZ & FERNANDA DOREA** (Atlantic Veterinary College, University of Prince Edward Island)

[Monday June 13 / lundi 13 juin, 14:00]

*Syndromic Surveillance: Can Large Data Sets Really Be "Unreasonably Effective"?*

*Surveillance syndromique : les grands ensembles de données peuvent-ils vraiment être « excessivement efficaces » ?*

Much of the literature around the growing field of systems for syndromic surveillance focuses on the capture and processing of large volumes of health-related data from a range of sources in near real-time. Quantity of data often appears to be more important than quality or analytical method. In some settings it seems this may indeed be the case, but how can the user of such systems be sure? This talk will explore tools and techniques from the fields of machine learning, statistics and data science that will play an increasingly important role in the development of effective syndromic surveillance.

Dans le domaine en pleine expansion des systèmes de surveillance syndromique, la littérature met largement l'accent sur la collecte et le traitement de gros volumes de données de santé, tirées de sources multiples, en temps presque réel. La quantité de données semble souvent l'emporter sur la qualité ou la méthode d'analyse. Dans certaines situations, c'est peut-être le cas, mais comment l'utilisateur de tels systèmes peut-il s'en assurer ? Cette présentation explorera des outils et techniques tirés de l'apprentissage machine, de la statistique et de la science de données qui joueront un rôle croissant dans le développement d'une surveillance syndromique efficace.

## 1C-I3: Nonparametric Function Estimation Estimation des fonctions non paramétriques

---

Organizer and Chair / Responsable et président: Alexandre Leblanc (University of Manitoba)

**Room/Salle: BAC237**

**Abstracts/Résumés**

---

**JEAN-FRANCOIS ANGERS** (Université de Montreal), **BRENDA MACGIBBON** (UQAM)

[Monday June 13 / lundi 13 juin, 13:00]

*Function Estimation with Non-negative "wavelets"*

*Estimation de fonction utilisant des "ondellettes" non négatives*

Wavelets have been successfully used for non-parametric function estimation, but a major challenge in density and hazard estimation is that the function must be nonnegative. As there does not exist a nonnegative scaling function for an orthonormal wavelet basis, we develop a quasi non-negative wavelet basis from Daubechies wavelets of order 4 with good approximation properties and use this for hazard function estimation from randomly right censored data. The proposed method is compared with other wavelet methods for hazard estimation and is also illustrated on a real medical data set methods.

Les ondellettes sont utilisées avec succès dans l'estimation non paramétrique de fonction. Cependant, un problème majeur en estimation de densité ou de fonction de hasard est que l'estimateur doit être non négatif. Comme il n'existe pas de fonction d'échelle non négative pour une base d'ondellettes orthonormale, nous développons une base de quasi-ondellettes non négatives à partir des ondellettes de Daubechies d'ordre 4. Cette base possède de bonnes propriétés d'approximation et elle est utilisée pour estimer une fonction de hasard pour des données censurées à droite. Cette méthode est comparée avec d'autres méthodes et elle est illustrée sur des données réelles.

---

**JOHN BRAUN** (University of Western Ontario)

[Monday June 13 / lundi 13 juin, 13:30]

*Kernel Density and Regression Function Estimation using Derivative Information*

*Estimation de densité et de fonctions de régression par la méthode du noyau utilisant l'information à propos des dérivées*

Local polynomial density estimation and regression function estimation can often have better bias properties than local constant estimation, but the increase in degree can lead to instability. In this talk, we explore the situation where derivative information can be used to reduce the number of parameters that need to be estimated locally.

L'estimation polynomiale locale de la densité et l'estimation de la fonction de régression peuvent souvent avoir de meilleures propriétés de biais que l'estimation locale constante, mais une augmentation de degré peut mener à de l'instabilité. Dans cet exposé, nous examinons la situation où de l'information à propos des dérivées peut être utilisée pour réduire le nombre de paramètres à estimer localement.

---

**ÉRIC MARCHAND** (Université de Sherbrooke), **MOHAMMAD JAFARI JOZANI & ALEXANDRE LEBLANC** (University of Manitoba)

[Monday June 13 / lundi 13 juin, 14:00]

*On Estimating a Cumulative Distribution Function with a Balanced Loss Function*

*Sur l'estimation d'une fonction de répartition avec une fonction de perte pondérée*

## 1C-I3: Nonparametric Function Estimation Estimation des fonctions non paramétriques

---

We review and discuss decision-theoretic approaches for estimating a continuous cumulative distribution function related to the choice of a loss function, to the determination and properties of a best invariant estimator (with respect to monotone transformations of the data), and to minimaxity. Integrated balanced loss functions are presented and we argue in favour of their flexibility. We exploit an elegant representation to derive best invariant and minimax estimators and some of their properties for such losses from known results for the “non-balanced” case. (joint work with Alexandre Leblanc and Mohammad Jafari Jozani)

Nous présentons le problème de l'estimation d'une fonction de répartition absolument continue dans un cadre décisionnel en mettant l'accent sur le choix de la fonction de perte, le critère du minimax et la détermination du meilleur estimateur équivariant (par rapport à des transformations monotones des données) et de ces propriétés. Une fonction de perte pondérée est proposée et motivée. Nous exploitons une jolie représentation pour trouver un meilleur estimateur équivariant et ses propriétés pour de telles pertes à partir de résultats existants pour leurs versions "non-pondérées". (travail conjoint avec Mohammad Jafari Jozani et Alexandre Leblanc)

## 1C-I4: Random Effects Modelling for Spatial Data Modélisation des effets aléatoires appliquée aux données spatiales

---

Organizer and Chair / Responsable et président: Tariqul Hasan (University of New Brunswick)

**Room/Salle: BAC239**

**Abstracts/Résumés**

---

**XIN FENG & CHARMAINE DEAN** (Simon Fraser University)

[Monday June 13 / lundi 13 juin, 13:00]

*An Overview of Random Effects Models for Spatial Analyses*

*Vue d'ensemble des modèles d'effets aléatoires pour analyses spatiales*

The development of methods for spatial analyses has seen tremendous growth over the last two decades, with considerable impact on disease monitoring and on exploratory analyses to investigate etiology. This talk discusses spatial and spatial-temporal models with specific emphasis on accommodating the dependence across space and time through random effects in the error components. Such a specification provides insights in the spatio-temporal risk surface through the random effects. Spatial random effects models are also useful in the joint analysis of outcomes in mapping studies where shared random effects accommodate correlations in outcomes. A variety of studies will illustrate the methods.

Le développement de méthodes pour les analyses spatiales a connu une progression importante ces deux dernières décennies, avec un impact considérable sur la surveillance de maladies et sur les analyses exploratoires pour étudier l'étiologie. Cet exposé traite des modèles spatiaux et spatio-temporels où l'accent est mis sur l'accommodation de la dépendance à travers l'espace et le temps via des effets aléatoires dans les composantes d'erreur. Une telle spécification aide à comprendre la surface de risque spatio-temporelle via les effets aléatoires. Les modèles d'effets aléatoires spatiaux sont aussi très utiles dans l'analyse conjointe des résultats d'études de cartographie où des effets aléatoires partagés incorporent de la corrélation dans les résultats. Plusieurs études illustreront les méthodes.

---

**RENJUN MA** (University of New Brunswick), **ED HUGHES** (Ed Hughes Consulting, Ottawa, Canada)

[Monday June 13 / lundi 13 juin, 13:30]

*Tweedie Mixed Models for Spatiotemporal Data.*

*Modèles mixtes de Tweedie pour les données spatio-temporelles*

Massive data sets with complex spatiotemporal structures are common in forestry, health and environmental studies. In order to account for such spatiotemporal structures, we incorporate spatially and temporally correlated random effects into Tweedie generalized linear models, accommodating a wide range of discrete, continuous and semi-continuous data. The estimation of these models often poses theoretical and computational challenges. We propose a unified estimation method for these models based on orthodox best linear unbiased predictors of random effects. Our approach is illustrated with application to analysis of environmental data.

Les ensembles de données massifs à structures spatio-temporelles complexes sont courants dans les études de sylviculture, de santé et d'environnement. Afin de tenir compte de telles structures spatio-temporelles, nous incorporons des effets aléatoires spatialement et temporellement corrélés dans des modèles linéaires généralisés de Tweedie, accommodant une large gamme de données discrètes, continues et semi-continues. L'estimation de ces modèles est souvent source de défis théoriques et calculatoires. Nous proposons une méthode d'estimation unifiée pour ces modèles basée sur les meilleurs prédicteurs linéaires sans biais orthodoxes des effets aléatoires. Notre approche est illustrée par une application à une analyse de données environnementales.

## 1C-I4: Random Effects Modelling for Spatial Data Modélisation des effets aléatoires appliquée aux données spatiales

---

**GARY SNEDDON** (Mount Saint Vincent University), **M. TARIQUL HASAN & RENJUN MA** (University of New Brunswick)

[Monday June 13 / lundi 13 juin, 14:00]

*Analysis of Cross-Classified Spatial Data*

*Analyse de données spatiales recoupées*

Our work is motivated by analysis of hospital admission data where patients are cross-classified by their residential areas and general practices. Patient outcomes tend to be correlated within geographical areas and general practices; however, the clustering effects are cross-classified rather than hierarchical. In addition, patient outcomes might be spatially correlated among geographical areas. In this study, we incorporate crossed random effects of geographical areas and general practices into Poisson models. We also consider spatial correlation among effects of geographical areas. An orthodox best linear unbiased predictor approach has been developed in the estimation of our model.

Notre travail est motivé par l'analyse des données d'admission hospitalière, où les patients sont recoupés selon les lieux de leur résidence et de leur pratique générale. Les résultats des patients tendent à être corrélés entre les zones géographiques et les pratiques générales ; néanmoins, les effets de classification sont recoupés plutôt que hiérarchiques. De plus, les résultats de patients peuvent être corrélés spatialement parmi les zones géographiques. Dans cette étude, nous incorporons des effets aléatoires croisés de la zone géographique par la pratique générale dans des modèles de Poisson. Nous considérons aussi une corrélation spatiale parmi les effets des zones géographiques. Une approche orthodoxe du meilleur prédicteur linéaire sans biais a été développée dans l'estimation de notre modèle.



## 1C-I5: Sampling & Biometry Sondage et biométrie

---

Organizer and Chair / Responsable et président: Louis-Paul Rivest (Université Laval)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

**Room/Salle: BAC142**

**Abstracts/Résumés**

---

**STEPHEN SMITH** (Bedford Institute)

[Monday June 13 / lundi 13 juin, 13:00]

*Application of Sampling with Partial Replacement to Scallop Surveys*

*Application de l'échantillonnage à remplacement partiel aux enquêtes sur les pétoncles*

Many of the population models used for fisheries stock assessment require long-term monitoring estimates of population abundance from annual field surveys. In cases where the survey series is new or of short duration, population models may not be able to provide accurate stock assessment advice. Surveys using partial replacement sampling designs for surveying sea scallops (*Placopecten magellanicus*) provide a survey-based means of evaluating annual changes in abundance with higher precision than the standard survey designs. Examples from scallop fishing areas off Nova Scotia illustrate the application of this sampling approach.

Plusieurs modèles de population utilisés pour évaluer les stocks de poissons nécessitent la surveillance à long terme des estimations d'abondance de la population d'enquêtes annuelles réalisées sur le terrain. Lorsque la série d'enquêtes est nouvelle ou de courte durée, les modèles de population ne fournissent pas toujours une évaluation précise des stocks. Les enquêtes utilisant des plans de sondage à remplacement partiel pour enquêter sur les pétoncles géants (*Placopecten magellanicus*) servent à évaluer les changements annuels d'abondance avec une plus grande précision que les plans de sondage habituels. Cette méthode d'échantillonnage est illustrée à l'aide d'exemples tirés de zones de pêche au pétoncle au large de la Nouvelle-Écosse.

---

**NOEL CADIGAN** (Fisheries and Oceans Canada)

[Monday June 13 / lundi 13 juin, 13:30]

*Accounting for Survey Variance Can Improve Fish Stock Assessment*

*Tenir compte de la variance expérimentale peut améliorer l'évaluation des inventaires de poissons*

I derive some basic statistics that describe the variability of a survey index of abundance derived from stratified random sampling for Northwest Atlantic fish stocks. I show how the survey variance component, in addition to "non-sampling" model variability, can be incorporated into stock assessment models. Quasi-likelihood methods are used to develop a model parameter estimation procedure that incorporates survey sample sizes and estimates of within-survey variability. This may lead to improved estimation of stock size, in terms of more precise

Je dérive quelques statistiques de base qui décrivent la variabilité d'un indice de précision d'abondance à partir d'un échantillon aléatoire stratifié des stocks de poissons du nord-ouest de l'Atlantique. Je montre comment les composantes de la variance expérimentale, en plus de la variabilité « non- expérimentale », peuvent être intégrées aux modèles d'évaluation des stocks. Des méthodes basées sur la quasi-vraisemblance sont utilisées pour développer une procédure d'estimation des paramètres du modèle qui intègre les tailles d'échantillon et les estimations de la variabilité à l'intérieur des enquêtes. Cela peut améliorer les estimations

## 1C-I5: Sampling & Biometry Sondage et biométrie

---

parameter estimates that are less sensitive to more uncertain model inputs.

de la taille des stocks, grâce à des estimations plus précises des paramètres et moins sensibles à des écarts aux présupposés du modèle.

---

**SIMON J BONNER** (University of Kentucky), **CARL J SCHWARZ** (Simon Fraser University)

[Monday June 13 / lundi 13 juin, 14:00]

*Advances in Modelling Time-Stratified Mark-Recapture Data from Salmon Outmigrations via Bayesian P-Splines*  
*Développements dans la modélisation de données stratifiées par le temps de marquage-recapture de la migration des saumons au moyen de p-splines bayésiennes*

At the 2008 Annual Meeting of the SSC, we described a method to estimate abundance from sparse, time-stratified, mark-recapture data. Modelling population size across strata as a smooth function, via a Bayesian P-spline, we obtained estimates of abundance that were unbiased and more precise than existing estimates. In this talk, we describe extensions of the method to model data in which the matrix of recoveries of marked fish is non-diagonal using either a parametric or non-parametric model of fish movements. We also demonstrate how the method can address problems with missing or contaminated data and present applications to salmon outmigrations.

Au congrès annuel 2008 de la Société statistique du Canada, nous avons décrit une méthode pour estimer l'abondance à partir de données rares et stratifiées par le temps de marquage-recapture. Modélisant la taille de la population selon les strates comme une fonction lisse, au moyen d'une p-spline bayésienne, nous avons obtenu des estimateurs d'abondance sans biais et plus précis que les estimateurs existants. Dans cet exposé, nous décrivons des extensions possibles de la méthode pour modéliser des données dont la matrice de récupérations des poissons marqués n'est pas diagonale, en utilisant un modèle soit paramétrique, soit non paramétrique, des mouvements de poissons. Nous montrons également comment la méthode peut résoudre les problèmes associés aux données manquantes ou contaminées et présentons des applications aux données de migrations de saumons.

## 1C-C1: Estimation in Linear Models Estimation dans les modèles linéaires

---

Chair/Président: Elif Acar (McGill University)

Room/Salle: BAC141

### Abstracts/Résumés

---

**ANNE MICHELE MILLAR** (Mount Saint Vincent University)

[Monday June 13 / lundi 13 juin, 13:00]

*Gain Scores and Regression to the Mean*  
*Scores de gain et régression à la moyenne*

Many studies attempt to assess the effectiveness of an intervention by comparing the initial and final value of a response variable, such as student grades or the effect of a drug. Often the gain score (post - pre) depends on the pre-score. Since the scores include measurement error, the true relationship between the gain and the intervention may be hard to estimate due to the phenomenon of regression to the mean. We explore methods to compensate for this, to better reveal the underlying relationship between the "true" gain and the "true" initial score.

Plusieurs études tentent d'évaluer l'efficacité d'une intervention en comparant la valeur initiale et finale d'une variable réponse, telles les notes des étudiants ou l'effet d'une drogue. Souvent le score de gain (après-avant) dépend du score-avant. Puisque les scores comprennent l'erreur de mesure, la réelle relation entre le gain et l'intervention peut être difficile à estimer à cause du phénomène de la régression à la moyenne. Nous explorons des méthodes pour compenser cette limite, pour mieux révéler la relation sous-jacente entre le « réel » gain et le « réel » score initial.

---

**XIAOYU CHEN & XIAOJIAN XU** (Brock University)

[Monday June 13 / lundi 13 juin, 13:15]

*Robust Estimation in Case of Asymmetry*  
*Estimation robuste dans le cas asymétrique*

In this project, we discuss the construction of a modified Huber function for asymmetry. This new function involves a measurement of asymmetry so that the outliers are given unequal weight depends on the degree of asymmetry. We investigate the effects of such a modified Huber function on robust estimation of mean in location, parameters in regression and ARCH models. We use bias, variance, MSE and relative efficiency as criteria in comparing with least squared estimates and Huber's estimates. A set of suitable tuning constants  $k$  for the modified Huber function is attained. The results demonstrate that our function is efficient.

Ce projet illustre la construction d'une fonction de Huber modifiée pour l'asymétrie. Cette nouvelle fonction nécessite une mesure d'asymétrie de telle façon que les valeurs aberrantes reçoivent un poids inégal selon leur niveau d'asymétrie. Nous étudions les effets de la fonction Huber modifiée sur l'estimation robuste de la moyenne, des paramètres de régression et des modèles ARCH. Les critères utilisés pour comparer les estimations des moindres carrés et ceux de Huber sont le biais, la variance, l'EQM et l'efficacité relative. Nous obtenons un ensemble d'ajustement adéquat du réglage de la constante  $k$  de la fonction de Huber modifiée. Les résultats démontrent que notre fonction est efficace.

## 1C-C1: Estimation in Linear Models Estimation dans les modèles linéaires

---

**HENSLEY HUBERT MARIATHAS** (Memorial University of Newfoundland), **BRAJENDRA SUTRADHAR & J CONCEPCIÓN LOREDO-OSTI** (Memorial University of Newfoundland)

[Monday June 13 / lundi 13 juin, 13:30]

*Inferences in Linear Mixed Models with Spatially Correlated Random Effects*

*Inférences pour modèles linéaires mixtes avec effets aléatoires spatialement corrélés*

In spatial data analysis, univariate or multivariate responses are collected from a large sequence of spatial clusters. This type of responses from a spatial cluster may be influenced by certain fixed regression variables as well as suitable latent effects from the members of the cluster. Also the responses may be subject to certain measurement errors. In this talk we demonstrate a simulation based unified moment estimation approach that produces consistent estimates for the regression effects, variance and correlation parameters of the random effects for the members of the cluster. Some simulation results will be presented.

En analyse de données spatiales, des réponses univariées ou multivariées sont recueillies à partir de grandes séquences de grappes spatiales. Ces types de réponse peuvent être influencés par des variables de régression fixes et des effets latents provenant des membres de la grappe. De plus, les réponses peuvent être exposées à certaines erreurs de mesure. Dans cet exposé, nous démontrons une approche d'estimation du moment unifié basée sur une simulation qui produit des estimateurs convergents pour les effets de régression, les paramètres de la variance et de la corrélation des effets aléatoires pour les membres de la grappe. Quelques résultats de la simulation seront présentés.

**BINGRUI SUN** (Memorial University), **BRAJENDRA SUTRADHAR** (Memorial University of Newfoundland)

[Monday June 13 / lundi 13 juin, 13:45]

*GQL Estimation in Linear Dynamic Models for Panel Data*

*Estimation du quasi maximum de vraisemblance généralisé pour les jeux de données provenant d'un modèle linéaire dynamique*

In the econometrics literature, it is standard to use the existing instrumental variables as well as generalized method of moments approaches, for the estimation of the parameters of a linear dynamic mixed model for panel data. In this talk, we introduce a generalized quaslikelihood estimation approach that produces estimates with smaller biases and standard errors as compared to the aforementioned and other existing approaches.

Dans la littérature économétrique, l'approche par l'existence de variables instrumentales et celle de la méthode des moments généralisés sont fréquentes pour l'estimation des paramètres des modèles linéaires dynamiques mixtes de jeux de données. Dans cet exposé, nous introduirons l'approche de l'estimation du quasi maximum de vraisemblance généralisé qui produit des estimations ayant un biais et un écart-type moindres que ceux des approches mentionnées précédemment et ceux provenant de d'autres approches connues.

**MAHMOUD TORABI** (University of Manitoba)

[Monday June 13 / lundi 13 juin, 14:00]

*Likelihood Inference in Generalized Linear Mixed Models with Two Components of Dispersion*

*Inférence de vraisemblance dans les modèles linéaires mixtes généralisés à deux composantes de dispersion*

In this talk, I will consider the generalized linear mixed models (GLMMs) with two components of dispersion. The frequentist analysis of GLMM is computationally difficult. On the other hand, the

Dans cet exposé, je considérerai les modèles linéaires mixtes généralisés (GLMM) à deux composantes de dispersion. L'analyse fréquentiste des GLMM est analytiquement difficile. Par contre, l'arrivée des algorithmes de chaînes

## 1C-C1: Estimation in Linear Models Estimation dans les modèles linéaires

---

advent of the Markov chain Monte Carlo algorithm has made the Bayesian analysis of GLMM computationally convenient. I use recently proposed data cloning approach, which is computationally convenient, to conduct frequentist analysis of GLMMs with two components of dispersion based on maximum likelihood estimation. I discuss the performance of the proposed estimators through simulation, and also illustrate their use with the well known salamander mating data.

de Markov par méthode de Monte Carlo a rendu l'analyse Bayésienne analytique des GLMM beaucoup plus pratique. J'utilise la nouvelle approche du clonage de données, qui est analytiquement pratique, pour l'analyse fréquentiste des GLMM à deux composantes de dispersion basée sur l'estimation du maximum de vraisemblance. Je discuterai des performances des estimateurs proposés à travers des simulations, et je montrerai aussi leur utilité avec le jeu de données bien connu d'accouplement salamandre.

---

**MARC SORMANY** (Université de Moncton), **VARTAN CHOULAKIAN** (Université de Moncton, Dépt. de Math/Statistique)

[Monday June 13 / lundi 13 juin, 14:15]

*Picture of 2 and 3 Parameters Least Squares Solutions in Rational Regression*  
*Régression rationnelle par moindres carrés : carte des solutions*

We consider the problem of estimating the parameters by least squares in some rational regression models, such as, Michaelis-Menten, Misra1b and Hailwood-Horrobin. The aim of this paper is to show how to pick the best solution from the complete solution set using computational algebraic methods.

Nous considérons le problème d'estimation des paramètres par moindres carrés dans certains modèles de régression rationnelle, comme ceux de Michaelis-Menten, Misra1b et Hailwood-Horrobin. Ces modèles ont 2 ou 3 paramètres. Le but de cette présentation est de montrer comment choisir la meilleure solution de l'ensemble des solutions à l'aide de méthodes algébriques.

## 1C-C2: Probability Applications in Finance and Economics Applications de la probabilité en finance et économie

---

Chair/Président: Boxin Tang (Simon Fraser University)

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

**Room/Salle: BAC138**

**Abstracts/Résumés**

---

**VICKNESWARY TAGORE** (University of Saskatchewan), **BRAJENDRA C SUTRADHAR** (Memorial University)

[Monday June 13 / lundi 13 juin, 13:00]

*Simple and Efficient Estimation Approaches for Stochastic Volatility Models*

*Approches simples et efficaces d'estimation des modèles de volatilité stochastique*

There is a long history on the estimation of the parameters of the standard stochastic volatility models. Among the existing approaches, the so-called generalized method of moments (GMM) and the quasi-maximum likelihood estimation techniques are widely used. We introduce a simpler method of moments (MM), which, unlike the existing GMM approach, does not require an arbitrarily large number of unbiased moment functions to construct moment estimating equations for the parameters involved. Furthermore, the small sample behaviour of the MM approach is examined through a simulation study. We also provide another simpler 'working' generalized quasi likelihood (WGQL) approach.

L'estimation des paramètres de modèles de volatilité stochastique standard possède une riche histoire. Parmi les approches existantes, les méthodes d'estimation généralisée des moments (GMM) et du quasi-maximum de vraisemblance sont très connues et couramment utilisées. Nous présentons une méthode simple des moments (MM) qui, contrairement à l'approche MGM existante, ne nécessite pas un nombre arbitrairement élevé de fonctions non biaisées des moments pour construire les équations d'estimation des moments des paramètres. De plus, nous examinons en détail la performance pour de petits échantillons de l'approche MM à l'aide d'une étude de simulation. Nous présentons aussi une autre approche simple généralisant l'approche basée sur la quasi-vraisemblance.

---

**ZHONGXIAN MEN** (University of Waterloo)

[Monday June 13 / lundi 13 juin, 13:15]

*Multivariate Stochastic Volatility Models: a Probabilistic Principal Component Analysis Approach*

*Les modèles de volatilité stochastique multivariés : une approche probabiliste de l'analyse en composantes principales*

Many financial time series exhibit changing variance and this can have important consequences in formulating economic or financial decisions. We propose a parsimonious multivariate stochastic volatility model to describe this time varying property. The model is formulated based on probabilistic principal component analysis. Non-zero correlations are permitted among innovations and between those of the asset returns and volatility dynamics. A Gibbs method is derived under a Bayesian framework to estimate the parameters

Plusieurs séries chronologiques financières comportent une variabilité non-homogène dans le temps et ceci peut avoir des conséquences importantes sur les prises de décision économique ou financière. Nous proposons un modèle de volatilité stochastique multivarié parcimonieux pour décrire cette variabilité. La formulation du modèle se base sur une analyse en composantes principale probabiliste. Des corrélations non-nulles sont permises entre les innovations et celles entre les retours sur les actifs et les dynamiques de volatilité. Une méthode de Gibbs est appliquée dans un cadre bayésien pour estimer les paramètres et les loga-

## 1C-C2: Probability Applications in Finance and Economics Applications de la probabilité en finance et économie

---

and log volatilities. Simulation studies and application to real data illustrate how the model and proposed approach work.

rithmes des volatilités. Le fonctionnement du modèle et de l'approche proposée sont illustrés à l'aide d'études par simulation et d'applications à des données réelles.

---

**ZHENYU CUI & CAROLE BERNARD** (University of Waterloo), **WILLIAM GORNALL** (Stanford University)

[Monday June 13 / lundi 13 juin, 13:30]

*Improved Continuity Correction for Discrete Barrier and Lookback Options*

*Correction de continuité améliorée pour des options discrètes de barrière et rétrospective*

Barrier and lookback options are the most popular path-dependent options traded in the derivatives market. Based on the higher order expansion of Brownian motion approximation to random walk given by Janssen and Van Leeuwaarden (2009), we extend the "continuity correction" in Broadie et al (1999) to develop a more accurate method for pricing discrete barrier options. We also extend the technique in Horfelt (2003) to the pricing of discrete lookback options and propose a highly accurate semi-analytical approximation to the covariance term to improve the continuity correction for discrete lookback options. A numerical study confirms the advantages of our method.

Les options de barrière et rétrospective sont les options de trajectoire dépendante les plus populaires échangées sur le marché dérivé. En se basant sur un développement d'ordre supérieur du mouvement brownien approchant la marche aléatoire donnée par Janssen et Van Leeuwaarden (2009), nous prolongeons la « correction de continuité » de Brodie et al (1999) pour développer une méthode plus précise pour fixer le prix des options discrètes de barrière. Nous prolongeons également la technique de Horfelt (2003) pour fixer les prix d'options discrètes rétroactives et proposons une approximation semi-analytique très précise du terme de covariance pour améliorer la correction de continuité des options discrètes rétroactives. Une étude numérique confirme les avantages de notre méthode.

---

**FRANCOIS WATIER & ALEXANDRE SCOTT** (University of Quebec in Montreal)

[Monday June 13 / lundi 13 juin, 13:45]

*Goal Achieving Probability of a Mean-Variance Strategy Without Short-Selling*

*Probabilité d'atteindre le but d'une stratégie moyenne-variance sans vente à découvert*

In a 2006 paper, Li and Zhou established that an investor, following an unconstrained mean-variance strategy, will achieve its discounted targeted wealth with a probability greater than 80%. Surprisingly, we will show that under short-selling restrictions (i.e. without the possibility of borrowing stocks) this lower bound probability still holds.

Li et Zhou ont établi, dans un article publié en 2006, qu'un investisseur, suivant une stratégie moyenne-variance sans contrainte, atteindra avec une probabilité supérieure à 80% sa richesse actualisée visée. Nous montrerons que, curieusement, cette probabilité de borne inférieure tient toujours sous des restrictions de vente à découvert (c.-à.-d. sans possibilité d'emprunter des actions).

---

**ALEXANDRE SCOTT & FRANCOIS WATIER** (University of Quebec in Montreal)

[Monday June 13 / lundi 13 juin, 14:00]

*A Stopping Time Rule for Mean-Variance Portfolios Without Bankruptcy*

*Un temps d'arrêt dans le temps pour des portefeuilles moyenne-variance sans faillite*

In this talk, we will evaluate the probability of reaching a discounted targeted wealth based on a stopping time rule introduced by Li and Zhou (2006) and applied to mean-variance strategies

Dans cet exposé, nous évaluerons la probabilité d'atteindre une richesse actualisée visée basée sur un temps d'arrêt introduit par Li et Zhou (2006) et appliquée à des stratégies moyenne-variance avec interdiction de faillite. À cet effet,

## 1C-C2: Probability Applications in Finance and Economics Applications de la probabilité en finance et économie

---

with bankruptcy prohibition. To this end, we will make use of results on first passage time of Gauss-Markov processes through time-varying boundaries.

nous utiliserons les résultats sur le premier temps de passage des processus de Gauss-Markov à travers des frontières dépendantes du temps.



## 1C-C3: Survival Models: Predictive and Causal Analysis Modèles de survie : analyse prévisionnelle et causale

---

Chair/Président: Tim Ramsay (University of Ottawa)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 234**

**Abstracts/Résumés**

---

**JAMES HANLEY & ZHIHUI LIU** (McGill University)

[Monday June 13 / lundi 13 juin, 13:00]

*Measuring the Performance of Prognostic Models for Survival Data*

*Mesure de la performance des modèles pronostiques pour des données de survie*

We wish to quantify how much prognostic performance is improved by adding items to the profile used for risk functions derived from survival data. Measures include re-classification probabilities, and improvements in indices based on ROC curves. With censored observations, the measures are challenging. We focus on the auc measure introduced by Heagerty and Zheng. Using worked examples, we provide heuristics to better explain and interpret the measure. We propose another measure, based on percentiles within each risk set, that may have a simpler and more direct interpretation. We illustrate the index using data from the Framingham Heart Study.

Nous voulons quantifier l'amélioration de la performance pronostique avec l'ajout d'items au profil utilisé pour les fonctions de risque dérivées des données de survie. Les mesures comprennent les probabilités de reclassification et les améliorations des indices basés sur des courbes ROC. Comme des données sont censurées, les mesures peuvent être problématiques. Nous nous concentrons sur la mesure de la surface sous la courbe ROC introduite par Heagerty et Zheng. À l'aide d'exemples pratiques, nous présentons des heuristiques afin de mieux expliquer et interpréter la mesure. Nous proposons une autre mesure basée sur les centiles de chacun des ensembles de risque dont l'interprétation pourrait être plus simple et plus directe. Nous illustrons l'indice à l'aide des données de l'étude coronarienne de Framingham.

---

**AMAY S.M. CHEAM** (University of Ottawa)

[Monday June 13 / lundi 13 juin, 13:15]

*ROC Curve Modeling with Pearson Distributions*

*Modélisation de la courbe ROC à partir des distributions de Pearson*

There exists two methods to model the ROC: the *direct* method, which is less appealing, consists of constructing the curve based on the observed data, while in the *indirect* method, a distribution is assumed for diseased and nondiseased patients. Because of the ease of its manipulation, the normal distribution has been used to model both types of patients. Instead, we use Pearsonian system of distributions in order to select the distribution for both subjects. An approach using Monte Carlo

La courbe ROC évalue l'efficacité d'un diagnostic. Il existe deux approches de modélisation : la méthode *directe*, moins favorisée, consiste à modéliser directement la courbe ROC par une forme fonctionnelle quelconque, alors que la méthode *indirecte* repose sur une hypothèse concernant la distribution des populations malades et non-malades. Pour des raisons de commodité, la distribution normale est généralement utilisée pour modéliser les deux types de patients. Notre étude propose d'utiliser plutôt la distribution de Pearson, étant donné sa flexibilité à caractériser des données

## 1C-C3: Survival Models: Predictive and Causal Analysis Modèles de survie : analyse prévisionnelle et causale

---

simulation provides a confidence band for the derived ROC. It is seen that the proposed method is at least as accurate as the binormal curve, and often superior.

jusqu'au quatrième moment. L'absence d'une écriture analytique nous amène à employer une technique de simulation Monte-Carlo permettant de construire un intervalle de confiance autour de la courbe ROC. Les résultats affichent une performance égale et souvent supérieure à l'approche binormale.

---

**OLLI SAARELA** (McGill University), **ELJA ARJAS** (University of Helsinki)

[Monday June 13 / lundi 13 juin, 13:30]

*On Non-parametric Bayesian Regression in Cardiovascular Disease Risk Assessment*

*Régression Bayésienne non-paramétrique dans l'évaluation du risque de maladies cardiovasculaires*

Assessing the absolute risk for a future disease event for presently healthy individuals has an important role in primary prevention of cardiovascular diseases (CVD). In addition to utilizing data on novel risk factors on top of well-established ones, the predictions can potentially be improved using more flexible statistical models. Here we concentrate on this latter task, using non-parametric Bayesian regression techniques. The computational tool utilized is our previously published Bayesian multivariate monotonic regression procedure, which we generalize here to a survival analysis setting. We apply the proposed methods to 10-year CVD prediction in Finnish population based cohorts.

L'évaluation du risque absolu d'un futur épisode de maladie pour les individus présentement en santé joue un rôle important dans la prévention primaire des maladies cardiovasculaires. En plus d'utiliser les données des nouveaux facteurs de risque sur ceux bien établis, les prédictions peuvent potentiellement être améliorées en utilisant des modèles statistiques plus flexibles. Ici, nous nous concentrons sur cette tâche en utilisant des méthodes de régression Bayésienne non-paramétriques. Nous utilisons ici, comme outil calculatoire, notre précédente procédure de régression Bayésienne monotone multivariée, que nous généralisons à l'analyse de survie. Nous appliquons les méthodes proposées à des prédictions de maladies cardiovasculaires sur 10 ans dans des cohortes d'une population finlandaise.

---

**JULIE HORROCKS & GREG PERSSON** (University of Guelph), **ANNE DUFFY & SARAH DOUCETTE** (Dalhousie University), **PAUL GROF** (University of Ottawa)

[Monday June 13 / lundi 13 juin, 13:45]

*Statistical Models for Staging*

*Modèles statistiques pour la stadification*

Accruing evidence suggests that childhood psychopathology predicts for adult psychiatric disorders, often of a qualitatively different nature. Recent research has proposed a clinical staging model for the development of bipolar disorder. The staging model is meant to be similar to staging used in other medical conditions (e.g. cancer). I will examine procedures for confirming or disconfirming the staging hypothesis, using methods of survival analysis. The methods will be illustrated on a rich data set which incorporates dates of various diagnoses of psychopathology in subjects at high risk of bipolar disorder and controls, followed

De nombreuses preuves suggèrent que la psychopathologie infantile permet de prédire les troubles mentaux chez les adultes, souvent de nature qualitativement différente. Une recherche récente propose un modèle clinique de stadification du développement du trouble bipolaire. Ce modèle est conçu de façon semblable à la stadification effectuée sous d'autres conditions médicales (p. ex. les cas de cancer). J'examine des procédures de confirmation et de réfutation de l'hypothèse de stadification au moyen de méthodes d'analyse de survie. Ces méthodes sont illustrées grâce à un riche jeu de données incorporant la date de plusieurs diagnostics de psychopathologie chez des sujets à haut risque pour le trouble bipolaire et des témoins qui ont

## 1C-C3: Survival Models: Predictive and Causal Analysis Modèles de survie : analyse prévisionnelle et causale

---

prospectively for up to 15 years.

été suivis de façon prospective pendant une période allant jusqu'à 15 ans.

---

**MEAGHAN CUERDEN, CECILIA COTTON & RICHARD COOK** (University of Waterloo)

[Monday June 13 / lundi 13 juin, 14:00]

*Statistical Models and Methods for Causal Analysis of the Effect of Storage Time of Blood in Patients Receiving Transfusions*

*Modèles statistiques et méthodes d'analyse causale des effets du temps de stockage du sang chez les transfusés*

Recent evidence suggests that the length of time red blood cells have been stored may affect outcomes such as survival among transfused hospitalized patients. We describe a model for simulating the need for transfusions over time, the response to transfusions, and survival times. Through this model we examine the effect of model misspecification on estimates of the effect of duration of storage on survival. We discuss issues of causal inference and critique a recent article on this topic which used a propensity score approach.

De récentes études suggèrent que le temps de stockage des globules rouges peut avoir un effet, entre autres, sur la survie chez les patients hospitalisés transfusés. Nous décrivons un modèle qui simule les besoins en transfusions dans le temps, les réactions aux transfusions et les temps de survie. Via ce modèle, nous examinons l'effet des erreurs de spécification du modèle sur les estimés de l'effet du temps de stockage sur la survie. Nous explorons divers problèmes concernant l'inférence causale et critiquons un récent article sur ce sujet qui utilise une approche fondée sur les cotes de propension.

---

**HUI SONG, YINGWEI PENG & DONGSHENG TU** (Queen's University)

[Monday June 13 / lundi 13 juin, 14:15]

*A New Approach for Joint Modeling of Longitudinal Measurements and Survival Times with a Cure Fraction*

*Nouvelle approche pour la modélisation combinée des mesures longitudinales et des temps de survie avec une fraction guérie*

Joint analysis of longitudinal measurements and survival data is useful in clinical trials and other medical studies. In this paper, we consider a joint model which assumes a linear mixed tt model for longitudinal measurements and promotion time cure model for survival data and links these two models through a latent variable. A semiparametric inference procedure with EM algorithm is developed for the parameters in the joint model. The proposed procedure is evaluated in a simulation study and also applied to analyze the quality of life and time to recurrence data from a clinical trial in women with early breast cancer.

L'analyse combinée des données longitudinales et des données de survie est très utile pour les essais cliniques et les autres études médicales. Dans cet exposé, nous considérerons un modèle combiné prenant pour acquis un modèle linéaire mixte tt pour les mesures longitudinales et un modèle promouvant le temps de traitement pour les données de survie, et nous relierons ces modèles par une variable latente. Une procédure d'inférence semi-paramétrique avec algorithme EM sera développée pour les paramètres du modèle combiné. La procédure proposée sera évaluée par simulation et sera appliquée pour analyser les données de qualité de vie et de temps de rechute d'un essai clinique sur des femmes prématurément atteintes de cancer du sein.

Room/Salle: Festival Theatre

Abstracts/Résumés

---

**LENGYI SPECTRUM HAN & WILLARD JOHN BRAUN** (University of Western Ontario)

[Monday June 13 / lundi 13 juin, 14:30]

*Prediction Contours for the Prometheus Fire Growth Model*

*Contours du modèle Prometheus de prédiction de la croissance des feux*

The Prometheus Fire Growth Model is a deterministic simulator used by the Alberta Government to predict forest fire spread, given weather, topographical and fuel information. Unpredictability of fire behaviour makes such deterministic predictions inaccurate. Probability mapping is needed. The Prometheus model equations are estimated from data on experimental fires and well-documented wildfires. These data can be used to model not only the mean rate of spread (ROS) but also the variance, after applying variance-stabilizing transformations to the data in a piecewise manner: surface fires require a different transformation from crown fires. Probability contours are then based on ROS prediction quantiles.

Le modèle de prédiction de la croissance des feux Prometheus est un simulateur déterministe utilisé par le gouvernement de l'Alberta pour prévoir la propagation d'un incendie de forêt, en fonction de données météorologiques, topographiques et relatives aux types de combustible. L'imprévisibilité du comportement des feux rend ces prévisions déterministes imprécises. Une application de probabilités est nécessaire. Les équations du modèle de Prometheus sont estimées à partir de données sur des feux expérimentaux ainsi que sur des feux de friches bien documentés. Ces données peuvent être utilisées pour modéliser non seulement le taux moyen de propagation, mais aussi la variance, après avoir appliqué les transformations stabilisatrices de la variance à des données en plusieurs groupes : les feux de surface nécessitent une transformation différente de celle des feux de cime. Les contours de probabilité sont ainsi basés sur les quantiles de prédiction du taux moyen de propagation.

---

**JONATHAN LEE, HAO YU & REG KULPERGER** (University of Western Ontario)

[Monday June 13 / lundi 13 juin, 14:30]

*Parallel Computation of Spatial Lattice Models*

*Calculs parallèles de modèles spatiaux en treillis*

Lattice models are a way of representing spatial locations in a grid where each cell is in a certain state and evolves according to transition rules and rates dependent on a surrounding neighbourhood. These models are capable of describing many phenomena such as the simulation and growth of a forest fire front. Lattice models have the potential to take advantage of parallel computing by dividing the lattice into smaller parts and having individual CPUs perform computations on a single sub-lattice while maintaining local interaction with neighbours using a message passing interface (MPI) framework in an efficient and scalable way.

Les modèles en treillis permettent de représenter des endroits sur une grille, où chaque cellule est dans un certain état, et évolue selon des règles et des taux de transition qui dépendent d'un environnement proche. Ces modèles peuvent décrire de nombreux phénomènes, comme la simulation et la propagation d'un front de feu de forêt. Les modèles en treillis permettent de tirer avantage du calcul parallèle par la division du treillis en petites parties et par le calcul effectué par des unités centrales (CU) sur un unique « sous-treillis ». Ces opérations sont effectuées tout en maintenant l'interaction locale avec les voisins au moyen d'un cadre d'interface de passage de messages d'une manière efficace et extensible. \_\_\_\_\_

**SHUDONG LI & ZILIN WANG** (Wilfrid Laurier University)

[Monday June 13 / lundi 13 juin, 14:30]

*Analysis of Generalized Linear Mixed Models for Complex Survey Data*

*Analyse des modèles linéaires mixtes généralisés pour des données d'enquêtes complexes*

We analyze a class of generalized linear mixed models based on the Tweedie exponential dispersion model distributions for complex survey data. To take consideration of sampling design, we extend the orthodox best linear unbiased predictor approach in Ma and Jørgensen (2007) by incorporating sampling weights in both the optimal estimating function for the regression parameters and the prediction of the random effects. Variance estimates of the resulting estimators are derived. Performance of the survey estimates of the model parameters are examined through simulation studies. Empirical examples are given using then Canada Workplace Employer Survey data.

Nous analysons une classe de modèles linéaires mixtes généralisés pour des données d'enquêtes complexes en fonction des distributions de Tweedie du modèle de dispersion exponentielle. Pour prendre en considération le plan d'échantillonnage, nous étendons l'approche orthodoxe du meilleur prédicteur linéaire sans biais de Ma et Jørgensen (2007). À cette fin, nous incorporons des poids d'échantillonnage dans la fonction d'estimation optimale pour les paramètres de régression et la prédiction des effets aléatoires. Nous obtenons les estimations de la variance des estimateurs résultants. Nous examinons également les résultats des estimations d'enquête sur les paramètres du modèle au moyen d'études de simulation. Nous présentons des exemples empiriques avec des données d'enquête sur le milieu de travail et les employés au Canada.

**WENKAI MA & WILLARD JOHN BRAUN** (University of Western Ontario)

[Monday June 13 / lundi 13 juin, 14:30]

*Fitting a Stochastic Spread Model to Experimental Fire Data*

*L'ajustement d'un modèle de propagation stochastique à des données sur des incendies expérimentaux*

We fit an interacting particle system model for fire spread to some fires burned at various slopes. The model has 5 parameters when the slope is nonzero and 2 when the slope is 0. We compare observed fire shapes and sizes with those computed from simulations of the model.

Nous ajustons un modèle de système de particules interactif pour la propagation d'un incendie sur diverses pentes. Le modèle a 5 paramètres quand la pente est différente de 0, et en a 2 lorsque la pente est de 0. Nous comparons les formes et les superficies des incendies avec celles calculées à partir des simulations du modèle.

**MATEEN SHAIKH, LUIZA ANTONIE & PAUL D. MCNICHOLAS** (University of Guelph)

[Monday June 13 / lundi 13 juin, 14:30]

*Modelling with Association Rules*

*Modélisation avec des règles d'association*

Association rules are a data mining technique for the analysis of transaction data. Though the concept of association rules is not novel, high-performance computing has recently made the discovery of interesting association rules more practical. Much of the work with association rules has been without statistical models. We present a statistical model-based approach to association rule

Les règles d'association sont une technique d'exploration de données qui permettent d'analyser des données de transaction. Bien que ce concept ne soit pas nouveau, un calcul très efficace a récemment permis de découvrir des règles d'association plus pratiques. La plupart des travaux avec ces règles ont été effectués sans modèle statistique. Nous présentons une approche basée sur un modèle statistique pour l'exploration de règles d'association. Notre approche

## 1D-P: Poster Session 1 Séance d'affichage 1

---

mining. Our approach is illustrated on real data and the results are discussed.

est illustrée de données réelles, et nous discutons des résultats.

---

**AZADEH (FATEMEH) SHOHOUDI** (Department of Mathematics and Statistics, McGill University), **MASOUD ASGHARIAN, ABBAS KHALILI & DAVID WOLFSON** (McGill University)

[Monday June 13 / lundi 13 juin, 14:30]

*Variable Selection in Multipath Change-point Problems*

*Sélection de variables dans les problèmes de point de rupture à voies multiples*

Follow-up studies are frequently carried out to study evolution of one or several measurements taken on some subjects through time. When a stimulus is administered on subjects, it is of interest to study the reaction times, change-points. One may want to select the covariates that accelerate reaction to the stimulus. Selecting effective covariates in this setting pose a challenge when the number of covariates is large. We develop such methodology and study the large sample behavior of the method. Small sample behavior is studied by the means of simulation. The method is applied to a Parkinson disease data set.

Des études de suivi sont fréquemment effectuées pour étudier l'évolution d'une ou plusieurs mesures prises sur quelques sujets à travers le temps. Quand un stimulus est administré à des sujets, il est intéressant d'étudier les points de rupture du temps de réaction. On peut vouloir choisir les covariables qui accélèrent la réaction au stimulus. Sélectionner les covariables efficaces dans ce cas est un véritable défi quand le nombre de covariables est grand. Nous avons développé de telles méthodologies et étudions les comportements de la méthode pour de larges échantillons. Les comportements pour de petits échantillons sont étudiés par la moyenne de simulations. La méthode est appliquée à des données sur la maladie de Parkinson.

---

**LU XIN & MAHESH PANDEY** (University of Waterloo)

[Monday June 13 / lundi 13 juin, 14:30]

*An Improved Statistical Model for Predicting the Deuterium Ingress in Zirconium Alloy Pressure Tubes*

*Un modèle statistique amélioré pour prévoir l'introduction de deutérium dans des tubes de force en alliage de zirconium*

In the CANDU pressurized heavy water reactor (PHWR), the corrosion of zirconium alloy produces deuterium that causes hydrogen embrittlement. An accurate prediction of deuterium accumulation over time is an important step for ensuring the fitness-for-service of pressure tubes. Deuterium ingress data exhibits heteroscedasticity. The currently used model by the nuclear industry does not deal with heteroscedasticity precisely, so it results in a conservative prediction of the deuterium ingress. We present a new approach for predicting deuterium ingress based on weighted least-squares regression that overcomes the limitations of the existing model, and it provides realistic prediction bounds of deuterium ingress.

Dans le réacteur à eau lourde sous pression (RELP) CANDU, la corrosion de l'alliage de zirconium produit du deutérium qui cause une fragilisation par l'hydrogène. Une prévision précise de l'accumulation de deutérium dans le temps est importante pour assurer l'aptitude fonctionnelle des tubes de force. Les données relatives à l'introduction de deutérium montrent l'hétéroscédasticité. Le modèle actuellement utilisé par l'industrie nucléaire ne porte pas précisément sur l'hétéroscédasticité, ce qui engendre une prévision conservatrice de l'introduction de deutérium. Nous présentons une nouvelle approche pour prévoir l'introduction de deutérium, en fonction de la régression pondérée des moindres carrés qui surmonte les limitations du modèle existant, et qui fournit des bornes de prévision réaliste d'introduction de deutérium.

**1E-A: SSC Impact Award Address**  
**Allocution du récipiendaire du prix pour impact de la SSC**

---

Chair/Président: Chris Field (Dalhousie University)

**Room/Salle: Festival Theatre**

**Abstract/Résumé**

---

**ABDEL EL-SHAARAWI** (National Water Research Institute, Environment Canada and McMaster University)

[Monday June 13 / lundi 13 juin, 15:00]

*An Overview of Environmetrics Research*

*La recherche en statistique appliquée aux sciences de l'environnement : une vue d'ensemble*

The International Environmetrics Society (TIES) was effective in promoting statistical research on the environment since 1989. I will summarize the process which led to creating TIES and describe the evolution of environmental problems from local to global scale and its impact on statistical research. The analysis, interpretation and prediction of space-time variation is an area of very rapid development. The efforts are directed towards: the integration of physical processes and data in statistical models, the analysis of massive data sets, and the setting of sampling designs. I will review some of these methods and use examples to illustrate their applications.

La Société Internationale Environmetrics (TIES) contribue, depuis 1989, à promouvoir la recherche en statistique appliquée aux sciences de l'environnement. Je résumerai le processus ayant conduit à la création de TIES avant de décrire l'évolution des problèmes environnementaux, du plan local jusqu'au niveau global, avec leurs implications dans la recherche en statistique. L'analyse, l'interprétation et la prédiction de la variation spatio-temporelle connaissent un développement très rapide. Les axes de recherche concernent : l'intégration de processus physiques dans les modèles statistiques, l'analyse de données massives et l'élaboration de plans d'expériences. J'examinerai certaines de ces méthodes en illustrant leurs applications par quelques exemples.

## 1F-I1: Developing Teaching Skills in Graduate Students Comment former les étudiants de cycle supérieur à l'enseignement

---

Chair/Président: Kevin Thorpe (University of Toronto)  
Organizer/Responsable: Alison Gibbs (University of Toronto)

Session sponsored by the Statistical Education Committee  
Session parrainée par le Comité d'éducation en statistique

**Room/Salle: BAC 237**

**Abstracts/Résumés**

---

**JOE BLITZSTEIN & XIAO-LI MENG** (Harvard University)

[Monday June 13 / lundi 13 juin, 16:00]

*Teaching the Teaching of Statistics*

*Formation relative à l'enseignement des statistiques*

We discuss experiences in teaching Stat 303 (The Art and Practice of Teaching Statistics) with Xiao-Li Meng and staff from the Derek Bok Center for Teaching and Learning. This course is required for all first year PhD students in the Harvard Statistics Department. Stat 303 teaches students to be effective teaching assistants, but also it serves a much broader purpose: helping students become good *communicators*, and showing that a deep understanding of statistical thinking – how to *connect the dots* rather than just memorize formulas – is crucial both in becoming successful teachers and in becoming successful researchers.

Nous avons parlé des expériences d'enseignement du cours intitulé « Stat 303 » (sur l'art et la pratique de l'enseignement de la statistique) avec Xiao-Li Meng et le personnel du Centre Derek Bok pour l'enseignement et de l'apprentissage. Tous les étudiants en première année de doctorat en statistique à l'Université Harvard sont obligés de suivre ce cours, dont l'objectif n'est pas seulement de les faire devenir des assistants d'enseignement efficaces. Il les aide à devenir de bons *communicateurs*. Il montre également qu'il est crucial de bien comprendre le raisonnement statistique, à savoir *d'établir des liens* au lieu d'apprendre par coeur des formules, autant pour devenir un bon enseignant qu'un bon chercheur.

---

**BETHANY WHITE** (University of Western Ontario)

[Monday June 13 / lundi 13 juin, 16:30]

*Training our Graduate Students to be Statistics Educators*

*Formation de nos étudiants diplômés pour qu'ils deviennent des pédagogues statisticiens*

Most of our graduate programs develop students' technical and computational skills and prepare students for research in the statistical sciences. Many graduate students gain little to no teaching experience but "Communication Skills" and "Teaching and Knowledge Transfer Skills" are two of the four professional skills identified by The Canadian Association for Graduate Studies as being key professional skills to develop during graduate studies. This talk will present two literature-based models to develop students' teaching skills. I will also share our experiences with departmental TA develop-

La plupart de nos programmes permettent aux étudiants diplômés d'améliorer leurs compétences techniques et en calcul, ainsi que de se préparer à la recherche en sciences statistiques. Un grand nombre de ces étudiants n'acquièrent peu ou pas d'expérience d'enseignement, mais les « aptitudes à communiquer » ainsi que les « aptitudes à enseigner et à transférer des connaissances » sont deux des quatre compétences professionnelles essentielles à développer pendant leurs études, d'après l'Association canadienne pour les études supérieures. Cet exposé présente deux modèles documentés pour élargir les compétences des étudiants. Je vais également partager nos expériences



## 1F-I1: Developing Teaching Skills in Graduate Students Comment former les étudiants de cycle supérieur à l'enseignement

---

ment and a "Teaching Statistics" graduate course in Western's Department of Statistical and Actuarial Science.

avec un cours sur l'évolution des assistants d'enseignement et sur « l'enseignement des statistiques » à la faculté des sciences statistiques et actuarielles de l'Université de l'ouest de l'Ontario.

---

**ALISON GIBBS, ZEYNEP BASKURT & RAMYA THINNIYAM** (University of Toronto)

[Monday June 13 / lundi 13 juin, 17:00]

*Developing Graduate Students' Supervisory Skills*

*L'amélioration des compétences des étudiants diplômés*

When today's graduate students become career statisticians in industry or in academia, supervising junior statisticians will be one of their responsibilities. In this talk we describe our experiences with a new undergraduate course in statistical collaboration at the University of Toronto. The course was designed, in part, to develop supervisory skills in graduate students by having them advise undergraduates in their work on collaborative research projects. Team meetings and feedback from a faculty member provide the graduate students with the opportunity to reflect on and develop their personal mentoring styles.

De nos jours, lorsque les étudiants diplômés deviennent des statisticiens de carrière, que ce soit dans les milieux industriel ou universitaire, une de leurs responsabilités consistera à superviser les nouveaux statisticiens. Dans cet exposé, nous décrivons nos expériences avec un nouveau cours pour les étudiants du premier cycle en collaboration statistique à l'Université de Toronto. Le cours a été conçu en partie pour améliorer les compétences en supervision des étudiants diplômés, en l'occurrence ils conseillent les étudiants du premier cycle dans leurs travaux sur des projets de recherche conjoints. Les réunions d'équipe et les commentaires d'un membre de la faculté donnent, aux étudiants diplômés, l'occasion de réfléchir sur leur style personnel d'encadrement, et de l'améliorer.

## 1F-I2: Measuring Water Quality Comment mesurer la qualité des eaux

---

Chair/Président: Lisa Lix (University of Saskatchewan)  
Organizer/Responsable: Abdel El-Shaarawi (National Water Research Institute, Environment  
Canada)

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

**Room/Salle: BAC132**

**Abstracts/Résumés**

---

**PIERRE GOSSELIN** (Institut national de santé publique du Québec, INRS-ETE and CHUQ)

[Monday June 13 / lundi 13 juin, 16:00]

*Public Water, Public Health, Public Concerns*

*L'eau comme préoccupation publique, bien public, et facteur de santé publique*

A brief review of the current status of water sources for human use in Canada is presented, as well as known threats to their sustainable use and availability, including demographic pressures, climate change and economic greed. These themes will be illustrated by various topics such as meteorological extremes, average warming, shale gas fracking, oil sands, human behavior and urban sprawl. The role of scientific actors in the protection of this basic resource is examined and discussed in the context of current political priorities.

Cette conférence présente l'état de la situation concernant les approvisionnements en eau pour l'utilisation humaine au Canada, ainsi que les menaces quant à son utilisation et sa disponibilité sur une base durable. Ceci inclut notamment les pressions démographiques, les changements climatiques et la cupidité économique. Les sujets discutés inclueront les extrêmes météorologiques, le réchauffement moyen, la fracturation hydraulique des gas de schiste, les sables bitumineux, les comportements humains et l'étalement urbain. Le rôle des acteurs scientifiques dans la protection de cette ressource de base sera examiné et discuté dans le contexte des priorités politiques actuelles.

---

**SYLVIA ESTERBY** (University of British Columbia, Okanagan), **ABDEL EL-SHAARAWI** (National Water Research Institute and McMaster University)

[Monday June 13 / lundi 13 juin, 16:30]

*Assessing the Performance of Composite Water Quality Indices*

*Évaluation de la performance des indices composites de la qualité de l'eau*

Monitoring programs which measure and report on individual water quality variables are the usual model. Recently, accounting for the health of natural systems as well as economic performance has received attention and various indicators have been developed. The Canadian Water Quality Index, which measures the state of water quality referenced to use-specific objectives, is an example. Since the WQI is an aggregated index, it is of interest to compare for uniqueness index values from different scenarios of water quality variables. A

Le modèle traditionnel de programme de surveillance mesure et décrit des variables individuelles sur la qualité de l'eau. Récemment, en tenant compte de l'économie et de la santé des écosystèmes naturels, la performance a retenu l'attention et plusieurs indicateurs ont été développés. L'indice canadien de la qualité de l'eau, qui mesure l'état de la qualité de l'eau à des fins et usages spécifiques, est un exemple. Comme cet indice est agrégé, il est intéressant de comparer l'unicité des valeurs de l'indice dans différents scénarios de variables de la qualité de l'eau. Une mesure de la variabilité de l'indice est également nécessaire. Des don-

## 1F-12: Measuring Water Quality Comment mesurer la qualité des eaux

---

measure of variability of the index is also necessary. Such issues are investigated using data from several provinces.

nées de plusieurs provinces sont utilisées pour examiner de tels problèmes.

---

**WILLIAM ROSS** (Health Canada)

[Monday June 13 / lundi 13 juin, 17:00]

*Assessing Human Health Risks from the Viral Contamination of Shellfish*

*Évaluer les risques de la contamination virale des mollusques sur la santé humaine*

In recent years, viruses have emerged as important pathogens in food safety. A particular example is the possible introduction of norovirus and Hepatitis A virus into the food supply through the contamination of bivalve molluscan shellfish. One source of this contamination is the release of contaminated waste water near harvest areas. Water quality standards are a key component to managing this risk. The analysis of risk and the development of standards is a complex mixture of science, mathematical modelling and statistical analysis. This presentation provides an overview of these challenges.

Ces dernières années, les virus se sont révélés des pathogènes importants pour la sécurité sanitaire des aliments. Par exemple, notons une introduction possible du norovirus et du virus de l'hépatite A dans l'approvisionnement alimentaire par l'entremise de la contamination de mollusques bivalves. Cette contamination peut être le fruit du rejet d'eau contaminée à proximité d'aires de récolte. Des normes de qualité de l'eau sont donc une partie intégrale de la gestion de ce risque. L'analyse du risque et la création de ces normes allie science, modélisation mathématique et analyse statistique. Cette présentation donne une vue d'ensemble de ces défis.

## 1F-13: New Developments in (Disease and) Syndromic Surveillance Du neuf en surveillance (des maladies et) syndromique

---

Chair/Président: Patrick Brown (Cancer Care Ontario)  
Organizer/Responsable: Gauri Datta (University of Georgia)

**Room/Salle: BAC234**

**Abstracts/Résumés**

---

**ANDREW LAWSON & ANA CORBERAN** (Medical University of South Carolina)

[Monday June 13 / lundi 13 juin, 16:00]

*A Multivariate Bayesian Model-Based Approach for On-Line Spatio-Temporal Disease Surveillance*  
*Une approche bayésienne multivariée pour la surveillance spatio-temporelle de maladies en ligne*

We introduce the surveillance conditional predictive ordinate as a general Bayesian on-line surveillance technique that allows us to detect any small area of increased disease incidence. As surveillance systems are often focused on more than one disease within a predefined area, we extend this technique to the analysis of multiple diseases, which allows for inclusion of correlation between diseases and, consequently, increases the outbreak detection capability of the methodology.

Nous présentons la surveillance de l'ordonnée prédictive conditionnelle comme une technique générale de surveillance bayésienne en ligne qui nous permet de détecter toute petite région où l'incidence d'une maladie augmente. Étant donné que les systèmes de surveillance se concentrent souvent sur plus d'une maladie dans une zone prédéfinie, nous étendons cette technique à l'analyse de plusieurs maladies, ce qui permet l'inclusion de corrélation entre les maladies, et par conséquent, augmente la capacité de cette méthodologie à détecter les épidémies.

---

**RONG WEI** (National Center for Health Statistics), **JIRAPHAN SUNTORCHOST & MELONIE HERON** (National Center for Health Statistics, CDC, USA), **ERIC SLUD** (University of Maryland, USA), **MEENA KHARE** (National Center for Health Statistics, CDC, USA)

[Monday June 13 / lundi 13 juin, 16:30]

*Changes in Age-Specific Mortality Rates from Six Leading Causes of Death in US Populations, 1970-2007*  
*Changements des taux de mortalité par âges pour six causes principales de décès aux États-Unis, 1970-2007*

Based on US mortality data from 1970 to 2007, this study demonstrates changes in age-specific mortality rates from six leading causes of death: heart disease, cancer, stroke, accidents, chronic obstructive pulmonary disease, and diabetes. A segmented Lee-Carter model is applied to smooth the rates in two dimensions: age and year of death, so the change patterns in mortality rates are clearly demonstrated. Mortality rates by leading causes are computed for subpopulations by gender and state. The study on differences across subpopulations provides information for health surveillance for US populations.

Basée sur les données de mortalité aux États-Unis de 1970 à 2007, cette étude montre les changements des taux de mortalité par âges des six causes principales de mortalité : les maladies cardiaques, le cancer, les accidents vasculaires cérébraux, les accidents, les maladies pulmonaires obstructives chroniques, et le diabète. Un modèle segmenté de Lee-Carter est appliqué pour lisser les taux en deux dimensions : l'âge et l'année de décès, pour montrer clairement la structure des changements au taux de mortalité. Les taux de mortalité par causes principales sont calculés pour les sous-populations par sexe et état. L'étude sur les différences à travers les sous-populations fournit des informations pour la surveillance liée à la santé pour la population américaine.

## 1F-13: New Developments in (Disease and) Syndromic Surveillance Du neuf en surveillance (des maladies et) syndromique

---

**GAURI DATTA** (University of Georgia), **DAVID BANKS & MATT HEATON** (Duke University), **ALAN KARR** (National Institute of Statistical Sciences), **JAMES LYNCH** (University of South Carolina), **FRANCISCO VERA** (University of Ecuador), **FRANK ZOU** (National Institute of Statistical Sciences)

[Monday June 13 / lundi 13 juin, 17:00]

*Hierarchical Bayesian Modeling in Syndromic Surveillance*

*Modélisation bayésienne hiérarchique dans la surveillance syndromique*

Syndromic surveillance plays an important role in quick detection of emerging geographical clusters of disease to provide swift intervention to prevent a pandemic. While many statistical models have been devised for surveillance purposes, none are able to simultaneously achieve important practical goals such as good sensitivity and specificity, proper use of domain information, inclusion of spatio-temporal dynamics, and transparent support to decision-makers. To this end we propose a spatio-temporal conditional autoregressive hidden Markov model with an absorbing state. We use a hierarchical Bayesian Poisson model for the disease counts to assess posterior probabilities of an epidemic at the locations.

La surveillance syndromique joue un rôle important dans la prompte détection de l'émergence de cellules géographiques de maladie pour assurer une intervention rapide afin d'éviter une pandémie. Quoique plusieurs modèles statistiques aient déjà été conçus pour la surveillance, aucun n'est capable d'atteindre simultanément les buts importants et pratiques comme une bonne sensibilité et une bonne spécificité, l'usage correct de l'information de domaine, l'inclusion des dynamiques spatio-temporelles, et le soutien transparent aux décideurs. À cette fin nous proposons un modèle spatio-temporel conditionnel de Markov latent autorégressif avec un état absorbant. Nous utilisons un modèle hiérarchique bayésien de Poisson pour le décompte de maladies afin d'évaluer les probabilités a posteriori d'une épidémie aux différents endroits.

**1F-14: Recent Advances in the Analysis of Correlated Data**  
**Récents progrès en analyse de données corrélées**

---

Organizer and Chair / Responsable et président: Wei Liu (York University)

**Room/Salle: BAC239**

**Abstracts/Résumés**

---

**RICHARD COOK & LIQUN DIAO** (University of Waterloo)

[Monday June 13 / lundi 13 juin, 16:00]

*Joint Analysis of Multiple Markov Processes under Interval Censoring*

*Analyse conjointe de processus de Markov multiples en présence de censure par intervalles*

Degenerative chronic disease processes are often naturally modelled using multi-state Markov processes. With paired organ systems two processes may be observed to degenerate over time. We develop models which allow two progressive Markov processes to be modelled simultaneously by using copula functions to characterize the association in transition times. An appealing aspect of this model is that the marginal processes retain a simple Markov structure and covariate effects are therefore easily interpreted. We use this model to analyse two interval-censored processes pertaining to damage of the sacro-iliac joints in patients with arthritis. Maximum likelihood and two-stage estimation procedures are investigated.

Les processus de maladies chroniques dégénératives sont souvent modélisés naturellement à l'aide de processus de Markov multi-états. Pour des systèmes d'organes pairs, deux processus peuvent montrer une dégénération au fil du temps. Nous développons des modèles qui permettent de modéliser simultanément deux processus de Markov progressifs en utilisant des fonctions copules pour caractériser l'association aux temps de transition. L'un des aspects intéressants de ce modèle est le fait que les processus marginaux conservent une structure de Markov simple et les effets des covariables sont ainsi facilement interprétés. Nous utilisons ce modèle pour analyser deux processus censurés par intervalles relatifs aux dommages à l'articulation sacro-iliaque chez les patients atteints d'arthrite. Nous examinons les méthodes d'estimation à deux étapes et du maximum de vraisemblance.

---

**MARIA LORENZI** (British Columbia Cancer Centre), **XIAOGIONG JOAN HU** (Simon Fraser University)

[Monday June 13 / lundi 13 juin, 16:30]

*Zero-inflated Longitudinal Data Analysis*

*Analyse de données longitudinales à surreprésentation de zéros*

Motivated by a study of long-term health care utilization patterns of childhood cancer survivors, we consider analysis of zero-inflated longitudinal counts or measurements. A simulation study is conducted to examine misleadingness and efficiency loss caused by ignoring the features of zero-inflation and/or correlation in the data structure. We propose likelihood-based and robust inference methods by extending the loglinear-logistic modeling strategy for cross-sectional count data (e.g. Hall and Shen, 2010) to longitudinal data, accounting for the inherent correlation. The proposed approaches are studied via simulation, and illustrated

Motivés par une étude sur les motifs d'utilisation des soins de santé de longue durée pour les survivants du cancer infantile, nous considérons l'analyse de mesures ou de dénombrements longitudinaux à surreprésentation de zéros. Une étude de simulation est menée pour examiner les résultats trompeurs et la perte d'efficacité causés par l'ignorance des caractéristiques de la surreprésentation de zéros ou de la corrélation dans la structure des données. Nous proposons des méthodes basées sur la vraisemblance et l'inférence robuste en adaptant la stratégie de modélisation logistique log linéaire de données de dénombrement transversales (p. ex. Hall et Shen, 2010) à des données longitudinales, en tenant compte de la corrélation inhérente. Les

## 1F-14: Recent Advances in the Analysis of Correlated Data Récents progrès en analyse de données corrélées

---

with the motivating dataset.

approches proposées sont étudiées au moyen de simulation et illustrées à l'aide du jeu de données d'intérêt.

---

**GUOHUA YAN, RENJUN MA & M. TARIQUL HASAN** (University of New Brunswick)

[Monday June 13 / lundi 13 juin, 17:00]

*Analysis of Longitudinal Semi-continuous Data*

*Analyse de données longitudinales semi-continues*

Semi-continuous data such as daily precipitation consists of positive continuous measurements and a significant portion of exact zeros. We introduce a mixed model for longitudinal semicontinuous data, in which the semicontinuous response is modeled by a compound Poisson distribution. The temporal and/or hierarchical structure of data is accommodated by multilevel random effects. An optimal estimating function for the regression parameters is developed using the orthodox best linear unbiased predictors of random effects. In addition, we obtain consistent estimators for both the regression and dispersion parameters. Our approach is illustrated with application to analysis of real data.

Les données semi-continues comme les précipitations quotidiennes se composent de mesures positives continues et d'une proportion importante de valeurs précisément égales à zéro. Nous présentons un modèle mixte pour des données longitudinales semi-continues dont la réponse semi-continue est modélisée au moyen d'une distribution de Poisson composée. La structure temporelle ou hiérarchique des données est traitée par des effets aléatoires à plusieurs niveaux. La fonction estimatrice optimale pour les paramètres de régression est développée à partir des meilleurs prédicteurs linéaires conventionnels sans biais des effets aléatoires. De plus, nous obtenons des estimateurs convergents pour les paramètres de régression et de dispersion. Notre approche est illustrée à l'aide d'une application à l'analyse de données réelles.

## 1F-I5: Robust Designs of Statistical Experiments Conception robuste d'expériences statistiques

---

Organizer and Chair / Responsable et président: Xiaojian Xu (Brock University)

Room/Salle: BAC 236

### Abstracts/Résumés

---

**JULIE ZHOU & MICHAEL WILMUT** (University of Victoria)

[Monday June 13 / lundi 13 juin, 16:00]

*D-Optimal Minimax Design Criterion for Fractional Factorial Designs*

*Critère de conception minimax D-optimal pour les plans factoriels fractionnés*

A D-optimal minimax design criterion is proposed to construct two-level fractional factorial designs, which can be used for a linear model with main effects and some specified interactions. D-optimal minimax designs are robust against model misspecification and have small biases if the linear model contains more interaction terms. Lower and upper bounds are derived for the loss functions of optimal designs, and they can be used to estimate the efficiencies of any design. Numerical algorithms are also discussed.

Nous proposons un critère de conception minimax D-optimal pour construire des plans factoriels fractionnés à deux niveaux qui peuvent être utilisés pour un modèle linéaire avec des effets principaux et certaines interactions spécifiées. Les plans minimax D-optimaux sont robustes aux erreurs de spécification du modèle et présentent de légers biais si le modèle linéaire contient davantage de termes d'interaction. Nous dérivons les limites inférieures et supérieures des fonctions de perte des plans optimaux; elles peuvent être utilisées pour estimer l'efficacité d'un plan quelconque. Nous discutons aussi des algorithmes numériques.

---

**SANJOY SINHA** (Carleton University)

[Monday June 13 / lundi 13 juin, 16:30]

*Robust Designs for Multivariate Logistic Regression*

*Plans d'expérience robustes pour la régression logistique multivariée*

In this talk, I will discuss the choice of optimal designs for multivariate binary logistic regression models. As the calculation of an exact likelihood for multivariate binary data is often tedious, here I propose and explore a pseudo-likelihood approach for choosing the optimal designs. I will demonstrate in small simulations that the proposed design is robust against misspecified correlation structure among the binary response variables, and can provide estimators that are almost as efficient as those obtained from the D-optimal designs with correctly specified joint likelihoods for the multivariate responses. I will also discuss some examples using real life data.

Dans cette présentation, je discute du choix de plans d'expérience optimaux pour les modèles de régression logistique binaire multivariée. Le calcul d'une vraisemblance exacte pour des données binaires multivariées étant fastidieux, je propose et explore une approche basée sur la pseudo-vraisemblance permettant de choisir les plans optimaux. Je démontre par de petites simulations que le plan proposé est robuste aux erreurs de spécification de la structure de corrélation des variables réponses binaires, et qu'il peut fournir des estimateurs qui sont presque aussi efficaces que ceux obtenus à partir des plans D-optimaux avec des vraisemblances jointes correctement spécifiées pour les réponses multivariées. Je présente aussi quelques exemples basés sur de vraies données.



## 1F-I5: Robust Designs of Statistical Experiments Conception robuste d'expériences statistiques

---

**DOUG WIENS** (University of Alberta), **ALAN WELSH** (Australian National University)

[Monday June 13 / lundi 13 juin, 17:00]

*Robust Model-Based Sampling Designs*

*Plans d'échantillonnage robustes basés sur des modèles*

Consider a finite population of  $N$  units in which a survey variable  $Y$  is related to a  $q$  dimensional auxiliary variable  $x$ . The values of  $x$  are known for all  $N$  population units, and we are to select a sample of  $n \leq N$  population units and then observe the  $n$  corresponding values of  $Y$ . The objective is to predict the population total  $T = \sum_{i=1}^N Y_i$ . We are to construct an optimal design, minimizing the prediction MSE, in the face of uncertainty about the form of the regression model, and of the variance and correlation structures.

Considérons une population finie de  $N$  unités dans laquelle une variable d'enquête  $Y$  est liée à une variable auxiliaire  $x$  de dimension  $q$ . Les valeurs de  $x$  sont connues pour toutes les  $N$  unités de population et nous devons sélectionner un échantillon de  $n \leq N$  unités de population, puis observer les  $n$  valeurs correspondantes de  $Y$ . L'objectif est de prévoir le total de population  $T = \sum_{i=1}^N Y_i$ . Nous devons construire un plan optimal qui minimise l'erreur quadratique moyenne de prévision tout en faisant face à de l'incertitude à propos de la forme du modèle de régression, et au sujet des structures de variance et de corrélation.

# 1F-C1: CANNeCTIN Methodology for Clinical Trials Méthodologie du CANNeCTIN pour les essais cliniques

---

Chair/Président: Lehana Thabane (McMaster University)  
Organizer/Responsable: Robert Platt (McGill University)

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

**Room/Salle: BAC 141**

**Abstracts/Résumés**

---

**LEHANA THABANE**, McMaster University

[Monday June 13 / lundi 13 juin, 16:00]

---

**HAOCHENG LI & GRACE YI** (University of Waterloo)

[Monday June 13 / lundi 13 juin, 16:15]

*Variable Selection for Longitudinal Data via the Composite Likelihood Method*

*Sélection de variables pour données longitudinales par la méthode de vraisemblance composite*

Longitudinal data arise commonly in health research fields. An important topic in longitudinal studies is variable selection for which many methods have been developed in the literature. In particular, penalized likelihood approaches have been commonly explored. However, the use of these methods is usually hampered by computational difficulties that are often caused by the involvement of high dimensional integrals. In this talk, we discuss a variable selection method that is based on the composite likelihood formulation. This strategy shows modeling tractability and computational cheapness in model selection. Empirical studies will be presented to assess the performance of the proposed method.

Dans le domaine de la recherche en santé, il est courant d'utiliser des données longitudinales. Un sujet important dans les études longitudinales concerne la sélection de variables dont plusieurs méthodes ont été développées dans la littérature. En particulier, les approches basées sur la vraisemblance pénalisée ont été couramment explorées. Cependant, ces méthodes rencontrent habituellement des problèmes de puissance de calculs causés par l'implication des intégrales à haute dimension. Dans cet exposé, nous abordons une méthode de sélection de variables basée sur la formulation composite de la vraisemblance. Cette stratégie montre la traçabilité du modèle et une économie de puissance de calculs lors de la sélection du modèle. Des études empiriques seront présentées afin d'évaluer la performance de la méthode proposée.

---

**JINHUI MA, PARMINDER RAINA, JOSEPH BEYENE & LEHANA THABANE** (McMaster University)

[Monday June 13 / lundi 13 juin, 16:30]

*Simulation Study using a Multi-State Markov Model to Estimate the Statistical Power of Population Based Longitudinal Studies*

*Étude par simulation utilisant un modèle de Markov multi-états pour estimer la puissance statistique de population basée sur des études longitudinales*

We design a simulation study to investigate the statistical power profile of longitudinal studies which aim to evaluate the impact of risk exposure on

Une étude par simulation est utilisée pour étudier la puissance statistique des études longitudinales qui visent évaluer l'impact du risque d'exposition sur les transitions des

## 1F-C1: CANNeCTIN Methodology for Clinical Trials Méthodologie du CANNeCTIN pour les essais cliniques

---

health state transitions. The simulation study is based on the illness-death model, a commonly used three-state Markov model. To capture the increasing hazard with time for aging related illness, we assume the transition time from state to state follows Weibull distribution with shape parameter greater than one. We also assess the impact of the prevalence of risk exposures and illness, progression of the illness, misclassification of risk exposures and outcomes, and effect size etc. on statistical power.

états de santé. L'étude par simulation est basée sur le modèle maladie-décès, un modèle de Markov à trois-états couramment utilisé. Pour capturer l'augmentation du risque avec le temps pour des maladies liées à l'âge, nous supposons que le temps de transition d'un état à un autre suit une distribution Weibull avec un paramètre de forme supérieur à un. Nous évaluons aussi l'impact de la prévalence du risque d'exposition et de maladie, de la progression de la maladie, de la mauvaise classification du risque d'exposition et des variables réponses, de l'effet de taille etc. sur la puissance statistique.

---

**HOI SUEN WONG, AARON CHILDS & ELEANOR PULLENAYEGUM** (McMaster University)

[Monday June 13 / lundi 13 juin, 16:45]

*Analysis of Health Utility Data*

*Analyse de données d'utilité sur la santé*

Health utilities are the quality weight used to calculate quality-adjusted life years, the preferred outcome in cost-effectiveness analysis. Cost-effectiveness analysis of different programs can help governments to decide which programs are more worthwhile to implement. But health utility data tends to be non-normal with a cluster of observations at the maximum of the measurement scale. This raises questions as to whether linear relationships are appropriate for regression models. Generalized additive models (GAMs) are proposed as an alternative method to overcome the non-linearity concerns. The bias produced by the GAMs is generally smaller than the OLS.

Les utilités de santé sont les poids utilisés pour calculer le nombre d'années ajusté selon la qualité de vie, un résultat préférable pour l'analyse du rapport coût-efficacité. L'analyse du rapport coût-efficacité de différents programmes aide le gouvernement à décider quels programmes sont plus intéressants à implémenter. Mais les données d'utilité de santé tendent vers la non-normalité avec un groupe d'observations au maximum de l'échelle de mesures. Ceci soulève des questions quand au fait que les relations linéaires sont appropriées comme modèles de régression ou non. Les modèles additifs généralisés (GAM) sont proposés comme méthode alternative pour surmonter les inquiétudes sur la non-linéarité. Le biais produit par les GAM est généralement plus petit que celui des OLS.

---

**MARIA ESTHER PEREZ TREJO & ROBERT WILLIAM PLATT** (McGill University)

[Monday June 13 / lundi 13 juin, 17:00]

*Modeling Data from Cluster Randomized Trials in Presence of Measurement Error Induced by Double Clustering*

*Modélisation de données d'essais aléatoires en grappes en présence de mesures d'erreurs induites par double classement*

Cluster randomized trials have a potential disadvantage that cannot be addressed at the design level: when the measurement of the outcome is naturally clustered, and that clustering is occurring within the same groups used as units for randomization (double clustering). The variation due to cluster measurement cannot be separated from the variation caused by inherent clustering when a standard analysis via mixed-effects models is con-

Les essais aléatoires en grappes ont un désavantage potentiel qui ne peut être considéré au niveau du plan : quand la mesure du résultat est naturellement en grappes, et que le classement apparaît dans le même groupe utilisé comme unité de randomisation (double classement). La variation due à la mesure de grappes ne peut être séparée de la variation causée par le classement inhérent quand l'analyse par les modèles à effets mixtes est conduite. En alternative, une approche Bayésienne basée sur les MCMC, reposant

## 1F-C1: CANNeCTIN Methodology for Clinical Trials Méthodologie du CANNeCTIN pour les essais cliniques

---

ducted. Alternatively, a Bayesian MCMC-based approach, relying on external data is proposed. A simulation-based sensitivity analysis for the accuracy of estimated treatment effects is conducted under different scenarios of missingness/double-clustering severity.

sur les données externes, est proposée. Une analyse de sensibilité basée sur une simulation pour la fiabilité des effets de traitements estimés est menée sous différents scénarios de sévérité d'information manquante/double classement.

## 1F-C2: Distribution Theory and Asymptotics Théorie de la distribution et asymptotique

---

Chair/Président: Wei Sun (Concordia University)

Session sponsored by the Probability Section  
Session parrainée par le Groupe de probabilité

**Room/Salle: BAC138**

**Abstracts/Résumés**

---

**LILIA LETICIA RAMÍREZ RAMÍREZ, MARY E. THOMPSON & YULIA GEL** (University of Waterloo)

[Monday June 13 / lundi 13 juin, 16:00]

*On the Estimation of the Degree Distribution of Random Networks*

*Estimation de la distribution de degrés des réseaux aléatoires*

Nowadays networks have become a key tool for modelling relationships of interest in various disciplines, from social science to computer engineering. Random graph models of networks allow representation and prediction of transmission of information, spread of infections, and resilience to interruption. Many employed techniques are restricted to parameter estimation within a family/group of degree distributions and treat all available information as reliable and complete, which is often unrealistic. Our work-in-progress aims to show how various resampling procedures can be employed in network inference, with a target of estimating general characteristics of degree distribution and assessing related uncertainty under limited information.

De nos jours, les réseaux sont devenus une clé pour la modélisation des relations d'intérêt dans plusieurs disciplines, allant des sciences sociales à la génie informatique. Les modèles graphiques aléatoires des réseaux permettent la représentation et la prédiction de la transmission de l'information, de la propagation d'infections et de la résilience à l'interruption. Plusieurs techniques employées sont restreintes à l'estimation de paramètres d'une famille/groupe des distributions de degrés et traitent toute l'information disponible comme fiable et complète, ce qui est fréquemment irréaliste. Notre travail actuel tente de montrer comment plusieurs procédures de ré-échantillonnage peuvent être employées dans l'inférence des réseaux, avec pour but d'estimer les caractéristiques générales des distributions de degrés et d'évaluer l'incertitude connexe sous une information limitée.

---

**A. AKBAR MOHSENIPOUR & SERGE PROVOST** (The University of Western Ontario)

[Monday June 13 / lundi 13 juin, 16:15]

*Approximating the Distributions of Singular Quadratic Expressions and their Ratios*

*Approximer les distributions des expressions quadratiques singulières et de leurs ratios*

Non-central indefinite quadratic expressions in possibly non-singular normal vectors are represented in terms of the difference of two positive definite quadratic forms and an independently distributed linear combination of standard normal random variables. This result also applies to quadratic forms in singular normal vectors for which no general representation is currently available. General ratios of quadratic forms and an approximation to the density of ratios of quadratic expressions in

Des formes quadratiques indéfinies décentrées pour des vecteurs de loi normale avec matrice de covariance possiblement non-singulière sont représentés par la différence de deux formes quadratiques définies positives et une combinaison linéaire distribuée de façon indépendante de variables aléatoires de loi normale centrée-réduite. Ce résultat peut également être appliqué aux formes quadratiques des vecteurs normaux singuliers pour lesquelles aucune représentation générale n'est présentement disponible. Des représentations et une approximation de la densité du ratio

## 1F-C2: Distribution Theory and Asymptotics Théorie de la distribution et asymptotique

---

singular normal vectors are being proposed. The results are applied to the Durbin-Watson statistic and Burg's estimator, both of which are expressible as ratios of quadratic forms.

de deux formes quadratiques avec vecteurs de loi normale sont proposés. Les résultats sont appliqués à la statistique de Durbin-Watson et à l'estimateur de Burg, lesquels s'expriment comme des ratios de formes quadratiques.

---

**JOHN SHERIFF** (University of Lethbridge)

[Monday June 13 / lundi 13 juin, 16:30]

*Central Limit Theorems for Additive Functionals of the Ginzburg-Landau Process*

*Théorèmes limite central pour des fonctionnelles additives du processus de Ginzburg-Landau*

We consider the Ginzburg-Landau process, and study additive functionals  $\int_0^t W(\eta_s) ds$  of this process. We obtain a central limit theorem for  $a^{-1}(t) \int_0^t W(\eta_s) ds$  with  $a(t) = \sqrt{t}$ ,  $d \geq 3$ ;  $\sqrt{t \log t}$ ,  $d = 2$ ;  $t^{3/4}$ ,  $d = 1$ . Standard arguments address the case  $d \geq 3$ . Martingale methods together with  $L^2$  decay estimates for the semigroup associated with the process are employed to establish the result in the cases  $d = 1$  and  $d = 2$ . This extends similar results for noninteracting random walks, the symmetric simple exclusion processes, and the zero-range process.

Nous considérons le processus de Ginzburg-Landau et étudions des fonctionnelles additives  $\int_0^t W(\eta_s) ds$  de ce processus. Nous obtenons un théorème limite central pour  $a^{-1}(t) \int_0^t W(\eta_s) ds$  avec  $a(t) = \sqrt{t}$ ,  $d \geq 3$ ;  $\sqrt{t \log t}$ ,  $d = 2$ ;  $t^{3/4}$ ,  $d = 1$ . Les arguments standards traitent le cas  $d \geq 3$ . Des propriétés de martingales avec des estimations décroissantes  $L^2$  pour le semi-groupe associé au processus sont utilisées pour obtenir le résultat dans les cas  $d = 1$  et  $d = 2$ . Des résultats similaires sont obtenus pour des marches aléatoires non-interactives, des processus d'exclusion symétrique simples et des processus dits «zero-range».

---

**WEIBIN JIANG, HAO YU & REG KULPERGER** (The University of Western Ontario)

[Monday June 13 / lundi 13 juin, 16:45]

*Asymptotic Theory for General Multivariate GARCH Models*

*Théorie asymptotique pour des modèles multivariés généraux GARCH*

This paper studies the asymptotic theory of the quasi-maximum likelihood estimator (QMLE) for multivariate generalized autoregressive conditional heteroscedasticity (GARCH) model parameters under mild conditions. We give some counterexamples for the parameter identifiability result in Jeantheau (1998) and provide a better necessary and sufficient condition. We prove the ergodicity of the conditional variance process on an application of theorems by Meyn and Tweedie (2009). Under those conditions, the consistency and asymptotic normality of the QMLE can be proved by the standard compactness argument and Taylor expansion of the logarithm of the score function.

Cet article étudie la théorie asymptotique de l'estimateur de quasi-maximum de vraisemblance (QMV) pour estimer les paramètres des modèles GARCH multivariés généralisés autorégressifs conditionnels hétéroscédastiques sous des conditions faibles. Nous donnons quelques contre-exemples du résultat d'unicité de paramétrisation de Jeantheau (1998) en plus d'une meilleure condition suffisante et nécessaire. Nous prouvons l'ergodicité du processus de variance conditionnelle en appliquant les théorèmes de Meyn et Tweedie (2009). Sous ces conditions, la convergence et la normalité asymptotique du QMV peuvent être prouvées au moyen d'un argument de compacité standard et d'un développement en série de Taylor du logarithme de la fonction score.

## 1F-C2: Distribution Theory and Asymptotics Théorie de la distribution et asymptotique

---

---

**DAVID BRILLINGER** (University of California, Berkeley)

[Monday June 13 / lundi 13 juin, 17:00]

*Background and Applications of Particle Process Systems*

*Contexte et applications des systèmes de processus des particules*

A particle process system will be defined generally and some properties indicated. Then three particular cases will be discussed and applied to animal movement data. Extensions and Discussion will be provided.

Un système de processus des particules sera défini de manière générale et quelques propriétés seront présentées. Par la suite, trois cas particuliers feront l'objet d'un examen et seront appliqués à des données de mouvements d'animaux. Enfin, des extensions et une discussion seront abordées.

## 1F-C3: Survey Methods: Precision and Analysis Méthodes d'enquête : précision et analyse

---

Chair/Président: David Haziza (Université de Montréal)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

**Room/Salle: BAC142**

**Abstracts/Résumés**

---

**ALASTAIR SCOTT** (The University of Auckland), **THOMAS LUMLEY** (University of Auckland)

[Monday June 13 / lundi 13 juin, 16:00]

*Pseudo-Likelihood Methods for Survey Data*

*Modèles de quasi-vraisemblance pour données d'enquête*

With the advent of large-scale health surveys such as NHANES, methods for the analysis of survey data have become increasingly important - PubMed lists almost 8000 papers published in the last 5 years with NHANES in the abstract, for example. We review the work of Rao and Scott (1981, 1984) on likelihood-ratio tests for contingency tables and show that the key results are valid in a much wider setting. We illustrate with applications to GLMs and to the Cox model.

Avec l'avènement d'enquêtes à grandes échelles sur la santé, telle que NHANES, les méthodes pour analyser les données d'enquête sont devenues de plus en plus importantes - PubMed répertorie près de 8 000 articles publiés au cours des 5 dernières années contenant par exemple, le terme NHANES dans le résumé. Nous revoyons le travail de Rao et Scott (1981, 1984) sur les tests du rapport des vraisemblances pour des tableaux de contingence et montrons que les principaux résultats sont valides dans un cadre beaucoup plus large. Nous illustrons ceci par l'entremise d'applications aux MLG et au modèle de Cox.

---

**ANNE-SOPHIE CHAREST** (Carnegie Mellon University)

[Monday June 13 / lundi 13 juin, 16:15]

*Inference from Differentially-Private Synthetic Datasets*

*Inférence avec des jeux de données satisfaisant le critère de confidentialité différentielle*

To respect their confidentiality agreement, statistical agencies sometimes generate synthetic datasets for publication. Although this heuristically provides privacy protection, the method has no formal such guarantee. We thus consider the generation of synthetic datasets which satisfy differential privacy, a rigorous criterion limiting what an intruder can learn about an individual from the released data. We show that the combining rules used with multiply-imputed synthetic datasets are not appropriate for differentially-private datasets and propose an alternate method to obtain valid inferences. We also discuss the loss in statistical efficiency incurred as an effect of the added privacy guarantee.

Certaines agences statistiques publient des données synthétiques afin de respecter leur engagement de confidentialité. Bien que cette approche semble appropriée, il n'existe aucune preuve formelle de ses garanties. Nous considérons la génération de données synthétiques satisfaisant la confidentialité différentielle, un critère rigoureux limitant l'information contenue sur un individu dans les données publiées. Nous montrons que les règles habituellement utilisées pour analyser les données synthétiques sont invalides si les données créées satisfont le critère de confidentialité différentielle et proposons une nouvelle méthode d'inférence. Nous abordons aussi l'effet de la confidentialité différentielle sur la qualité de l'inférence.



## 1F-C3: Survey Methods: Precision and Analysis Méthodes d'enquête : précision et analyse

---

**MARTIN KLEIN** (U.S. Census Bureau), **TOMMY WRIGHT**

[Monday June 13 / lundi 13 juin, 16:30]

*Ranking Procedures for Several Normal Populations: An Empirical Investigation*

*Procédures de classement de plusieurs population normales : une étude empirique*

Statistical agencies commonly release ranking tables in which a collection of entities such as states are ranked based on an estimate of some quantity of interest. Taking such ranking tables as our motivation, this paper presents several methods for obtaining a complete ranking of several normal populations with respect to the unknown means. Many of the methods are derived from a Bayesian perspective. We explore computational aspects of each ranking method, and use simulation to compare the various methods under several scenarios. The procedures are then applied to a dataset from the American Community Survey of the U.S. Census Bureau.

Les agences de statistique publient souvent des tables dans lesquelles des entités telles des états sont classés selon une estimation d'une quantité d'intérêt. En utilisant ces tables de classement comme motif, cet article présente plusieurs méthodes permettant d'obtenir un classement complet de plusieurs distributions normales par rapport aux moyennes qui sont inconnues. Plusieurs de ces méthodes proviennent d'une perspective bayésienne. Nous étudions les aspects computationnels de chaque méthode de classement et comparons les différentes méthodes, selon plusieurs scénarios, à l'aide de simulations. Les procédures sont par la suite appliquées à un jeu de données de l'American Community Survey du U.S. Census Bureau.

**JOANNE CHARLEBOIS & GUYLAINE DUBREUIL** (Statistics Canada)

[Monday June 13 / lundi 13 juin, 16:45]

*Variance Estimation for the Redesigned Survey of Household Spending*

*Estimation de la variance pour l'Enquête sur les dépenses des ménages remaniée*

A major redesign of the Survey of Household Spending conducted annually by Statistics Canada has been undertaken over the last few years. A new continuous data collection model combining the use of a diary and a recall interview was developed. This presentation will give an overview of the variance estimation method used, which is based on the Rao-Wu bootstrap for stratified multi-stage sample designs. The changes brought about by the redesign and their effects on the estimated coefficients of variation of survey variables, which may now have components from both the interview and diary sources, will be highlighted.

L'Enquête sur les dépenses des ménages, menée chaque année par Statistique Canada, a récemment fait l'objet d'un remaniement majeur. Un nouveau modèle de collection continue utilisant un journal ainsi qu'une entrevue rétrospective a été développé. Cette présentation sera axée sur la méthode d'estimation de la variance utilisée pour l'enquête, méthode qui est basée sur le bootstrap de Rao-Wu pour les sondages stratifiés à degrés multiples. Les changements entraînés par le remaniement seront mis en évidence ainsi que leurs effets sur les coefficients de variation estimées de variables d'enquête qui peuvent maintenant comporter des composantes de l'entrevue et du journal.

**ZEINAB MASHREGHI, CHRISTIAN LÉGER & DAVID HAZIZA** (Université de Montréal)

[Monday June 13 / lundi 13 juin, 17:00]

*Bootstrap Variance Estimation for Imputed Data from Regression, Ratio and Hot Deck Imputation Methods*

*Sur l'estimateur bootstrap de la variance pour des données imputées par la régression, le ratio ou le hot-deck*

Estimating consistently the variance of a statistic based on imputed data is a difficult task, especially

L'estimation de la variance d'une statistique basée sur les données imputées est une tâche difficile, particulièrement

## 1F-C3: Survey Methods: Precision and Analysis Méthodes d'enquête : précision et analyse

---

when imputing from the commonly used methods of the ratio, regression or random hot deck. Using the Rao-Wu (1988) bootstrap method and the response rate, we obtain a valid variance estimator, even for large sampling fraction, by computing an imputation method based rescaling factor  $C$ , resampling the data and the response status independently, and reimputing the resample using the same method. If the response status of each case is available, applying the Shao-Sitter (1996) method with a different  $C$  leads to another valid estimator.

lors de l'imputation par le ratio, la régression ou le hot-deck aléatoire. En utilisant la méthode bootstrap de Rao-Wu (1988) et le taux de réponse, nous obtenons un estimateur valide même pour une grande fraction de sondage en prenant un ajustement  $C$  dépendant de la méthode d'imputation, en rééchantillonnant les données indépendamment des indicateurs de réponse, et en réimputant le rééchantillon. En présence des statuts de réponse, une modification de  $C$  mène à un autre estimateur valide via la méthode de Shao-Sitter (1996).

---

**WEI QIAN & ABDELLATIF DEMNATI** (Statistics Canada)

[Monday June 13 / lundi 13 juin, 17:15]

*Variance Estimation for Survey Data with Multiple Cycle Imputation*

*Estimation de la variance pour les données d'enquête ayant des cycles d'imputation multiples*

Imputation is a common method for handling item non response in surveys. Two or more imputation methods are often used, for example, direct substitution when values of variable from external files are available; otherwise, multiple cycle ratio imputation is implemented based on different auxiliary variables. In each cycle, the preceding imputed values are treated as observed ones. In this study, we apply a unified linearization approach to obtain the variance estimator for population total estimator under the above set-up. Simulation results are presented to show the performance of the variance estimator in tracking the MSE.

L'imputation est une méthode commune pour le traitement de la non-réponse partielle dans les enquêtes. Deux méthodes ou plus d'imputation sont souvent utilisées. Par exemple, la substitution directe lorsque les valeurs des variables provenant de fichiers externes sont disponibles. Une autre solution consiste à faire l'imputation par le ratio à cycles multiples en se fondant sur différentes variables auxiliaires. Dans chaque cycle, les variables imputées qui précèdent sont traitées comme des variables observées. Pour cette étude, nous utilisons une approche de linéarisation unifiée pour obtenir l'estimateur de la variance pour l'estimateur de la population totale selon le système ci-dessus.

**2A-A: Gold Medal Address**  
**Allocution de la récipiendaire de la médaille d'or**

---

Chair/Président: Christian Genest (McGill University)

**Room/Salle: Festival Theatre**

**Abstract/Résumé**

---

**LOUIS-PAUL RIVEST** (Université Laval)

[Tuesday June 14 / mardi 14 juin, 8:30]

*Looking for Directions: from Directional Statistics to the Analysis of Object Data*

*A la recherche de directions : de la statistique directionnelle à l'analyse de données-objets*

This talk is about statistical problems for which the data points belong to some mildly non Euclidean manifold such as  $S^2$ , the 3D unit sphere, and  $SO(3)$  and  $SE(3)$ , the Lie groups of  $3 \times 3$  rotations and of rigid body motions in 3D. Scientific problems in geophysics and in biomechanics that have motivated the development of statistical methods for this type of data over the last 30 years will be reviewed. A unified presentation for the statistical methodology will be emphasized. It involves the embedding of the manifold in a Euclidean vector space  $R^k$ , for some  $k$ , and the application of the Euclidean distance, or of a weighted Euclidean distance, to create statistical models and estimation criteria for the manifold data. The statistical properties of these methods will be shown to depend on the properties of the Euclidean distance in the embedded manifold.

Cet exposé porte sur des problèmes statistiques où les données appartiennent à une variété faiblement non euclidienne telle que  $S^2$ , la sphère unité en 3D, et  $SO(3)$  et  $SE(3)$ , les groupes de Lie des rotations et des mouvements de corps solides en 3D. Des problèmes scientifiques en géophysique et en biomécanique qui ont motivé le développement de méthodes statistiques pour ce type de données au cours des 30 dernières années seront rappelés. On mettra l'accent sur une présentation unifiée de la méthodologie statistique. Elle implique le plongement de la variété dans un espace vectoriel euclidien  $R^k$ , pour un certain  $k$ , et l'utilisation de la distance euclidienne, ou d'une distance euclidienne pondérée, pour créer des modèles statistiques et des critères d'estimation pour les données de variété. On montrera que les propriétés statistiques de ces méthodes dépendent des propriétés de la distance euclidienne dans la variété.

## 2B-11: Recent Advances in MCMC Récents progrès en MCMC

---

Chair/Président: Jeff Rosenthal (University of Toronto)  
Organizer/Responsable: Radu Craiu (University of Toronto)

Session sponsored by the Probability Section  
Session parrainée par le Groupe de probabilité

**Room/Salle: BAC138**

**Abstracts/Résumés**

---

**GALIN JONES** (University of Minnesota), **KUO-JUNG LEE** (University of Rochester), **BRIAN CAFFO & SUSAN SPEAR BASSETT** (Johns Hopkins University)

[Tuesday June 14 / mardi 14 juin, 10:15]

*Spatial Bayesian Variable Selection Models on fMRI Time-Series Data*

*Modèles bayésiens de sélection des variables spatiales sur des séries chronologiques d'imagerie par résonance magnétique fonctionnelle (IRMf)*

A major objective of fMRI studies is determining areas of the brain which are activated in response to a stimulus or task. The Bayesian paradigm provides an attractive framework for making inference using complex models and bypassing the multiple comparison problem that encumbers standard frequentist approaches. We propose a Bayesian model which not only takes into account the spatio-temporal relationships in the data while still being computationally feasible, but gives a framework for addressing other interesting questions related to how the human brain works. We study the properties of this approach and demonstrate its performance on simulated and real examples.

Un des objectifs principaux des études d'IRMf consiste à déterminer les zones du cerveau qui sont activées en réponse à un stimulus ou une tâche. La perspective bayésienne offre un cadre propice pour faire des inférences avec des modèles complexes et permet d'outrepasser de nombreux problèmes de comparaison multiples avec les approches fréquentistes standards. Nous proposons un modèle bayésien qui ne tient pas seulement compte de la relation spatio-temporelle des données tout en restant calculable, mais qui offre un cadre pour soulever d'autres questions intéressantes qui concernent la façon dont fonctionne le cerveau humain. Nous étudions les propriétés de cette approche et démontrons son efficacité avec des exemples simulés et réels.

---

**DAVID STEPHENS** (McGill University)

[Tuesday June 14 / mardi 14 juin, 10:45]

*Particle Methods in Markov Chain Monte Carlo*

*Méthodes à particules dans la chaîne de Markov Monte-Carlo*

Recent results of Andrieu, Doucet et al. have established that using particle approximations to terms necessary for the implementation of the Metropolis-Hastings algorithm does not disrupt the convergence of the algorithm to the target posterior distribution. Such methods are especially useful in latent or state-space models, which are typically of high dimension. In this talk I will give a review of key results in Particle MCMC, give illustrations

Les résultats récents d'Andrieu, Doucet et coll. ont permis d'établir que l'utilisation d'approximations à particules nécessaires pour l'implémentation de l'algorithme de Metropolis-Hastings ne freine pas la convergence de l'algorithme à la distribution cible a posteriori. Ces méthodes sont particulièrement utiles dans les modèles latents ou d'espace-état, qui sont normalement de haute dimension. Dans cet exposé, je vais revoir les résultats clés de la chaîne de Markov Monte-Carlo (MCMC), je donnerai une illustration à partir de

## 2B-11: Recent Advances in MCMC Récents progrès en MCMC

---

from financial time series modelling, and discuss practical advantages and limitations.

modèles financiers de séries chronologiques, et je discuterai des avantages et des limitations pratiques.

---

**DAWN WOODARD** (Cornell University), **MOISES GOLDSZMIDT** (Microsoft Research)

[Tuesday June 14 / mardi 14 juin, 11:15]

*Online Inference for Model-Based Clustering in Distributed Computing*

*Inférence en ligne pour classification basée sur des modèles en calcul réparti*

We develop a method for online inference in Dirichlet process mixture models, motivated by management of distributed computing systems. These systems can suffer from occasional violation of performance goals ("crises"); rapid automatic recognition of a problem can aid decision-making. We frame this as an online clustering problem, and give a solution using a Dirichlet process mixture model. In the periods between crises we perform inference using a Markov chain, and as a new crisis occurs we apply fast Bayesian updating. We make the Markov chain on clusters as efficient as possible by combining parallel tempering with a collapsed-space split-merge method.

Nous élaborons une méthode d'inférence en ligne pour les modèles de mélanges de processus de Dirichlet, motivés par la gestion des systèmes de calcul réparti. Les objectifs de rendement de ces derniers peuvent occasionnellement ne pas être atteints (on parle de « crise » dans ce cas), et la reconnaissance automatique rapide d'un problème peut aider à prendre des décisions. Nous considérons cela comme un problème de classification en ligne, et donnons une solution en utilisant un modèle de mélanges du processus de Dirichlet. Entre les crises, nous faisons des inférences en utilisant une chaîne de Markov, et lorsqu'une nouvelle crise a lieu, nous appliquons les mises à jour bayésiennes rapides. Nous rendons la chaîne de Markov sur des grappes aussi efficace que possible en combinant la méthode d'atténuation parallèle avec une méthode de division et de jonction de l'espace écoulé.

## 2B-12: Recent Developments in Multiple Comparison Procedures Du neuf dans les procédures de comparaisons multiples

---

Organizer and Chair / Responsable et président: Jianan Peng (Acadia University)

Room/Salle: BAC 237

### Abstracts/Résumés

---

**JASON HSU** (Ohio State University), **SZU-YU TANG** (The Ohio State University), **YI LIU** (Millennium)

[Tuesday June 14 / mardi 14 juin, 10:15]

*Statistical Principles and Methods in Quantitative Biomarker Usage*

*Principes et méthodes statistiques pour l'utilisation de biomarqueurs quantitatifs*

Personalized medicine involves a drug, and a companion biomarker test identifying the patient subgroup that the drug targets. This presentation is toward marketable Personalized Medicine. Biomarkers can be disease prognostic, predictive of efficacy, or predictive of safety. We first suggest designs for biomarker studies, applying the principles of randomization, replication, control, and blocking. The problem of simultaneous testing for efficacy in the population and in genomic subgroups is then formulated in terms of Partition Testing. Error rate control issues, subtle assumption implicit in permutation testing, as well as advantage of the strong partition principle, will be discussed.

La médecine personnalisée comprend un médicament et un test avec un biomarqueur associé identifiant le sous-groupe du patient ciblé par le médicament. Cette présentation traite de la médecine personnalisée commercialisable. Il existe des biomarqueurs pronostiques de la maladie, prédictifs de l'efficacité ou prédictifs de la toxicité. Nous proposons d'abord des plans pour des études de biomarqueurs en appliquant les principes de randomisation, de réplication, de contrôle et de blocage. Ainsi, le problème de tester l'efficacité simultanément dans la population et dans les sous-groupes génomiques est formulé en termes de tests de partition. Nous abordons les questions de contrôle du taux d'erreur, les hypothèses subtiles implicites aux tests de permutation, et les avantages du principe de partition forte.

---

**WEI LIU** (University of Southampton, UK)

[Tuesday June 14 / mardi 14 juin, 10:45]

*Simultaneous Confidence Bands for Regression Analysis*

*Bandes de confiance simultanées pour l'analyse de régression*

This talk will provide an overview of the methodology of simultaneous confidence bands for parametric regression analysis. We will look at how simultaneous confidence bands can be used in regression analysis to make sensible and informative inference, the key in the construction of simultaneous confidence bands, the two optimality criteria for comparison of simultaneous confidence bands, and some unsolved problems. The talk should be accessible to anyone who has done linear regression models.

Cette présentation propose un aperçu de la méthodologie des bandes de confiance simultanées pour l'analyse de régression paramétrique. Nous regardons la façon d'utiliser les bandes de confiance simultanées pour l'analyse de régression afin de produire une inférence sensible et informative, ainsi que l'élément clé dans la construction de ces bandes, les deux critères d'optimalité pour leur comparaison et quelques problèmes non résolus. La présentation s'adresse à quiconque a déjà étudié les modèles de régression linéaire.

## 2B-I2: Recent Developments in Multiple Comparison Procedures Du neuf dans les procédures de comparaisons multiples

---

**WENGE GUO** (New Jersey Institute of Technology), **SANAT SARKAR** (Temple University)

[Tuesday June 14 / mardi 14 juin, 11:15]

*Adaptive FWER and FDR Control under Block Dependence*

*Contrôle adapté de l'erreur de type global (ETG) et du taux de fausses découvertes (TFD) sous la dépendance en blocs*

Recently, a number of adaptive multiple testing procedures have been proposed. However, in a non-asymptotic setting, the FWER or FDR control of these adaptive procedures is proved under independence, although simulation studies have suggested that these procedures perform well under certain dependence. In this talk, several variants of the conventional adaptive Bonferroni and BH methods will be presented, along with proofs that these procedures provide ultimate control of the FWER or FDR under block dependence. Results of simulation studies comparing the performances of these adaptive procedures with the conventional FWER and FDR controlling procedures under dependence will also be presented.

Récemment, plusieurs procédures adaptées de tests multiples ont été proposées. Cependant, dans des conditions non asymptotiques, le contrôle de l'ETG ou du TFD de ces procédures adaptées a été prouvé sous l'indépendance, même si des études de simulation ont montré que ces procédures performant bien sous une certaine dépendance. Dans cette présentation, nous présentons plusieurs variantes des méthodes conventionnelles adaptées de Bonferroni et de Benjamini et Hochberg, ainsi que des preuves que ces procédures procurent le meilleur contrôle de l'ETG et du TFD sous la dépendance en blocs. Nous présentons aussi des résultats d'études de simulation comparant la performance de ces procédures adaptées aux procédures conventionnelles de contrôle de l'ETG et du TFD sous la dépendance.

**2B-13: Recent Developments in Shrinkage Estimation with Related Implications for Predictive Density Estimation, Loss Estimation and Model Selection**  
**Du neuf dans l'estimation par rétrécissement et implications quant à l'estimation de la densité prédictive, l'estimation des pertes et la sélection de modèles**

Organizer and Chair / Responsable et président: Eric Marchand (Université de Sherbrooke)

**Room/Salle: BAC 239**

**Abstracts/Résumés**

---

**BILL STRAWDERMAN** (Rutgers University), **JOHN KENT & REMAN ABU-SHANAB** (University of Leeds)

[Tuesday June 14 / mardi 14 juin, 10:15]

*Shrinkage Estimation for Matrix Loss*

*Estimation par rétrécissement avec perte matricielle*

We study shrinkage estimation in location problems with matrix loss, partially ordered by positive semi-definiteness of matrix differences. The basic problem is to estimate the  $n \times p$  mean matrix of a random  $n \times p$  matrix of independent normal variables with a  $p \times p$  loss matrix, whose  $(i,j)$  entry is the inner product of the errors vectors of the  $i$ th and  $j$ th column respectively. We consider improved Stein-type shrinkage estimators which shrink each column individually. We show that the allowable shrinkage factors are reduced by a factor of  $1/p$  over the individual  $n \times 1$  vector problems.

Nous étudions l'estimation par rétrécissement dans les problèmes de localisation avec perte matricielle, ordonnée partiellement par le degré de semi-définition positive des différences de la matrice. Le problème de base consiste à estimer la matrice  $n \times p$  des moyennes d'une matrice aléatoire  $n \times p$  de variables normales indépendantes à l'aide d'une matrice de pertes  $p \times p$  dont l'élément  $(i,j)$  est le produit scalaire des vecteurs d'erreurs des colonnes  $i$  et  $j$  respectivement. Nous explorons des estimateurs améliorés par un rétrécissement à la Stein qui permettent de rétrécir chaque colonne individuellement. Nous montrons que les facteurs de rétrécissement possibles sont réduits d'un facteur de  $1/p$  par rapport aux problèmes des vecteurs individuels  $n \times 1$ .

---

**DOMINIQUE FOURDRINIER & AURÉLIE BOISBUNON** (Université de Rouen), **STÉPHANE CANU** (INSA de Rouen), **WILLIAM STRAWDERMAN** (Rutgers University), **MARTY WELLS** (Cornell University)

[Tuesday June 14 / mardi 14 juin, 10:45]

*Variable Selection: a Decision Theory Approach*

*Sélection de variable : une approche décisionnelle*

We propose a new procedure for variable selection based on loss estimation. One big advantage of this procedure is that it can be applied to a distributional context more general than the usual Gaussian model: spherically symmetric distributions. We apply here our procedure to the lasso estimator and we give simulation results for various examples of the error distribution. We compare our results to AIC and BIC criteria. AMS 2010 subject classification: 62C20, 62F07, 62F10, 62H12, 62J05, 62J07. Keywords: variable selection, linear model, loss estimation, spherically symmetric distribution.

Nous proposons une nouvelle procédure de sélection de variable fondée sur l'estimation de coût. Un grand avantage de cette procédure est qu'elle peut être mise en oeuvre dans un contexte plus général que le cadre usuel gaussien : celui des lois à symétrie sphérique. Nous appliquons notre procédure à l'estimateur du lasso et nous donnons des résultats de simulation pour divers exemples de lois d'erreur. Nous comparons nos résultats aux critères AIC et BIC. Classification soumise de mathématiques 2010 : 62C20, 62F07, 62F10, 62H12, 62J05, 62J07. Mots clefs : sélection de variable, modèle linéaire, estimation de coût, loi à symétrie sphérique.



**2B-I3: Recent Developments in Shrinkage Estimation with Related Implications for Predictive Density Estimation, Loss Estimation and Model Selection**  
**Du neuf dans l'estimation par rétrécissement et implications quant à l'estimation de la densité prédictive, l'estimation des pertes et la sélection de modèles**

---

**ANDREW L. RUKHIN** (National Institute of Standards and Technology)

[Tuesday June 14 / mardi 14 juin, 11:15]

*Reducing Data Nonconformity in Linear Models*

*Réduction de la non-conformité des données dans les modèles linéaires*

Procedures to reduce nonconformity in interlaboratory studies by shrinking multivariate data toward a consensus mean are discussed. Some of them are shown to have a smaller quadratic risk than the ordinary least squares rule. Bayes procedures and shrinkage estimators in random effects models are also considered. The results are illustrated by an example of collaborative studies.

Des procédures pour réduire la non-conformité dans les études inter-laboratoires en réduisant les données multivariées vers une moyenne consensuelle sont discutées. Certaines d'entre elles semblent avoir un risque quadratique plus petit que la règle usuelle des moindres carrés. Les procédures bayésiennes et les estimateurs par rétrécissement pour les modèles à effets aléatoires sont aussi considérés. Les résultats sont illustrés par un exemple d'études collaboratives.

## 2B-I4: Sampling Theory and Practice: In Honour of Mary Thompson's Retirement Théorie et pratique de l'échantillonnage : À l'occasion du départ à la retraite de Mary Thompson

---

Organizer and Chair / Responsable et président: Changbao Wu (University of Waterloo)

**Room/Salle: BAC132**

**Abstracts/Résumés**

---

**JNK RAO** (Carleton University), **CHANGBAO WU** (University of Waterloo)

[Tuesday June 14 / mardi 14 juin, 10:15]

*Some Applications of Estimating Equations Approach to Inference in Complex Survey Data*

*Quelques applications de l'approche d'inférence par équations estimatrices aux données provenant de plans de sondage complexes*

Mary Thompson has done pioneering work, jointly with Godambe, on using survey weighted estimating equations to make inference from complex survey data, and also established some optimality results. In this talk, I will present some recent applications of the weighted estimating equations approach and the use of re-sampling methods for variance estimation and hypothesis testing. In particular, I will give results on using post-stratification information in constructing point estimators, variance estimators and quasi score tests. Extensions to longitudinal survey data and constructing pseudo empirical likelihood intervals will also be presented. (Joint work with Changao Wu, University of Waterloo).

Mary Thompson a réalisé un travail de pionnière, en collaboration avec Godambe, sur l'utilisation d'équations estimatrices pondérées pour faire de l'inférence à partir de données provenant de plans de sondage complexes, en plus d'obtenir des résultats associés d'optimalité. Dans cet exposé, je présenterai quelques applications récentes de l'approche par équations estimatrices et de l'utilisation des méthodes de ré-échantillonnage pour l'estimation de la variance et pour les tests d'hypothèses. Notamment, je présenterai des résultats de l'utilisation de l'information de post-stratification dans la construction d'estimateurs ponctuels, de variance et de quasi-tests de scores. Des extensions aux données d'enquêtes longitudinales et à la construction d'intervalles de confiance basées sur la fonction de vraisemblance pseudo-empirique seront également présentées. (Travail effectué conjointement avec Changao Wu de l'Université de Waterloo).

---

**DAVID BINDER** (Retired (Statistics Canada))

[Tuesday June 14 / mardi 14 juin, 10:45]

*Causal Inference for Observational Data Obtained from a Complex Survey*

*Inférence causale pour des études observationnelles provenant d'un plan de sondage complexe*

Mary Thompson's 2004 Gold Medal address included a discussion of causation and causality in observational studies. One of Mary Thompson's passions has been on making analytic inferences from surveys with a complex design. Since population-based surveys are observational studies, it seems natural to ask what the impact of the survey design might be for making causal inferences. First, we give a brief review of the literature on causal inference from observational studies. Using Cox and Wermuth's (2004) delineation

En 2004, l'allocution de Mary Thompson, la récipiendaire de la médaille d'or incluait une discussion sur la causalité pour des études observationnelles. L'une des passions de Mary Thompson fut de mener des analyses inférentielles d'enquêtes à plan de sondage complexe. Puisque les enquêtes sur la population sont des études observationnelles, il semble naturel de s'interroger sur l'impact du plan de sondage dans les inférences causales. Nous débiterons par une courte revue de littérature de l'inférence causale d'études par observation. En utilisant la délimitation de Cox et Wermuth (2004) pour différents niveaux de cau-

## 2B-I4: Sampling Theory and Practice: In Honour of Mary Thompson's Retirement Théorie et pratique de l'échantillonnage : À l'occasion du départ à la retraite de Mary Thompson

---

of various levels of causality, we investigate which assumptions are suitable for the ignorability of the survey design.

salité, nous étudierons les hypothèses valables permettant d'ignorer le plan de sondage.

---

**DANNY PFEFFERMANN** (University of Southampton, UK)

[Tuesday June 14 / mardi 14 juin, 11:15]

*Fitting Models to Complex Survey Data; Why is it a problem? How should we approach it?*

*Ajuster des modèles aux données d'enquêtes complexes; pourquoi est-ce un problème? Comment devons-nous l'aborder?*

Survey data are frequently used for analytic inference about statistical models which are assumed to hold for the population from which the sample is taken. I shall start my talk by discussing four features of complex survey data which makes it different from other typical data sets. Next I shall review various approaches proposed in the literature for dealing with these features, discussing their merits and limitations. I will present also simulation results, which compare the approaches when estimating regression models from a single stage sample in terms of bias, variance, and coverage rates.

Nous utilisons fréquemment les données d'enquête pour l'analyse inférentielle de modèles statistiques dont les pré-supposés sont respectés pour la population sous-jacente. Je débiterai mon exposé en présentant quatre caractéristiques des données d'enquêtes complexes qui les distinguent des autres types d'ensembles de données. Par la suite, je passerai en revue différentes approches proposées dans la littérature pour traiter ces caractéristiques en discutant de leurs bien-fondés et limites. Je présenterai également les résultats de simulations comparant les approches en terme de biais, de variance et de taux de recouvrement, et en estimant des modèles de régression d'un plan de sondage à une phase.

## 2B-15: Statistical Consulting: Communication, Interpersonal, Project Management Skills Required

### Conseil statistique : compétences requises en communications, interpersonnel et gestion de projet

---

Chair/Président: Boxin Tang (Simon Fraser University)  
Organizer/Responsable: John Amrhein (McDougall Scientific Ltd.)

Session sponsored by the Business and Industrial Statistics Section

Session parrainée par le Groupe de statistique industrielle et de gestion

**Room/Salle: BAC141**

**Abstracts/Résumés**

---

**LORNE ROTHMAN** (SAS Institute Canada Inc.)

[Tuesday June 14 / mardi 14 juin, 10:15]

*How Many Consultants Does It Take to Fit a Regression? Lessons From a Large Business Practice.*  
*Combien de consultants faut-il pour ajuster une régression ? Leçons d'une grande pratique.*

Successful consulting services provided by large commercial operations require more than good statisticians. Sales people, project managers, business experts, and data modellers and programmers are important members of a consulting team. Even when the singular goal of a project is statistical inference, or as is far more common—prediction, a surprisingly small percentage of engagement time is spent building statistical models (e.g. 10-20%). The majority of time is devoted to data preparation. Once data are available for analysis, familiar algorithms are most often applied, but embedded within methodologies geared to prediction and the analysis of large datasets.

Il faut plus que de bons statisticiens pour que le service de consultation issu d'une grande opération commerciale connaisse du succès. Les représentants, les directeurs de projet, les experts en entreprise, ainsi que les modélisateurs de données et les programmeurs sont des membres importants de l'équipe de consultation. Même si le but unique d'un projet est l'inférence statistique, ou comme c'est beaucoup plus courant – la prédiction, un très petit pourcentage du temps facturable est investi dans la construction des modèles statistiques (e.g. 10-20%). La majorité du temps est dédié à la préparation des données. Dès que les données sont disponibles pour l'analyse, des algorithmes familiers sont souvent utilisés, mais à l'intérieur de méthodologies orientées vers la prédiction et l'analyse de larges ensembles de données.

---

**FERNANDO CAMACHO** (Damos Inc.)

[Tuesday June 14 / mardi 14 juin, 10:45]

*Statistical Consulting within an Engineering Environment*  
*Consultation statistique dans un environnement d'ingénierie*

This talk will discuss several consulting experiences to illustrate some skills the author considers important to succeed as a statistical consultant within an engineering environment. Some of these skills include: understanding why the work is needed, cooperating within a multidisciplinary group, learning about the problem and providing an acceptable solution, educating the client on the contributions of statistics, communicating with the client and other stakeholders, selling the statistical

Cet exposé parlera de plusieurs expériences de consultation pour illustrer certaines compétences que l'auteur juge importantes pour réussir en tant que consultant statistique dans un environnement d'ingénierie. Certaines de ces compétences sont de : comprendre pourquoi le travail est nécessaire, coopérer à l'intérieur d'un groupe multi-disciplinaire, comprendre le problème et fournir une solution acceptable, instruire le client au sujet des contributions de la statistique, communiquer avec le client et les autres parties concernées, vendre le travail statistique, répondre dans les temps

## 2B-15: Statistical Consulting: Communication, Interpersonal, Project Management Skills Required

### Conseil statistique : compétences requises en communications, interpersonnel et gestion de projet

---

work, responding on time and on budget, and earning the trust of the client and other stakeholders.

en respectant le budget, et gagner la confiance du client et des autres parties concernées.

---

**RHONDA ROSYCHUK** (University of Alberta)

[Tuesday June 14 / mardi 14 juin, 11:15]

*Statistical Consulting in an Academic Medical Setting: If I Only Knew Then What I Know Now*

*Consultation statistique dans un milieu académique médical : Si j'avais pu savoir à l'époque ce que je sais maintenant*

Effective statistical consulting is a blend of multiple skills and many of these skills are often honed through trial and error. The academic medical setting can provide additional challenges. The clients have varying levels of statistical sophistication and diverse expectations for the consulting relationship. Personalities, expectations, activities, and timelines need to be carefully managed. Statistical consulting units may be cost-recovery, and issues of charging and recovering fees can be thorny. Grant proposals figure prominently in this setting, and critical appraisal and problem solving skills are key. This talk will highlight key skills for success in this rewarding setting.

La consultation statistique efficace est un mélange de nombreuses compétences, dont plusieurs sont souvent affinées par essais et erreurs. Le milieu académique médical peut présenter des défis supplémentaires. Les clients ont des niveaux variés de connaissances statistique et des attentes diverses au niveau de la relation avec le consultant. Les personnalités, les attentes, les activités et les échéanciers doivent être gérés prudemment. Les unités de consultation statistique peuvent récupérer leurs frais, et les problèmes de facturation et de collection peuvent être délicats. Les demandes de subvention jouent un rôle important dans ce milieu, et les aptitudes pour évaluer et résoudre les problèmes sont capitales. Cet exposé soulignera les compétences nécessaires au succès dans ce cadre gratifiant.

## 2B-C1: High-Dimensional Data in Biology Données de grande dimension en biologie

---

Chair/Président: Yutaka Yasui (University of Alberta)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 234**

**Abstracts/Résumés**

---

**KAREN A KOPCIUK** (Alberta Health Services - Cancer Care), **YUAN WANG** (University of Calgary)

[Tuesday June 14 / mardi 14 juin, 10:15]

*Sample Size Estimation Methods for NMR Metabolomics Data*

*Méthodes d'estimation de la taille d'échantillon pour données métabolomiques de la résonance magnétique nucléaire (RMN)*

Typical projection-based methods used to analyze metabolomics data, such as Partial Least Squares Regression, do not have statistical inference methods or the corresponding sample size methods. To overcome this study planning limitation, we investigated sample size estimation approaches for high dimensional microarray data for their suitability with NMR metabolomics data. Simulation studies compared three methods across a number of relevant data features. Simulation study results and an application to a pancreatic cancer data set suggested two methods could be used for sample size estimation of NMR metabolomics data. Future work will investigate these methods with mass spectrometry metabolomics data.

Les méthodes typiques basées sur la projection pour l'analyse de données métabolomiques, telle la régression des moindres carrés partiels, n'ont pas de méthodes d'inférence statistique ou de méthodes de la taille d'échantillon correspondantes. Pour surmonter cette limite dans la planification de l'étude, nous avons étudié des approches d'estimation de la taille d'échantillon pour données de micropuces à haute dimension afin de tester leur pertinence avec des données métabolomiques RMN. Des études par simulation ont comparé trois méthodes selon différentes caractéristiques pertinentes. Les résultats de l'étude par simulation et une application aux données sur le cancer du pancréas ont suggéré deux méthodes à utiliser pour l'estimation de la taille d'échantillon pour données métabolomiques RMN. A l'avenir, nous étudierons ces méthodes avec des données métabolomiques par spectrométrie de masse.

---

**TOBY KENNEY & HONG GU** (Dalhousie University)

[Tuesday June 14 / mardi 14 juin, 10:30]

*A Generalized Codon-based Model of Nucleotide Substitution for Protein-coding DNA Sequences*

*Modèle généralisé à base de codons pour la substitution de nucléotide dans des séquences d'ADN codant pour des protéines*

A generalized codon-based model for protein-coding DNA sequences is proposed for phylogenetic analysis. This model framework provides a unified framework for existing codon (or DNA or amino acid) models. Furthermore, it offers greater flexibility in the choice of rate matrix, allowing existing models to be easily extended to incorporate more of the possible driving forces in molecular

Nous proposons pour l'analyse phylogénétique un modèle généralisé à base de codons pour les séquences d'ADN codant pour des protéines. Ce modèle fournit un cadre unifié pour les modèles de codons (ou d'ADN ou d'acides aminés) existants. Il offre également une grande flexibilité pour le choix de la matrice de transition, permettant de facilement étendre les modèles existants pour incorporer un plus grand nombre de forces motrices de l'évolution moléculaire, les in-

## 2B-C1: High-Dimensional Data in Biology Données de grande dimension en biologie

---

evolution, such as structure information and amino acid properties. We provide a software package called Codon Optimal Likelihood Discoverer (COLD) to implement these proposed generalized codon models. We demonstrate how our model framework allows model selection based on standard likelihood theory.

formations sur la structure et les propriétés des acides aminés. Nous offrons un progiciel appelé Codon Optimal Likelihood Discoverer (COLD) pour mettre en oeuvre les modèles généralisés à base de codons proposés. Nous démontrons avec la théorie habituelle de la vraisemblance le fonctionnement du cadre de notre modèle permettant la sélection de modèle.

---

**HONG GU, TOBY KENNEY & JOSEPH BIELAWSKI** (Dalhousie University)

[Tuesday June 14 / mardi 14 juin, 10:45]

*Estimation of Darwinian Positive Selection Using Generalized Codon-based Models*

*Estimation de la sélection positive de Darwin à l'aide de modèles généralisés à base de codons*

The generalized codon-based model for protein-coding DNA sequences offers great flexibility in the choice of rate matrix and significantly improves the goodness-of-fit. We further develop the mixed model approach for this type of model by allowing random effects on one or more parameters across the amino acid sites. By allowing random effects on the parameter that estimates the nonsynonymous change rate on different sites, generalized codon-based models include as special cases the M-series models in Yang et al (2000). We demonstrate the more accurate estimation of heterogeneous selection pressure at amino acid sites, comparing to the estimation by M-series models.

Le modèle généralisé à base de codons pour les séquences d'ADN codant pour des protéines offre une grande flexibilité pour le choix de la matrice de transition et améliore significativement la qualité de l'ajustement. Nous développons davantage l'approche du modèle mixte pour ce type de modèle en permettant les effets aléatoires pour un ou plusieurs paramètres sur des sites d'acides aminés. En permettant les effets aléatoires pour le paramètre estimant le taux de mutation non synonyme sur différents sites, les modèles généralisés à base de codons incluent la série de modèles M de Yang et al. (2000) à titre de cas particuliers. Nous démontrons l'estimation plus précise de la pression sélective hétérogène aux sites d'acides aminés, comparativement à l'estimation par la série de modèles M.

---

**THIERRY CHEKOUO TEKOUANG & ALEJANDRO MURUA** (Université de Montreal)

[Tuesday June 14 / mardi 14 juin, 11:00]

*Statistical View of Popular Biclustering Algorithms*

*Approche statistique des algorithmes populaires de grappes bi-dimensionnelles*

Recently, several biclustering algorithms have been proposed to reveal submatrices of the data matrix whose rows exhibit similar behaviour across a set of columns, and conversely. These have important applications to gene expression analysis, for example, to find genes that are co-regulated across a subset of conditions. Few of these algorithms are based on explicit models. This work proposes some underlying statistical models associated with some of the most popular biclustering algorithms. It shows that these algorithms can be justified within a Bayesian framework and can be derived from Bayesian computational techniques.

Récemment, plusieurs algorithmes de recherche de grappes bi-dimensionnelles ont été proposés. Ces algorithmes ont pour but de déterminer une sous-matrice de la matrice de données dont les lignes exhibent un comportement similaire à travers les colonnes, et vice versa. Peu de ces algorithmes sont basés sur des modèles statistiques explicites. Ce travail propose des modèles statistiques sous-jacents aux algorithmes les plus populaires. Il montre que ces algorithmes peuvent être justifiés dans un cadre bayésien et peuvent être dérivés à travers des techniques computationnelles bayésiennes.

## 2B-C1: High-Dimensional Data in Biology Données de grande dimension en biologie

---

**DENA GIVARI, PAUL MCNICHOLAS & SANJEENA SUBEDI** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 11:15]

*On the Application of a Bayesian Infinite Mixture Modelling*

*De l'application d'un modèle de mélange bayésien infini*

In microarray analyses, genes are expressed to varying levels across different cellular conditions. It is believed that the identification of gene with similar expression patterns across conditions demonstrating variable temporal, developmental, topographical, histological and physiological patterns will give insight to their biological functions. The aim is to create clusters which hold genes whose expression levels are inter-related at various conditions. We further develop a recent algorithm based on the Bayesian infinite mixture model; in particular we allow for non-isotropic covariance matrices in order to better reflect reality. Our algorithm is discussed and illustrated on simulated and real data.

En analyse de micropuces, les gènes sont exprimés à différents niveaux selon les conditions cellulaires. L'identification de gènes est supposée renseigner sur leurs fonctions biologiques lorsque leurs formes d'expression sont similaires dans des conditions démontrant des régimes variables de temps, de croissance, de topographie, d'histologie et de physiologie. L'objectif est de créer des grappes rassemblant les gènes dont les niveaux d'expression sont liés entre eux sous différentes conditions. Nous avons développé un algorithme récent basé sur un modèle de mélange bayésien infini, et nous permettons en particulier des matrices de covariance non isotropes afin de mieux refléter la réalité. Notre algorithme est présenté et illustré au moyen de données réelles et simulées.

**VAHID PARTOVI NIA & VAHID PARTOVI NIA** (McGill University)

[Tuesday June 14 / mardi 14 juin, 11:30]

*Agglomerative Bayesian Biclustering*

*Classification double agglomérative bayésienne*

In many biological domains such as metabolomics, proteomics, and genetics a data matrix with subjects in rows and variables in columns is produced. For such data simultaneous clustering of subjects and variables, called biclustering, is of interest. For instance, in gene expression data, clustering subjects reflects which subjects have similar genetic make up, and clustering genes reflects which genes might function similarly on the measured subjects. We suggest a fully-automatic biclustering algorithm using a Bayesian model which does not require knowledge about the number of partitions. Furthermore, agglomerative method produces a graphical representation of biclusters through dendrogram.

Dans plusieurs domaines biologiques tels que la métabolomique, la protéomique et la génétique, une matrice de données dont les sujets correspondent aux lignes et les variables aux colonnes, est produite. Pour ce type de données, la classification simultanée des sujets et des variables, appelée classification double, est intéressante. Par exemple, avec des données d'expression génétique, la classification des sujets groupe les sujets ayant des similarités génétiques, tandis que la classification des gènes groupe les gènes ayant des fonctions similaires. Nous suggérons un algorithme de classification double complètement automatisé en utilisant un modèle bayésien dont il n'est pas nécessaire de connaître le nombre de partitions. De plus, la méthode agglomérative produit une représentation graphique des classes doubles à l'aide d'un dendrogramme.



## 2B-C2: Methods in Mixture Models and Spatial Statistics Méthodes dans les modèles de mélange et spatiale statistique

---

Chair/Président: Eleanor Pullenayegum (McMaster University)

Room/Salle: BAC 142

Abstracts/Résumés

---

**ABBAS KHALILI & DAVID A. STEPHENS** (McGill University)

[Tuesday June 14 / mardi 14 juin, 10:15]

*A New Order Estimation Method in Mixture of Autoregressive Models*

*Nouvelle méthode d'estimation de l'ordre dans les modèles de mélanges autorégressifs*

Finite mixture of autoregressive (MAR) models are relatively new in time series analysis. A distinguished feature of a MAR model is that it allows for combining stationary and non-stationary AR processes to capture heterogeneity in a time series. In an application one needs to first decide on the proper order of the AR components of the MAR model. We present a new computationally efficient method for simultaneous order and parameter estimation in MAR models. Martingale theory is used to study asymptotic behaviour of the new method. We also present some simulation study, and a real data analysis.

Les modèles de mélanges autorégressifs (MAR) finis sont relativement récents en analyse de séries chronologiques. Le modèle MAR a une caractéristique distinctive qui permet de combiner des processus autorégressifs stationnaires et non stationnaires afin de capter l'hétérogénéité dans les séries chronologiques. Dans une application, il importe d'établir d'abord l'ordre approprié des composantes autorégressives du modèle MAR. Nous proposons une nouvelle méthode efficace en matière de temps de calcul pour l'estimation simultanée de l'ordre et du paramètre dans les modèles MAR. La théorie de Martingale est utilisée pour étudier le comportement asymptotique de la nouvelle méthode. Nous présentons aussi des études de simulation et une analyse de données réelles.

---

**ANDREW RAIM** (University of Maryland), **MATTHIAS GOBBERT & NAGARAJ NEERCHAL** (University of Maryland, Baltimore County), **JORGE MOREL** (Procter and Gamble Company)

[Tuesday June 14 / mardi 14 juin, 10:30]

*Maximum Likelihood Estimation of the Random-Clumped Multinomial Model using High Performance Computing*

*Estimation du maximum de vraisemblance du modèle multinomial aléatoire groupé par le calcul de haute performance*

For many practical problems, when maximum likelihood estimates (MLEs) are desired they must be computed numerically with the aid of a computer. Occasionally long computing times can become a serious issue. We will consider one such MLE problem for the case of the Random-Clumped Multinomial distribution, a mixture which is useful for modeling overdispersion. We apply parallel computing to this problem with a very general approach, making use of a high performance computing cluster and optimization toolkit. We show that computing time for larger problems can be improved significantly by scaling the number of parallel processes.

Dans plusieurs problèmes pratiques, quand les estimateurs du maximum de vraisemblance (EMV) sont nécessaires, ils doivent être calculés numériquement par ordinateur. Occasionnellement, le temps de calcul peut devenir un sérieux problème. Nous considérerons un tel problème d'EMV pour le cas de la distribution multinomiale aléatoire groupée, une combinaison utile pour la modélisation de la surdispersion. On appliquera le calcul en parallèle à ce problème avec une approche très générale, utilisant le calcul de haute performance par grappe de serveur et des outils d'optimisation. Nous montrerons que le temps de calcul pour les problèmes d'envergure peut être amélioré significativement en normalisant le nombre de processus parallèles.

## 2B-C2: Methods in Mixture Models and Spatial Statistics Méthodes dans les modèles de mélange et spatiale statistique

---

**SHAHEENA BASHIR** (University Health Network, Toronto), **E. M. CARTER** (University of Guelph, Canada)

[Tuesday June 14 / mardi 14 juin, 10:45]

*Robust Covariate Density Defined (CDD) Mixture of Logistic Regression Models*

*Mélange robuste de modèles de régression logistique défini par la densité des covariables*

Logistic regression is widely used in biological sciences in modeling data with categorical response variable. For heterogeneous populations, mixture of logistic regression is used instead of an overall logistic regression model. CDD mixtures of logistic regression model use the marginal densities of the continuous covariates as a mixture of different latent components. When there are potential outliers, then the estimates of CDD mixture of logistic regression model are affected resulting in unreliable inference. The current work proposes to use robust regression estimators for the continuous predictors. Simulations show this robust CDD mixture of logistic regression models produce robust estimates.

La régression logistique est largement utilisée en sciences biologiques pour modéliser des données comportant des variables-réponses catégorielles. Pour les populations hétérogènes, on utilise un mélange de régressions logistiques plutôt qu'un modèle de régression logistique global. Les mélanges de régressions logistiques définis par la densité des covariables considèrent les densités marginales des covariables continues comme des mélanges de variables latentes. En présence de valeurs aberrantes, les estimés du mélange de modèles de régression logistique défini par densité des covariables sont affectés, si bien que l'inférence n'est plus fiable. Nos travaux actuels proposent d'utiliser des estimateurs de régression robustes pour les variables explicatives continues. Les simulations montrent que ce mélange robuste de modèles de régression logistique défini par densité des covariables produit des estimés robustes.

**MÉLINA MAILHOT, HÉLÈNE COSSETTE & ÉTIENNE MARCEAU** (Université Laval)

[Tuesday June 14 / mardi 14 juin, 11:00]

*TVaR-Based Capital Allocation for Bivariate Compound Distributions*

*Allocation de capital basée sur la mesure TVaR pour distributions composées bivariées*

Considering an insurance portfolio of dependent risks, the objective is to evaluate the capital allocation for the overall portfolio and the contribution of each risk, using the Tail Value-at-Risk (TVaR) risk measure. Considering multivariate compound distributions, models based on common mixtures and models based on copulas, exact expressions are presented for the TVaR of the sum of the risks and for the TVaR-based capital allocation. A numerical method is proposed for any proposed dependence model in order to approximate the TVaR of the aggregate risk and the contribution of each risk of the portfolio. Numerical examples are presented.

La présente étude considère un portefeuille d'assurance constitué de plusieurs risques dépendants et a pour objectif d'évaluer le capital alloué au portefeuille ainsi qu'à chaque risque faisant partie de ce dernier. La mesure TVaR (Tail Value-at-Risk) est utilisée. Les modèles de dépendance considérés sont des distributions composées multivariées, des modèles basés sur des mixtures et des modèles basés sur les copules. Des cas particulier seront présentés, où les valeurs exactes sont trouvées par rapport à l'allocation de capital basée sur la TVaR pour le portefeuille, ainsi que chaque risque le composant. Des exemples numériques seront présentés.

**LINGLING FAN** (University of New Brunswick)

[Tuesday June 14 / mardi 14 juin, 11:15]

*Classification and Spatial Point Pattern Recognition*

*Classification et reconnaissance de motifs ponctuels spatiaux*

## 2B-C2: Methods in Mixture Models and Spatial Statistics

### Méthodes dans les modèles de mélange et spatiale statistique

---

Spatial point patterns exist in many areas such as materials science, engineering, biology, and ecology. Focusing on the question of how to spot differences in patterns that look very similar, we will discuss how to construct different classification models based on some summarizing statistics to recognize the spatial patterns, compare different types of models. This would be very useful especially for collecting data.

Les motifs ponctuels spatiaux existent dans plusieurs domaines, notamment en science des matériaux, en ingénierie, en biologie et en écologie. En nous concentrant sur les façons de repérer les différences dans des motifs très similaires, nous examinons des méthodes pour construire différents modèles de classification basés sur des statistiques de synthèse afin de reconnaître les motifs spatiaux, et nous comparons différents types de modèles. Ces méthodes pourraient s'avérer très utiles, particulièrement pour la collecte de données.

---

**JEFFREY PICKA** (University of New Brunswick)

[Tuesday June 14 / mardi 14 juin, 11:30]

*Spatial Statistical Analysis of Powder Flow Data*

*Analyse statistique spatiale de données d'écoulement des poudres*

Powders with the same general composition can settle in two very different ways when subjected to tapping. Spatial statistics can be used to describe differences between particle arrangements that are related to the bulk settling properties of the powder. These methods are demonstrated on data from simulated tapping experiments, and the relationship between these statistics and model checking will be discussed.

Les poudres ayant la même composition générale peuvent s'arranger de deux différentes manières lorsqu'elles sont tapotées. Les statistiques spatiales peuvent être utilisées pour décrire les différences entre les arrangements de particules liées aux propriétés d'arrangement en vrac de la poudre. Ces méthodes sont démontrées par des données issues d'expériences simulées de tapotements, et la relation entre ces statistiques et la vérification du modèle sera discutée.

## 2B-C3: Spatial-temporal Modelling in the Natural World Modélisation spatio-temporelle dans le monde naturel

---

Chair/Président: Fahim Ashkar (Université de Moncton)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 236**

**Abstracts/Résumés**

---

**MIKE DOWD** (Dept. of Mathematics and Statistics, Dalhousie University), **RUTH JOY** (Dept. of Statistics, Simon Fraser University)

[Tuesday June 14 / mardi 14 juin, 10:15]

*Estimating Behavioural Parameters in Stochastic Animal Movement Models*

*Estimation de paramètres comportementaux dans des modèles stochastiques de déplacements d'animaux*

Data on fine-scale animal movement is being collected worldwide. Here, a general methodology is proposed for analysing and modelling animal behaviour. We apply a state space model comprised of a movement model and high-resolution vertical movement data for northern fur seals tagged on the Pribilof Islands, Alaska. Parameters of the movement model are estimated on an appropriate time scale, providing a direct link to animal behaviour. This is based on a multiple iterated particle filter with state augmentation. The fitted parameters show distinct time-evolving changes in fur seal behavior over time, matching well what is observed for the fur seals.

Des données sur les déplacements d'animaux à petite échelle sont collectées dans le monde entier. Nous proposons ici une méthodologie générale pour analyser et modéliser le comportement animal. Nous appliquons un modèle d'espace d'états composé d'un modèle de déplacement et de données de déplacement vertical de haute résolution provenant d'otaries à fourrure marquées sur les îles Pribilof en Alaska. Les paramètres du modèle de déplacement sont estimés sur une échelle temporelle appropriée, fournissant un lien direct avec le comportement animal. Ce processus est basé sur un filtre particulaire itéré multiple à augmentation d'état. Les paramètres ajustés montrent des changements temporels distincts dans le comportement des otaries à fourrure évoluant dans le temps, ce qui correspond aux observations pour ces animaux.

---

**ANGUS ARGYLE** (University of Victoria)

[Tuesday June 14 / mardi 14 juin, 10:30]

*Estimating the Number of Species using Abundance-Based Data from Multiple Sites within a Region*

*Estimation du nombre d'espèces à l'aide de données d'abondance provenant de différents sites d'une même région*

Estimating the number of species in a region (e.g. woody plant species in a tropical rainforest) is useful as one quantitative measure of the region's diversity. The estimates indicate the potential number of species new to science. I will present a frequentist estimator and a bayesian estimator of the number of species. The two estimators are developed for use when multiple spatial-based sites within a region are surveyed and abundance-based data are recorded. The estimators are ap-

L'estimation du nombre d'espèces qu'abrite une région (p. ex., espèces de plantes ligneuses dans une forêt tropicale humide) offre une mesure utile de sa diversité. De tels estimés indiquent le nombre potentiel d'espèces restant à découvrir. Je présenterai des estimateurs fréquentiste et un bayésien du nombre d'espèces. Ces deux estimateurs sont construits pour accommoder des données d'abondance récoltées en faisant l'inventaire de nombreux sites d'une même région. Je les appliquerai à deux ensembles de données en provenance du Panama. Je comparerai leur

## 2B-C3: Spatial-temporal Modelling in the Natural World Modélisation spatio-temporelle dans le monde naturel

---

plied to two data sets from Panama. The performances of the two estimators are compared with other estimators from simpler sampling scenarios.

performance à celle d'autres estimateurs construits sous l'hypothèse de scénarios d'échantillonnage plus simples.

---

**MARYLÈNE TROUPÉ** (Université des Antilles et de la Guadeloupe), **JEAN VAILLANT** (Université des Antilles et de la Guyane), **LÉONIDE CELINY** (Université Paris-Est Créteil)

[Tuesday June 14 / mardi 14 juin, 10:45]

*Dispersion Indexes and Thinned Point Processes*  
*Indices de dispersion et processus ponctuels amincis*

We are interested in the effect of some types of events on dispersion indexes such as the mean crowding. From a mathematical point of view, we consider that a distribution of individuals is a realization of a point process and we study the consequences of a stochastic point thinning on the model distributional properties. A case study is provided from counting of aphids of the cotton plant carried out before and after thunderstorm. The mortality rate due to the thunderstorm is analyzed according to characteristics of thunderstorms, leaf strata and larvae stages.

Nous nous intéressons à l'effet de certains types d'événements sur les valeurs de quelques indices de dispersion tels que l'attroupelement moyen. D'un point de vue mathématique, on assimile une répartition d'individus à une réalisation d'un processus ponctuel et on étudie les conséquences d'un mécanisme stochastique de suppression de points sur les propriétés distributionnelles du modèle. Une illustration est fournie à partir de comptages de pucerons du cotonnier effectués avant et après orage. Le taux de mortalité dû à l'orage est analysé en fonction des caractéristiques des orages, des strates foliaires et des stades larvaires.

---

**NOEL CADIGAN** (Fisheries and Oceans Canada)

[Tuesday June 14 / mardi 14 juin, 11:00]

*Measurement Error Bias in Fish Stock Spawner-Recruitment Models*  
*Biais dû aux erreurs de mesure dans les modèles de recrutement du stock de poissons reproducteur*

Understanding the relationship between spawning stock size and subsequent reproduction and recruitment of juveniles to a fishery is a fundamental component of sustainable fisheries management. I use local influence diagnostics to show that the effect of stock size measurement errors when fitting spawner-recruitment models is to overestimate the slope at the origin and underestimate the maximum recruitment. This will often lead to poor estimates of maximum sustainable yield reference points, which have been adopted by many national and international fisheries management agencies to guide fisheries management. This measurement error bias can lead to unsustainable over-exploitation of fish stocks.

Pour une gestion durable des pêches, il est essentiel de bien comprendre la relation entre la biomasse du stock reproducteur et sa reproduction et le recrutement des alevins vers un lieu de pêche. J'utilise un diagnostic d'influence locale pour montrer que les erreurs de mesure de la biomasse du stock ont pour effet, lors de l'ajustement des modèles de recrutement du stock reproducteur, une surestimation de la pente à l'origine et une sous-estimation du recrutement maximum. Cela conduit souvent à de mauvaises estimations des repères de rendement équilibré maximum, repères adoptés par de nombreux organismes de gestion des pêches au Canada et ailleurs dans le monde pour diriger la gestion des pêches. Ce biais dû aux erreurs de mesure risque de conduire à une surexploitation insoutenable des stocks de poissons.

## 2B-C3: Spatial-temporal Modelling in the Natural World Modélisation spatio-temporelle dans le monde naturel

---

**LARISSA VALMY & JEAN VAILLANT** (Université des Antilles et de la Guyane)

[Tuesday June 14 / mardi 14 juin, 11:15]

*Statistical Study of Caribbean Seismic Events*

*Etude statistique d'événements sismiques caribéens*

Calculation of seismic risks in a zone are based on stochastic models describing at best the seismic activity. Our focus is the spatial-temporal evolution of the probability of earthquake occurrences. The main modelling tool is the theory of the point processes. We discuss about the spatial-temporal extension of ETAS (Epidemic Type Aftershock Sequence) with marks and study the properties of the log-likelihood and the maximum likelihood estimators. Seismic data from the Lesser Antilles Arc carried out between 1999 and 2004 are analyzed.

Le calcul de risques sismiques dans une zone s'appuie sur des modèles stochastiques décrivant au mieux l'activité sismique. Il s'agit de calculer les probabilités d'occurrences de séismes et d'analyser l'évolution spatio-temporelle de ces probabilités. L'outil principal de modélisation est la théorie des processus ponctuels. On y retrouve le modèle ETAS (Epidemic Type Aftershock Sequence) qui intègre l'aspect spatio-temporel et les marques. Nous étudions les propriétés de la log-vraisemblance et des estimateurs de maximum de vraisemblance. Des données de tremblements de terre de l'arc des Petites Antilles relevées entre 1999 et 2004 sont analysées.

**KHADIJA BOUZAACHANE** (Université Cadi Ayyad Ecole Supérieure de Technologie d'Essaouira), **YOUSSEF BENG-HABRIT** (ENSAM)

[Tuesday June 14 / mardi 14 juin, 11:30]

*Spatial-temporal Model for Rainfall Data of Morocco*

*Modélisation spatio-temporelle des précipitations au Maroc*

In Morocco, no operational system actually exists for the early prediction of the drought year. The prediction in the face of a hypothesized global warming scenario is very important, particularly for agricultural economy in Morocco. The need for accurate modelling of the rainfall data set is vital. In order to forecast a drought year or a wet year, we need rainfall forecasting models that can accurately capture the main characteristics of the rainfall across the space and time. The aim of this work is to choose the adequate spatial-temporal model for monthly rainfall data in Morocco.

Au Maroc, aucun système opérationnel n'existe réellement pour la prévision hâtive d'années de sécheresse. La prévision face à un scénario présumé de réchauffement global est très importante, en particulier pour l'économie agricole au Maroc. Le besoin d'une modélisation adéquate des données de précipitation est essentiel. Afin de prévoir une année de sécheresse ou de précipitations fréquentes, nous avons besoin de modèles de prévision des précipitations qui peuvent correctement capturer les caractéristiques principales des précipitations au niveau de l'espace et du temps. Le but de ce travail est de choisir le modèle spatio-temporel adéquat pour les données mensuelles des précipitations au Maroc.

## 2C-I1: Designs for Generalized Linear Models Conception de modèles linéaires généralisés

---

Chair/Président: Sue Lewis (University of Southampton, UK)  
Organizer/Responsable: Abhyuday Mandal (University of Georgia)

**Room/Salle: BAC239**

**Abstracts/Résumés**

---

**ABHYUDAY MANDAL** (University of Georgia), **JIE YANG & DIBYEN MAJUMDAR** (University of Illinois at Chicago)

[Tuesday June 14 / mardi 14 juin, 13:00]

*Optimal Designs for Two-Level Factorial Experiments with Binary Response*

*Plans optimaux pour expériences factorielles à deux niveaux avec réponse binaire*

We consider the problem of obtaining locally D-optimal designs for factorial experiments with qualitative factors at two levels each with binary response. For the  $2^2$  factorial experiment with main effects model we obtain optimal designs analytically in special cases. The optimal designs are shown to be robust to the choice of the assumed values of the prior and when there is no basis to make an informed choice of the assumed values we recommend the use of the *uniform* design. For the general  $2^k$  case we show that the uniform design has a *maximin* property.

Nous considérons le problème de l'élaboration de plans localement D-optimaux pour des expériences factorielles avec facteurs qualitatifs à deux niveaux, chacun avec réponse binaire. Pour l'expérience factorielle  $2^2$  avec modèle à effets principaux, nous obtenons des solutions analytiques dans des cas particuliers. Les plans optimaux sont robustes au choix des lois a priori, et lorsqu'aucune base n'est disponible pour procéder au choix de loi a priori de manière informée, nous conseillons l'utilisation d'un plan *uniforme*. Quant au cas général  $2^k$ , nous montrons que le plan uniforme a une propriété *maximin*.

---

**KARABI NANDY** (University of California at Los Angeles)

[Tuesday June 14 / mardi 14 juin, 13:22]

*Optimal Designs for Non-Monotone Dose-Response Models*

*Plans optimaux pour modèles non-monotones de la relation dose-effet*

Dose response studies arise in many medical/nursing applications and are an important counterpart to efficient development of clinical intervention. Often such dose-response studies can be considered within the framework of binary response experiments where popular choices for modeling the probability of response are logistic or probit models. We consider a special type of dose-response model such as hormesis where the dose-response relationship is characterized by low dose stimulation (beneficial effect) and high dose inhibition (destructive intoxication). We develop some optimal design strategies for this situation by considering a logistic dose-response model with a quadratic term in the linear predictor.

Les études sur la relation dose-effet surviennent dans plusieurs applications en médecine et en soins infirmiers, et sont un complément important au développement efficace de l'intervention clinique. De telles études sur la relation dose-effet sont souvent considérées dans le cadre d'expériences dont les réponses binaires sont modélisées à l'aide des fonctions logistique ou probit. Nous considérons un modèle de type particulier, comme hormesis, où la relation dose-effet est caractérisée par une stimulation à faible dose (l'effet favorable) et par une inhibition à forte dose (intoxication destructive). Nous développons quelques stratégies pour l'obtention de plans optimaux dans cette situation en considérant un modèle de relation dose-effet logistique avec un terme quadratique dans les prédicteurs linéaires.

## 2C-I1: Designs for Generalized Linear Models Conception de modèles linéaires généralisés

---

---

**JOHN STUFKEN** (University of Georgia)

[Tuesday June 14 / mardi 14 juin, 13:45]

*Tools to Identify Optimal Designs for Generalized Linear Models*

*Outils pour identifier les plans optimaux de modèles linéaires généralisés*

The literature on optimal design for linear models is very well developed. In spite of seminal advances over the past decades, tackling the optimal design problem for non-linear models has proven to be much more difficult. In recent years, new tools that yield very general results have been developed. Many models are covered by this new approach, including generalized linear models for binary and count data, such as logistic, probit, and loglinear models. We discuss the basic idea underlying the new tools, present key results obtained by their use, discuss limitations, and offer some open problems in the area.

La littérature sur les plans optimaux pour modèles linéaires est très bien développée. Malgré les avancées déterminantes de ces dernières décennies, il demeure très difficile de s'attaquer au problème de plan optimal de modèles non-linéaires. Au cours des dernières années, de nouveaux outils fournissant des résultats très généraux ont été développés. Plusieurs modèles sont traités par cette nouvelle approche, y compris les modèles linéaires généralisés pour des données binaires ou de dénombrement, tels que les modèles logistiques, probit et loglinéaires. Nous discutons l'idée de base qui sous-tend ces nouveaux outils, présentons des résultats clés obtenus par leur usage, discutons leurs limites, et offrons quelques problèmes ouverts dans le domaine.

---

**DAVE WOODS** (University of Southampton, UK), **TIM WAITE** (University of Southampton), **TIM WATERHOUSE** (Eli Lilly)

[Tuesday June 14 / mardi 14 juin, 14:07]

*Optimal Block Designs for Non-Normal Data*

*Plans optimaux en blocs pour données non-normales*

A generalized linear mixed model (GLMM) may describe data from a range of experiments in science, technology and medicine where experimental units are arranged in blocks. Conditional on the random effects, the data may be described by a variety of distributions including Bernoulli, Poisson, and gamma. Methods for the construction of optimal designs for GLMMs are described that make use of approximations to the information matrix. In common with designs for other nonlinear models, efficient designs for GLMMs depend on the unknown model parameter values, and hence a Bayesian approach is adopted for design selection.

Un modèle linéaire mixte généralisé (MLMG) peut décrire les données d'une gamme d'expériences en science, en technologie, et en médecine, où les unités expérimentales sont rangées en blocs. Conditionnellement aux effets aléatoires, les données peuvent être décrites par une variété de distributions, y compris Bernoulli, Poisson et Gamma. Les méthodes pour la construction de plans optimaux pour les MLMG sont décrites pour utiliser les approximations dans la matrice d'information. Tout comme les plans pour d'autres modèles non-linéaires, les plans efficaces pour les MLMG dépendent de la valeur inconnue des paramètres du modèle, d'où le choix de la méthode bayésienne pour la sélection des plans.



## 2C-I2: New Developments in Mixture Models & Their Applications Du neuf dans les modèles de mélange et leurs applications

---

Organizer and Chair / Responsable et président: Abbas Khalili (McGill University)

**Room/Salle: BAC132**

**Abstracts/Résumés**

---

**GEOFF MCLACHLAN** (University of Queensland, Australia)

[Tuesday June 14 / mardi 14 juin, 13:00]

*Clustering via Mixtures of Skew Normal or Skew t-Distributions*

*Classification via des mélanges de distributions asymétriques normales et de Student*

In the typical application of normal mixture models to clustering, the number of clusters corresponds to the number of normal components in the mixture model. But in cases where the clusters are not elliptically symmetric, this correspondence will not hold if additional normal components are needed to cover the asymmetry in the data and the presence of outliers. We therefore consider the fitting of mixtures of skew normal or skew t-components. Examples of this approach are given in the modeling of flow cytometric data.

Dans une application typique d'usage de modèles de mélanges normaux pour la classification, le nombre de grappes correspond au nombre de composantes normales dans le modèle de mélanges. Mais dans les cas où les grappes ne sont pas elliptiquement symétriques, cette correspondance ne sera pas valide s'il faut d'autres composantes normales pour couvrir l'asymétrie dans les données et la présence des valeurs aberrantes. Nous considérons donc l'ajustement de mélanges de composantes asymétriques normales ou de Student. Quelques exemples de cette approche sont présentés par la modélisation d'un flux de données cytométriques.

---

**JIAHUA CHEN** (University of British Columbia), **PENGFEI LI** (University of Alberta)

[Tuesday June 14 / mardi 14 juin, 13:30]

*Tuning the EM-test for Finite Mixture Models*

*Réglage du test EM pour modèles de mélanges finis*

The EM-test is a recently developed effective and easy-to-use test of the order of a finite mixture model. It has a relatively simple limiting distribution and enjoys broad applicability. The  $p$ -value of the EM-test is approximated via its limiting distribution. The built-in tuning parameter has an important influence on the approximation precision. Thus, choosing an appropriate value for this parameter is an important issue. We develop a computer-experiment approach to address this issue. Through designed experiments, we derive a number of empirical formulas for the tuning parameter. Extensive validation simulation shows that these formulas work well.

Récemment développé, le test EM est un test efficace et facile à utiliser pour l'ordre d'un modèle de mélanges finis. Il a une distribution limite relativement simple et jouit d'une large applicabilité. Le seuil observé du test EM est approximé via sa distribution limite. Le paramètre de réglage intégré à la méthode a une influence importante sur la précision de l'approximation. Le choix d'une valeur appropriée pour ce paramètre est donc important. Nous développons une approche d'expérience informatique pour répondre au problème. À travers les expériences planifiées, nous tirons des formules empiriques pour le paramètre de réglage. Une simulation approfondie valide ces formules en montrant qu'elles marchent bien.

## 2C-I2: New Developments in Mixture Models & Their Applications Du neuf dans les modèles de mélange et leurs applications

---

**PAUL MCNICHOLAS & SANJEENA SUBEDI** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 14:00]

*Parameter Estimation in Model-Based Clustering*

*Estimation paramétrique dans l'analyse de regroupements basée sur des modèles*

The most common approach to parameter estimation in model-based clustering, and its semi-supervised analogue model-based classification, is via the expectation-maximization (EM) algorithm or some variant thereof. When using EM algorithms for mixture model parameter estimation, problems arise due to the nature of the likelihood surface; yet it remains the approach of choice within the literature. Variational Bayes approximations have recently emerged as an alternative to the EM algorithm for mixture model parameter estimation. We consider a number of families of mixture models, comparing the performance of variational Bayes approximations to the EM algorithm in each case.

L'approche la plus courante pour l'estimation paramétrique dans l'analyse de regroupements basée sur des modèles, et son analogue semi-supervisée, la classification à base de modèles, consiste à utiliser l'algorithme d'espérance-maximisation (EM), ou l'une de ses variantes. Lors de l'utilisation de l'algorithme EM pour l'estimation du paramètre des modèles de mélanges, des problèmes surviennent de par la nature de la surface de vraisemblance. Pourtant il s'agit de l'approche la plus courante dans la littérature. Les approximations variationnelles bayésiennes sont récemment apparues comme alternatives à l'algorithme EM pour l'estimation du paramètre des modèles de mélanges. Nous considérons un nombre de familles de modèles de mélange, et comparons la performance des approximations variationnelles de Bayes à l'algorithme EM dans chaque cas.

## 2C-13: On the Issues Raised by Environmental Surveys at Statistics Canada Au sujet des questions soulevées par les études environnementales chez Statistique Canada

---

Chair/Président: Pierre Lavallée (Statistique Canada)  
Organizer/Responsable: Nathalie Hamel (Statistique Canada)

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

**Room/Salle: BAC142**

### **Abstracts/Résumés**

---

**JOHN MARSHALL** (Statistics Canada)

[Tuesday June 14 / mardi 14 juin, 13:00]

*The Diverse Needs and Challenges of Environmental Surveys : an Analytical Perspective*  
*La diversité des besoins et des défis des enquêtes sur l'environnement : Une perspective analytique*

Surveys on the environment are a relatively new addition to Statistics Canada. The methodological and analytic challenges presented by these surveys are unlike those faced in the traditional social and business fields. This presentation will explain a little about the history of environmental surveys at Statistics Canada as well as specific challenges posed by two of them: the Survey of Environmental Protection Expenditures, and the Industrial Water Survey.

Les enquêtes sur l'environnement sont un ajout relativement récent à la gamme d'enquêtes menées par Statistique Canada. Les défis méthodologiques et analytiques que présentent ces enquêtes sont différents de ceux rencontrés dans les domaines sociaux traditionnels et d'affaires. Cette présentation abordera l'histoire des enquêtes sur l'environnement à Statistique Canada et discutera des défis spécifiques présentés par deux d'entre elles : l'Enquête sur les dépenses de protection de l'environnement et l'Enquête sur l'eau dans les industries.

---

**MARTIN HAMEL, CLAUDE GIRARD, NATHALIE HAMEL & SÉBASTIEN LABELLE-BLANCHET** (Statistique Canada)

[Tuesday June 14 / mardi 14 juin, 13:30]

*Application of the Methodology Commonly used in Economics Surveys to Environmental Surveys*  
*Application de la méthodologie couramment utilisée pour les enquêtes économiques aux enquêtes environnementales*

During the past six years, the Business Survey Methods Division (BSMD) of Statistics Canada has actively been involved in the redesign or in the development of new environmental surveys. For the implementation of these surveys, the BSMD team used the knowledge and the experience acquired over the years in the development of economic surveys. This presentation will discuss some of the steps involved in the implementation of a business survey at Statistics Canada.

Au cours des six dernières années, la division des méthodes d'enquêtes auprès des entreprises (DMEE) de Statistique Canada a été activement impliquée dans le remaniement ou le développement de nouvelles enquêtes environnementales. Pour la mise en oeuvre de celles-ci, la DMEE a utilisé l'expertise ainsi que l'expérience acquise au fil des ans dans le développement des enquêtes économiques. Cette présentation portera sur certaines étapes nécessaires à la réalisation d'une enquête à Statistique Canada.

## 2C-13: On the Issues Raised by Environmental Surveys at Statistics Canada Au sujet des questions soulevées par les études environnementales chez Statistique Canada

---

**CLAUDE GIRARD** (Statistique Canada)

[Tuesday June 14 / mardi 14 juin, 14:00]

*Addressing Some of the Methodological Issues Raised by Environment Surveys*

*Aborder certains des enjeux méthodologiques soulevés par les enquêtes environnementales*

There is a growing need to measure the impact industrial activities have on the environment, which explains the emergence of environment surveys. The experience gathered recently conducting environment surveys has shown that they differ significantly from economic surveys. In order to carry out environment surveys successfully, we need to rethink some of the methodologies and expand the systems in place as to accommodate their specific needs. In this presentation, we will present some of the key elements that set these surveys apart from the usual economic surveys and discuss some of the methodologies used so far to address their needs.

Le besoin de mesurer l'impact qu'ont les activités industrielles sur l'environnement est grandissant, d'où l'émergence d'enquêtes sur ce sujet. L'expérience acquise ces dernières années avec les enquêtes sur l'environnement montre qu'elles diffèrent significativement des enquêtes économiques. Afin qu'elles soient couronnées de succès, les enquêtes sur l'environnement requièrent le développement de nouvelles méthodes et l'expansion des systèmes en place afin que leurs besoins spécifiques soient pris en compte. Dans cet exposé, nous présenterons certains des éléments qui distinguent ces enquêtes des enquêtes économiques usuelles, ainsi que les méthodologies mises en place jusqu'à maintenant pour subvenir à leurs besoins particuliers.

## 2C-14: Panel on Online Teaching of Statistics Table ronde sur l'enseignement de la statistique en ligne

---

Chair/Président: Bethany White (University of Western Ontario)

Session sponsored by the Statistical Education Committee

Session parrainée par le Comité d'éducation en statistique

**Room/Salle: BAC 237**

**Abstracts/Résumés**

---

**MOHAMED ABDOLELL & JENNIFER PAYNE** (Dalhousie University)

[Tuesday June 14 / mardi 14 juin, 13:00]

*Teaching Statistics in Academic Medical Research Institutions*

*L'enseignement des méthodes de biostatistique dans les institutions académiques de recherche médicale*

Introductory courses in biostatistics aimed at medical residents have traditionally had limited success. In this talk we will discuss the Online Modular Research Methods Course (<http://ResearchMethods.dal.ca>), which employs video podcasts, downloadable PDF lecture notes, embedded content mastery quizzes, custom GUI-driven graphical tools, as well as synchronous and asynchronous methods of teaching/learning to leverage the inherent synergistic relationships between content, pedagogy, and technology. The course is developed and delivered using open source software (R, LaTeX, Moodle) and is freely accessible under the Creative Commons Canada 2.5 License, and is customizable for the needs of individual medical residency programs.

Le cours en ligne "Online Modular Research Methods Course" (<http://ResearchMethods.dal.ca>) emploie des podcasts vidéo, des notes de cours téléchargeables au format PDF, des questionnaires d'évaluation des apprentissages, une interface graphique personnalisée, ainsi que des méthodes synchrones et asynchrones pour l'enseignement et l'apprentissage. Le but est de renforcer la synergie entre le contenu, la pédagogie et la technologie. Le cours est développé et offert avec l'aide d'un logiciel "open source" (R, LaTeX, Moodle) et est accessible gratuitement. En outre, il peut être personnalisé afin de répondre aux besoins des programmes de résidence en médecine.

---

**JOEL DUBIN** (University of Waterloo)

[Tuesday June 14 / mardi 14 juin, 13:00]

*Experiences Teaching a Distance Education Biostatistics Course for Masters in Public Health Students*

*Expérience d'enseignement à distance d'un cours de biostatistique pour les étudiants de maîtrise en santé publique*

Teaching biostatistics to those not enrolled/majoring in statistics/biostatistics can be difficult in its own right. Doing so in a distance format creates additional challenges. In this talk, I will focus on experiences (including our use of R software) teaching a distance education biostatistics course in the University of Waterloo Masters in Public Health program. I will compare and contrast teaching such a course via distance education versus in a traditional lecture format, and provide some take-home messages that hopefully will be useful to those of

Enseigner de la biostatistique à ceux qui ne sont pas encore inscrits qui étudient en statistique et en biostatistique peut être une tâche difficile en soi. Enseigner à distance pose des défis supplémentaires. Cet exposé est axé sur les expériences d'enseignement à distance de biostatistique (y compris sur notre utilisation du logiciel R), dans le programme de maîtrise en santé publique à l'Université de Waterloo. Je vais comparer ce cours, enseigné à distance, avec le cours traditionnel en classe, et les opposer. Je vais fournir quelques messages à retenir, qui j'espère, seront utiles à ceux d'entre vous qui enseigneront peut-être

## 2C-I4: Panel on Online Teaching of Statistics

### Table ronde sur l'enseignement de la statistique en ligne

---

you who may be teaching a similar course in the future.

un cours similaire dans le futur.

---

**ALISON WEIR** (University of Toronto at Mississauga)

[Tuesday June 14 / mardi 14 juin, 13:00]

*The Instructor Dimension of e-Learning in Undergraduate Statistics*

*La dimension de l'enseignant dans la formation en ligne des étudiants de premier cycle en statistique*

e-Learning options proliferate. The technological component of online education has been well studied. Other aspects of e-learning are emerging fields of study. A growing body of work is defining the dimensions that underpin e-learner satisfaction and e-learner success. These dimensions include course flexibility, e-learner computer competency, student diversity and interaction, and the instructor. This talk focuses on the instructor dimension of e-learning. We will discuss instructor competencies, skills, and attitudes necessary for successful e-learning in an undergraduate level course in Applied Regression Analysis. We will contrast these attributes with those necessary for learning in a traditional classroom setting.

Les options de formation en ligne prolifèrent. L'élément technologique de l'enseignement en ligne a été bien étudié. D'autres aspects de la formation en ligne sont des domaines d'étude émergents. Un ensemble de travaux croissant définit les dimensions qui assurent la satisfaction et le succès des apprenants en ligne. Ces dimensions comprennent la flexibilité des cours, les compétences informatiques, la diversité et l'interaction des apprenants en ligne ainsi que l'enseignant. Cet article met l'emphase sur la dimension de l'enseignant à donner des cours en ligne. Nous parlerons des compétences de l'enseignant, des aptitudes et des comportements nécessaires pour enseigner en ligne avec succès à des étudiants de premier cycle en analyse de régression appliquée. Nous allons opposer ces attributs à ceux qui sont nécessaires pour l'enseignement traditionnel en classe.

## 2C-15: Sequential Designs for Computer Experiments Conception séquentielle d'expériences informatiques

---

Chair/Président: Duncan Murdoch (University of Western Ontario)  
Organizer/Responsable: Pritam Ranjan (Acadia University)

**Room/Salle: BAC 138**

**Abstracts/Résumés**

---

**DEREK BINGHAM & JOSLIN GOH** (Simon Fraser University), **JAMES HOLLOWAY** (University of Michigan)

[Tuesday June 14 / mardi 14 juin, 13:00]

*An Application of Sequential Design for Multi-fidelity Computer Models*

*Une application du plan séquentiel aux modèles informatiques de multi-fidélité*

The design and analysis of experiments on simulators of physical systems have become common practice in many physics-based applications. For many of these applications, the computer model can be run in different levels of fidelity. The question of how one should allocate the experimental resources is one of primary importance. It turns out that sequentially allocating the trials can improve the predictive ability of the final model. In this work, we consider an application where experimenters have a multi-fidelity simulator of radiative shocks and also real-world experimenters to consider. Allocating the simulator and experimental trials are the main issues.

La conception et l'analyse d'expériences sur des simulateurs de systèmes physiques sont devenues une pratique courante de plusieurs applications en physique. Le modèle informatique de plusieurs de ces applications peut être exécuté selon différents niveaux de fidélité. La question à savoir comment les ressources expérimentales devraient être allouées est d'une importance primordiale. La répartition séquentielle des essais peut améliorer la capacité de prédiction du modèle final. Nous considérons, dans ce travail, une application où les expérimentateurs ont un simulateur multi-fidélité des chocs radiatifs en plus d'expérimentateurs dans le monde réel. Les principaux enjeux sont la répartition des essais simulés et expérimentaux.

---

**JASON LOEPPKY** (University of British Columbia, Okanagan), **BRIAN WILLIAMS & LESLIE MOORE** (Los Alamos National Laboratory)

[Tuesday June 14 / mardi 14 juin, 13:30]

*Gaussian Process Models for Mixture Experiments*

*Modèles de processus gaussiens pour des expériences de mélanges*

Computer models are often used to simulate outputs of complex physical systems when physical experiments are not possible. In many areas of science the input variables are expressed as a percentage contribution to a total amount of material, resulting in a mixture experiment. In such situations the input variables are correlated because they must sum to one. In the context of a deterministic computer model additional care must be used in building an emulator. We propose use of a Gaussian process with transformed input variables and illustrate the resulting methodology on a chemical compositional simulator and a physical experiment.

Lorsque les expériences physiques sont impossibles, des modèles informatiques sont souvent utilisés pour simuler des sorties de systèmes physiques complexes. Dans plusieurs domaines de la science, les variables prédictives sont exprimées en pourcentage de contribution à une quantité totale de matériel, produisant une expérience de mélange. Dans ces cas, les variables prédictives sont corrélées puisqu'elles doivent totaliser un. Dans le contexte d'un modèle informatique déterministe, un soin particulier doit être accordé à la construction d'un émulateur. Nous proposons l'utilisation d'un processus gaussien avec des variables prédictives transformées, et nous illustrons la méthodologie résultante au moyen d'un simulateur compositionnel chimique et d'une expérience physique.

## 2C-I5: Sequential Designs for Computer Experiments Conception séquentielle d'expériences informatiques

---

**ROBERT B GRAMACY** (University of Chicago), **HERBIE LEE** (University of California, Santa Cruz)

[Tuesday June 14 / mardi 14 juin, 14:00]

*Optimization under Unknown Constraints*

*Optimisation dans le cadre de contraintes inconnues*

We consider optimization for computer experiments under unknown constraints, i.e., when simulation is required to determine real-valued responses and check constraints. We develop GP surrogates to approximate both simulator outputs. A new integrated improvement criterion recognizes that responses that violate the constraint are still informative about the function, and thus potentially useful in the optimization. We illustrate our approach on a problem from health care policy.

Nous considérons l'optimisation pour des expériences informatiques dans le cadre de contraintes inconnues, c'est-à-dire lorsqu'une simulation est requise afin de déterminer les réponses de valeurs réelles et vérifier les contraintes. Nous développons des processus gaussien substitués pour approximer les deux sorties du simulateur. Un nouveau critère intégré d'amélioration établit que les réponses qui ne respectent pas la contrainte permettent quand même d'en apprendre sur la fonction, et sont ainsi potentiellement utiles pour l'optimisation. Nous illustrons notre approche à l'aide d'un problème dans la politique des soins de santé.



## 2C-C1: Distributions: Theory and Application Distributions : théorie et application

---

Chair/Président: Paramjit Gill (University of British Columbia Okanagan)

Room/Salle: BAC 236

### Abstracts/Résumés

---

**IMAN MABROUK** (University of Western Ontario), **SERGE PROVOST** (University of Western Ontario)

[Tuesday June 14 / mardi 14 juin, 13:00]

*An Extended Inverse Gaussian Model*

*Un modèle gaussien inverse étendu*

The generalized inverse Gaussian distribution popularized in the late seventies by Ole Barndorff-Neilsen is extended by incorporating an additional parameter in its density function. A reduced version of the resulting model is also being considered. The effects of the parameters are pointed out for both distributions. Additionally, several of their statistical functions are given explicitly. Data sets pertaining to maximum flood levels and snowfall precipitations were fitted with various statistical models for comparison purposes.

La distribution gaussienne inverse généralisée rendue populaire à la fin des années soixante-dix par Ole Barndorff-Neilsen est étendue en incorporant un paramètre additionnel dans sa fonction de densité. Une version réduite du modèle résultant est aussi considérée. Les effets des paramètres sont discutés pour les deux distributions. De plus, leurs fonctions statistiques sont présentées explicitement. Des données sur les niveaux de crue maximale et de précipitations de chute de neige sont ajustées avec différents modèles statistiques pour comparaison.

---

**WEIYU QIU** (School of Public Health, University of Alberta), **YUTAKA YASUI** (Department of Public Health Sciences, School of Public Health, University of Alberta)

[Tuesday June 14 / mardi 14 juin, 13:15]

*Elliptical Distributions with Linear Conditional Second Moments: Multivariate Pearson Type VII Family and its Applications*

*Distributions elliptiques avec seconds moments conditionnels linéaires: la famille de Pearson multivariée de Type VII et ces applications*

Multivariate normal (MVN) distributions have constant conditional second moments which may not represent real underlying stochastic processes. In contrast, we show that, among elliptical distributions, conditional second moments are linear in the empirical second moment of the conditioning vector if and only if the distribution belongs to the symmetric multivariate Pearson Type VII (MP VII) family. We introduce a graphical approach to assess MP VII vs. MVN distributions, and show, by simulation, under-coverage of MVN-based conditional prediction intervals under the MP VII family. The application of the MP VII family is illustrated using real multivariate data.

Les distributions normales multivariées (NMV) ont des seconds moments conditionnels constants qui peuvent ne pas représenter les réels processus stochastiques sous-jacents. En revanche, nous montrons que, parmi les distributions elliptiques, les seconds moments conditionnels sont linéaires aux seconds moments empiriques du vecteur conditionnel si et seulement si la distribution fait partie de la famille de Pearson multivariée symétrique de type VII (PM VII). Nous présentons une approche graphique pour évaluer les distributions PM VII contre NMV, et montrons par simulation la sous-couverture des intervalles de prédiction conditionnels basés sur les distributions NMV sous la famille PM VII. L'application de la famille PM VII est illustrée en utilisant de réelles données multivariées.

## 2C-C1: Distributions: Theory and Application Distributions : théorie et application

---

**HANNA JANKOWSKI** (York University), **JON WELLNER** (University of Washington)

[Tuesday June 14 / mardi 14 juin, 13:30]

*Estimation of a Discrete Monotone Distribution*

*Estimation d'une distribution discrète monotone*

We study and compare three estimators of a discrete monotone distribution: (a) the (raw) empirical estimator; (b) the "method of rearrangements" estimator; and (c) the maximum likelihood estimator. We show that the MLE strictly dominates both the rearrangement and empirical estimators in cases when the distribution has intervals of constancy. For example, when the distribution is uniform on  $\{0, \dots, y\}$ , the asymptotic risk of the method of rearrangements estimator (in squared  $\ell_2$  norm) is  $y/(y+1)$ , while the asymptotic risk of the MLE is of order  $(\log y)/(y+1)$ . For strictly decreasing distributions, the estimators are asymptotically equivalent.

Nous étudions et comparons trois estimateurs d'une distribution discrète monotone : (a) l'estimateur empirique (brute); (b) l'estimateur de la « méthode des réarrangements » ; et (c) l'estimateur du maximum de vraisemblance (EMV). Nous montrons que l'EMV domine strictement les estimateurs empirique et de réarrangements dans les cas où la distribution est constante par intervalles. Par exemple, lorsque la distribution est uniforme sur  $\{0, \dots, y\}$ , le risque asymptotique de l'estimateur de la méthode des réarrangements (norme carrée  $\ell_2$ ) est  $y/(y+1)$ , et celui de l'EMV est de l'ordre de  $(\log y)/(y+1)$ . Pour les distributions strictement décroissantes, les estimateurs sont asymptotiquement équivalents.

**ZHENG SUN** (Simon Fraser University)

[Tuesday June 14 / mardi 14 juin, 13:45]

*K sample Cramer-von Mises Tests for Grouped Data*

*Test de Cramer-von Mises à k échantillons pour les données groupées*

$K$  sample Cramer-von Mises tests are developed for grouped data, including the Cramer-von Mises statistic  $W^2$ , the Watson's statistic  $U^2$ , and the Anderson-Darling statistic  $A^2$ . Asymptotic percentage points may be calculated and are shown to give good approximations to points for finite samples. Cramer-von Mises statistics are shown to be powerful against a range of alternatives. Comparisons are made with some grouping corrections procedures. Some real data sets are analyzed.

Les tests de Cramer-von Mises à  $K$  échantillons ont été développés pour les données groupées, incluant la statistique de Cramer-von Mises  $W^2$ , la statistique de Watson  $U^2$ , et la statistique d'Anderson-Darling  $A^2$ . Les points de pourcentage asymptotiques peuvent être calculés et donnent de bonnes approximations aux points pour les échantillons de taille finie. Les statistiques de Cramer-von Mises seront montrées comme puissantes contre une gamme d'alternatives. Des comparaisons seront faites avec quelques procédures de corrections dues aux groupes. Quelques données réelles seront étudiées.

**VAL ANDREI FAJARDO** (University of Waterloo), **MEI LING HUANG** (Brock University)

[Tuesday June 14 / mardi 14 juin, 14:00]

*On Confidence Regions for High Quantiles of a Heavy Tailed Distribution*

*À propos des régions de confiance des quantiles supérieures d'une distribution à ailes relevées*

High quantile estimation of a heavy-tailed distribution is important in real life applications. Traditional maximum likelihood and moment estimators sometimes fail to make sense, or are nonexistent.

L'estimation des quantiles supérieurs d'une distribution à ailes relevées est importante dans des applications réelles. Les estimateurs traditionnels du maximum de vraisemblance et du moment sont parfois illogiques ou inexis-

## 2C-C1: Distributions: Theory and Application

### Distributions : théorie et application

---

tent. The Hill-Weissman estimator also has problems. This paper proposes a new estimator of high quantile that aims to reduce the bias and mean square error. The new estimator contributes narrower, more efficient confidence intervals for high quantiles. Computational simulations show that the new estimator is more efficient relative to the Hill-Weissman estimator. Finally, the paper shows that the new estimator applying to a real world example is superior to the Hill-Weissman method.

tants. L'estimateur de Hill-Weissman comporte également des problèmes. Cet article propose un nouvel estimateur des quantiles supérieurs qui vise à réduire le biais et l'erreur quadratique moyenne. Ce nouvel estimateur présente des intervalles de confiance plus étroits et plus efficaces pour les quantiles supérieurs. Des simulations informatiques montrent que le nouvel estimateur est plus efficace comparativement à l'estimateur de Hill-Weissman. Enfin, l'article montre que le nouvel estimateur appliqué à un exemple réel est supérieur à la méthode Hill-Weissman.

## 2C-C2: Genetic Association and Analysis of Families Association génétique et analyse des familles

---

Chair/Président: Angelo Canty (McMaster University)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 234**

**Abstracts/Résumés**

---

**KARIM OUALKACHA** (McGill University, Department of Epidemiology, Biostatistics and Occupational Health), **AURÉLIE LABBE & ANTONIO CIAMPI** (McGill University Department of Epidemiology, Biostatistics and Occupational Health), **MARC-ANDRÉ ROY & MICHEL MAZIADE** (Centre de Recherche Université Laval Robert Giffard)

[Tuesday June 14 / mardi 14 juin, 13:00]

*Principal Components of Heritability for High Dimension Quantitative Traits and General Pedigrees*

*Des composantes principales d'héritabilité pour des traits quantitatifs de grande dimension et pour des pedigrees quelconques*

For many complex disorders, genetically relevant disease definition is still unclear. Therefore, researchers tend to collect large numbers of items related directly or indirectly to the disease diagnostic. To reduce dimensionality, Ott and Rabinowitz (1999) introduced the principal components of heritability (PCH). The PCHs' calculation is based on the variance components' estimation (VC). In the genetic context, the VC' standard estimators are Lange's MLE, which require complex calculations. Here we propose ANOVA estimators of the VC, which permit the PCH's approach to be applied to general pedigrees and high dimensional family data. Simulations and real data analysis illustrate our methodology.

Pour plusieurs maladies complexes, la définition des caractéristiques génétiques reste ambiguë. Pour cette raison, les chercheurs tentent de collecter un grand nombre de traits reliés à la maladie. Pour réduire la dimensionnalité, Ott et Rabinowitz (1999) ont introduit les composantes principales d'héritabilité (CPH). Le calcul des CPHs est basé sur l'estimation des composantes de variance (CV). En génétique, les estimateurs standards des CV sont les EMV de Lange qui nécessitent des calculs complexes. Ici, nous proposons des estimateurs d'ANOVA des CV. Ceux-ci peuvent permettre à l'approche des PCH d'être appliquée aux pedigrees quelconques et aux données familiales de grande dimension. Des simulations et des analyses avec les données réelles sont utilisées pour illustrer la méthodologie.

---

**LUCIA MIREA** (Maternal-Infant Care Research Centre (MiCare), Mount Sinai Hospital), **SHELLEY BULL** (Samuel Lunenfeld Research Institute, Mount Sinai Hospital), **LEI SUN** (Dalla Lana School of Public Health, University of Toronto)

[Tuesday June 14 / mardi 14 juin, 13:15]

*Incorporating Evidence of Population Stratification Bias in Genetic Association Analyses Combining Individual and Family Data*

*Incorporer la preuve d'un biais de stratification population dans les analyses d'association génétique combinant des données individuelles et familiales*

A novel weighted framework is proposed to combine estimates from robust family-based analysis and individual-level genetic association subject to population stratification bias. The weighted

Un nouveau cadre pondéré est proposé pour combiner les estimateurs issus d'analyse robuste basée sur la famille et d'analyse d'association génétique au niveau de l'individu exposés au biais de stratification population. L'approche pon-

## 2C-C2: Genetic Association and Analysis of Families Association génétique et analyse des familles

---

approach generalizes existing methods by using a continuous function that depends only on the observed p-value comparing individual-level and family-based estimates, instead of a binary one that also depends on specification of an arbitrary testing level. Simulations of quantitative trait and case-control data show that in comparison to existing methods, the weighted approach has 5% type I error closer to the nominal level, increased (decreased) accuracy for larger (smaller) bias, and overall increased positive predictive value.

dérée généralise les méthodes existantes en utilisant une fonction continue qui dépend seulement de la valeur-p observée en comparant les estimateurs au niveau de l'individu et de la famille, au lieu d'une fonction binaire qui dépend également de la spécification d'un niveau de test arbitraire. Des simulations sur des données de trait quantitatif et de cas-témoins montrent que comparée aux méthodes existantes, l'approche pondérée a une erreur de type I à 5% plus proche du niveau nominal, faisait accroître (décroître) la précision pour un biais plus grand (plus petit), et globalement augmentait la valeur positive de prédiction.

---

**ELIF FIDAN ACAR** (McGill University), **LEI SUN** (University of Toronto)

[Tuesday June 14 / mardi 14 juin, 13:30]

*A Generalized Kruskal-Wallis Test Incorporating Group Uncertainty with Application to Genome-Wide Association Studies*

*Test de Kruskal-Wallis généralisé avec incertitude au niveau du groupe et ses applications aux études d'association sur le génome entier*

Motivated by genome-wide association studies of imputed SNPs, we propose a generalization of the Kruskal-Wallis test incorporating group uncertainty when comparing  $k$  samples. The extended test statistic is based on probability-weighted rank-sums and follows an approximate chi-square distribution with  $k - 1$  degrees of freedom under the null hypothesis. Simulation studies demonstrate the proposed test's superiority over commonly used imputation-based parametric predecessors. Using real data from a genome-wide association study of Type I diabetic complications, we attain robust performance in detecting putative associations under genotype uncertainty.

Motivés par les études d'association des SNP imputés sur le génome entier, nous proposons une généralisation du test de Kruskal-Wallis qui tient compte de l'incertitude au niveau du groupe lors de la comparaison de  $k$  échantillons. La statistique du test est fondée sur des sommes de rangs pondérées par des probabilités et suit approximativement une chi-deux avec  $k - 1$  degrés de liberté sous l'hypothèse nulle. Des études de simulation démontrent la supériorité du test proposé par rapport à ses prédécesseurs paramétriques basés sur l'imputation. À l'aide de données réelles tirées d'une étude d'association des complications du diabète type 1 sur le génome entier, nous obtenons une performance robuste dans la détection d'associations présumées en présence d'incertitudes au niveau du génotype.

---

**LAURENT BRIOLLAIS** (Samuel Lunenfeld Research Institute)

[Tuesday June 14 / mardi 14 juin, 13:45]

*Bayesian Graphical Models (BGMs) for Genetic Association Studies*

*Modèles graphiques bayésiens (MGB) pour études d'association génétique*

The application of BGMs to genetic association studies is not new but the emerging development of high-throughput technologies in this field have raised important statistical challenges for this methodology, including variable selection, the study of complex correlation structures and statistical inference about the genetic associations. Through various applications, we review

L'application des MGB aux études d'association génétique n'est pas nouvelle mais le développement émergent des technologies à haut débit dans ce domaine a soulevé d'importants défis statistiques concernant cette méthodologie, telle la sélection de variables, l'étude des structures de corrélations complexes et l'inférence statistique concernant les associations génétiques. À travers plusieurs applications, nous faisons la revue de l'intérêt des MGB pour la sélection

## 2C-C2: Genetic Association and Analysis of Families Association génétique et analyse des familles

---

the interest of BGMs to perform variable selection in genome-wide association studies (GWAS), to study complex patterns of linkage disequilibrium in candidate gene studies and to infer complex gene networks in pathway-based analyses. Finally, we discuss future promising applications of BGMs to genetic and genomic studies.

de variables dans les études d'association pangénomique (EAP), afin d'étudier les structures complexes du déséquilibre de liaison dans les études du gène candidat et d'inférer les réseaux complexes de gène dans les analyses basées sur les chemins. Finalement, nous discutons des applications promettantes pour l'avenir des MGB pour les études génétiques et génomiques.

---

**ANTONIO CIAMPI, LIN YANG & AURÉLIE LABBE** (McGill University), **CHANTAL MERETTE** (Laval University)

[Tuesday June 14 / mardi 14 juin, 14:00]

*PLS Regression and Hybrid Methods in Genomics Association Studies*

*Régression PLS et méthodes hybrides dans les études d'associations en génomiques*

Single Nucleotide Polymorphisms (SNPs) are studied in genomics to detect associations between genes and diseases. Partial Least Square (PLS) regression is used as a tool to extract a few orthogonal components from correlated SNP variables. We propose two hybrid methods for assessing the impact of SNP variables on case/control status. Method 1 uses the most informative PLS components to build a classification tree, using the case/control variable as outcome. Method 2 performs a hierarchical classification on these PLS components. We show on a real example that PLS outperforms several competing approaches. Moreover, both Method 1 and Method 2 outperform PLS.

Les SNP sont étudiés en génomique afin de découvrir les associations entre gènes et maladie. La régression PLS réduit efficacement la dimensionnalité de l'espace des variables SNP. Nous proposons deux méthodes hybrides pour jauger l'effet de ces variables sur la variable binaire cas/témoin. La méthode 1 construit un arbre de classification à partir des composantes PLS les plus importantes ; la méthode 2 construit une classification hiérarchique ascendante à partir de ces composantes. Appliqué à un exemple réel, la régression PLS est plus performante que plusieurs approches compétitives. De surcroît, les deux méthodes proposées performant mieux que la régression PLS.

## 2C-C3: Optimizing Products and Processes Optimisation des produits et processus

---

Chair/Président: Vickneswary Tagore (University of Saskatchewan)

Session sponsored by the Business and Industrial Statistics Section

Session parrainée par le Groupe de statistique industrielle et de gestion

**Room/Salle: BAC141**

**Abstracts/Résumés**

---

**SHAUN S. WULFF** (University of Wyoming), **JOHN BREWSTER** (University of Manitoba)

[Tuesday June 14 / mardi 14 juin, 13:00]

*Objective Bayesian Inference of Fixed Effects in Industrial Split-Plot Experiments*

*Inférence bayésienne objective des effets fixes dans des expériences par parcelles divisées*

In industrial split-plot experiments, inferences for the whole-plot and subplot effects rely on estimates of the variance components in the model. In some situations, classical ANOVA estimates of some of the variance components may be negative and the resulting inferences for the whole-plot and subplot effects may have poor post-data coherence properties. To avoid these difficulties, an alternative is to use Bayesian procedures under a non-informative prior. Bayesian posterior intervals for the whole-plot and subplot effects are obtained directly along with Satterthwaite-type approximations. The approximate posterior intervals can be acquired using frequentist confidence intervals based on an adjusted ANOVA table.

Dans les expériences industrielles par parcelles divisées, les inférences sur les effets des parcelles entières ou des sous-parcelles se fondent sur les estimations des composantes de la variance dans le modèle. Dans certaines situations, les estimations de quelques composantes de la variance d'une ANOVA classique peuvent être négatives et la cohérence a posteriori des inférences des effets des parcelles entières ou des sous-parcelles qui en découlent peuvent être mauvaise. Des procédures bayésiennes non-informatives peuvent être employées pour éviter ces difficultés. Des intervalles bayésiens a posteriori pour les effets des parcelles entières ou des sous-parcelles peuvent être obtenus directement en plus des approximations de type Satterthwaite. Ces intervalles approximatifs a posteriori peuvent également être obtenus en employant des intervalles de confiance fréquentistes à partir d'une table d'ANOVA ajustée.

---

**LIN FANG & ROMAN VIVEROS-AGUILERA** (McMaster University)

[Tuesday June 14 / mardi 14 juin, 13:15]

*Statistical Quality Control Charts to Monitor Processes Under Competing Risks*

*Les cartes statistique de contrôle de la qualité pour surveiller des processus à risques compétitifs*

Exponentially weighted moving average control charts based on cause-specific conditional expected life have been used to detect cause-specific mean-to-failure changes under competing risks. They work well for normal failure times with moderate sample sizes. However, for small samples (i.e.  $n=1$ ) or skewed underlying distributions the method underperforms. We study cumulative sum (CUSUM) control charts based on the log-

Les cartes de contrôle à moyenne mobile avec poids exponentiels basées sur une durée de vie liée à une cause spécifique sont utilisées pour détecter des changements du temps moyen de défaillance spécifique à la cause sous des risques compétitifs. Ces cartes fonctionnent bien lorsque les temps de défaillance sont distribués normalement avec des échantillons de taille moyenne. Cependant, la méthode est peu efficace lorsque les échantillons sont de petite taille (ex.  $n=1$ ) ou que les distributions sous-jacentes sont asy-

## 2C-C3: Optimizing Products and Processes Optimisation des produits et processus

---

likelihood ratio statistic to detect changes in cause-specific mean-to-failure in a general setting under competing risks. The performance of both approaches has been evaluated by simulation studies showing better performance for the CUSUM in small samples. An illustrative example is provided.

métriques. Nous étudions les cartes de contrôle de sommes cumulées (CUSUM) basées sur la statistique du rapport de vraisemblances pour détecter des changements du temps moyen de défaillance spécifique à la cause dans un cadre général sous des risques compétitifs. Des études par simulation ont permis d'évaluer la performance des deux approches et ont montré que la performance de la CUSUM pour de petites tailles d'échantillon est meilleure. Un exemple illustre le tout.

---

**LU XIN & MAHESH PANDEY** (University of Waterloo)

[Tuesday June 14 / mardi 14 juin, 13:30]

*Using Mixed-Effects Regression Model to Analyze the Deformation in Nuclear Reactor Components*

*L'utilisation d'un modèle de régression à effets mixtes pour analyser la déformation des composantes d'un réacteur nucléaire*

Nuclear reactor core consists of pressurized components, also known as pressure tubes, which are exposed to high radiation, temperature and pressure. This results in deformation of tubes, as measured by the creep strain. The creep strain data recorded over time exhibit heteroscedasticity and inter-cluster correlation. This paper presents a comprehensive analysis of data using mixed-effects regression model in combination with a variance function to model the heteroscedasticity. Several interesting aspects are discussed.

Le coeur d'un réacteur nucléaire est formé de composantes pressurisées, également connues sous le nom de tubes de pression, qui sont exposées à des radiations, températures et pressions élevées. Ceci entraîne la déformation des tubes qui est mesurée par la déformation due au fluage. La déformation due au fluage enregistrée au fil du temps démontre de l'hétéroscédasticité et une corrélation inter-grappe. Cet article comporte une analyse complète des données en utilisant un modèle de régression à effets mixtes combiné à une fonction de variance pour modéliser l'hétéroscédasticité. Nous discutons de plusieurs aspects intéressants.

---

**MEI TANG** (Sanofi Pasteur Canada)

[Tuesday June 14 / mardi 14 juin, 13:45]

*Statistical Power Analysis in Vaccine Development in Non-Clinical Areas*

*Analyse de la puissance statistique dans le développement de vaccins pour les domaines non cliniques*

The studies at assay development or early development stage in vaccine development include animal studies and experiments in assay development with objectives of assessing the performance of assays and developing the initial safety and efficacy profiles. Although the requirement of sample size determinations from regulatory agencies for studies in non-clinical area is not as strict as the ones in full development, the power analysis is still essential. The initiatives of the power analysis and the statistical considerations including the Generalized Linear Mixed Model (GLMM) will be discussed in detail.

Les études à l'étape du développement des dosages ou au tout début du développement d'un vaccin incluent des études sur les animaux et des expériences de développement du dosage ayant pour but l'évaluation de la performance des dosages et le développement des profils initiaux de sécurité et d'efficacité. Bien que les exigences provenant des agences à vocation réglementaire quant au choix de la taille d'échantillon pour les études dans les domaines non cliniques ne soient pas aussi strictes que celles à plein déploiement, l'analyse de la puissance est toujours essentielle. Nous discuterons en détail d'initiatives sur l'analyse de la puissance et de considérations statistiques, incluant le modèle linéaire mixte généralisé (MLMG).



## 2C-C3: Optimizing Products and Processes Optimisation des produits et processus

---

**ANDREY FENDYUR & GIOVANI J.C. DA SILVEIRA** (University of Calgary)

[Tuesday June 14 / mardi 14 juin, 14:00]

*Statistical Detection of Moderation and Mediation Effects among Predictors in Industrial Productivity Research*

*Détection statistique des effets de modération et de médiation entre les variables explicatives dans la recherche sur la productivité industrielle*

The industrial productivity growth rate has been an important and multi-facet phenomenon. Many factors of economy affect it. The current research produced three contributions: (1) identified factors influential during the 2009 recession; (2) statistically detected their moderation and mediation effects on the industrial productivity growth; (3) added to understanding of the mechanism how various economic factors affect the industrial productivity growth. The findings may be used in (i) industry: for investment and operations forecasting and planning; (ii) public policy: for industrial and financial policy development; (iii) academia: for further research into the industrial productivity growth.

Le taux de croissance de la productivité industrielle s'avère être un phénomène important à plusieurs facettes. Il a été affecté par plusieurs facteurs de l'économie. Cette recherche a amené trois apports : (1) l'identification des facteurs ayant eu une influence lors de la récession de 2009 ; (2) la détection de manière statistique des effets de modération et de médiation sur la croissance de la productivité industrielle ; (3) une compréhension accrue du mécanisme expliquant comment différents facteurs économiques affectent la croissance de la productivité industrielle. Les résultats peuvent être utilisés pour (i) l'industrie : l'investissement et les opérations de prévision et de planification ; (ii) la politique publique : le développement de politiques industrielle et financière ; (iii) le domaine académique : l'avancement de la recherche sur la croissance de la productivité industrielle.

## 2C-C4: Survival Models: Correlated Data and Non-parametric Estimation Modèles de survie : données corrélées et estimation non paramétrique

---

Chair/Président: Julie Atherton (McGill University)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 241**

**Abstracts/Résumés**

---

**HENRIK STRYHN** (Atlantic Veterinary College, University of Prince Edward Island), **ADEL ELGHAFGHUF** (Atlantic Veterinary College, UPEI)

[Tuesday June 14 / mardi 14 juin, 13:00]

*Estimating Correlated Frailties in a Cox Model*

*Estimer les fragilités partagées corrélées dans le modèle de Cox*

The inclusion of random effects (frailties) in survival analysis has become a standard approach to account for hierarchical data structures. It is less common to include random regression coefficients ("random slopes"), perhaps due to both the computational complexity and the scarcity of datasets that support such modelling. Using four estimation methods available in standard software and previously published data from veterinary epidemiology, with animals clustered in herds, we demonstrate how the inclusion of herd-level correlated frailty terms may substantially affect fixed effect estimates and inferences. An additional simulation study provided a thought-provoking comparison of the performance of the four methods.

L'inclusion des effets aléatoires (fragilité partagée) dans l'analyse de survie est devenue une approche standard pour tenir compte des structures de données hiérarchiques. Il est par contre beaucoup plus rare d'inclure les coefficients de régression aléatoires (« pente aléatoire ») dû à la complexité calculatoire et au manque de données supportant ce genre de modèle. En utilisant quatre méthodes d'estimation disponibles dans les logiciels standards et des données provenant d'épidémiologie vétérinaire sur des animaux groupés en troupeaux, nous démontrerons comment l'inclusion de fragilité partagée corrélée en termes de niveaux de troupeau peut substantiellement affecter les estimations des effets fixes et leurs inférences. Une simulation additionnelle apportera une comparaison étonnante sur la performance des quatre méthodes.

---

**YI NIU & YINGWEI PENG** (Queen's University)

[Tuesday June 14 / mardi 14 juin, 13:15]

*A Marginal Mixture Cure Model for Clustered Survival Data*

*Un modèle marginal de mélange avec fraction non à risque pour données de survie en grappes*

Marginal models have been widely used to analyze correlated survival data. Most of the studies assume that the event of interest will eventually occur given sufficiently long follow-up. We consider a marginal mixture cure model for clustered failure time data with a possible surviving fraction, and propose novel generalized estimating equations to incorporate the correlation within clusters in the marginal model. A simulation study demonstrates the substantial efficiency gain over the con-

Les modèles marginaux ont été couramment utilisés pour l'analyse de données de survie corrélées. La plupart des études suppose que l'évènement d'intérêt se réalisera éventuellement si la suivie est suffisamment longue. Nous considérons un modèle marginal de mélange avec fraction non à risque pour données de temps de défaillance en grappes avec une possible fraction de survie, et proposons des équations d'estimation généralisées pour incorporer la corrélation intra-classe dans le modèle marginal. Une étude par simulation démontre le gain substantiel d'efficacité par rap-

## 2C-C4: Survival Models: Correlated Data and Non-parametric Estimation Modèles de survie : données corrélées et estimation non paramétrique

---

ventional EM method. The model and the proposed method are applied to a data set of failure times from a cancer study.

port à la méthode conventionnelle espérance-maximisation. Le modèle et la méthode proposée sont appliqués à un jeu de données de temps de défaillance provenant d'une étude sur le cancer.

---

**HEDY JIANG & PATRICK BROWN** (Cancer Care Ontario and McMaster University)

[Tuesday June 14 / mardi 14 juin, 13:30]

*Survival Models for A Large Cohort Study in Pickering*

*Modèles de survie pour une vaste étude de cohorte à Pickering*

In population-based cancer incidence studies, the majority of people are not affected by cancer. The cancer-free fraction is of interest and a useful measure to monitor trends in cancer occurrences. In a Pickering cohort study regarding the risk of radiation sensitive cancers in relation to residential proximity to a large nuclear plant, we propose a cure rate model with spatial frailties to estimate the cancer-free fraction and the spatial dependence of census tracts. A Gaussian random field with a Matern correlation function is added to the baseline hazard to model the intracluster correlation in census tracts.

Dans les études sur l'incidence du cancer basées sur la population, la majorité des individus ne sont pas affectés par le cancer. La proportion de la population non atteinte est d'intérêt puisqu'elle constitue une mesure utile pour surveiller les tendances d'apparition de cancers. Dans une étude de cohorte de Pickering sur les risques de cancers sensibles aux radiations en lien avec la proximité résidentielle d'une centrale nucléaire, nous proposons un modèle du taux de guérison avec fragilités spatiales pour estimer la proportion de la population non atteinte et la dépendance spatiale des secteurs de recensement. Un champ aléatoire gaussien avec une fonction de corrélation de Matérn est ajouté au risque de base pour modéliser la corrélation intragroupe dans les secteurs de recensement.

---

**DAVID E. MATTHEWS** (University of Waterloo)

[Tuesday June 14 / mardi 14 juin, 13:45]

*Exact Nonparametric Likelihood Confidence Bands for the Survivor Function*

*Bornes de confiance par vraisemblance exacte non-paramétrique de la fonction de survie*

A method to produce exact simultaneous confidence bands for the empirical cdf that was first described in Owen (1995), and later modified by Jager and Wellner, is the starting point for deriving simultaneous confidence bands for the survivor function,  $\mathcal{F}(\cdot)$ , of any positive random variable. We invert a nonparametric likelihood test of uniformity to obtain simultaneous lower and upper bands for  $\mathcal{F}(\cdot)$  with global confidence level  $1 - \alpha$ . Noé's recursion provides the computational engine. Various aspects of these exact bands are investigated, using as an illustration survival times for non-Hodgkin's lymphoma patients with Stage 4, advanced disease.

Une méthode pour produire les bornes de confiance simultanées exactes de la fonction de répartition (f.d.r.) empirique fut premièrement décrite par Owen (1995), puis modifiée par Jager et Wellner, est notre point de départ pour obtenir les bornes de confiance simultanées de la fonction de survie,  $\mathcal{F}(\cdot)$ , de n'importe quelle variable aléatoire. Nous inversons un test de vraisemblance non-paramétrique d'uniformité pour obtenir les bornes inférieures et supérieures simultanées de  $\mathcal{F}(\cdot)$  avec un niveau de confiance global  $1 - \alpha$ . Nous utilisons la récursion de Noé. Différents aspects de ces bornes exactes sont étudiés, utilisant des données de temps de survie provenant de patients atteints du stage 4 du lymphome non-Hodgkin.

## 2C-C4: Survival Models: Correlated Data and Non-parametric Estimation Modèles de survie : données corrélées et estimation non paramétrique

---

**MOHAMMAD DEGHAN & THIERRY DUCHESNE** (Laval University)

[Tuesday June 14 / mardi 14 juin, 14:00]

*Nonparametric Estimation of the Conditional Survival Function with Time-Varying Covariate and Interval Censoring*  
*Estimation non-paramétrique de la fonction de survie conditionnelle avec covariable qui dépend du temps et censure par intervalle*

In this talk we propose a nonparametric approach to estimate the conditional survival function of failure time given a time-varying covariate,  $S(t|z(y); 0 \leq y \leq t)$ , when time is subject to interval censoring. In this case, observations appear as  $\{(L_i, R_i, Z(t_{i,j})); j = 1, \dots, K_i, i = 1, \dots, n\}$ . We assume that  $Z(t_{i,j}) = \alpha_i + \beta_i t_{i,j} + \varepsilon_{ij}; j = 1, \dots, k_i$ , where  $\alpha_i$  and  $\beta_i$  are random effects. We treat the estimate of the slope  $\beta_i$  of the  $i$ -th path as a positive covariate. We propose a weighted average estimator based on generalized Turnbull and Kaplan-Meier estimators to estimate  $P(T_i > t|z(y_i); 0 \leq y_i \leq t)$ ,  $P(T_i > t|T_i > t_{i0}, z(y_i); 0 \leq y_i \leq t)$  and the  $\gamma$ -th quantile of the distribution of  $\{T_i|z_i(y_i); 0 \leq y_i \leq t\}$ ,  $q_{\gamma i}$ .

Dans cet exposé, nous proposons une approche non-paramétrique pour estimer la fonction de survie conditionnelle du temps de défaillance étant donné une covariable qui dépend du temps,  $S(t|z(y); 0 \leq y \leq t)$ , lorsque ce temps de défaillance est soumis à une censure par intervalle. Dans ce cas, les observations sont de la forme  $\{(L_i, R_i, Z(t_{i,j})); j = 1, \dots, K_i, i = 1, \dots, n\}$ . On suppose que  $Z(t_{i,j}) = \alpha_i + \beta_i t_{i,j} + \varepsilon_{ij}; j = 1, \dots, k_i$ , où  $\alpha_i$  et  $\beta_i$  sont des effets aléatoires. Nous traitons l'estimation de la pente  $\beta_i$  de la  $i$ -ème trajectoire comme une covariable positive. Nous proposons une méthode basée sur une moyenne pondérée des estimateurs généralisés de Turnbull et de Kaplan-Meier pour estimer  $P(T_i > t|z(y_i); 0 \leq y_i \leq t)$ ,  $P(T_i > t|T_i > t_{i0}, z(y_i); 0 \leq y_i \leq t)$  et le  $\gamma$ -ème percentile de la distribution de  $\{T_i|z_i(y_i); 0 \leq y_i \leq t\}$ ,  $q_{\gamma i}$ .

**ZHIHUI (AMY) LIU, BENJAMIN RICH & JAMES HANLEY** (McGill University)

[Tuesday June 14 / mardi 14 juin, 14:15]

*Recovering the Raw Data behind a Kaplan-Meier or Nelson-Aalen Survival Curve*  
*Retrouver les données brutes à partir d'une courbe de survie de Kaplan-Meier ou de Nelson-Aalen*

It is preferable but not always possible to obtain from the authors the raw data that were used to create published Kaplan-Meier and Nelson-Aalen survival curves. Using examples, we show that much of the original information can be recovered from the curve itself, by measuring the lengths of the vertical steps – by digitization or accessing the postscript files. These reconstruction techniques are useful for retrieving the raw data when there is no censoring; when either the numbers at risk or the numbers of deaths at various follow-up times are given; and when neither of them are available.

Il est préférable mais pas toujours possible d'obtenir des auteurs les données brutes utilisées dans les publications présentant des courbes de survie de Kaplan-Meier et de Nelson-Aalen. À l'aide d'exemples, nous montrons qu'une grande partie de l'information d'origine peut être retrouvée à partir de la courbe elle-même, en mesurant la longueur des marches d'escalier verticales - par digitalisation ou par accès aux fichiers postscript. Ces techniques de reconstruction sont utiles pour récupérer les données brutes lorsqu'il n'y a pas de censure; lorsque le nombre à risque ou le nombre de décès à divers temps de suivi sont donnés; et lorsqu'aucune de ces informations n'est disponible.

## 2D-P: Poster Session 2 Séance d'affichage 2

---

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

**Room/Salle: Festival Theatre**

**Abstracts/Résumés**

---

**SHAHEENA BASHIR** (University Health Network), **CELIA GREENWOOD & QAZI AZAD** (McGill University)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Multivariate Meta Analysis in Gene Expression Microarray Studies*

*Méta-analyse multivariée des études d'expression génique sur micro-réseaux*

This study addresses an issue in meta-analysis of gene expression studies, when each has multiple treatment groups. Several different pairwise comparisons can be made within each study, but here we look at finding a summary F-statistic that will test collectively for inter-group differences. Empirical Bayes methods have been proposed for microarray gene expression analysis that shrink the probe-specific variances, within one study, towards a common value. We examine how to combine the empirical-Bayes shrunken estimates across studies to obtain an appropriate summary F-statistic.

Cette étude soulève une question en méta-analyse des études d'expression génique sur micro-réseaux lorsqu'elle comporte de multiples groupes de traitement. Nous pourrions effectuer plusieurs comparaisons par paire dans chaque étude, mais nous cherchons une statistique sommaire F qui permet de tester collectivement les différences entre les groupes. Les méthodes empiriques de Bayes ont été proposées pour l'analyse de l'expression génique sur micro-réseaux qui, lors d'une étude, réduit les variances d'une sonde vers une valeur commune. Nous examinons la manière de combiner les estimations réduites empiriques de Bayes dans les études pour obtenir une statistique sommaire F appropriée.

---

**NADIA BIFOLCHI, ROB DEARDON & ZENY FENG** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 14:30]

*The Interplay Between Spatial and Network-based Infectious Disease Models*

*La relation entre les modèles de maladies infectieuses spatiaux et de réseaux*

Individual-level models (ILMs) are used to model disease spread when it depends on various individual-level risk factors. Infectious diseases are generally spread via complex individual-level interactions which are impossible to record in practice. Since this data is often unavailable, we sometimes use information at the spatial level as a proxy. Here we use simulated epidemics that are propagated through networks of varying complexity, to compare results from fitted network and spatial ILMs. The primary interest concerns how far the spatial proxy can deviate from the true underlying network before conclusions made from a fitted spatial model become unreliable.

Les modèles au niveau de l'individu permettent de modéliser la propagation de la maladie lorsque celle-ci dépend de divers facteurs de risques, en fonction de chaque individu. Les maladies se propagent généralement au moyen d'interactions complexes au niveau de l'individu, qui ne peuvent être enregistrées dans la pratique. Vu que ces données ne sont souvent pas disponibles, nous utilisons parfois cette information à l'échelle spatiale, comme une approximation. Dans cette étude, nous simulons des épidémies propagées par des réseaux de complexité variée, pour comparer les résultats des modèles spatiaux et de réseaux ajustés au niveau de l'individu. L'objectif principal consiste à savoir jusqu'où l'approximation spatiale peut dévier des réseaux sous-jacents réels avant que les conclusions d'un modèle spatial ajusté ne soient plus fiables.

## 2D-P: Poster Session 2 Séance d'affichage 2

---

**SABRINA DICINTIO** (University of Guelph), **SABRINA DICINTIO, PAUL MCNICHOLAS & BRIAN FRAN CZAK**

[Tuesday June 14 / mardi 14 juin, 14:30]

*Early Disease Detection of Veterinary Data*

*La détection précoce de maladies au moyen de données vétérinaires*

The problem of analyzing veterinary surveillance data with the objective of developing an early detection system for cattle diseases is discussed. The data are described along with a model that will classify observations into groups so that unusual patterns in illnesses can be identified swiftly and addressed in order to help prevent an outbreak. Once the best model is determined, a discriminant rule is developed and tested.

Nous soulevons le problème de l'analyse des données de surveillance vétérinaire, l'objectif étant de mettre au point un système de détection précoce des maladies du bétail. Les données sont décrites d'après un modèle de classement des observations dans des groupes, qui permet d'identifier rapidement et d'exposer les profils pathologiques anormaux pour aider à prévenir une épidémie. Après avoir déterminé le meilleur modèle, une règle discriminante est élaborée et testée.

**ANGIE CHRISTINA DOBBS & ROB DEARDON** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Improving MCMC Convergence in Individual-Level Infectious Disease Models*

*Améliorer la convergence MCMC dans les modèles de maladies infectieuses à l'échelle de l'individu*

Monte Carlo Markov Chain (MCMC) convergence can be an issue when fitting spatial infectious disease models to data due to the correlation between parameters. Here a model with just two parameters, an infectivity constant and spatial rate of decay, are estimated using MCMC methods in such a model. Introducing a normalization constant to alleviate the correlation results in MCMC convergence over fewer iterations. However, this occurs at the expense of greater computation time per MCMC iteration. This trade-off is examined via the effective sample size to run time ratio.

La convergence de "Monte Carlo Markov Chain" (MCMC) peut s'avérer problématique lorsque l'on tente d'adapter les modèles de maladies infectieuses virales aux données en raison de la corrélation entre les différents paramètres. Voici un modèle avec seulement deux paramètres évalués en appliquant les méthodes MCMC à un tel modèle. L'introduction d'une variable de normalisation afin de réduire cette corrélation aboutit à une convergence de MCMC sur une fréquence plus faible. Toutefois, cela se produit au détriment de meilleurs temps de calcul par fréquence de MCMC. On observe ce compromis à travers le pourcentage d'échantillons réels pour un temps d'exécution donné.

**ERIC FU** (BC Centre for Excellence in HIV/AIDS), **JOHN PETKAU** (University of British Columbia)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Relative Efficiency of Modified Poisson Regression and Log-Binomial Regression*

*L'efficacité relative de la régression de Poisson et de la régression log-binomiale modifiées*

When two exposure groups are compared on a binary response, an estimate of the covariate-adjusted relative risk, often the actual target of inference rather than the odds ratio, can be obtained by log-binomial regression. However, convergence problems with log-binomial regression have been reported in the literature especially when the outcome is a common event. Recently, Poisson re-

Lorsque deux groupes de sujets exposés sont comparés en fonction d'une réponse binaire, une estimation du risque relatif ajusté pour les covariables, plus souvent la cible de l'inférence actuelle que le risque relatif approché, peut être obtenue par la régression log-binomiale. Toutefois, les problèmes de convergence avec la régression log-binomiale ont été mentionnés dans la littérature, particulièrement lorsque le résultat est un événement commun.

## 2D-P: Poster Session 2 Séance d'affichage 2

---

gression with robust variance estimator was proposed as an alternative to circumvent these problems. We examine the computational issues and compare the statistical efficiency of log-binomial regression and Poisson regression under several simple scenarios.

Récemment, la régression de Poisson avec un estimateur robuste de la variance a été proposée comme alternative pour pallier à ces problèmes. Nous examinons les problèmes de calcul et comparons l'efficacité statistique des régressions log-binomiale et de Poisson d'après plusieurs scénarios simples.

---

**RIHONG HUI & HANNA JANKOWSKI** (York University)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Cox Proportional Hazards with Shape-Constraints*  
*Risques proportionnels de Cox avec contraintes de forme*

We consider maximum likelihood estimation of the baseline hazard and the "beta" parameters in the Cox proportional hazards model under the assumption that the baseline hazard satisfies a uni-model shape-constraint. We present R code which produces the estimators, and apply our methods to analyze the mortality of adult flour beetles under different dose levels of DDT (Chen, 2007; Hewlett, 1974; Diggle and Gratton, 1984; Pack and Morgan, 1990)

Nous considérons l'estimation par maximum de vraisemblance du risque de base et les paramètres « beta » dans le modèle de risque proportionnel de Cox. Nous supposons également que le risque de base satisfait un modèle unique de contraintes de forme. Nous présentons un code R qui produit les estimateurs, et nous appliquons nos méthodes pour analyser la mortalité des triboliums de la farine adultes exposés à différents niveaux de dose de DDT (Chen, 2007 ; Hewlett, 1974 ; Diggle et Gratton, 1984 ; Pack et Morgan, 1990)

---

**XIANG JI & HANNA JANKOWSKI** (York University)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Simultaneous Confidence Regions for Effective Dose*  
*Régions de confiance simultanée pour dose efficace*

The effective dose is defined as the set of covariates such that the response reaches a predetermined level. We consider the problem of finding a confidence region for the so called plug-in estimator of the effective dose, under the logistic model. We also compare different methods, including theoretical bounds, bootstrap methods, and random set quantiles.

La dose efficace est définie comme la série de covariables qui permet à la réponse d'atteindre un niveau prédéterminé. Nous considérons le problème qui consiste à trouver une région de confiance pour un estimateur d'insertion (« plug-in ») de la dose efficace, en fonction d'un modèle logistique. Nous comparons également différentes méthodes, y compris les limites théoriques, les méthodes bootstrap et les ensembles aléatoires de quantiles.

---

**RAJAT MALIK, ROB DEARDON & GRACE KWONG** (University of Guelph), **BEN COWLING** (The University of Hong Kong)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Modeling the Spread of Influenza within Households in Hong Kong*  
*Modélisation de la propagation de l'influenza dans les maisons de Hong Kong*

Individual-level models (ILMs) can be used to model the spread of infectious diseases in discrete time. The key feature of these ILMs is that they can

Des modèles au niveau individuel peuvent être employés pour modéliser la propagation des maladies infectieuses à temps discret. La caractéristique principale de ces modèles

## 2D-P: Poster Session 2 Séance d'affichage 2

---

take into account covariate information on susceptible and infectious individuals as well as shared covariate information such as geography or contact measures. Here such ILMs are fitted in a Bayesian framework using Markov Chain Monte Carlo techniques to data from a study on influenza within households in Hong Kong during 2009. The goal is to estimate the effect of vaccination on infection risk.

tient compte des informations covariables sur les individus prédisposés et infectieux ainsi que les informations covariables partagées telles que les mesures de géographie ou de contact. Ici, ces modèles sont adaptés dans un cadre bayésien à l'aide des méthodes Markov Chain Monte Carlo aux données obtenues de l'étude sur l'influenza dans les maisons de Hong Kong pendant 2009. Le but de cette étude est d'estimer les effets de la vaccination sur le risque d'infection.

---

**RACHEL O'REILLY, PAUL D. MCNICHOLAS, SHANNA WILSON & CHEVONNE CARLOW** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Supervised Gene Selection in Gene Expression Microarray Data*

*Sélection de gènes supervisée dans les données d'expression génique sur micro-réseaux*

When working with genetic microarray data, gene selection is a necessary step in virtually all types of analyses, and clearly has an impact on the performance and accuracy of methods applied thereafter. The proposed gene selection procedure will make use of the known tissue classes in order to fit a variant of the mixture of factor analyzers model. The 'value' of each gene will be assessed based on the fit of a two-component mixture model when compared to a one-component model. This procedure will be illustrated via a cluster analysis of real gene expression microarray data.

Lorsqu'on travaille avec des données d'expression génique sur micro-réseaux, la sélection de gènes est une étape nécessaire dans presque tous les types d'analyse. Cette sélection a clairement des conséquences sur les résultats et la précision des méthodes appliquées par la suite. La procédure de sélection des gènes proposée fonctionnera avec des classes de tissus connus pour s'ajuster à une variance du modèle de mélange des analyseurs de facteurs. La « valeur » de chaque gène sera analysée en comparant l'ajustement d'un modèle de mélange à deux composantes avec un modèle à une composante. Cette procédure sera illustrée d'une analyse en classification automatique de données réelles d'expression génique sur micro-réseaux.

---

**IRENE VRBIK & PAUL MCNICHOLAS** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Model-based Classification of Time Course Data Using the Skew t Distribution*

*Classification basée sur des modèles de données de décours temporel au moyen de la distribution t asymétrique*

There is a great need for effective clustering techniques to analyze gene expression time course data. Lin (2007) illustrated a robust mixture modelling approach using the skew t distribution and McNicholas and Murphy (2010) introduced a family of models that effectively deal with time course data by using Cholesky decomposition on the covariance matrix. Combining these two adaptations results in a new family of models for the analysis of time course data. Parameter estimates are found via the EM-algorithm and models are selected according to the BIC. Our approach is illustrated on two real gene expression time course data sets.

On a sérieusement besoin de techniques de classification efficaces pour analyser les données de décours temporel d'expression génique. En 2007, Lin a illustré une approche de modélisation progressive de mélanges robustes par l'utilisation d'une distribution asymétrique t. En 2010, McNicholas et Murphy ont présenté une famille de modèles qui traite efficacement les données de décours temporel au moyen de la décomposition Cholesky sur la matrice de covariance. La combinaison de ces deux adaptations résulte en une nouvelle famille de modèles qui permet d'analyser les données de décours temporel. Les estimations des paramètres sont déterminées avec l'algorithme d'espérance-maximisation (EM) et les modèles sont sélectionnés en



## 2D-P: Poster Session 2 Séance d'affichage 2

---

fonction du critère d'information bayésien (CIB). Notre approche est illustrée de deux jeux de données réelles de dé-cours temporel d'expression génique.

---

**MONICA WONG, PAUL D. MCNICHOLAS & DAVID M. MUTCH** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Application of Clustering in Nutrigenomics to Identify Diet Intervention Sub-populations*

*Application de classification en nutriginomique pour déterminer l'intervention alimentaire chez les sous-populations*

Nutrigenomics is burgeoning into an important discipline in modern nutritional sciences. Nutrigenomics is the field of research that studies diet-gene interactions and how these interactions can influence health. Within nutrigenomics, clustering is used to identify sub-populations that respond differently to a given diet intervention. The aim of this study was to examine the benefits of clustering on an adipose tissue gene expression data set obtained during a low-calorie diet intervention carried out on human subjects. The data used for the current study was generated in the Framework 6 EU Project Diogenes (<http://www.diogenes-eu.org/>).

La nutriginomique est en train de devenir une discipline importante en science de la nutrition. La nutriginomique est le domaine de recherche qui étudie les interactions entre le régime alimentaire et les gènes ainsi que la façon dont ces interactions peuvent influencer la santé. En nutriginomique, la classification est utilisée pour déterminer les sous-populations qui répondent différemment à une intervention alimentaire donnée. L'objectif de cette étude consistait à examiner les avantages de la classification sur un jeu de données d'expression génique de tissu adipeux obtenu lors d'une intervention alimentaire hypocalorique chez des êtres humains. Les données utilisées pour l'étude actuelle ont été générées dans le cadre du projet européen Diogenes 6 (Diet Obesity Genes - alimentation, obésité, gènes). Veuillez vous référer au site : <http://www.diogenes-eu.org/>.

---

**YUAN XU** (University of Victoria)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Use of Closed Population Models to Estimate the Number of Injection Drug Users in Victoria*

*L'utilisation de modèles de population fermée pour estimer le nombre d'utilisateurs de drogues par injection à Victoria*

Closed mark-recapture experiments are widely used to estimate population size due to their simplicity compared to open population models. However, the assumption that a population is "closed" is often violated because of birth, death, immigrations and emigrations. This work investigates various methods in closed population models (Petersen estimator, MLE, and Huggins) to estimate the injection drug users' population in Victoria. The model selection methods used to choose a final model, as well as the goodness-of-fit options for the final model, are discussed. The mark-recapture data of the investigations in 2003 and 2005 in this study are from I-TRACK survey program.

Les expériences fermées de marquage et de recapture sont largement utilisées pour estimer la taille de la population en raison de sa simplicité comparativement à d'autres modèles de population ouverte. Cependant, l'hypothèse selon laquelle la population est « fermée » est souvent erronée en raison des naissances, de la mort, de l'immigration et de l'émigration. Ces travaux permettent d'étudier les diverses méthodes dans les modèles de population fermée (estimateur de Petersen, estimateur du maximum de vraisemblance -EMV -, méthodologie de Huggins) pour estimer la population d'utilisateurs de drogues par injection à Victoria. Nous discutons de ces méthodes de sélection de modèles qui permettent de choisir un modèle final ainsi que des options d'adéquation connexes. Les données de marquage et de recapture des enquêtes de 2003 et 2005 dans cette étude proviennent du programme d'enquête de surveillance améliorée (appelé « I-TRACK »).

## 2D-P: Poster Session 2 Séance d'affichage 2

---

**LIN ZHANG** (University of Guelph), **ROB DEARDON** (Department of Mathematics and Statistics, University of Guelph), **DONGMEI CHEN** (Department of Geography, Queen's University)

[Tuesday June 14 / mardi 14 juin, 14:30]

*Modeling 2009 H1N1 Epidemic in Southern Ontario over Time and Space*

*Modélisation temporelle et spatiale de l'épidémie H1N1 de 2009 dans le sud de l'Ontario*

Individual Level Models (ILMs) are being applied to 2009 H1N1 data from Southern Ontario to examine the spatial-temporal dynamics of the epidemic. The disease is modelled at the level of census regions, taking into account risk factors such as: distance between regions and population size. Meanwhile, some problems in the data need to be taken into account: recorded cases will only be a small subset of those infected; second, individual infection times are not known. We attempt to account for such uncertainty by allowing infectivity of regions to vary over time.

Les modèles au niveau de l'individu ont été appliqués aux données relatives au H1N1 de 2009 qui proviennent du sud de l'Ontario, pour examiner la dynamique spatio-temporelle de l'épidémie. La maladie est modélisée à l'échelle des régions recensées, prend en compte les facteurs de risques, comme la distance entre les régions et la taille de la population. Toutefois, certains problèmes dans les données doivent être considérés : certains cas enregistrés ne seront qu'un petit sous-ensemble des cas infectés ; deuxièmement, les moments où l'infection a été contractée ne seront pas connus. Nous tentons de tenir compte de cette incertitude en permettant à l'infectivité des régions de varier sur le temps.

**2E-A: CJS Award Address**  
**Allocution du récipiendaire du Prix de la RCS**

---

Chair/Président: Duncan Murdoch (University of Western Ontario)  
Organizer/Responsable: Mary Lesperance (University of Victoria)

**Room/Salle: Festival Theatre**

**Abstract/Résumé**

---

**ISABEL MOLINA** (Carlos III University-Madrid), **J. N. K. RAO** (Carleton University)

[Tuesday June 14 / mardi 14 juin, 15:00]

*Small Area Estimation of Poverty Indicators*

*Estimation des mesures de la pauvreté pour petits domaines*

We propose to use the empirical Bayes (EB) method to obtain efficient EB estimators of small area poverty measures, based on a nested error linear regression model. Mean squared errors of the EB estimators are estimated through a parametric bootstrap method. We study the performance of the proposed estimators relative to direct area-specific estimators and widely used synthetic estimators, based on a "simulated census" approach, using model-based and design-based simulation studies. We also apply the proposed method to estimate poverty incidences and poverty gaps in Spanish provinces by gender. The proposed methodology is applicable to general non-linear parameters.

Nous proposons la méthode bayésienne empirique pour estimer les mesures de la pauvreté pour petits domaines, fondée sur un modèle de couplage par régression linéaire avec erreurs emboîtées. L'erreur quadratique moyenne est estimée grâce à la méthode du bootstrap paramétrique. Nous étudions le rendement des estimateurs proposés relatifs à des estimateurs de domaines particuliers directs et les estimateurs fondés sur une approche de "simulation du recensement". Nous appliquons la méthode proposée pour estimer les incidences de la pauvreté et les écarts de pauvreté dans les provinces espagnoles, par genre. Cette méthodologie peut être appliquée à des paramètres non linéaires généraux.

## 2C-A: Isobel Loutit Invited Address Allocution Isobel Loutit

---

Organizer and Chair / Responsable et président: Will Welch (University of British Columbia)

Session sponsored by the Business and Industrial Statistics Section

Session parrainée par le Groupe de statistique industrielle et de gestion

**Room/Salle: BAC 241**

**Abstract/Résumé**

---

**DENNIS KJ LIN** (Pennsylvania State University)

[Tuesday June 14 / mardi 14 juin, 15:00]

*BIG Statistics*

*Statistiques BIG*

In the past decades, we have witnessed the revolution of information technology. Its impact to statistical research is enormous. This talk attempts to address recent developments and some potential research issues in Business, Industry and Government (BIG) Statistics, with special focus on computer experiment and information systems. An overall introduction and review will be given, followed by specific research potentials. For each subject, the problem will be introduced, some initial results will be presented, and future research problems will be suggested. If time permits, I will also discuss some recent advances in Search Engine and RFID study.

Durant ces dernières décennies, nous avons constaté la révolution de la technologie d'information. L'impact sur la recherche statistique est énorme. Cet exposé a pour but d'adresser les récents développements et quelques problèmes potentiels de recherche dans les statistiques de commerce, d'industrie et de gouvernement, avec une focalisation spéciale sur les expérimentations par ordinateur and les systèmes d'information. Pour chaque sujet, le problème sera introduit, quelques résultats préliminaires seront présentés, et des problèmes de recherche futurs seront suggérés. Si le temps le permet, je parlerai aussi de certaines avancées récentes dans l'étude de Moteur de Recherche et de Radio-Identification (RFID).

**2F-A: Why a Bayesian be?**  
**Conférence de vue d'ensemble : Pourquoi être Bayésien?**

---

Chair/Président: John Petkau (University of British Columbia)  
Organizer/Responsable: Shelley Bull (University of Toronto)

**Room/Salle: BAC241**

**Abstract/Résumé**

---

**ANDREW WILLAN** (SickKids Research Institute and University of Toronto)

[Tuesday June 14 / mardi 14 juin, 16:00]

*Why a Bayesian Be?*

*Pourquoi être Bayésien?*

When making statistical inference the Bayesian definition of probability provides important advantages over the frequentist definition. Firstly, the Bayesian definition permits simple, intuitive and relevant statements of statistical inference specifically about the parameters of interest. Also, a Bayesian approach provides a transparent framework for combining new information with current knowledge. Lastly, and more importantly, the Bayesian framework facilitates decision theory for optimal decision-making and research design, answering questions such as: Is the current information sufficient for decision-making, and if not, What is the optimal design for future research? These advantages will be illustrated in the context of clinical research.

Pour produire de l'inférence statistique, la définition bayésienne de la probabilité offre d'importants avantages par rapport à la définition fréquentiste. D'abord, la définition bayésienne permet des énoncés d'inférence statistique simples, intuitifs et pertinents axés spécifiquement sur les paramètres d'intérêt. De plus, une approche bayésienne fournit un cadre transparent pour la combinaison de nouvelle information avec des connaissances existantes. Enfin, et plus important encore, le cadre bayésien facilite la théorie de la décision pour une prise de décision et un plan d'expérience optimaux, répondant à des questions comme : Les renseignements actuels sont-ils suffisants pour la prise de décision ? Dans le cas contraire, quel est le plan d'expérience optimal pour la recherche future ? Ces avantages sont illustrés dans le contexte de la recherche clinique.

## 2F-I1: Financial Time Series Séries chronologiques financières

---

Organizer and Chair / Responsable et président: Rafal Kulic (University of Ottawa)

Session sponsored by the Probability Section  
Session parrainée par le Groupe de probabilité

**Room/Salle: BAC138**

**Abstracts/Résumés**

---

**PHILIPPE SOULIER** (Université Paris X), **RAFAL KULIK** (Université d'Ottawa), **HURVICH CLIFFORD** (New York University)

[Tuesday June 14 / mardi 14 juin, 16:00]

*Extremal Properties of Stochastic Volatility Models*

*Propriétés extrémales de processus à volatilité stochastique*

Stochastic volatility models have been introduced in order to account for some stylised facts of financial time series; in particular, heavy tail of the marginal distribution and long memory in volatility. A lot of empirical and theoretical work has been devoted to study long memory in volatility, but much less to the extremal properties of these processes. This talk reviews recent results concerning these properties.

Les processus à volatilité stochastique ont été introduits en économétrie financière pour rendre compte de deux propriétés empiriquement constatées des séries financières : queues de distribution épaisses et longue mémoire de la volatilité. De nombreux travaux théoriques ont été consacrés aux propriétés de longue mémoire, mais très peu aux propriétés extrémales de ces processus. Le présent exposé fait le panorama de résultats récents concernant ces propriétés.

---

**CLIFFORD HURVICH** (New York University.), **ALEXANDER AUE** (University of California, Davis), **LAJOS HORVATH** (University of Utah), **PHILIPPE SOULIER** (Université Paris X)

[Tuesday June 14 / mardi 14 juin, 16:30]

*Limit Laws in Transaction-Level Asset Price Models*

*Lois limites dans les modèles de niveaux de transaction des prix des actifs*

We consider pure jump transaction-level models for asset prices in continuous time, driven by point processes. In a bivariate model that admits cointegration, we allow for time deformations to account for seasonality in volatility and non-trading periods that may be different for the two assets. We also allow for leverage effects. We obtain the asymptotic distribution of the log-price process, and of the ordinary least-squares and tapered estimators of the cointegrating parameter. We show that the persistence of intertrade durations (hence of volatility) can affect the degree of cointegration.

Nous considérons des processus de saut pour modéliser des niveaux de transaction pour les prix d'actifs en temps continu. Dans un modèle bivarié admettant de la cointégration, nous permettons aux déformations du temps de prendre en compte la saisonnalité de la volatilité et les périodes de non-vente qui peuvent être différentes pour les deux actifs. Nous permettons aussi les effets de levier financier. Nous obtenons la distribution asymptotique du processus log-prix, et, pour les moindres carrés réguliers et les estimateurs dilatés du paramètre de cointégration. Nous montrons que la persistance des durées de l'inter-échange (donc, de volatilité) peut affecter le degré de cointégration.

## 2F-I1: Financial Time Series Séries chronologiques financières

---

**ZHOU ZHOU** (University of Toronto)

[Tuesday June 14 / mardi 14 juin, 17:00]

*Generalized Likelihood Ratio Test for Nonstationary Time Series*

*Test du rapport de vraisemblance généralisé pour les séries chronologiques non stationnaires*

We investigate the theoretical properties of the Generalized Likelihood Ratio (GLR) test (Fan, Zhang and Zhang, 2001) for the time varying coefficient models when the regressors and error are (possibly) nonstationary time series. It is found that, in general, the Wilks phenomenon is vulnerable to either non-stationarity or temporal dependence. An alternative bootstrap method is proposed for the inference.

Nous étudions les propriétés théoriques du test du rapport de vraisemblance généralisé (Fan, Zhang et Zhang, 2001) pour les modèles à coefficients variables dans le temps lorsque les régresseurs et l'erreur peuvent être des séries chronologiques non stationnaires. En général, le phénomène de Wilks est vulnérable à la dépendance non stationnaire ou temporelle. Une méthode alternative de bootstrap est proposée pour l'inférence.

## 2F-I2: Gene-environment Effects in Pharmacogenetic and Related Settings Effets génotype – environnement en pharmacogénétique et domaines connexes

---

Organizer and Chair / Responsable et président: Judy-Anne Chapman (Queen's University)

Room/Salle: BAC 234

### Abstracts/Résumés

---

**BROOKE FRIDLEY** (Mayo Clinic, Minnesota)

[Tuesday June 14 / mardi 14 juin, 16:00]

*Analysis of Pharmacogenomics Studies: Past, Present, Future*

*Analyse d'études pharmacogénétiques : passé, présent, futur*

Pharmacogenetics is the study of the role of inheritance in individual genetic variation in response to drugs. In this post-genomic era, pharmacogenetics has evolved into pharmacogenomics, the study of the influence of genetic variation across the entire genome on drug-response. Pharmacogenomics has been heralded as one of the first major clinical applications of the striking advances that have occurred and continue to occur in human genomic science. In this talk, presentation of the past, present and future of pharmacogenomics is discussed with respect to study design and statistical analysis.

La pharmacogénétique est l'étude du rôle de la transmission dans la variation génétique individuelle en réponse aux drogues. Dans cette ère post-génomique, la pharmacogénétique s'est transformée en pharmacogénomique, c'est-à-dire l'étude de l'influence de la variation génétique sur tout le génome concernant la drogue-réponse. La pharmacogénomique a été présentée comme une des premières applications cliniques majeures découlant des avancées qui se sont produites, et qui continuent de se produire, en science du génome humain. Dans cet entretien, on discutera du passé, du présent et du futur de la pharmacogénomique en ce qui concerne la conception d'études et l'analyse statistique.

---

**JOSÉE DUPUIS** (Boston University), **ALISA K. MANNING** & **MICHAEL LAVALLEY** (Boston University School of Public Health), **JAMES B. MEIGS** (Massachusetts General Hospital and Harvard Medical School), **L. ADRIENNE CUPPLES** (Boston University School of Public Health)

[Tuesday June 14 / mardi 14 juin, 16:30]

*Meta-Analysis of Genome-Wide Association Results Allowing for Gene-by-Environment Interactions*

*Méta-analyse d'études d'associations pangénomiques permettant l'incorporation d'un facteur modifiant environnemental*

Despite the success of genome-wide association scans (GWAS) in identifying trait associated loci, most discoveries to date explain only a small portion of the total heritability. Part of the missing heritability may be due to gene-by-environment effects that are not included in typical GWAS. I introduce a method to jointly meta-analyze the main and environmental interaction effects of a genetic variant on a trait of interest. I also present results from application of the method to mapping genes influencing fasting insulin levels, a trait related to type 2 diabetes, by incorporating gene-by-body mass index interaction.

Malgré le succès de l'analyse d'association pangénomique (AAP) dans l'identification de gènes liés à certains phénotypes, la plupart des découvertes n'expliquent qu'une partie de l'héritabilité totale. L'héritabilité non expliquée peut être due, en partie, à l'interaction entre un facteur environnemental et un ou plusieurs gènes, interactions qui ne sont pas typiquement incluses dans l'AAP. Je présente une méthode de méta-analyse qui incorpore l'interaction possible entre un gène et un facteur modifiant environnemental. Je présente également des résultats de l'application de la méthode pour identifier des gènes qui influencent les niveaux d'insuline en incorporant des interactions de gène-indice de masse corporelle.



## 2F-I2: Gene-environment Effects in Pharmacogenetic and Related Settings Effets génotype – environnement en pharmacogénétique et domaines connexes

---

**FEI ZOU** (University of North Carolina at Chapel Hill)

[Tuesday June 14 / mardi 14 juin, 17:00]

*Bayesian Semiparametric QTL Mapping for Gene-Environment Interactions*

*Cartographie bayésienne semi-paramétrique de loci à caractère quantitatif dans les interactions gènes-environnement*

In linkage analysis, it is often necessary to include covariates and environmental factors to increase power or avoid spurious false positive findings. If a covariate term in the model is specified incorrectly, the inclusion of the covariate may adversely affect power and accuracy of the identification of Quantitative Trait Loci (QTL). Furthermore, some covariates may interact with genetic factors in complicated fashions. We implement semiparametric models which include an unspecified function of any covariate found or suspected to have a more complex than linear but unknown relationship with the response variable. They allow for interactions among genetic factors and covariates.

En analyse de liaisons génétiques, il est souvent nécessaire d'inclure des covariables et des facteurs environnementaux pour augmenter la puissance ou éviter un surplus de faux positifs. Si un terme de covariable est mal spécifié dans le modèle, l'inclusion de la covariable peut nuire à la puissance et à la précision de l'identification des loci à caractère quantitatif. De plus, certaines covariables peuvent interagir de façon complexe avec les facteurs génétiques. Nous proposons des modèles semi-paramétriques qui comprennent une fonction indéterminée pour toute covariable ayant ou susceptible d'avoir avec la variable réponse une relation inconnue et plus complexe que simplement linéaire. Ces modèles permettent des interactions entre les facteurs génétiques et les covariables.

## 2F-I3: Statistics Literacy Littératie en statistique

---

Organizer and Chair / Responsable et président: Bovas Abraham (University of Waterloo)

**Room/Salle: BAC132**

### Abstracts/Résumés

---

**NEVILLE DAVIES** (Royal Statistical Society Centre for Statistical Education)

[Tuesday June 14 / mardi 14 juin, 16:00]

*The Royal Statistical Society getstats Campaign: Ten years to Statistical Literacy?*

*La campagne « getstats » de la Royal Statistical Society : littératie statistique d'ici dix ans?*

The Royal Statistical Society (RSS) 10-year statistical literacy campaign, getstats, aims to help build a society in which our lives and choices are enriched by an understanding of statistics in: - education; - the media; - the wider public; - employers; - elected representatives. I will: - outline activities the RSS plans for each; - report engagement with outside bodies; - describe a citizen's charter for statistics, new ways of teaching statistics and the production of an employee's course in statistical awareness.

La campagne décennale d'éducation statistique de la Royal Statistical Society (RSS), « getstats », vise à contribuer à la création d'une société dans laquelle nos vies et nos choix sont nourris par de solides notions statistiques dans le contexte : - du système d'éducation ; - des médias ; - du grand public ; - des employeurs ; - des élus. Je vais : - présenter les activités prévues par la RSS dans chaque catégorie ; - faire le point sur sa collaboration avec d'autres organismes ; - décrire la charte statistique du citoyen, de nouvelles façons d'enseigner la statistique et la création d'un programme de sensibilisation des employés à la statistique.

---

**ROBERT RODRIGUEZ** (President-Elect American Statistical Association)

[Tuesday June 14 / mardi 14 juin, 16:30]

*Making the Case for Statistical Literacy: Recent ASA Initiatives*

*Le bien-fondé de l'instruction statistique : les récentes initiatives de l'ASA*

The American Statistical Association is championing statistical literacy at multiple levels. At the K-12 level, the GAISE Report is influencing curriculum standards in several states and has served to improve the statistics content of mathematics standards. The ASA is also developing a voice in statistics education policy by meeting with legislators to promote statistical literacy, and it is educating policy makers about the role of statisticians in areas such as climate change and election auditing. While many activities are carried out by its committees and staff, the ASA is also partnering with other associations and working directly with teachers.

L'American Statistical Association soutient l'instruction statistique à plusieurs niveaux. Au niveau K-12, le rapport GAISE influence les normes des programmes d'études dans plusieurs états et a contribué à améliorer le contenu statistique des normes mathématiques. L'ASA travaille à développer son influence sur la politique d'éducation des statistiques en rencontrant des législateurs pour promouvoir la littératie statistique, et elle instruit les décideurs sur le rôle des statisticiens dans des domaines comme les changements climatiques et la vérification des élections. Alors que plusieurs activités sont effectuées par ses comités et son personnel, l'ASA établit également des partenariats avec d'autres associations et travaille directement avec les professeurs.

## 2F-I3: Statistics Literacy Littératie en statistique

---

**MARY TOWNSEND** (Statistics Canada)

[Tuesday June 14 / mardi 14 juin, 17:00]

*Statistical Literacy in Youth*

*Littératie statistique chez les jeunes*

In Canada, statistical literacy is promoted not only by academics and ministries of education, but also by Statistics Canada, the national statistical office. Through its award-winning Education Outreach Program, Statistics Canada offers students and teachers support in statistical learning with online resources that include E-STAT, Statistics: Power from Data! and Census at School. Census at School develops critical-thinking skills and an understanding of the statistical enquiry process. Students become 'little statisticians', completing the survey cycle from data collection to analysis and exploring data concepts with analytical software.

Au Canada, la littératie statistique est promue non seulement par les universitaires et les ministères de l'éducation, mais aussi par Statistique Canada, le bureau de statistique national. Par le biais du Programme de soutien à l'éducation, une initiative primée, Statistique Canada offre aux élèves et enseignants diverses ressources en ligne qui les aideront dans l'apprentissage de la statistique : « E-STAT », « Les statistiques : le pouvoir des données » et « Recensement à l'école ». « Recensement à l'école » aide les élèves à développer une pensée critique et à mieux comprendre le processus d'enquête statistique. Les élèves deviennent des « statisticiens en herbe » en complétant un cycle d'enquête de la collecte de données à l'analyse et en explorant divers concepts statistiques à l'aide de logiciels d'analyse.

## 2F-C1: Clustering and Classification Mise en grappes et classification

---

Chair/Président: Wayne Olford (University of Waterloo)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 141**

**Abstracts/Résumés**

---

**ADRIAN WADDELL & R. WAYNE OLDFORD** (University of Waterloo)

[Tuesday June 14 / mardi 14 juin, 16:00]

*Interactive Visual Clustering of High Dimensional Data using Navigation Graphs*

*Classement interactif visuel des données de haute dimension par l'utilisation de graphiques de navigation*

Based on a series of practical examples, we show how navigation graphs are useful to visually cluster and explore high dimensional data using user determined low dimensional trajectories through the data space. The navigation graph's nodes represent plots and its edges represent transitions from one plot into another (see Hurley and Oldford, 2011, *Comp. Stat.*). The working examples are demonstrated using the R package RnavGraph (Waddell and Oldford, 2011). The data visualization shown will be a 2d scatterplot of the high dimensional data objects (points, images, text, glyphs) which are interactively clustered using selection, zooming, and brushing.

En se basant sur une série d'exemples pratiques, nous montrons comment les graphiques de navigation sont utiles pour regrouper visuellement et explorer les données de haute dimension en utilisant des trajectoires de faible dimension déterminées par l'utilisateur à travers l'espace des données. Les noeuds du graphique de navigation représentent des graphiques et ces côtés représentent des transitions d'un graphique vers un autre (voir Hurley and Oldford, 2011, *Comp. Stat.*). Les exemples sur lesquels nous travaillerons sont démontrés à l'aide du package R RnavGraph (Waddell and Oldford, 2011). Nous visualiserons les données à l'aide d'un nuage de points 2D des objets de données de haute dimension (points, images, texte, glyphes) qui sont interactivement regroupés par sélection, zoom et balayage.

---

**JEFFREY L. ANDREWS & PAUL D. MCNICHOLAS** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 16:15]

*Semi-Supervised Variable Selection*

*Sélection de variables semi-supervisée*

A versatile variable selection technique based upon within-group variance is introduced as a fast alternative to more established methods. Developed with high-dimensional data in mind, this technique is versatile in the sense that it can be implemented with both classification and clustering paradigms, and can even be used as a catalyst for more complicated variable selection procedures. After the method is introduced, semi-supervised variable selection (SSVS) will be illustrated in a model-based clustering framework using the MCLUST family of Gaussian mixture models.

Nous présentons une technique polyvalente de sélection de variables fondée sur la variance intra-groupe en guise d'alternative rapide à d'autres méthodes mieux établies. Mise au point pour les données de grande dimension, cette technique est polyvalente car elle peut être utilisée dans le cadre d'un paradigme de classification ou de regroupement. Elle peut même servir de catalyseur pour des procédures plus compliquées de sélection de variables. Après avoir présenté la technique en question, nous illustrerons la sélection de variables semi-supervisée (SVSS) dans un cadre d'analyse de regroupement basée sur un modèle à l'aide des familles de mélanges gaussiens de la bibliothèque MCLUST. Nous

## 2F-C1: Clustering and Classification

### Mise en grappes et classification

---

The performance of SSVS will be compared with that of the popular *clustvarsel* package for the R software.

comparerons la performance de la technique SVSS avec celle de la populaire bibliothèque *clustvarsel* sous R.

---

**BRIAN FRAN CZAK, PAUL MCNICHOLAS & RYAN P. BROWNE** (University of Guelph)

[Tuesday June 14 / mardi 14 juin, 16:30]

*Model-based Clustering Using Mixtures of Asymmetric Laplace Distributions*

*Classification à base de modèles utilisant des mélanges de distributions de Laplace asymétriques*

A model-based clustering technique is introduced, based on mixtures of asymmetric Laplace distributions. Maximum likelihood estimates for the parameters are calculated using an expectation-maximization algorithm, where Aitken's acceleration is used to determine convergence and the BIC is used to select the number of mixture components. Our approach is demonstrated using both a simulation study and a real data analysis and its performance is compared to other model-based clustering methods.

Nous proposons une technique de classification à base de modèles basée sur des mélanges de distributions de Laplace asymétriques. Les estimateurs du maximum de vraisemblance pour les paramètres sont calculés grâce à un algorithme d'espérance-maximisation, alors que l'accélération d'Aitken est utilisée pour déterminer la convergence et le critère d'information bayésien permet de choisir le nombre de composantes du mélange. Notre approche est démontrée au moyen d'une étude de simulation et de l'analyse de données réelles, et sa performance est comparée à d'autres méthodes de classification à base de modèles.

---

**ANTONIO CIAMPI, LUC VILLANDRÉ & BENJAMIN RICH** (McGill University)

[Tuesday June 14 / mardi 14 juin, 16:45]

*Classification and Regression Trees with Soft Nodes*

*Arbres de classification et de régression à noeuds tendres*

We generalize the previously developed notion of tree with soft nodes for a binary outcome to define Classification and Regression Trees with soft nodes (SoftCART), i.e. tree-structured prediction rules for k-level categorical and continuous outcomes. We present an EM-based tree-growing algorithm for SoftCART. The new prediction rule is evaluated on simulated and real data and compared with the performance of hard-node CART. We show that the soft tree outperforms uniformly the hard tree and the gain may be substantial as the size of the learning sample increases.

Nous généralisons la notion, développée antérieurement, d'arbres à noeuds tendres pour réponse binaire, au cas de réponse qualitative et quantitative (CART). Nous présentons un algorithme de construction d'arbres à noeuds tendres basé sur l'approche EM. Le nouvel algorithme est évalué sur des données simulées et sur des données réelles, et il est comparé avec l'algorithme CART classique. Nous concluons que l'arbre à noeuds tendre donne des prédictions plus correcte que l'arbre classique, et que l'avantage augmente avec la taille de l'échantillon d'apprentissage.

---

**JESSY DONELLE & XU WANG** (St. Francis Xavier University), **BARBARA CAFFERY** (Yorkville Eye Clinic)

[Tuesday June 14 / mardi 14 juin, 17:00]

*Differentiating Between Sjogren's Syndrome and Dry Eye Disease: An Analysis using Random Forests*

*Différenciation du syndrome de Sjögren et du syndrome des yeux secs : Analyse au moyen de forêts aléatoires*

Determining which non-invasive clinical tests easily differentiate between primary Sjogren's syndrome

Il est utile d'établir les tests cliniques non invasifs permettant de facilement différencier le syndrome primaire de Sjögren

## 2F-C1: Clustering and Classification

### Mise en grappes et classification

---

(pSS) dry eye and non-autoimmune aqueous deficient dry eye (DE) is useful. The non-parametric statistical technique, Random Forests (RF), was applied and these results were compared to previous research results obtained by single classification trees. RF reduces the misclassification rate of prediction and provides more accurate identification of medical non-invasive tests. Rose bengal staining of the conjunctiva and severity of dry eye and dry mouth symptoms were the most important non-invasive variables in differentiating pSS from DE. Random Forest analysis confirms the previous results using single classification trees.

gren (SpS), le syndrome des yeux secs et le syndrome non auto-immun des yeux secs par déficience lacrymale (SYS). Nous avons appliqué la technique statistique non paramétrique des forêts aléatoires (FA) et comparé ces résultats à des résultats de recherches antérieures obtenus au moyen d'arbres de classification simples. Les FA réduisent le taux d'erreurs de classement des prévisions et fournissent une identification plus précise des tests médicaux non invasifs. L'épreuve de la coloration de la conjonctive au rose bengale et la gravité des symptômes de sécheresse des yeux et de la bouche constituent les variables non invasives les plus importantes pour différencier le SpS du SYS. L'analyse des FA confirme les résultats précédents obtenus au moyen d'arbres de classification simples.

## 2F-C2: Longitudinal Models: Measurement Error, Missing Data, and Misspecification Modèles longitudinaux : erreurs de mesure, données manquantes et erreurs de spécification

---

Chair/Président: Joel Dubin (University of Waterloo)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 239**

**Abstracts/Résumés**

---

**WEI LIU** (York University, Canada), **GRACE Y. YI** (University of Waterloo, Canada), **LANG WU** (University of British Columbia, Canada)

[Tuesday June 14 / mardi 14 juin, 16:00]

*Simultaneous Inference and Bias Analysis for Longitudinal Data with Covariate Measurement Error and Missing Responses*

*Inférence simultanée et analyse de biais pour données longitudinales avec erreur de mesure des covariables et réponses manquantes*

Longitudinal data arise frequently in medical studies and it is common practice to analyze such data with generalized linear mixed models. Such models enable us to account for between- and within-subjects heterogeneities. Inferential procedures may be dramatically more complicated when missing observations or measurement error arise. In this paper our objectives are to study simultaneous impact of missingness and covariate measurement error on inferential procedures and to develop a method that is both computationally feasible and theoretically valid. Simulation studies are conducted to assess the performance of the proposed method, and a real example is analyzed with the proposed method.

Il est courant d'utiliser des données longitudinales dans les études médicales et de les analyser avec des modèles linéaires généralisés mixtes. Ces modèles nous permettent de tenir compte des hétérogénéités inter- et intra-sujets. En présence d'observations manquantes ou d'erreur de mesure, les procédures inférentielles deviennent plus complexes. Dans cet article, nos objectifs sont d'étudier l'impact simultané des données manquantes et d'erreur de mesure des covariables sur les procédures inférentielles et de développer une méthode qui sera à la fois faisable computationnellement et valide théoriquement. Des études par simulation sont effectuées pour évaluer la performance de la méthode proposée, et un exemple réel utilisant la méthode proposée est analysé.

---

**YUNQI JI & FAN ZHAOZHI** (Memorial University of Newfoundland)

[Tuesday June 14 / mardi 14 juin, 16:15]

*Analysis of Longitudinal Categorical Data Subject to Misclassification and Informative "Unsure" Responses*

*Analyse de données catégorielles longitudinales sujettes à la mauvaise classification et aux réponses informatives « Incertain »*

In epidemiological studies, respondents are often required to answer some questions from pre-tested questionnaires using a "Yes", "No" or "Unsure" as the response. An "Unsure" answer leads to loss of information about the respondent's inherent status. Even a "Yes" or "No" response may misclassify the respondent's true status. An unbalanced mis-

Dans les études épidémiologiques, les participants doivent souvent répondre à des questions provenant de questionnaires prétestés en répondant par « Oui », « Non » ou « Incertain ». La réponse « Incertain » mène à une perte d'information sur l'état inhérent du participant. Même les réponses « Oui » et « Non » peuvent entraîner une mauvaise classification de l'état véritable du participant. Nous présentons

## 2F-C2: Longitudinal Models: Measurement Error, Missing Data, and Misspecification Modèles longitudinaux : erreurs de mesure, données manquantes et erreurs de spécification

---

classification model is presented to describe the misclassification and "Unsure" responses. We examined the impact of misclassification and "Unsure" responses in analysis. An approach is proposed to correct the attenuation and improve the efficiency of model estimation taking into account both misclassification and "Unsure" responses.

un modèle de mauvaise classification non compensée pour décrire les problèmes de mauvaise classification et des réponses « Incertain ». Dans l'analyse, nous examinons l'impact de ces problèmes. Nous proposons une approche pour corriger l'atténuation et améliorer l'efficacité de l'estimation du modèle en tenant compte de ces problèmes.

---

**ELEANOR PULLENAYEGUM** (McMaster University)

[Tuesday June 14 / mardi 14 juin, 16:45]

*Longitudinal Data Subject to Intermittent Measurement*

*Données longitudinales exposées à des mesures intermittentes*

Longitudinal data arising from routine follow-up of patients will often have irregular measurement times, and be subject to censoring. Existing methods for analysis include joint modelling of the outcome and measurement processes, and inverse-probability weighting. This work extends Diggle et al.'s analysis of increments to the case of irregular follow-up, and also considers binning in combination with weighting, imputation, and multiple outputation.

Les données longitudinales provenant du suivi routinier de patients auront souvent des temps de mesures irréguliers, et seront exposées à la censure. Les méthodes existantes d'analyse incluent la modélisation conjointe de la variable réponse et des processus de mesure, et la pondération par probabilité inverse. Ce travail généralise l'analyse des incréments de Diggle et al. au cas du suivi irrégulier, et considère le groupement par classe (binning) en combinaison avec la pondération, l'imputation, et « l'outputation » multiple.

---

**MIREILLE SCHNITZER & ERICA MOODIE** (McGill University), **ROBERT PLATT** (McGill University, Montreal Children's Hospital Research Institute)

[Tuesday June 14 / mardi 14 juin, 17:00]

*Targeted Maximum Likelihood for Longitudinal Marginal Structural Models Under Density Misspecification*

*Les méthodes de maximum de vraisemblance ciblé pour modèles structuraux marginaux où la densité est mal déterminée*

Targeted maximum likelihood (tML) methods have been proposed to estimate the parameters of marginal structural models for longitudinal data with binary intermediary variables and outcome. This class of methods has been mathematically shown to improve estimating efficiency and reduce small-sample bias. We show that the methods proposed by Rosenblum and van der Laan (2010) generalize for any exponential family member outcome. We evaluate these methods in simulations studies with exponential family outcomes. Specifically, we compare tML G-computation to regular G-computation and inverse-probability-of-treatment-weighting (IPTW) under such scenarios as rare treatment, misspecified data generating density and unmeasured confounding.

Les méthodes de maximum de vraisemblance ciblé (tML) ont été suggérées pour estimer les paramètres liés aux modèles structuraux marginaux pour des données longitudinales avec des variables intermédiaires et résultantes binaires. Il a été démontré mathématiquement que cette méthode améliore l'efficacité des estimations et réduit les biais liés aux petits échantillons. Nous montrons que les méthodes proposées par Rosenblum et van der Laan (2010) se généralisent pour toute variable résultante membre d'une famille exponentielle. Nous évaluons les méthodes tML avec des méthodes de maximum de vraisemblance par des études de simulation avec plusieurs scénarios où la densité est mal déterminée.



## 2F-C2: Longitudinal Models: Measurement Error, Missing Data, and Misspecification Modèles longitudinaux : erreurs de mesure, données manquantes et erreurs de spécification

---

**ABDERAZZAK MOUIHA** (Centre de Recherche Université Laval Robert Giffard), **SIMON DUCHESNE** (Radiology Department, and Centre de Recherche Robert-Giffard, Université Laval, Canada)

[Tuesday June 14 / mardi 14 juin, 17:15]

*Alzheimer's Disease Biomarkers Dynamics: Best Fit Models*

*Dynamique des biomarqueurs de la maladie d'Alzheimer: les modèles les mieux ajustés*

We investigated the dynamics of some biomarkers for Alzheimer's disease (*AD*), namely, Cerebrospinal fluid  $A\beta$  1-42, *tau*, *hippocampal* volumes, and *FDG-PET* (fluorodeoxyglucose-positron emission tomography), in patients at different stages of *AD*. We selected 229 healthy subjects, 154 with mild cognitive impairment and 193 *AD* patients from the Alzheimer's Disease Neuroimaging Initiative database. Linear, quadratic, sigmoid, loess and Pbspline models were fitted for each biomarker. The independent variable was the subject *AD* assessment scale-cognitive score. Akaike criterion (*AIC*) was used to select, for each biomarker, the correct model. We tested the selected model fit to the data.

Nous avons étudié la dynamique de quelques biomarqueurs de la maladie d'Alzheimer (*MA*), à savoir, le liquide cérébrospinal  $A\beta$  1-42, *tau*, des volumes *hippocampiques*, et *FDG-PET* (tomographie par émission de positrons fluorodésoxyglucose), sur des patients atteints à différents stages de la maladie d'Alzheimer. Nous avons sélectionné 229 sujets en bonne santé, 154 patients atteints d'un déficit cognitif léger et 193 patients atteints de *MA* à partir de la base de données de l'initiative en imagerie cérébrale de la maladie d'Alzheimer. Des modèles linéaires, quadratiques, sigmoïdes, de loess et de Pbspline ont été ajustés pour chaque biomarqueur. La variable indépendante était le score d'évaluation du sujet atteint de *MA* sur une échelle cognitive.

## 2F-C3: Survey Methods: Estimation Méthodes d'enquête : estimation

---

Chair/Président: Geneviève Demers (Statistique Canada)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

**Room/Salle: BAC142**

**Abstracts/Résumés**

---

**PIERRE LAVALLÉE & SÉBASTIEN LABELLE-BLANCHET** (Statistique Canada)

[Tuesday June 14 / mardi 14 juin, 16:00]

*Indirect Sampling Applied to Skewed Populations*

*Le sondage indirect appliqué aux populations asymétriques*

It is possible to produce estimates at the enterprise level using data from a survey conducted at the establishment level by using the Generalised Weight Share Method. However, it is known that establishments belong to an asymmetrical population. By using the classical approach for weight sharing to such a population, large variances can be obtained. The purpose of this presentation is to propose alternate approaches for weight sharing in order to reduce the variance. Nine different methods are proposed. These nine approaches will be compared using a real population extracted from Statistics Canada's Business Register.

Il est possible de produire des estimations au niveau entreprise à partir d'une enquête au niveau établissement en utilisant la Méthode généralisée du partage des poids. On sait toutefois que les établissements constituent une population asymétrique. En appliquant l'approche classique du partage des poids à une telle population, on peut obtenir des variances élevées. Le but de cette présentation est de proposer une alternative à la façon de partager les poids de façon à réduire la variance. On propose neuf approches différentes. On comparera ces neuf approches à partir d'une population réelle extraite du Registre des entreprises de Statistique Canada.

---

**STEVE MATTHEWS & TYLER KIRKLAND** (Statistics Canada)

[Tuesday June 14 / mardi 14 juin, 16:15]

*Statistics Canada's Survey on Commercial and Institutional Energy Use: An Application of Indirect Sampling*

*L'Enquête sur la consommation d'énergie du secteur commercial et institutionnel de Statistique Canada : Une application du sondage indirect*

The Survey on Commercial and Institutional Energy Consumption measures various aspects of the commercial and institutional sector in Canada. An annual survey of establishments is conducted to support establishment-based analysis, and a building component is included periodically to support analysis of building-level data. The building component consists of a parallel survey that uses the establishment survey to produce a partial list of buildings from which a sample is selected. This presentation summarizes the project, and describes the methodology for several impor-

L'Enquête sur la consommation d'énergie du secteur commercial et institutionnel mesure différents aspects des secteurs commercial et institutionnel au Canada. Une enquête annuelle auprès d'établissements est menée pour l'analyse au niveau des établissements et une composante bâtiment est incluse périodiquement pour permettre l'analyse des données au niveau des bâtiments. La composante bâtiment se veut une enquête parallèle qui utilise l'échantillon des établissements pour produire une liste partielle des bâtiments de laquelle l'échantillon est tiré. Cette présentation résume le projet et décrit la méthodologie de certaines étapes importantes de l'enquête, à savoir le plan d'échan-

## 2F-C3: Survey Methods: Estimation Méthodes d'enquête : estimation

---

tant steps in the survey within the context of indirect sampling, namely the sample design, data collection, and estimation.

tillonnage, la collecte des données et l'estimation, dans le contexte du sondage indirect.

---

**KYLE VINCENT** (Simon Fraser University)

[Tuesday June 14 / mardi 14 juin, 16:30]

*On the Use of Adaptive Sampling for Estimating Population Sizes*

*Sur l'utilisation de l'échantillonnage adaptatif pour estimer les tailles de population*

Frank and Snijders (1994) introduced design-based moment unbiased estimators for population sizes based on a one-wave snowball sample. This talk extends on the aforementioned strategy while controlling for sample size as well as the sampling effort allocated towards recruiting individuals through social links. For some hard-to-reach drug using populations, auxiliary information in the form of hospitalisation records or data from safe-injection sites may be available. Ratio estimators and a Rao-Blackwellization technique for improving estimates based on a sampling design dependent on such available auxiliary information is also outlined in this talk.

Frank et Snijders (1994) ont introduit des estimateurs non biaisés par rapport au plan d'échantillonnage des moments basés pour les tailles de population en se basant sur un échantillon dit «one-wave snowball». Cet exposé étend cette stratégie tout en contrôlant la taille de l'échantillon et l'effort échantillonnal alloué dans le recrutement des individus par l'entremise des liens sociaux. Pour certaines populations d'utilisateurs de drogues difficiles à rejoindre, l'information auxiliaire peut être disponible par l'entremise de dossiers hospitaliers ou de données de sites sécuritaires d'injection. Les estimateurs par le ratio et l'utilisation de la technique de Rao-Blackwell pour améliorer les estimations basées sur un plan de sondage qui dépend de l'information auxiliaire disponible sont également présentés dans cet exposé.

---

**ABDELLATIF DEMNATI** (Statistique Canada)

[Tuesday June 14 / mardi 14 juin, 16:45]

*On a Calibrated Estimator for Two-phase Sampling Design in the Presence of Nonresponse*

*Estimateur calé pour plan de sondage à deux phases en présence de non-réponse*

Suppose a main sample is first selected for self-enumeration, and then from the main sample, we select in advance a sub-sample for follow-up. The final sample consists of respondents from the main sample and from the sub-sample. The design weights of the final sample are adjusted to compensate for nonresponse. We studied point and variance estimation under the above set-up. Results of a simulation study illustrate the effect of the weight adjustments on the mean square errors.

Supposons qu'un échantillon principal est d'abord sélectionné pour l'autodénombrement et qu'ensuite à partir de ce même échantillon, nous sélectionnons à l'avance un sous-échantillon pour le suivi. L'échantillon final se compose de répondants provenant de l'échantillon principal et du sous-échantillon. Les poids de sondage de l'échantillon final sont ajustés pour compenser la non-réponse. Nous étudions l'estimation ponctuelle et l'estimation de la variance en fonction du système ci-dessus. Les résultats de l'étude de simulation montrent les effets de la pondération sur les erreurs quadratiques moyennes.

---

**GENEVIÈVE DEMERS** (Statistique Canada)

[Tuesday June 14 / mardi 14 juin, 17:00]

*Ever-alive Units in the Unified Enterprise Survey: Is a Special Treatment Worth it?*

*Les unités "ever-alive" dans l'Enquête unifiée auprès des entreprises : Un traitement particulier vaut-il la peine ?*

## 2F-C3: Survey Methods: Estimation Méthodes d'enquête : estimation

---

The "ever-alive" units in business surveys are special units that have been active for only part of the year. In the Unified Enterprise Survey (UES) at Statistics Canada, these units are given a special treatment at the sampling stage; but in all subsequent survey steps, they receive the same treatment as other units. The goal of this analysis was to first identify these records throughout the UES process, to study what happened to them, and finally, to see how they contributed to the estimates.

Les unités "ever-alive" dans les enquêtes entreprises sont des unités ayant été actives une partie de l'année seulement. Dans l'Enquête unifiée auprès des entreprises (EUE) à Statistique Canada, ces unités reçoivent un traitement particulier à l'échantillonnage, mais toutes les étapes suivantes de l'enquête ne leur portent aucun traitement spécial. Le but de cette analyse était de tout d'abord identifier ces enregistrements tout au long du processus de l'EUE, d'observer ce qui leur est arrivé, et de finalement déterminer comment elles contribuent aux estimations.

---

**KATE WILDER** (Statistics Canada), **STEVEN THOMAS**

[Tuesday June 14 / mardi 14 juin, 17:15]

*Combining Data from Multiple Surveys*

*Combiner les données de plusieurs enquêtes*

Several options exist for combining data collected from multiple surveys. In the case of the Canadian Community Health Survey (CCHS) and the Labour Force Survey (LFS), the samples are selected from the same area frame, with a large overlap of primary sampling units (PSUs), creating dependence. This paper explores the options for estimating a population parameter using two such samples and gives the results of a study in which sample weights were created for a combined sample. Variance estimation is also explored.

Il existe plusieurs options pour l'intégration des données recueillies d'enquêtes multiples. Pour l'Enquête sur la santé dans les collectivités canadiennes (ESCC) et l'Enquête sur la population active (EPA), les échantillons sont sélectionnés d'un même cadre géographique, avec un chevauchement important pour les unités primaires d'échantillonnage, ce qui crée un effet de dépendance. Cet article explore différentes options pour estimer un paramètre en utilisant deux échantillons dépendants, et présente les résultats d'une étude dans laquelle les poids ont été créés pour l'échantillon combiné. L'estimation de la variance sera aussi discutée.

**3A-A: Pierre Robillard Award Address**  
**Allocution du récipiendaire du Prix Pierre-Robillard**

---

Chair/Président: Keumhee Carriere Chough (University of Alberta)

**Room/Salle: Festival Theatre**

**Abstract/Résumé**

---

**GUN HO JANG** (University of Pennsylvania)

[Wednesday June 15 / mercredi 15 juin, 9:00]

*Invariant Procedures for Model Checking, Checking for Prior-Data Conflict and Bayesian Inference*

*Procédures invariantes pour la vérification d'un modèle, la détection de conflits données-a priori et l'inférence bayésienne*

We consider a statistical theory as being *invariant* when the results of two statisticians' independent data analyses, based upon the same statistical theory and using effectively the same statistical ingredients, are the same. We propose invariant assessments of single null hypothesis without any specific alternative hypothesis which are applicable for both model checking and checking for prior-data conflict. Then we explore several properties of relative surprise inference, an invariant Bayesian inference which compares the belief changes from *a priori* to *a posteriori*, such as consistency and asymptotic normality.

Nous considérons une théorie statistique comme *invariante* lorsque les résultats d'analyses statistiques menées indépendamment par deux statisticiens, sur la base de cette théorie statistique et en utilisant les mêmes ingrédients statistiques, sont les mêmes. Nous proposons des évaluations invariantes d'une hypothèse nulle simple sans aucune hypothèse alternative spécifique qui sont applicables à la fois pour la vérification du modèle et la détection de conflits données- *a priori*. Puis, nous explorons plusieurs propriétés, notamment la cohérence et la normalité asymptotique, de l'inférence de surprise relative, une inférence bayésienne invariante qui compare les croyances *a priori* et *a posteriori* telles que la cohérence et la normalité asymptotique.

## 3B-A: Approaches to Pathway Modeling in Genetic Epidemiology Approches de la modélisation des voies en épidémiologie génétique

---

Organizer and Chair / Responsable et président: Shelley B Bull (University of Toronto)

**Room/Salle: BAC241**

**Abstract/Résumé**

---

**DUNCAN THOMAS** (University of Southern California)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*Approaches to Pathway Modeling in Genetic Epidemiology*

*Approches de modélisation de processus réactionnels en épidémiologie génétique*

Molecular epidemiology is concerned with the joint effects of genetic and environmental factors on disease risks. I will discuss approaches to incorporating biological knowledge into the analysis of gene-environment interactions, using Bayesian hierarchical and mechanistic modeling. Both will be illustrated on epidemiologic studies of folate metabolism in colorectal cancer and DNA repair following ionizing radiation exposure in second breast cancers. Issues to be considered include model selection and two-phase sampling for biomarker measurement or sequencing a candidate region. I will conclude with some thoughts on the prospects for extending these approaches to genome-wide scale.

L'épidémiologie moléculaire s'intéresse aux effets conjugués des facteurs génétiques et environnementaux sur les risques de maladie. Je discuterai des approches d'intégration des connaissances biologiques dans l'analyse des interactions gènes-environnement, à l'aide de modélisations hiérarchique bayésienne et mécaniste. Les deux modèles sont illustrés par des études épidémiologiques sur le métabolisme du folate dans le cancer colorectal et la réparation de l'ADN à la suite d'une exposition aux rayonnements ionisants en cas de second cancer du sein. Les problèmes abordés comprennent la sélection du modèle et l'échantillonnage à deux phases pour la mesure des biomarqueurs ou le séquençage d'une région d'intérêt. Je concluerai avec quelques réflexions sur les perspectives d'étendre ces approches à l'échelle du génome entier.

## 3B-I1: Ecological Sampling Échantillonnage écologique

---

Organizer and Chair / Responsable et président: Chris Field (Dalhousie University)

**Room/Salle: BAC 236**

### Abstracts/Résumés

---

**ALAN WELSH** (Australian National University)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*A Sample Survey View of Occupancy/Detection Surveys*

*Une perspective d'échantillonnage d'enquête pour les enquêtes d'occupation et de détection*

The approach to the problem of adjusting for imperfect detection in occupancy surveys advocated by Mckenzie et al (2005, Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence) involves carrying out repeated surveys and modelling both the probabilities of occupancy and detection. We examine this approach from a sample survey perspective, using the likelihood methodology of Breckling et al (1994, Intl. Statist. Reviews). This provides a useful perspective on the approach and enables us to bring in ideas from and make comparisons with other sample survey approaches.

L'approche au problème de la détection incomplète dans les enquêtes d'occupation recommandées par Mckenzie et al (2005, Occupancy estimation and modeling : inferring patterns and dynamics of species occurrence) implique des enquêtes répétées de réalisation et modélise autant les chances d'occupation que la détection. Nous examinons cette approche d'une perspective d'échantillonnage d'enquête, en utilisant la méthodologie de vraisemblance de Breckling et al (1994, Intl. Statist. Révisions). Cela fournit une perspective utile sur cette approche et permet d'introduire des idées et de faire des comparaisons avec d'autres approches d'échantillonnage d'enquête.

---

**JOANNA MILLS FLEMMING** (Dalhousie University)

[Wednesday June 15 / mercredi 15 juin, 10:45]

*The Ocean Tracking Network: A Data Perspective*

*Le « Ocean Tracking Network » : les données en vedette*

The Ocean Tracking Network is an unprecedented global project that aims to provide a permanent platform with which to monitor the movements and interactions among numerous acoustically tagged marine species. On the Scotian Shelf of Atlantic Canada, grey seals (*Halichoerus grypus*) are being used as roving acoustic receivers (bioprobes) that record encounters with other acoustic tagged species. These spatial encounter data are entirely novel. We describe the development of visualization tools and spatial point process methods for the analysis of spatial encounter data. Conclusions detail the kinds of questions that may be answered and the challenges that must be overcome.

Le « Ocean Tracking Network » (Réseau de suivi des océans) est un projet mondial sans précédent visant à la création d'une plateforme permanente permettant le suivi des mouvements et interactions parmi diverses espèces marines équipées de marques acoustiques. Sur la plateforme Scotian au Canada atlantique, les phoques gris (*Halichoerus grypus*) servent de récepteurs acoustiques itinérants (sondes biologiques) qui enregistrent leurs rencontres avec d'autres espèces équipées de marques acoustiques. Ces données de rencontre spatiales n'existent que depuis peu. Nous décrivons la mise au point d'outils de visualisation et de méthodes de processus de points spatiaux permettant l'analyse de ces données. Nos conclusions explorent les types de questions auxquelles ces données pourront répondre et les défis qui restent à surmonter.

## 3B-I1: Ecological Sampling Échantillonnage écologique

---

---

**RICK ROUTLEDGE, DISCUSSANT**, Simon Fraser University  
[Wednesday June 15 / mercredi 15 juin, 11:15]



**3B-I2: Order Statistics and Related Inference, Ranked Set Sampling and Censoring  
Methodology**  
**Statistiques d'ordre et inférence connexe, échantillonnage d'ensembles ordonnés et  
méthodologie de censurement**

---

Chair/Président: Alexandre Leblanc (University of Manitoba)  
Organizer/Responsable: Katherine Davies (University of Manitoba)

**Room/Salle: BAC 234**

**Abstracts/Résumés**

---

**MOHAMMAD JAFARI JOZANI** (University of Manitoba), **SAYED JAMAL MIRKAMALI** (Allameh Tabatabaie University)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*Maxima Nomination Sampling with Application in Constructing Quality Control Charts for Attributes*

*Échantillonnage par nomination de maxima avec application dans la construction de cartes de contrôle pour attributs*

We develop quality control charts for attributes using the maxima nomination sampling (MNS) method and study their performance using different criteria such as Average Run Length (ARL), the required sample size in detecting quality improvement, and non-existence region for control limits. We study the effect of the sample size, the set size, and nonconformity proportion on the performance of MNS control charts. These charts can be used as better benchmarks for indicating quality improvement or quality deterioration. We study MNS charts from a cost perspective and explain how to obtain the optimal control limits for an MNS  $p$ -chart.

Nous développons des cartes de contrôle pour attributs en utilisant la méthode d'échantillonnage par nomination de maxima et étudions leur performance avec des critères différents tels que la longueur moyenne d'exécution (LME), la taille de l'échantillon requise pour la détection de l'amélioration de la qualité, et la région de non-existence des limites de contrôle. Nous étudions les effets de la taille d'échantillon, de la taille de l'ensemble nominé, et de la proportion de non-conformité sur la performance des cartes de contrôle de l'échantillonnage par nomination de maxima. Ces cartes peuvent être utilisées comme un meilleur point de référence pour indiquer l'amélioration ou la détérioration de la qualité. Nous étudions les coûts des cartes de contrôle par nomination de maxima, et expliquons comment obtenir les limites optimales de contrôle pour la  $p$ -carte d'un échantillonnage par nomination de maxima.

---

**KATHERINE DAVIES** (University of Manitoba), **N. BALAKRISHNAN** (McMaster University), **JEROME P. KEATING** (University of Texas at San Antonio), **ROBERT L. MASON** (Southwest Research Institute)

[Wednesday June 15 / mercredi 15 juin, 10:45]

*Optimal Plotting Points Based on Pitman Closeness*

*Choix optimal du tracé de points selon la proximité de Pitman*

Plotting points of order statistics are often used in the determination of goodness-of-fit of observed data to theoretical quantiles. They are often determined using nonparametric methods which produce, for example, the mean- and median-ranks commonly used in practice. In this talk I will describe a distribution-based approach which selects plotting points based on the simultaneous-closeness of order statistics to population quantiles. I will then show that the plotting points so

Le tracé de points des statistiques d'ordre est souvent utilisé pour établir la qualité de l'ajustement des données observées à des quantiles théoriques. Les points à tracer sont souvent déterminés à l'aide de méthodes non-paramétriques qui produisent, par exemple, les rangs moyens ou médians couramment utilisés en pratique. Dans cette présentation, je vais décrire une méthode basée sur la distribution qui sélectionne les points à tracer en fonction de la proximité de l'ensemble des statistiques d'ordre aux quantiles de population. Puis, je montrerai que les tracés

### 3B-I2: Order Statistics and Related Inference, Ranked Set Sampling and Censoring Methodology

#### Statistiques d'ordre et inférence connexe, échantillonnage d'ensembles ordonnés et méthodologie de censurement

---

determined are robust over a multitude of symmetric distributions. I will finish with a demonstration of their usefulness by examining the power properties of a correlation goodness-of-fit test for normality.

de points ainsi choisis sont robustes pour une multitude de distributions symétriques. Je terminerai par une démonstration de leur utilité en examinant la puissance d'un test de normalité basé sur la corrélation.

---

**BARRY ARNOLD** (University of California at Riverside)

[Wednesday June 15 / mercredi 15 juin, 11:15]

*Cantor Order Statistics Without Applications*

*Statistiques d'ordre de Cantor sans applications*

Building on ideas and concepts introduced by Lad, Taylor and Hosking, a generalized Cantor distribution and a corresponding skew generalized Cantor distribution are described. Associated inverse distributions are also introduced. Some results concerning the corresponding order statistics will be discussed.

Poursuivant les idées et les concepts proposés par Lad, Taylor et Hosking, une distribution généralisée de Cantor et une distribution asymétrique généralisée de Cantor correspondante sont décrites. Les distributions réciproques sont aussi introduites. Quelques résultats concernant les statistiques d'ordre correspondantes seront discutés.

### **3B-I3: Psychographic Surveys and Behavioural Surveys - Can information from one be used to accurately predict characteristics of the other?**

#### **Enquêtes psychographiques et enquêtes sur le comportement – peut-on utiliser des informations provenant des unes pour prévoir avec précision les caractéristiques des autres?**

---

Organizer and Chair / Responsable et président: Delbert Stewart (IMS Brogan)

Session sponsored by the Survey Methods Section

Session parrainée par le Groupe de méthodes d'enquête

**Room/Salle: BAC142**

**Abstracts/Résumés**

---

**JOHN S. LYONS** (University of Ottawa)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*Measurement as Communication: Communimetric Theory and Survey Design*

*Mesures comme communication : Théorie du communimetric et plan d'enquête*

Measurement in human services is designed either without theory or from psychometric theories. This presentation introduces a new theory of measurement for human service enterprises called communimetrics. In science the role of communication rests among scientists. In human services, more are involved in communication. Maximizing the communication value of measurement in the system enhances the reliability/validity of the measurement process and makes it a key component of the transaction. By making the measurement consistent with the work, the utility of the measures is dramatically enhanced. Characteristics of a communimetric measurement will be presented and discussed and examples provided.

Les mesures dans les services humains sont construites soit sans théorie, soit à partir des théories psychométriques. Cet exposé introduit une nouvelle théorie de mesures pour les entreprises de services humains, appelé la communimetric. En science, le rôle de communicateur repose surtout avec les scientifiques. Dans les services humains, plusieurs personnes sont impliquées dans la communication. Maximisant la valeur de la communication de la mesure dans un système améliore la fiabilité/validité de la procédure de mesure et la rend cruciale dans la transaction. En rendant la mesure concordante avec le travail, l'utilité de la mesure est considérablement améliorée. Les caractéristiques de la mesure communimetric seront présentées et discutées, et des exemples seront fournis.

---

**JANE E. LEDINGHAM, JESSICA KURTA, TIFFANY PURSOO, LEAH PUDDESTER & LAURIE CLARK** (University of Ottawa)

[Wednesday June 15 / mercredi 15 juin, 10:35]

*Children's Choices of Where to Play on the Playground as a Function of Emotional State*

*Les choix des enfants d'où jouer sur le terrain de jeu en fonction de leur état émotif*

Behavioral mapping procedures suggest that different children use their school playground differently: better adjusted children spread their play more widely over the playground, while aggressive and/or withdrawn children had more circumscribed areas of play. We subsequently asked children where they would choose to play and why when in different emotional states. Regardless of familiarity with the playground, for positive emotions children chose public areas with high social densities, giv-

Les procédures de cartographie comportementale suggèrent que des enfants différents utilisent leur terrain de jeu à l'école différemment : les enfants mieux ajustés utilisent plusieurs endroits, alors que les enfants agressifs et/ou renfermés jouent dans les secteurs plus circonscrits. Nous avons ensuite demandé aux enfants où ils choisiraient de jouer et pourquoi selon leurs différents états émotifs. Indépendamment de leur bonne connaissance du terrain de jeu, les enfants choisissaient des endroits publics avec des densités sociales élevées pour les émotions positives,

### 3B-I3: Psychographic Surveys and Behavioural Surveys - Can information from one be used to accurately predict characteristics of the other?

#### Enquêtes psychographiques et enquêtes sur le comportement – peut-on utiliser des informations provenant des unes pour prévoir avec précision les caractéristiques des autres?

---

ing social and activity reasons, while for negative emotions they chose private, low social density areas giving emotion regulation reasons. Aggressive children chose more public, high density areas for negative emotion states.

mentionnant des raisons d'ordre social et de niveau d'activité, alors que pour des émotions négatives, ils choisissaient les endroits de densité faible, mentionnant des raisons de régulation émotionnelle. Les enfants plus agressifs choisissaient plus d'endroits publics, à haute densité, pour les états d'émotion négatifs.

---

**DAVID W. JAMIESON** (Environics Research Group), **DELBERT E. STEWART & DAVID STEWART** (IMS Brogan)  
[Wednesday June 15 / mercredi 15 juin, 10:55]

*The Use of Demographic and Behavioural Characteristics to Predict Psychographic Profiles*

*L'emploi de caractéristiques démographiques et comportementales pour prédire les profils psychographiques*

The typically weak “signals” observed between psychological variables and demographic/behavioural variables can be amplified to acceptable levels if a suitable segmentation based on the psychological variables is created, and if the demographic/behavioural variables aggregated to predict it are cleverly conceptualized before-the-fact. We conducted a latent class analysis of Canadian primary care physicians (PCPs) based on self-reported social values, and we then modeled the classes via Decision Analysis and Knowledge Seeker using these same doctors' demographics and aggregated scripting information as predictors. The result is a psychographic profiling system for all Canadian PCPs, scored at over 80 percent mean classification accuracy.

Les signaux habituellement faibles observés entre les variables psychologiques et celles démographiques/comportementales peuvent être amplifiés à des niveaux satisfaisants si une segmentation convenable basée sur les variables psychologiques est réalisée et si l'ensemble des variables démographiques/comportementales explicatives sont adroitement conceptualisées a priori. Nous avons mené une analyse en classes latentes des médecins généralistes (MG) canadiens basée sur des valeurs sociales rapportées par le médecin et nous avons ensuite modélisé les classes par une analyse décisionnelle à l'aide de Knowledge Seeker en utilisant ces mêmes informations démographiques et des bilans globaux comme variables explicatives. Grâce à cette méthode, nous avons obtenu un système psycho graphique de profilage de tous les MG canadiens ayant obtenu une précision moyenne de classification de plus de 80%.

## 3B-I4: Randomness in Discrete Mathematics and Biology Stochast c t  en math matiques discr tes et biologie

---

Organizer and Chair / Responsable et pr sident: Ed Perkins (University of British Columbia)

Session sponsored by the Probability Section

Session parrain e par le Groupe de probabilit 

**Room/Salle: BAC138**

**Abstracts/R sum s**

---

**RICHARD LIANG** (University of British Columbia), **STEVE EVANS** (University of California, Berkeley)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*Stepping-Stone Models with Continuous Geography and Delayed Coalescence*

*Mod les du « stepping-stone » avec zone g ographique continue et coalescence retard e*

Stepping-stone models of population genetics are models which consider separate populations on discrete islands, with migration between the islands. For example, each island might have a population of birds; occasionally, a bird flies across the water to a neighbouring island. In this talk we will look at a model of population genetics with *continuous* geography. Rather than constructing it directly as a forwards-in-time model of evolution, we instead build the process from its associated dual process. We will also consider some properties of this model. This is joint work with Steve Evans.

Les mod les du « stepping-stone » de g n tique des populations consid rent des populations sur des  les s par es, avec des migrations entre ces  les. Par exemple, chaque  le aurait une population d'oiseaux, et occasionnellement, un oiseau volerait sur l' le voisine. Dans cet expos , nous  tudierons un mod le de g n tique des populations sur une zone g ographique *continue*. Au lieu de construire directement un mod le d' volution pr visible, nous en construisons un   partir du processus dual connexe. Nous allons  galement consid rer certaines propri t s de ce mod le. Ces travaux ont  t  effectu s conjointement avec Steve Evans.

---

**ORI GUREL-GUREVICH** (University of British Columbia), **ITAI BENJAMINI** (Weizmann Institute of Science), **RUSSELL LYONS** (Indiana University), **ODED SCHRAMM** (Microsoft Research)

[Wednesday June 15 / mercredi 15 juin, 10:45]

*Recurrence of the Simple Random Walk Path*

*R currence de la trajectoire de la marche al atoire simple*

A simple random walk (SRW) on a graph is a Markov chain whose state space is the vertex set and the next state distribution is uniform among the neighbors of the current state. A graph is called recurrent if a SRW on it returns to the starting vertex with probability 1, and called transient otherwise. The path of a walk on a graph is simply the set of edges this walk has traversed. Our main result is that the path of a SRW on any graph is a recurrent graph. Joint works with Itai Benjamini, Russell Lyons and Oded Schramm.

Une marche al atoire simple sur un graphe est une cha ne de Markov dont l'espace d' tat est l'ensemble de sommets et dont la prochaine transition est distribu e de fa on uniforme parmi les voisins de l' tat actuel. Un graphe est qualifi  de r curren t si une marche al atoire retourne au sommet de d part avec probabilit  un, sinon il est transitoire. La trajectoire d'une marche sur un graphe est simplement l'ensemble d'ar tes que cette marche a travers es. Le r sultat principal que nous avons obtenu, est que la trajectoire d'une marche al atoire simple sur tout graphe constitue un graphe r curren t. Travaux conjoints avec Itai Benjamini, Russell Lyons et Oded Schramm.

## 3B-I4: Randomness in Discrete Mathematics and Biology Stochasticité en mathématiques discrètes et biologie

---

**XIAOWEN ZHOU** (Concordia University)

[Wednesday June 15 / mercredi 15 juin, 11:15]

*A Generalized Fleming-Viot Process*

*Un processus de Fleming-Viot généralisé*

The Fleming-Viot superprocess is a probability-measure-valued process for population genetics. It describes the evolution of the relative frequencies of different alleles in a large population. When only mutation and re-sampling are involved, its moments are determined by the dual process of Kingman coalescent together with the mutation semi-group. In this talk we introduce a generalized Fleming-Viot process obtained by replacing the Kingman coalescent with the coalescent of simultaneous multiple collisions. We will discuss its existence, its reversibility and the associated SPDE. This talk is based on work in progress with Zenghu Li, Huili Liu and Jie Xiong.

Le superprocessus de Fleming-Viot est un processus prenant des mesures de probabilité comme valeurs pour la génétique des populations. Il décrit l'évolution des fréquences relatives de différents allèles dans une grande population. Lorsqu'il implique seulement une mutation et un ré-échantillonnage, les moments de ce processus sont déterminés par le processus dual de coalescence de Kingman avec le semi-groupe de mutation. Dans cet exposé, nous présenterons un processus de Fleming-Viot généralisé obtenu en remplaçant le coalescent de Kingman par le coalescent à collisions multiples simultanées. Nous discuterons de l'existence, de la réversibilité et des équations différentielles stochastiques aux dérivés partielles connexes de ce processus. Cet exposé est basé sur les travaux en cours avec Zenghu Li, Huili Liu et Jie Xiong.

## 3B-15: Statistical Problems Arising in the Lumber Industry La statistique en assurance de la qualité du bois d'œuvre

---

Chair/Président: Will Welch (University of British Columbia)  
Organizer/Responsable: Jim Zidek (University of British Columbia)

Session sponsored by the Business and Industrial Statistics Section

Session parrainée par le Groupe de statistique industrielle et de gestion

**Room/Salle: BAC141**

**Abstracts/Résumés**

---

**CONROY LUM** (FP Innovations, Vancouver), **JAMES ZIDEK & CAROLYN TAYLOR** (University of British Columbia)  
[Wednesday June 15 / mercredi 15 juin, 10:15]

*An Overview of the use of Statistics and Stochastic Modeling in Forest Products Research*

*Un aperçu de l'utilisation de la modélisation statistique et stochastique dans la recherche sur les produits de la forêt*

The Forest Products Stochastic Modeling Group is currently focused on the strength values of single pieces of lumber. Next steps include developing an understanding of how lumber strength is related to more basic wood properties, and how an assembly of lumber with varying properties is able to safely resist extreme loads. In this presentation, examples with an emphasis on the role of stochastic modeling will be provided to link the knowledge of forest, wood product manufacturing, wood product performance, and assembly performance to procedures that engineers use to specify wood products in building applications.

Le groupe de modélisation stochastique des produits de la forêt se concentre présentement sur les valeurs de résistance de pièces simples de bois d'œuvre. Les étapes subséquentes visent à comprendre comment la résistance du bois d'œuvre est liée à des propriétés de base du bois et comment un assemblage de bois d'œuvre comportant différentes propriétés peut résister de façon sécuritaire à des chargements extrêmes. Dans cet exposé, nous donnerons des exemples en mettant l'emphase sur le rôle de la modélisation stochastique pour relier les connaissances sur la forêt, les produits manufacturés du bois et la performance des produits du bois aux procédures qu'emploient les ingénieurs pour décrire les spécifications des produits du bois dans les applications en bâtiment.

---

**JUN ZHU** (University of Wisconsin, Madison), **YANBING ZHENG** (University of Kentucky), **BRIAN AUKEMA** (University of Minnesota), **ALLAN CARROLL** (University of British Columbia), **KISHAN SAMBARAJU** (University of Northern British Columbia)

[Wednesday June 15 / mercredi 15 juin, 10:45]

*Spatial-Temporal Statistics for Studying Mountain Pine Beetle Infestation of Pine Forests in Western Canada*

*Des statistiques spatio-temporelles pour étudier l'infestation de dendroctone du pin ponderosa dans les forêts de pin de l'ouest canadien*

Insects and diseases exert immense economic loss to forests on a landscape-level scale through reduced tree growth, vigour and increased mortality. Mountain pine beetle is a landscape-altering insect that infests forests of North America. There has been a plethora of work on models examining population dynamics of insects, but few link space, time, and environmental factors simultaneously.

Les insectes et les maladies entraînent des pertes économiques énormes à l'échelle du paysage des forêts en réduisant la croissance, la vigueur et en augmentant la mortalité des arbres. La dendroctone du pin ponderosa est un insecte qui modifie le paysage en infestant des forêts de l'Amérique du Nord. Plusieurs travaux ont porté sur des modèles examinant la dynamique de la population d'insectes, mais peu l'ont fait en étudiant simultanément le lien entre

### 3B-15: Statistical Problems Arising in the Lumber Industry La statistique en assurance de la qualité du bois d'œuvre

---

We review existing approaches and propose new spatial-temporal landscape models. We highlight some of the modeling difficulties and offer possible solutions. Furthermore, we study individual associations of several biologically-relevant cold temperature variables, and other temperature/degree-day terms, with outbreak occurrences in western Canada.

les facteurs espace, temps et environnement. Nous passons en revue les approches actuelles et proposons des modèles spatio-temporels du paysage. Nous mettons l'emphasis sur certaines difficultés de modélisation et offrons des solutions potentielles. De plus, nous étudions les associations individuelles entre plusieurs variables de température froide qui sont pertinentes au niveau biologique et d'autres de type température/degré-jour, lors de l'émergence d'épidémies dans l'ouest canadien.

---

**JIGUO CAO & JING CAI** (Simon Fraser University), **JIM ZIDEK** (University of British Columbia)

[Wednesday June 15 / mercredi 15 juin, 11:15]

*Comparing and Selecting Damage Accumulation Models*

*Comparer et choisir des modèles d'accumulation des dommages*

Duration of load, which is the time during which a load is placed on a wood member, is one of the most significant factors in wood structural design. Several damage accumulation models have been developed to quantify the phenomenon of creep-rupture in wood. These models describe the duration-of-load effects and are used to predict the time-to-failure distributions. We compare and select these damage accumulation models based on their prediction accuracy in the framework of Bayesian hierarchical models.

La durée du chargement, qui correspond au temps pendant lequel un chargement est placé sur un élément en bois, est l'un des facteurs les plus importants du modèle structurel du bois. Plusieurs modèles d'accumulation des dommages ont été développés pour quantifier le phénomène de fluage-rupture du bois. Ces modèles décrivent les effets de la durée du chargement et sont utilisés pour prédire les distributions du temps de défaillance. Nous comparons et sélectionnons ces modèles d'accumulation des dommages en se basant sur la puissance prédictive dans le cadre de modèles hiérarchiques bayésiens.



## 3B-C1: Design and Analysis of Clinical Studies and Trials Conception et analyse d'études et d'essais cliniques

---

Chair/Président: James Hanley (McGill University)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 132**

**Abstracts/Résumés**

---

**STEPHEN D WALTER** (McMaster University)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*Optimal Treatment Allocation in Two-Stage Randomised Clinical Trials*

*Optimisation de l'allocation aux traitements dans les essais cliniques aléatoires à deux phases*

In clinical trials, patient outcomes may be affected by the choice of treatment that patients (if so permitted) would make (a selection effect), and by whether patients actually receive their preferred treatment (a preference effect). Selection and preference effects can be estimated in two-stage randomised trials, where patients are first randomly divided into two subgroups. In one subgroup, participants are randomised to treatments, while in the other, patients may choose their treatment. We derive the optimal proportions of patients to allocate to the random or choice subgroups, in order to maximise the precision of the effects of greatest interest.

En essais clinique, les résultats du patient peuvent être affectés par le choix de traitement que le patient (si permis) pourrait faire (effet de sélection), et par le fait que le patient reçoive actuellement son traitement favorisé (effet de préférence). Les effets de préférence et de sélection peuvent être estimés en essais aléatoires à deux phases, où les patients sont d'abord divisés en deux sous-groupes. Dans un sous-groupe, les patients se voient assigner un traitement de façon aléatoire et, dans l'autre, les patients choisissent eux-mêmes leur traitement. Nous calculons la proportion optimale de patients à allouer aux sous-groupes choix et aléatoire, de façon à maximiser la précision des effets de plus grand intérêt.

---

**QING GUO, GEOFF HALL, MARGARET MCKINNON, LEHANA THABANE & ELEANOR PULLENAYEGUM** (McMaster University)

[Wednesday June 15 / mercredi 15 juin, 10:30]

*Setting Sample Size Using Cost-efficiency in fMRI Studies*

*Établissement de la taille d'échantillon selon le ratio coût-efficacité en études d'IRMf*

Functional MRI (fMRI) studies are expensive, and cost is often the influential component in setting sample size. However, standard sample size and power calculations overlook cost. Here, we describe a recently proposed alternative that incorporates cost, namely a cost efficiency approach. Cost-efficiency is the ratio of a study's projected value to its cost and the sample size is chosen to maximize value per unit cost. Cost-efficiency provides a simple way to justify a sample size that cannot be charged as inadequate because the pro-

Les études d'IRM fonctionnelles (IRMf) sont dispendieuses, et le coût est souvent un facteur important dans l'établissement de la taille d'échantillon. Cependant, la taille d'échantillon standard et les calculs de la puissance ne tiennent pas compte du coût. Nous décrivons ici une alternative proposée récemment qui tient compte du coût, soit une approche axée sur le ratio coût-efficacité. Ce ratio est composé de la valeur projetée d'une étude par rapport à son coût, et la taille de l'échantillon est choisie de façon à maximiser la valeur par coût unitaire. Cette approche fournit une façon simple de s'assurer que la taille de l'échantillon ne soit pas considérée

### 3B-C1: Design and Analysis of Clinical Studies and Trials Conception et analyse d'études et d'essais cliniques

---

posed sample size is more cost-efficient than any larger choice. This approach is illustrated though an example.

inadéquate, puisqu'elle offre un meilleur ratio coût-efficacité que tout autre taille plus grande. L'approche est illustrée à l'aide d'un exemple.

---

**XUAN LI & XIKUI WANG** (University of Manitoba)

[Wednesday June 15 / mercredi 15 juin, 10:45]

*Response Adaptive Designs with Misclassified Responses*

*Plans de réponse adaptative avec réponses mal classées*

We consider response adaptive designs when the binary response may be misclassified and extend relevant results in the literature. We derive the optimal allocations under various objectives and examine the relationship between the power of statistical test and the variability of treatment allocation. Asymptotically best response adaptive randomization procedures and effects of misclassification on the optimal allocations are investigated. A real-life clinical trial is also discussed to illustrate our proposed approach.

Nous considérons les plans de réponse adaptative lorsque la réponse binaire est mal classée et nous présentons les résultats importants dans la littérature. Nous obtenons les répartitions optimales en fonction de divers objectifs, et nous examinons la relation entre la puissance du test statistique et la variabilité de la répartition des traitements. Nous analysons les meilleures procédures de randomisation de réponse adaptative asymptotiquement ainsi que les effets du mauvais classement des répartitions optimales. Nous discutons également d'un essai clinique réel pour illustrer notre approche.

---

**YANQING YI** (Memorial University of Newfoundland)

[Wednesday June 15 / mercredi 15 juin, 11:00]

*Small Sample Inference for Response Adaptive Designs*

*Inférence sur des échantillons de petite taille pour les plans à réponses adaptées*

Response adaptive designs use the available information of treatment effect to modify treatment allocation probabilities to assign more patients to the potential treatment. The modification treatment allocation probabilities create a dependency in the data collected from a response adaptive clinical trial. This study explores the statistical inference for the designs when sample sizes are small. Exact distribution of allocation proportion will be presented and exact statistical power function will be compared for different designs.

Les plans à réponses adaptées utilisent l'information disponible sur l'effet de traitement pour modifier les probabilités d'allocation à un traitement et ainsi assigner plus de patients à un traitement potentiel. La modification des probabilités d'allocation à un traitement crée une dépendance dans les données recueillies pour un essai clinique à réponses adaptées. Cette étude explore l'inférence statistique pour les plans de petites tailles échantillonales. La distribution exacte des proportions d'allocation sera présentée et nous comparerons la fonction exacte de puissance statistique pour différents plans.

---

**ERICA MOODIE** (McGill University), **BIBHAS CHAKRABORTY** (Columbia University)

[Wednesday June 15 / mercredi 15 juin, 11:15]

*Estimation of Optimal Dynamic Treatment Rules with Shared Parameters and Non-Regularity*

*Estimation pour les régimes de traitement dynamique optimaux avec des paramètres partagés et de la non-régularité*

The area of dynamic treatment regimes (DTR) aims to make inference about adaptive, multistage decision-making in clinical practice. A DTR is a

Les régimes de traitement dynamique (RTD) visent à faire de l'inférence au sujet de décisions adaptatives prises en plusieurs étapes. Un RTD est un ensemble de règles, un par

### 3B-C1: Design and Analysis of Clinical Studies and Trials Conception et analyse d'études et d'essais cliniques

---

set of decision rules, one per interval of treatment, where each decision is a function of treatment and covariate history which returns a recommended treatment. In some studies, the decision rule at each stage is the same function of the history and parameters are shared across intervals. Because of this, many of the usual, recursive methods of estimation are not appropriate. We propose an estimation procedure for the optimal DTR with shared decision rules which adjusts for non-regularity.

intervalle de traitement, où chaque décision est une fonction des traitements et de covariables historiques qui renvoie le traitement recommandé. Dans certaines études, la règle de décision à chaque étape est la même fonction de l'histoire, et les paramètres sont partagés entre les intervalles. Dans ces cas, les méthodes d'estimation récursive ne sont pas appropriées. Nous proposons une procédure d'estimation pour le RTD optimal avec des paramètres partagés et de la non-régularité.

## 3B-C2: Harmonic Analysis and Spectral Methods for Time Series Analyse harmonique et méthodes spectrales pour les séries chronologiques

---

Chair/Président: Mary Thompson (University of Waterloo)  
Organizer/Responsable: Wesley Burr (Queen's University)

**Room/Salle: BAC237**

### Abstracts/Résumés

---

**DAVID J. THOMSON** (Queen's University)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*Crossing Problems for Spectrum Estimates*

*Problèmes croisés pour les estimateurs du spectre*

I give estimates of the level crossing properties of nonparametric estimate of the power spectrum made from a sample of a purely non-deterministic random process. Level crossing problems have been extensively studied in the time domain but the procedures do not translate directly to the frequency domain. This talk introduces the problem, summarizes results to date, and shows that even when the original data is significantly non-Gaussian, they accurately describe the spacing and shapes of random peaks observed in spectrum estimates and so can be used to quantify coincidence detectors and false-discovery rates.

Je donne des estimateurs pour le niveau croisé de propriétés d'un estimateur non-paramétrique de la puissance spectrale par un échantillon d'un processus aléatoire purement non-déterministe. Les problèmes à niveaux croisés ont été considérablement étudié dans le domaine du temps, mais les procédures ne se transfèrent pas directement au domaine de la fréquence. Cet exposé introduit le problème, résume les résultats actuels et montre que même si les données originales sont significativement non-normales, elles décrivent précisément l'espace et la forme des sommets aléatoires observés dans l'estimation du spectre et peuvent donc être utiles pour quantifier les détecteurs de coïncidence et les taux de fausse découverte.

---

**WESLEY BURR** (Queen's University Dept. of Mathematics and Statistics)

[Wednesday June 15 / mercredi 15 juin, 10:30]

*Cross-Spectral Method for Gapfilling Time Series*

*Méthode interspectrale de complétion des séries chronologiques*

A common problem in time series analysis is the filling of 'gaps' in the data record. We consider a new technique for generating interpolated values using external data records, especially in the context of multi-site observations of physical phenomena using estimated auto- and cross-correlations. The test case is the daily 2695 GHz noon flux record from the Sagamore Hill, Mass. observatory, with external site the DRAO in Penticton, BC observing at 2802 GHz. Daily contiguous records from Penticton are used in conjunction with time-lagged records from Sagamore to fill missing data gaps due to instrumental failure.

L'analyse des séries chronologiques est souvent compliquée par le besoin de compléter les « données manquantes ». Nous explorons une nouvelle technique permettant de générer des valeurs interpolées à l'aide de données externes, notamment dans le contexte d'observations de phénomènes physiques sur plusieurs sites en utilisant des autocorrélations et intercorrélations estimées. Le cas d'essai est l'enregistrement quotidien, à midi, du flux 2695 GHz par l'observatoire de Sagamore Hill, Mass., le site externe étant celui de l'OFR à Penticton (C.-B.), qui observe à 2802 GHz. Les enregistrements quotidiens contigus de Penticton sont utilisés en liaison avec les enregistrements temporisés de Sagamore pour combler les données manquantes en cas de panne d'instrument.

## 3B-C2: Harmonic Analysis and Spectral Methods for Time Series Analyse harmonique et méthodes spectrales pour les séries chronologiques

---

**AZADEH MOGHADERI & GLEN TAKAHARA** (Queen's University)

[Wednesday June 15 / mercredi 15 juin, 10:45]

*Tests for Stationarity Based on Minors and Singular Values of the Nonstationary Spectrum*

*Tests pour la stationnarité basés sur les mineurs et valeurs singulières du spectre non stationnaire*

We consider the Evolutionary Spectrum (ES) of nonstationary processes. Considering the ES and its log-transformation on any 2-dimensional finite grid of time-frequency values, we obtain two matrices. The spectral matrix has rank 1 for stationary and Uniformly Modulated (UM) processes and greater than 1 otherwise. The log-spectral matrix has rank 1 for stationary processes, 2 for UM processes, and greater than 2 otherwise. We develop graphical and statistical tests based on the singular values and 2 by 2 minors of the estimates of these matrices to decide on their ranks. We illustrate the performance of our test via simulations.

Nous considérons un spectre évolutif (SE) de processus non stationnaires. Sachant SE et sa transformation logarithmique sur n'importe quelle grille finie de deux dimensions de valeurs temps-fréquence, nous obtenons deux matrices. La matrice spectrale est de rang 1 pour les processus stationnaires et uniformément modulés (UM) et de rang supérieur à 1 autrement. La matrice spectrale logarithmique est de rang 1 pour les processus stationnaires, de rang 2 pour les processus UM, et de rang supérieur à 2 autrement. Pour décider du rang de ces matrices, nous développons des tests graphiques et statistiques basés sur les valeurs singulières et les mineurs 2 par 2 de leurs estimateurs. Des simulations sont utilisées pour illustrer la performance de nos tests.

**CHARLOTTE LISA HALEY** (Queen's University)

[Wednesday June 15 / mercredi 15 juin, 11:00]

*Detection of Changes in Activation in ECG Time Series Modeling*

*Détection des changements d'activation dans la modélisation de séries chronologiques d'électrocardiogrammes (ECG)*

It is of interest to cardiologists studying intracardiac ECGs to determine whether or not signal activity has changed over the course of a recording. These changes may reflect life-threatening cardiac events and appear in the electrogram as transitions between quiet and active signals. Atrial fibrillation ECGs have been previously described as sub-Gaussian processes, while changes in the underlying signal distribution can occur even in short-term recordings. We seek to characterize these changes from a time series theoretical basis. In order to do this, we propose that the data can be sectioned, sorted and correlated in successive pairs of data sections.

Les cardiologues s'intéressent à l'étude des électrocardiogrammes (ECG) endocavitaires pour déterminer si l'activité a changé au cours d'un enregistrement. Ces changements peuvent signaler des événements cardiaques constituant un danger de mort et peuvent apparaître sur l'ECG comme des transitions entre les signaux de stabilité et d'activité. Les ECG de fibrillation auriculaire ont été décrits précédemment par des processus sous-gaussiens, alors que les changements dans la distribution des signaux sous-jacents peuvent se produire même dans des enregistrements de courte durée. Nous cherchons à caractériser ces changements sur la base théorique de séries chronologiques. Nous proposons ainsi de fractionner, de trier et de corrélérer les données en paires successives de sections de données.

**KARIM J. RAHIM & DAVID J. THOMSON** (Queen's University)

[Wednesday June 15 / mercredi 15 juin, 11:15]

*Spectral Analysis of Burgundy Pinot Noir Grape Harvest Dates*

*Analyse spectral des dates de récoltes des raisins Pinot noirs de Bourgogne*

We perform an analysis of Burgundy pinot noir grape harvest dates. This annual record re-

Nous effectuons une analyse des dates de récoltes des raisins Pinot noirs de Bourgogne. Les données annuelles rap-

### 3B-C2: Harmonic Analysis and Spectral Methods for Time Series

#### Analyse harmonique et méthodes spectrales pour les séries chronologiques

---

ports the annual grape harvest date in days after September 1st, which has been used in summer temperature reconstructions. We perform spectral analysis and dynamic spectral analysis of the data where we find evidence of coherence indicating solar influence. Specifically we find structure in the dynamic spectra indicating solar influence, and there is evidence of the non-stationary structure associated with Seuss cycles as seen in other historical temperature data.

portent les dates de récoltes en jours depuis le premier septembre, et ont été utilisées pour la reconstruction des températures d'été. Nous effectuons une analyse spectrale et une analyse spectrale dynamique des données et trouvons preuve d'une cohérence indiquant l'influence solaire. Spécifiquement, nous trouvons une structure dans le spectre dynamique indiquant l'influence solaire, et la preuve d'une structure non stationnaire associée aux cycles de Seuss comme observée par d'autres données historiques de températures.

---

**WESLEY BURR** (Queen's University Dept. of Mathematics and Statistics), **AZADEH MOGHADERI & GLEN TAKAHARA** (Queen's University), **HWASHIN H. SHIN** (Health Canada)

[Wednesday June 15 / mercredi 15 juin, 11:30]

*Estimation and Extraction of Harmonic Features from Pollution and Mortality Time Series*

*Estimation et extraction des caractéristiques harmoniques de séries chronologiques sur la pollution et la mortalité*

20 years of daily pollution and mortality data from a number of Canadian urban centres was analyzed as part of a Health Canada project to study the time-varying health effects of air pollution. For analysis, the time series must be decomposed into deterministic and stochastic parts. The challenge lies in the estimation of the deterministic portion, especially with presence of nonstationarity. Empirical Mode Decomposition was applied to data for slowly-varying structure (trend) estimation. Additional time- and frequency-domain techniques were applied to the residual (detrended) time series for line component estimation and extraction. The deterministic and harmonic structure found is discussed.

20 années de données quotidiennes sur la pollution et la mortalité en provenance de plusieurs centres urbains au Canada ont été analysées dans le cadre d'un projet de Santé Canada visant à étudier les effets sur la santé de la pollution atmosphérique dans le temps. Aux fins de l'analyse, les séries chronologiques doivent être décomposées en parties déterministes et stochastiques. La difficulté réside dans l'estimation de la portion déterministe, surtout en présence de non-stationnarité. La décomposition en modes empiriques est appliquée aux données pour estimer la structure à variation lente (tendance). D'autres techniques en domaine temporel et fréquentiel sont appliquées à la série chronologique résiduelle (filtrée) pour estimer et extraire la composante linéaire. Nous discutons de la structure déterministe et harmonique ainsi découverte.

## 3B-C3: Modelling Correlation Structure in Repeated Measures and Multivariate Data Modélisation de la structure de corrélation dans les mesures répétées et les données multivariées

---

Chair/Président: Sohee Kang (Trent University)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 239**

**Abstracts/Résumés**

---

**TOLULOPE T. SAJOBI & LISA M. LIX** (University of Saskatchewan), **BOLANLE M. DANSU** (University of Agriculture, Abeokuta, Nigeria)

[Wednesday June 15 / mercredi 15 juin, 10:15]

*Robust Descriptive Discriminant Analysis for Repeated Measures Data*

*Analyse discriminante descriptive robuste pour les mesures répétées*

Robust repeated measures discriminant analysis (RMDA) procedures based on parsimonious covariance structures were developed using trimmed estimators. The effects of non-normality, covariance structure, and mean configuration on bias and root mean square error (RMSE) of RMDA coefficients were studied using Monte Carlo techniques. The bias and RMSE values of robust RMDA coefficients were at least 10% and 5% smaller than those of coefficients for DA procedures based on least squares/maximum likelihood estimators when data were non-normal and the covariance structure was correctly specified. The proposed procedures are useful to identify the repeated measurements that describe group separation for non-normal data.

Des procédures robustes d'analyse discriminante des mesures répétées (ADMR) basées sur les structures de covariance parcimonieuses ont été développées à partir d'estimateurs tronqués. Les effets de la non-normalité, de la structure de covariance et de la configuration de la moyenne sur le biais et la racine carrée de l'erreur quadratique moyenne des coefficients RMDA sont étudiés par les méthodes de Monte Carlo. Le biais et les valeurs du RMSE des coefficients de la RMDA sont d'au moins 10% et 5% plus petits que ceux des coefficients des procédures d'analyse discriminante (AD) basées sur les estimateurs des moindres carrés et du maximum de vraisemblance, quand les données n'étaient pas normales et que la structure de covariance était correctement spécifiée. Les procédures proposées sont utiles pour identifier les mesures répétées décrivant la séparation du groupe pour les données non-normales.

---

**VINEETHA K.V. WARRIYAR & SUTRADHAR BRAJENDRA** (Memorial University of Newfoundland)

[Wednesday June 15 / mercredi 15 juin, 10:30]

*Generalized Least Square Estimation in Semiparametric Linear Models with Correlated Errors*

*Estimation des moindres carrés généralisée en modèles linéaires semi-paramétriques avec erreurs corrélées*

As opposed to the parametric longitudinal set up, the repeated responses in a semiparametric longitudinal model are influenced by a partially specified regression function and certain correlated errors. It is known that the efficiency of the regression estimates is adversely affected by the use of misspecified correlation structure. In this paper, we use a general auto-correlation structure based semi-

À l'opposé de la configuration longitudinale paramétrique, les réponses répétées dans un modèle longitudinal semi-paramétrique sont influencées par une fonction de régression partielle spécifiée et certaines erreurs corrélées. On sait que l'efficacité des estimateurs de régression est inversement affectée par l'utilisation de structures de corrélation mal spécifiées. Dans cet exposé, nous utilisons une structure d'auto-corrélation générale basée sur une approche

### 3B-C3: Modelling Correlation Structure in Repeated Measures and Multivariate Data Modélisation de la structure de corrélation dans les mesures répétées et les données multivariées

---

parametric generalized least square (SGLS) estimation approach that provides consistent and efficient regression estimates. We have compared the efficiency of this approach with other existing approaches including an estimating equation approach where the correlation structure is estimated non-parametrically.

d'estimation des moindres carrés semi-paramétrique généralisée qui nous procure des estimateurs de régression cohérents et efficaces. Nous avons comparé l'efficacité de cette approche avec d'autres approches existantes incluant une approche d'équation estimatrice, où la structure de corrélation est estimée de façon non-paramétrique.

---

**ALINA DYACHENKO** (McGill University), **ANTONIO CIAMPI**, **MARTIN COLE** & **JANE MCCUSKER** (McGill University)

[Wednesday June 15 / mercredi 15 juin, 10:45]

*Latent Classes and Hidden Markov Chains to Describe State and Course of a Disease: Delirium as an Example*  
*Classes latentes et chaînes de Markov cachées pour décrire l'état et le cours d'une maladie : l'exemple du Délirium*

Disease state and disease course are fundamental concept in Clinical Epidemiology. Suppose we measure a multivariate disease index at a number of time points on a group of patients. To define state and disease course directly from this type of multivariate longitudinal data, we propose a model in which a latent class at each time represents the state of the disease, while the course of the disease is represented by transitions from one state to the other over time, modeled by a Hidden Markov Chain. We show results for Delirium, a disorder commonly observed in elderly hospitalized patients.

L'état et le cours d'une maladie sont des concepts fondamentaux en Épidémiologie Clinique. Nous considérons un index multivarié de maladie, mesuré à intervalles de temps réguliers sur un groupe de patients. Afin de définir l'état et le cours d'une maladie à partir de ce type de données, nous proposons un modèle représentant un état par une classe latente, et un cours par une transition d'un état à l'autre, décrit par chaîne de Markov cachée. Nous donnons un exemple d'analyse pour le Délirium, un trouble cognitif observé couramment chez les patients âgés hospitalisés.

---

**HÉLA ROMDHANI**, **LOUIS-PAUL RIVEST** & **LAJMI LAKHAL CHAIEB** (Département de mathématiques et de statistique, Université Laval)

[Wednesday June 15 / mercredi 15 juin, 11:00]

*A Kendall's  $\tau$  Estimator for Exchangeable Data with an Application to One Random Factor Analysis of Variance.*  
*Un estimateur  $\tau$  de Kendall pour les données permutables et une application à l'analyse de variance à un facteur aléatoire*

We propose an estimator  $\tau_n^*$  of Kendall's  $\tau$  adapted to exchangeable data. Under a bivariate copula model, we give an expression of the finite sample size variance of  $\tau_n^*$  in terms of the copula. In the multivariate case, we derive the asymptotic distribution of  $\tau_n^*$  and give a nonparametric estimator for its variance. The relationship between the correlation coefficient  $\rho$  and Kendall's  $\tau$  has been investigated and found to be explicit for some multivariate models such as the elliptical distributions. These results allow us to perform an analysis of variance with one random effect when the observations are

Nous proposons un estimateur  $\tau_n^*$  pour le  $\tau$  de Kendall adapté aux données permutables. Sous un modèle de copule bivariable, nous donnons une expression de la variance de  $\tau_n^*$  d'un échantillon de taille finie en terme de la copule. Dans le cas multivarié, nous calculons la distribution asymptotique de  $\tau_n^*$  et donnons un estimateur non-paramétrique pour sa variance. La relation entre le coefficient de corrélation  $\rho$  et le  $\tau$  de Kendall a été recherchée et montrée comme étant explicite pour quelques modèles multivariés comme les distributions elliptiques. Ces résultats nous permettent d'effectuer l'analyse de variance à un effet aléatoire quand les observations sont distribuées elliptiquement.



### 3B-C3: Modelling Correlation Structure in Repeated Measures and Multivariate Data Modélisation de la structure de corrélation dans les mesures répétées et les données multivariées

---

elliptically distributed.

---

**SWARNA WEERASINGHE** (Dalhousie University)

[Wednesday June 15 / mercredi 15 juin, 11:15]

*Multilevel Time Series Decomposition Methods to Explore the Association between Asthma Hospital Admissions and Ozone Levels*

*Méthodes de décomposition à plusieurs niveaux de séries chronologiques pour explorer l'association entre les admissions d'asthmatiques dans les hôpitaux et les niveaux d'ozone*

Many air pollution time series demonstrates three significant components, long term annual trends, seasonal variations and irregular residual variations. In this presentation, I propose a multilevel time series decomposition method to estimate annual trends as level 1 marginal estimates, monthly seasonality as level 2 marginal estimates and the third level capture hourly variations. The components are estimated using residual components in both asthma and ozone series. Traditional methods filter out both level 1 and level 2 variations and focused only on the affect of irregular variations. Merits of the method are also discussed.

Plusieurs séries chronologiques sur la pollution de l'air démontrent trois principaux composants : une tendance annuelle à long terme, des variations saisonnière et une variation résiduelle irrégulière. Dans cette présentation, je propose une méthode de décomposition à plusieurs niveaux de séries chronologiques pour estimer la tendance annuelle comme estimateur marginal de niveau 1, la saisonnalité mensuelle comme estimateur marginal de niveau 2 et le niveau 3 introduit les variations horaires. Les composantes sont estimées par les composants résiduels dans les séries sur l'asthme et l'ozone. Les méthodes traditionnelles font ressortir les variations des niveaux 1 et 2 et visent seulement les effets des variations irrégulières. Les vertus de la méthode sont également abordées.

**3C-A: CRM-SSC Prize in Statistics Address  
Allocution du récipiendaire du Prix CRM-SSC**

---

Chair/Président: David Stephens (McGill University)

**Room/Salle: Festival Theatre**

**Abstract/Résumé**

---

**EDWARD SUSKO** (Dalhousie University)

[Wednesday June 15 / mercredi 15 juin, 13:00]

*Statistical Issues in Molecular Evolution*

*Problèmes statistiques dans le domaine de l'évolution moléculaire*

I review methods for evolutionary inference from sequence data and present some of our recent strategies for realistically modeling biological processes giving rise to such data. Uncertainty about evolutionary relationships is usually assessed through bootstrap support (BP) for splits, large values indicating strong support. A difficulty, however, has been deciding how large is large enough. We discuss ways of adjusting BP so that it is similar to a p-value. A related issue, method performance for data generated from a star tree, gives rise to an unusual statistical situation in which Bayesian and ML methods give strikingly different results, even asymptotically.

Je passe en revue différentes méthodes pour l'inférence évolutive à partir de données de séquences et présente nos nouvelles stratégies pour modéliser de façon réaliste les processus biologiques à l'origine de telles données. L'incertitude face aux relations évolutives est habituellement évaluée à l'aide du support bootstrap (BP) pour les divisions, une grande valeur de BP soutenant la division. Il est toutefois difficile de décider à partir de quel seuil le BP est suffisamment grand. Nous présentons des méthodes pour ajuster BP de sorte qu'il soit semblable à une valeur de p. Un problème connexe, la performance de méthodes pour données générées à partir d'un arbre étoile, donne lieu à une situation inhabituelle où les méthodes bayésiennes et du maximum de vraisemblance donnent des résultats remarquablement différents, même asymptotiquement.

## **3C-I1: Consultations for Statistical Consultants: Lessons Learned from the Field Consultations pour les consultants en statistique : leçons apprises sur le terrain (discussion en groupe)**

---

Organizer and Chair / Responsable et président: Depeng Jiang (University of Manitoba)

Session sponsored by the Biostatistics Section

Session parrainée par le Groupe de biostatistique

**Room/Salle: BAC 132**

**Abstracts/Résumés**

---

**JOHN AMRHEIN**, McDougall Scientific Ltd  
[Wednesday June 15 / mercredi 15 juin, 13:00]

---

**DEPENG JIANG** (University of Manitoba)  
[Wednesday June 15 / mercredi 15 juin, 13:00]

*Round-Table Discussion - Consultations for Statistical Consultants: Lessons Learned from the Field*  
*Table ronde - consultations pour les conseillers en statistique: leçons apprises sur le terrain*

Statistical consulting, often a major component of a biostatistician's job, is a complex activity that requires statistical and non-statistical skills (interpersonal and negotiation skills). Consultants must know how to: work effectively with research scientists, summarize the statistical aspects of substantive problems, provide appropriate technical assistance, and effectively communicate statistical results. This session will feature three panelists sharing their advice for running a successful biostatistical consulting service. The aim is to provide a discussion forum for the nature of statistical consulting, the preparation and education background for a good consultant, and the need for practical(hands-on)consultancy in training biostatisticians for success.

La consultation statistique, souvent une composante importante du travail d'un biostatisticien, est une activité complexe qui exige des compétences statistiques et non statistiques (compétences interpersonnelles et de négociation). Les consultants doivent savoir comment travailler efficacement avec des scientifiques de recherche, résumer les aspects statistiques des problèmes de fond afin de fournir une assistance technique appropriée, et communiquer efficacement les résultats statistiques. Cette session mettra en vedette trois experts qui partageront leurs trucs et conseils pour la mise en oeuvre d'un service de consultation en biostatistique réussi. L'objectif de cette table ronde est de fournir un forum de discussion sur la nature de la consultation statistique, sur les pré-requis pour devenir un bon consultant, et sur le besoin de formation pratique pour assurer le succès des futurs biostatisticiens.

---

**TIM RAMSAY**, Ottawa Hospital Research Institute and University of Ottawa  
[Wednesday June 15 / mercredi 15 juin, 13:00]

---

**LEHANA THABANE**, McMaster University  
[Wednesday June 15 / mercredi 15 juin, 13:00]

## 3C-I2: Rank Set Sampling: New Developments and Applications Échantillonnage d'ensembles ordonnés : récents développements et applications

---

Chair/Président: Sylvia Esterby (University of British Columbia)  
Organizer/Responsable: Abdulkadir Hussein (University of Windsor)

**Room/Salle: BAC 138**

### Abstracts/Résumés

---

**S. EJAZ AHMED** (University of Windsor)

[Wednesday June 15 / mercredi 15 juin, 13:00]

*Meta Analysis and Shrinkage Estimation Strategies in Ranked Set Sampling*

*Méta-analyse et stratégies d'estimation par rétrécissement dans l'échantillonnage d'ensembles ordonnés*

I consider the implementation and application of shrinkage and pretest estimation for population means from arbitrary populations based on multiple ranked set samples (RSS). I will discuss the basis for optimally combining sample information emerging from several data sources. Asymptotic properties of the proposed estimators are appraised and compared with the baseline estimators. We demonstrate that the shrinkage estimator will have a substantial quadratic risk reduction relative to the benchmark estimator based on RSS. A simulation study is conducted and the results strongly corroborates with the asymptotic analysis. We shed some light on related open problems for further research.

J'examine la mise en oeuvre et l'application de l'estimation par rétrécissement et par essais préliminaires pour les moyennes de populations arbitraires fondées sur des échantillons d'ensembles ordonnés multiples. Je discute de la façon optimale de combiner des informations sur les échantillons en provenance de plusieurs sources de données. J'évalue les propriétés asymptotiques des estimateurs proposés et je les compare aux estimateurs de référence. Nous démontrons que l'estimateur par rétrécissement présente une réduction du risque quadratique considérable par rapport à l'estimateur de référence fondé sur les échantillons d'ensembles ordonnés. Nous effectuons une étude en simulation dont les résultats confirment largement l'analyse asymptotique. Nous éclaircissons certains problèmes apparentés qui pourraient être étudiés par cette approche.

---

**ABDULKADIR HUSSEIN** (University of Windsor), **HASSEN MUTTLAK & ESSAM AL-SAWI** (King Fahd University of Petroleum and Minerals, Saudi Arabia)

[Wednesday June 15 / mercredi 15 juin, 13:30]

*Group Sequential Methods Based on Ranked Set Samples*

*Méthodes séquentielles groupées fondées sur des échantillons d'ensembles ordonnés*

Ranked set sampling schemes were originally proposed to increase efficiency in estimation. Group sequential methods provide substantial savings in sample and enable us to make decisions as early as possible. In this manuscript, we intend to combine the benefits of the two methodologies. We propose group sequential tests for one and two population means under ranked set sampling. We compare the power, average sample sizes and type I errors of the proposed tests to those of group sequential tests based on simple random sampling schemes. We illustrate the utility of the method by

Les schémas d'échantillonnage d'ensembles ordonnés ont été proposés au départ pour améliorer l'efficacité de l'estimation. Les méthodes séquentielles groupées offrent d'importantes économies en ce qui concerne les échantillons et nous permettent de prendre des décisions aussitôt que possible. Dans cet article, nous cherchons à combiner les avantages des deux méthodologies. Nous proposons des tests séquentiels groupés pour une et deux moyennes de populations dans un cadre d'échantillonnage d'ensembles ordonnés. Nous comparons la puissance, les tailles moyennes d'échantillons et les erreurs de type I des tests proposés à ceux de tests séquentiels groupés fondés sur des sché-

## **3C-I2: Rank Set Sampling: New Developments and Applications**

### **Échantillonnage d'ensembles ordonnés : récents développements et applications**

---

using data from HIV clinical trial.

mas de sondage aléatoires simples. Nous illustrons l'utilité de cette méthode par des données tirées d'un essai clinique sur le VIH.

### 3C-13: Recent Advances in Multivariate Control Charts Limites de contrôle multivariées

---

Chair/Président: David Matthews (University of Waterloo)  
Organizer/Responsable: Shoja'eddin Chenouri (University of Waterloo)

Session sponsored by the Business and Industrial Statistics Section  
Session parrainée par le Groupe de statistique industrielle et de gestion

**Room/Salle: BAC141**

**Abstracts/Résumés**

---

**SHOJA'EDDIN CHENOURI** (University of Waterloo), **BITA ROUSHANAEI**

[Wednesday June 15 / mercredi 15 juin, 13:00]

*Depth Based Multivariate Nonparametric Control Charts*

*Cartes de contrôle non paramétriques multivariées basées sur la profondeur*

In this talk, we discuss several control charts for monitoring multivariate processes. These control charts are based on the concept of data depth. We begin with reviewing control charts introduced in Liu (1995). We further construct several new depth based multivariate control charts. Unlike the commonly used multivariate control charts, the charts discussed in this talk are all distribution free, and do not need the assumption of the multivariate normality.

Dans cet exposé, nous nous intéressons à plusieurs cartes de contrôle pour la surveillance de processus multivariés. Ces cartes de contrôle sont basées sur le concept de profondeur des données. Nous commençons par faire la revue des cartes de contrôle introduites par Liu (1995). Nous nous intéressons d'avantage à la construction de nouvelles cartes de contrôle multivariées basées sur la profondeur. Contrairement aux cartes de contrôle multivariées fréquemment utilisées, celles présentées dans cet exposé sont sans hypothèse distributionnelle, et ne nécessitent pas de supposer la normalité multivariée.

---

**ASOKAN VARIYATH** (Memorial University of Newfoundland), **SHOJAEDDIN CHENOURI** (University of Waterloo)

[Wednesday June 15 / mercredi 15 juin, 13:30]

*Robust Multivariate Control Charts for Monitoring Process Mean and Covariance Matrix*

*Cartes de contrôle multivariées pour surveiller la moyenne d'un processus et la matrice de covariance*

To monitor a multivariate process, a classical Hotelling's T<sub>2</sub> chart is often used. Since sample mean and sample covariance matrix are sensitive to outliers, classical T<sub>2</sub> charts are not effective when Phase I data contains outliers. We propose robust Hotelling's T<sub>2</sub> type control charts for monitoring multivariate individual observations based on the highly robust and efficient estimates of mean and covariance matrix. Our Monte Carlo simulation studies are in favor of re-weighted MCD based T<sub>2</sub> charts irrespective of sample size, dimensionality and outlier patterns of Phase I data. We also propose multivariate robust control charts

Une carte de contrôle T<sub>2</sub> de Hotelling est souvent employée pour surveiller un processus multivarié. Puisque la moyenne et la matrice de covariance expérimentales sont sensibles aux valeurs aberrantes, les cartes de contrôle T<sub>2</sub> ne sont pas efficaces lorsque les données de phase I contiennent des valeurs aberrantes. Nous proposons des cartes de contrôle robustes T<sub>2</sub> de Hotelling pour surveiller les observations individuelles multivariées basées sur les estimations robustes et efficaces de la moyenne et de la matrice de covariance. Nos études par simulation de Monte Carlo sont favorables à une repondération CCM basée sur les cartes T<sub>2</sub> peu importe la taille d'échantillon, la dimension et les patrons de données aberrantes de phase I. Nous

### 3C-I3: Recent Advances in Multivariate Control Charts Limites de contrôle multivariées

---

for monitoring the covariance matrix.

proposons également des cartes de contrôle multivariées robustes pour surveiller la matrice de covariance.

---

**ROMAN VIVEROS-AGUILERA** (McMaster University), **JOCK MACKAY & STEFAN STEINER** (University of Waterloo)  
[Wednesday June 15 / mercredi 15 juin, 14:00]

*Monitoring Profiles*

*Profils de surveillance*

Dimensions, surface texture and geometric specifications are all important features of products. Thus keeping an eye on these characteristics as items are manufactured can be a useful effort to identify and correct deviations from course in the manufacturing process. Profile data typically consist of the coordinates of hundreds of points along the edge of the product. The main feature is high short-term correlation and circularity. Smoothing followed by analysis of scores from the leading principal components of smoothed values lead to our proposed multivariate Shewhart and multivariate exponentially weighted moving average charts. Numerical evidence, comparisons and illustrations are presented

Les dimensions, la texture de la surface et les spécifications géométriques sont toutes des caractéristiques importantes des produits. Ainsi, la surveillance de ces caractéristiques en cours de fabrication peut s'avérer utile pour identifier et corriger des déviations par rapport aux cibles durant le processus manufacturier. Les données de profils sont habituellement composées de coordonnées de centaines de points sur la bordure du produit. La caractéristique principale se traduit par de fortes corrélations et de la circularité à court terme. Notre suggestion de cartes de Shewart multivariées et de cartes de contrôle multivariées avec poids exponentiels proviennent d'une analyse des scores des composantes principales prépondérantes des valeurs lissées. Des simulations, des comparaisons et des illustrations sont présentées.

## Author List • Liste des auteurs

- Abdolell, Mohamed, 38, 141 (2C-I4)  
 Abu-Shanab, Reman, 34, 120 (2B-I3)  
 Acar, Elif Fidan, 39, 149 (2C-C2)  
 Ahmed, S. Ejaz, 50, 204 (3C-I2)  
 Al-Sawi, Essam, 50, 204 (3C-I2)  
 Amrhein, John, 50, 203 (3C-I1)  
 Andrews, Jeffrey L., 44, 172 (2F-C1)  
 Angers, Jean-Francois, 25, 77 (1C-I3)  
 Antonie, Luiza, 28, 93 (1D-P)  
 Argyle, Angus, 36, 132 (2B-C3)  
 Arjas, Elja, 27, 90 (1C-C3)  
 Arnold, Barry, 47, 186 (3B-I2)  
 Asgharian, Masoud, 28, 94 (1D-P)  
 Atherton, Juli, 21, 58 (1B-I3)  
 Aue, Alexander, 43, 166 (2F-I1)  
 Aukema, Brian, 48, 191 (3B-I5)  
 Azad, Qazi, 41, 157 (2D-P)
- Balakrishnan, N., 47, 185 (3B-I2)  
 Banerjee, Sudipto, 19  
 Bankole, Julius, 27 (1D-D)  
 Banks, David, 29, 100 (1F-I3)  
 Bashir, Shaheena, 35 (2B-C2), 41 (2D-P), 129 (2B-C2),  
 157 (2D-P)  
 Baskurt, Zeynep, 22 (1B-C1), 29 (1F-I1), 64 (1B-C1), 97  
 (1F-I1)  
 Bataineh, Osama, 22, 64 (1B-C1)  
 Benghabrit, Youssef, 36, 134 (2B-C3)  
 Benjamini, Itai, 47, 189 (3B-I4)  
 Bernard, Carole, 26, 87 (1C-C2)  
 Beyene, Joseph, 30, 106 (1F-C1)  
 Bickis, Mikelis, 22, 64 (1B-C1)  
 Bielawski, Joseph, 35, 127 (2B-C1)  
 Bifolchi, Nadia, 41, 157 (2D-P)  
 Binder, David, 34, 122 (2B-I4)  
 Bingham, Derek, 38, 143 (2C-I5)  
 Blitzstein, Joe, 29, 96 (1F-I1)  
 Blume, Jeffrey, 24, 73 (1C-I1)
- Boisbunon, Aurélie, 34, 120 (2B-I3)  
 Bonner, Simon J, 25, 82 (1C-I5)  
 Bouzaachane, Khadija, 36, 134 (2B-C3)  
 Brajendra, Sutradhar, 49, 199 (3B-C3)  
 Braun, Willard John, 25 (1C-I3), 28 (1D-P), 77 (1C-I3),  
 92, 93 (1D-P)  
 Brewster, John, 39, 151 (2C-C3)  
 Brillinger, David, 31, 110 (1F-C2)  
 Briollais, Laurent, 39, 149 (2C-C2)  
 Brown, Patrick, 23 (1B-C2), 40 (2C-C4), 67 (1B-C2), 155  
 (2C-C4)  
 Browne, Ryan P., 44, 173 (2F-C1)  
 Bull, Shelley, 39, 148 (2C-C2)  
 Burr, Wesley, 49, 196, 198 (3B-C2)  
 Buss, Jason, 27 (1D-D)
- Cadigan, Noel, 25 (1C-I5), 36 (2B-C3), 81 (1C-I5), 133  
 (2B-C3)  
 Caffery, Barbara, 44, 173 (2F-C1)  
 Caffo, Brian, 33, 116 (2B-I1)  
 Cai, Jing, 48, 192 (3B-I5)  
 Camacho, Fernando, 34, 124 (2B-I5)  
 Canu, Stéphane, 34, 120 (2B-I3)  
 Cao, Jiguo, 48, 192 (3B-I5)  
 Carlow, Chevonne, 42, 160 (2D-P)  
 Carroll, Allan, 48, 191 (3B-I5)  
 Carter, E. M., 35, 129 (2B-C2)  
 Celiny, Léonide, 36, 133 (2B-C3)  
 Chakraborty, Bibhas, 48, 194 (3B-C1)  
 Charest, Anne-Sophie, 31, 112 (1F-C3)  
 Charlebois, Joanne, 31, 113 (1F-C3)  
 Chauvet, Guillaume, 23, 71 (1B-C3)  
 Cheam, Amay S.M., 27, 89 (1C-C3)  
 Chen, Dongmei, 42, 161 (2D-P)  
 Chen, Jiahua, 37, 137 (2C-I2)  
 Chen, Jigyuan, 27 (1D-D)  
 Chen, Min, 41 (2D-D)  
 Chen, Xiaoyu, 26, 83 (1C-C1)



- Chenouri, Shoja'eddin, 51, 206 (3C-I3)  
 Childs, Aaron, 30, 107 (1F-C1)  
 Chipman, Hugh, 20, 20 (1A-A1), 54 (1B-I1)  
 Choi, Bernard, 24, 75 (1C-I2)  
 Choulakian, Vartan, 26, 85 (1C-C1)  
 Ciampi, Antonio, 39 (2C-C2), 44 (2F-C1), 49 (3B-C3),  
 148, 150 (2C-C2), 173 (2F-C1), 200 (3B-C3)  
 Clark, Laurie, 47, 187 (3B-I3)  
 Clifford, Hurvich, 43, 166 (2F-I1)  
 Cole, Martin, 49, 200 (3B-C3)  
 Coleman, Kimberley, 23, 70 (1B-C3)  
 Collins, Kayla, 21, 57 (1B-I2)  
 Contreras, Alberto, 22, 64 (1B-C1)  
 Cook, Richard, 27 (1C-C3), 30 (1F-I4), 91 (1C-C3), 102  
 (1F-I4)  
 Corberan, Ana, 29, 100 (1F-I3)  
 Cossette, Hélène, 35, 130 (2B-C2)  
 Cotton, Cecilia, 21 (1B-I3), 27 (1C-C3), 58 (1B-I3), 91  
 (1C-C3)  
 Coulthart, Michael, 24, 75 (1C-I2)  
 Cowling, Ben, 42, 159 (2D-P)  
 Cuerden, Meaghan, 27, 91 (1C-C3)  
 Cui, Zhenyu, 26, 87 (1C-C2)  
 Cupples, L. Adrienne, 43, 168 (2F-I2)
- Dansu, Bolanle M., 49, 199 (3B-C3)  
 Darlington, Gerarda, 23, 69 (1B-C2)  
 da Silveira, Giovanni J.C., 40, 152 (2C-C3)  
 Datta, Gauri, 29, 100 (1F-I3)  
 Davies, Katherine, 47, 185 (3B-I2)  
 Davies, Neville, 44, 170 (2F-I3)  
 Davis, Michael, 21, 60 (1B-I4)  
 Dean, Charmaine, 25, 79 (1C-I4)  
 Deardon, Rob, 23 (1B-C2), 41, 42 (2D-P), 68, 69 (1B-  
 C2), 157–159, 161 (2D-P)  
 Deeth, Lorna, 23, 69 (1B-C2)  
 Dehghan, Mohammad, 40, 155 (2C-C4)  
 Demers, Geneviève, 45, 179 (2F-C3)  
 Demnati, Abdellatif, 23 (1B-C3), 32 (1F-C3), 45 (2F-C3),  
 71 (1B-C3), 114 (1F-C3), 179 (2F-C3)  
 Desmond, Tony, 23, 67 (1B-C2)  
 Diao, Liqun, 30, 102 (1F-I4)  
 Dicintio, Sabrina, 41, 157 (2D-P)  
 Dobbs, Angie Christina, 41, 158 (2D-P)  
 Donelle, Jessy, 44, 173 (2F-C1)  
 Dorea, Fernanda, 24, 76 (1C-I2)  
 Doucette, Sarah, 27, 90 (1C-C3)  
 Dowd, Mike, 36, 132 (2B-C3)
- Dowden, Jeffrey, 21, 57 (1B-I2)  
 Dubin, Joel, 38, 141 (2C-I4)  
 Dubreuil, Guylaine, 31, 113 (1F-C3)  
 Duchesne, Simon, 45, 176 (2F-C2)  
 Duchesne, Thierry, 40, 155 (2C-C4)  
 Duffy, Anne, 27, 90 (1C-C3)  
 Dupuis, Josée, 43, 168 (2F-I2)  
 Dyachenko, Alina, 49, 200 (3B-C3)
- Elghafghuf, Adel, 40, 154 (2C-C4)  
 El-Shaarawi, Abdel, 28 (1E-A), 29 (1F-I2), 95 (1E-A), 98  
 (1F-I2)  
 Esterby, Sylvia, 29, 98 (1F-I2)  
 Evans, Michael, 22, 64 (1B-C1)  
 Evans, Steve, 47, 189 (3B-I4)
- Fajardo, Val Andrei, 39, 146 (2C-C1)  
 Fan, Lingling, 27 (1D-D), 36, 130 (2B-C2)  
 Fang, Lin, 39, 151 (2C-C3)  
 Fellegi, Ivan, 20, 53 (1A-A2)  
 Fendyur, Andrey, 40, 152 (2C-C3)  
 Feng, Shui, 22, 62 (1B-I5)  
 Feng, Xin, 25, 79 (1C-I4)  
 Feng, Zeny, 23 (1B-C2), 41 (2D-P), 68 (1B-C2), 157 (2D-  
 P)  
 Flemming, Joanna Mills, 46, 183 (3B-I1)  
 Fourdrinier, Dominique, 34, 120 (2B-I3)  
 Franczak, Brian, 27 (1D-D), 41 (2D-P), 44 (2F-C1), 157  
 (2D-P), 173 (2F-C1)  
 Fridley, Brooke, 43, 168 (2F-I2)  
 Fu, Eric, 41, 158 (2D-P)
- Gardner, Abbie, 23, 69 (1B-C2)  
 Gel, Yulia, 31, 109 (1F-C2)  
 Gibbs, Alison, 29, 97 (1F-I1)  
 Gill, Paramjit, 21, 60 (1B-I4)  
 Gillis, Daniel, 23, 67 (1B-C2)  
 Girard, Claude, 37, 139 (2C-I3)  
 Givari, Dena, 35, 127 (2B-C1)  
 Gobbert, Matthias, 35, 129 (2B-C2)  
 Goh, Joslin, 38, 143 (2C-I5)  
 Gold, Jourdan, 23, 68 (1B-C2)  
 Goldszmidt, Moises, 33, 117 (2B-I1)  
 Gornall, William, 26, 87 (1C-C2)  
 Gosselin, Pierre, 29, 98 (1F-I2)  
 Gouzi, Naima, 23, 71 (1B-C3)  
 Gramacy, Robert B, 38, 143 (2C-I5)  
 Greenwood, Celia, 41, 157 (2D-P)  
 Grof, Paul, 27, 90 (1C-C3)

- Gu, Hong, 35, 126, 127 (2B-C1)  
 Guimond, Tim, 28 (1D-D)  
 Guo, Qing, 48, 193 (3B-C1)  
 Guo, Wenge, 33, 118 (2B-I2)  
 Gurel-Gurevich, Ori, 47, 189 (3B-I4)
- Haley, Charlotte Lisa, 49, 197 (3B-C2)  
 Hall, Geoff, 48, 193 (3B-C1)  
 Hamel, Martin, 37, 139 (2C-I3)  
 Hamel, Nathalie, 37, 139 (2C-I3)  
 Han, Lengyi Spectrum, 28, 92 (1D-P)  
 Hanley, James, 27 (1C-C3), 40 (2C-C4), 89 (1C-C3), 156 (2C-C4)  
 Hasan, M. Tariqul, 25 (1C-I4), 30 (1F-I4), 79 (1C-I4), 103 (1F-I4)  
 Haziza, David, 23 (1B-C3), 31 (1F-C3), 71 (1B-C3), 113 (1F-C3)  
 He, Yulei, 20, 56 (1B-I2)  
 Heaton, Matt, 29, 100 (1F-I3)  
 Heron, Melonie, 29, 100 (1F-I3)  
 Holloway, James, 38, 143 (2C-I5)  
 Holt, John, 23, 67 (1B-C2)  
 Horrocks, Julie, 27, 90 (1C-C3)  
 Horvath, Lajos, 43, 166 (2F-I1)  
 Hsu, Jason, 33, 118 (2B-I2)  
 Hu, Xiaogiong Joan, 30, 102 (1F-I4)  
 Hu, Zechun, 22, 62 (1B-I5)  
 Huang, Mei Ling, 39, 146 (2C-C1)  
 Huang, Yuhui, 27 (1D-D)  
 Hughes, Ed, 25, 79 (1C-I4)  
 Hui, Rihong, 41, 159 (2D-P)  
 Hurley, Bill, 21, 60 (1B-I4)  
 Hurvich, Clifford, 43, 166 (2F-I1)  
 Hussein, Abdulkadir, 50, 204 (3C-I2)  
 Hyndman, Cody, 22, 62 (1B-I5)
- Ivany, Ray, 20 (1A-A1)
- Jafari Jozani, Mohammad, 23 (1B-C3), 25 (1C-I3), 71 (1B-C3), 77 (1C-I3)  
 Jamieson, David W., 47, 188 (3B-I3)  
 Jang, Gun Ho, 46, 181 (3A-A)  
 Jankowski, Hanna, 38 (2C-C1), 41, 42 (2D-P), 145 (2C-C1), 159 (2D-P)  
 Ji, Xiang, 42, 159 (2D-P)  
 Ji, Yunqi, 45, 175 (2F-C2)  
 Jiang, Depeng, 50, 203 (3C-I1)  
 Jiang, Haiyan, 41 (2D-D)  
 Jiang, Hedy, 40, 155 (2C-C4)
- Jiang, Weibin, 31, 110 (1F-C2)  
 Johnson, Brad, 23, 71 (1B-C3)  
 Jones, Galin, 33, 116 (2B-I1)  
 Joseph, Lawrence, 21, 56 (1B-I2)  
 Joy, Ruth, 36, 132 (2B-C3)  
 Jozani, Mohammad Jafari, 46, 185 (3B-I2)
- Karr, Alan, 29, 100 (1F-I3)  
 Keating, Jerome P., 47, 185 (3B-I2)  
 Kennedy, Christine, 24, 75 (1C-I2)  
 Kenney, Toby, 35, 126, 127 (2B-C1)  
 Kent, John, 34, 120 (2B-I3)  
 Khalili, Abbas, 28 (1D-P), 35 (2B-C2), 94 (1D-P), 129 (2B-C2)  
 Khare, Meena, 29, 100 (1F-I3)  
 Kirkland, Tyler, 45, 178 (2F-C3)  
 Klein, Martin, 31, 112 (1F-C3)  
 Kolenikov, Stanislav, 19  
 Kondo, Yumi, 28 (1D-D)  
 Kopciuk, Karen A, 35, 126 (2B-C1)  
 Kulik, Rafal, 43, 166 (2F-I1)  
 Kulkarni Woodstock, Manisha, 24, 75 (1C-I2)  
 Kulperger, Reg, 28 (1D-P), 31 (1F-C2), 92 (1D-P), 110 (1F-C2)  
 Kurta, Jessica, 47, 187 (3B-I3)  
 Kwong, Grace, 42, 159 (2D-P)  
 Kwong, Grace Pui Sze, 23, 68 (1B-C2)
- Labbe, Aurélie, 39, 148, 150 (2C-C2)  
 Labelle-Blanchet, Sébastien, 37 (2C-I3), 45 (2F-C3), 139 (2C-I3), 178 (2F-C3)  
 Ladouceur, Martin, 21, 56 (1B-I2)  
 Lai, Anky, 41 (2D-D)  
 Lakhali Chaieb, Lajmi, 50, 200 (3B-C3)  
 Landrum, Mary Beth, 20, 56 (1B-I2)  
 Laroche, Danielle, 23, 70 (1B-C3)  
 Lavallée, Pierre, 45, 178 (2F-C3)  
 LaValley, Michael, 43, 168 (2F-I2)  
 Lawson, Andrew, 29, 100 (1F-I3)  
 Leblanc, Alexandre, 25, 77 (1C-I3)  
 Ledingham, Jane E., 47, 187 (3B-I3)  
 Lee, Herbie, 38, 143 (2C-I5)  
 Lee, Jonathan, 28, 92 (1D-P)  
 Lee, Kuo-Jung, 33, 116 (2B-I1)  
 Léger, Christian, 31, 113 (1F-C3)  
 León, Carlos, 23, 72 (1B-C3)  
 Li, Haocheng, 30, 106 (1F-C1)  
 Li, Pengfei, 37, 137 (2C-I2)  
 Li, Shudong, 28, 92 (1D-P)

- Li, Xuan, 48, 194 (3B-C1)  
 Li, Ye, 23, 67 (1B-C2)  
 Liang, Richard, 47, 189 (3B-I4)  
 Lin, Dennis KJ, 42, 164 (2C-A)  
 Liu, Lucy Yao, 41 (2D-D)  
 Liu, Meng, 21, 58 (1B-I3)  
 Liu, Wei, 33 (2B-I2), 44 (2F-C2), 118 (2B-I2), 175 (2F-C2)  
 Liu, Yang, 27 (1D-D)  
 Liu, Yaqing, 27 (1D-D)  
 Liu, Yi, 33, 118 (2B-I2)  
 Liu, Zhihui (Amy), 27 (1C-C3), 28 (1D-D), 40 (2C-C4), 89 (1C-C3), 156 (2C-C4)  
 Lix, Lisa M., 49, 199 (3B-C3)  
 Lockhart, Richard, 22, 64 (1B-C1)  
 Loeppky, Jason, 38, 143 (2C-I5)  
 Loredó-Osti, J Concepción, 26, 83 (1C-C1)  
 Lorenzi, Maria, 30, 102 (1F-I4)  
 Lum, Conroy, 48, 191 (3B-I5)  
 Lumley, Thomas, 31, 112 (1F-C3)  
 Luo, Hao, 27 (1D-D)  
 Lynch, James, 29, 100 (1F-I3)  
 Lyons, John S., 47, 187 (3B-I3)  
 Lyons, Russell, 47, 189 (3B-I4)
- Ma, Jinhui, 30, 106 (1F-C1)  
 Ma, Renjun, 25 (1C-I4), 30 (1F-I4), 79 (1C-I4), 103 (1F-I4)  
 Ma, Wenkai, 28, 93 (1D-P)  
 Mabrouk, Iman, 38, 145 (2C-C1)  
 MacGibbon, Brenda, 25, 77 (1C-I3)  
 MacKay, Jock, 51, 207 (3C-I3)  
 Mailhot, Mélina, 35, 130 (2B-C2)  
 Majumdar, Dibyen, 37, 135 (2C-I1)  
 Maldonado, Sandra Gonzalez, 27 (1D-D)  
 Malik, Rajat, 42, 159 (2D-P)  
 Mandal, Abhyuday, 37, 135 (2C-I1)  
 Manning, Alisa K., 43, 168 (2F-I2)  
 Marceau, Étienne, 35, 130 (2B-C2)  
 Marchand, Éric, 25, 77 (1C-I3)  
 Mariathas, Hensley Hubert, 26, 83 (1C-C1)  
 Marshall, John, 37, 139 (2C-I3)  
 Mashreghi, Zeinab, 31, 113 (1F-C3)  
 Mason, Robert L., 47, 185 (3B-I2)  
 Matthews, David E., 40, 155 (2C-C4)  
 Matthews, Steve, 45, 178 (2F-C3)  
 Maziade, Michel, 39, 148 (2C-C2)  
 McCusker, Jane, 49, 200 (3B-C3)
- McKinnon, Margaret, 48, 193 (3B-C1)  
 McLachlan, Geoff, 37, 137 (2C-I2)  
 McLeish, Don, 20 (1A-A1)  
 McNicholas, Paul, 28 (1D-P), 35 (2B-C1), 37 (2C-I2), 41, 42 (2D-P), 44 (2F-C1), 93 (1D-P), 127 (2B-C1), 137 (2C-I2), 157, 160, 161 (2D-P), 172, 173 (2F-C1)  
 Meigs, James B., 43, 168 (2F-I2)  
 Men, Zhongxian, 26, 86 (1C-C2)  
 Meng, Xiao-Li, 29, 96 (1F-I1)  
 Merette, Chantal, 39, 150 (2C-C2)  
 Millar, Anne Michele, 26, 83 (1C-C1)  
 Mirea, Lucia, 39, 148 (2C-C2)  
 Mirkamali, Sayed Jamal, 46, 185 (3B-I2)  
 Moghtaderi, Azadeh, 49, 196, 198 (3B-C2)  
 Mohsenipour, A. Akbar, 31, 109 (1F-C2)  
 Molina, Isabel, 42, 163 (2E-A)  
 Moodie, Erica, 45 (2F-C2), 48 (3B-C1), 176 (2F-C2), 194 (3B-C1)  
 Moore, Leslie, 38, 143 (2C-I5)  
 Morel, Jorge, 35, 129 (2B-C2)  
 Morrison, Laura May, 41 (2D-D)  
 Mosesova, Sofia, 20, 54 (1B-I1)  
 Mouiha, Abderazzak, 45, 176 (2F-C2)  
 Mueller, Hans George, 20, 54 (1B-I1)  
 Mulayath Variyath, Asokan, 22, 65 (1B-C1)  
 Murdoch, Duncan, 19  
 Murray, Joshua, 28 (1D-D)  
 Murua, Alejandro, 35, 127 (2B-C1)  
 Mutch, David M., 42, 161 (2D-P)  
 Muttlak, Hassen, 50, 204 (3C-I2)
- Nadarajah, Tharshanna, 22, 65 (1B-C1)  
 Nambou, Christian Olivier, 23, 71 (1B-C3)  
 Nandy, Karabi, 37, 135 (2C-I1)  
 Narayanan, Purni, 24, 75 (1C-I2)  
 Neerchal, Nagaraj, 35, 129 (2B-C2)  
 Nia, Vahid Partovi, 35, 128 (2B-C1)  
 Nielsen, Jason D, 20, 54 (1B-I1)  
 Niu, Yi, 40, 154 (2C-C4)
- O'Brien, Kristen, 28 (1D-D)  
 Ogunnaike-Cooke, Susanna, 24, 75 (1C-I2)  
 Oldford, R. Wayne, 44, 172 (2F-C1)  
 O'Reilly, Rachel, 42, 160 (2D-P)  
 Oualkacha, Karim, 39, 148 (2C-C2)
- Pandey, Mahesh, 28 (1D-P), 40 (2C-C3), 94 (1D-P), 152 (2C-C3)

- Partovi Nia, Vahid, 35, 128 (2B-C1)  
Pavlov, Andrey, 21, 60 (1B-I4)  
Payne, Jennifer, 38, 141 (2C-I4)  
Peng, Yingwei, 27 (1C-C3), 40 (2C-C4), 91 (1C-C3), 154 (2C-C4)  
Persson, Greg, 27, 90 (1C-C3)  
Petkau, John, 41, 158 (2D-P)  
Pfeffermann, Danny, 34, 123 (2B-I4)  
Pfeiffer, Dirk, 24, 75 (1C-I2)  
Picka, Jeffrey, 36, 131 (2B-C2)  
Pineau, Christian, 21, 56 (1B-I2)  
Platt, Robert, 45, 176 (2F-C2)  
Platt, Robert William, 31, 107 (1F-C1)  
Polansky, Alan, 24, 73 (1C-I1)  
Proudfoot, James, 27 (1D-D)  
Provençal, Jean-Sébastien, 23, 70 (1B-C3)  
Provost, Serge, 31 (1F-C2), 38 (2C-C1), 109 (1F-C2), 145 (2C-C1)  
Puddester, Leah, 47, 187 (3B-I3)  
Pullenayegum, Eleanor, 30 (1F-C1), 45 (2F-C2), 48 (3B-C1), 107 (1F-C1), 176 (2F-C2), 193 (3B-C1)  
Pursoo, Tiffany, 47, 187 (3B-I3)
- Qian, Wei, 32, 114 (1F-C3)  
Qin, Lang Sirius, 28 (1D-D)  
Qiu, Weiyu, 38, 145 (2C-C1)  
Qiu, Xin, 27 (1D-D)  
Quan, Joanne, 28 (1D-D)
- Rahim, Karim J., 49, 197 (3B-C2)  
Rahme, Elham, 21, 56 (1B-I2)  
Raim, Andrew, 35, 129 (2B-C2)  
Raina, Parminder, 30, 106 (1F-C1)  
Ramírez, Lilia Leticia Ramírez, 31, 109 (1F-C2)  
Ramsay, Tim, 50, 203 (3C-I1)  
Ramunno, Daniel, 28 (1D-D)  
Rana, Masud, 27 (1D-D)  
Rao, J. N. K., 34 (2B-I4), 42 (2E-A), 122 (2B-I4), 163 (2E-A)  
Reid, Nancy, 22, 65 (1B-C1)  
Revie, Crawford, 24, 76 (1C-I2)  
Rich, Benjamin, 40 (2C-C4), 44 (2F-C1), 156 (2C-C4), 173 (2F-C1)  
Rivest, Louis-Paul, 33 (2A-A), 50 (3B-C3), 115 (2A-A), 200 (3B-C3)  
Rodriguez, Robert, 44, 170 (2F-I3)  
Romdhani, Héra, 50, 200 (3B-C3)  
Ross, William, 29, 99 (1F-I2)  
Rosychuk, Rhonda, 35, 125 (2B-I5)
- Rothman, Lorne, 34, 124 (2B-I5)  
Roushanaei, Bitá, 51, 206 (3C-I3)  
Routledge, Rick, 46, 183 (3B-I1)  
Roy, Marc-André, 39, 148 (2C-C2)  
Rukhin, Andrew L., 34, 120 (2B-I3)  
Run, Havard, 23, 67 (1B-C2)
- Saarela, Olli, 27, 90 (1C-C3)  
Sajobi, Tolulope T., 49, 199 (3B-C3)  
Sambaraju, Kishan, 48, 191 (3B-I5)  
Sanchez, Javier, 24, 76 (1C-I2)  
Sarkar, Sanat, 33, 118 (2B-I2)  
Schnitzer, Mireille, 28 (1D-D), 45, 176 (2F-C2)  
Schramm, Oded, 47, 189 (3B-I4)  
Schwarz, Carl J, 25, 82 (1C-I5)  
Scott, Alastair, 31, 112 (1F-C3)  
Scott, Alexandre, 26, 87 (1C-C2)  
Shaikh, Mateen, 28, 93 (1D-P)  
Sheriff, John, 31, 110 (1F-C2)  
Shin, Hwashin H., 49, 198 (3B-C2)  
Shohoudi, Azadeh (Fatemeh), 28, 94 (1D-P)  
Siddo, Radcliffe A., 28 (1D-D)  
Sikdar, Khokan, 21, 57 (1B-I2)  
Sinha, Sanjoy, 30, 104 (1F-I5)  
Slud, Eric, 29, 100 (1F-I3)  
Smith, Stephen, 25, 81 (1C-I5)  
Sneddon, Gary, 25, 79 (1C-I4)  
Song, Hui, 27, 91 (1C-C3)  
Song, Melissa, 28 (1D-D)  
Sormany, Marc, 26, 85 (1C-C1)  
Soulie, Philippe, 43, 166 (2F-I1)  
Spear Bassett, Susan, 33, 116 (2B-I1)  
Squires, Josh, 21, 57 (1B-I2)  
Stares, Joanne, 21, 57 (1B-I2)  
Steiner, Stefan, 51, 207 (3C-I3)  
Stephens, David, 33 (2B-I1), 35 (2B-C2), 116 (2B-I1), 129 (2B-C2)  
Stephens, Michael, 22, 64 (1B-C1)  
Stewart, David, 47, 188 (3B-I3)  
Stewart, Delbert E., 47, 188 (3B-I3)  
Strawderman, William, 34, 120 (2B-I3)  
Stryhn, Henrik, 40, 154 (2C-C4)  
Stufken, John, 37, 135 (2C-I1)  
Subedi, Sanjeena, 27 (1D-D), 35 (2B-C1), 37 (2C-I2), 127 (2B-C1), 137 (2C-I2)  
Sun, Bingrui, 26, 84 (1C-C1)  
Sun, Lei, 39, 148, 149 (2C-C2)  
Sun, Wei, 22, 62 (1B-I5)

- Sun, Zheng, 39, 146 (2C-C1)  
 Suntornchost, Jiraphan, 29, 100 (1F-I3)  
 Susko, Edward, 50, 202 (3C-A)  
 Sutradhar, Brajendra, 26 (1C-C1), 26 (1C-C2), 83, 84 (1C-C1), 86 (1C-C2)  
 Swartz, Tim, 21, 60 (1B-I4)
- Tagore, Vickneswary, 26, 86 (1C-C2)  
 Takahara, Glen, 49, 196, 198 (3B-C2)  
 Tang, Mei, 40, 152 (2C-C3)  
 Tang, Szu-Yu, 33, 118 (2B-I2)  
 Taylor, Carolyn, 48, 191 (3B-I5)  
 Tekougang, Thierry Chekouo, 35, 127 (2B-C1)  
 Thabane, Lehana, 30 (1F-C1), 48 (3B-C1), 50 (3C-I1), 106 (1F-C1), 193 (3B-C1), 203 (3C-I1)  
 Thaddeus, Edens, 21, 58 (1B-I3)  
 Thinniyam, Ramya, 29, 97 (1F-I1)  
 Thomas, Duncan, 46, 182 (3B-A)  
 Thomas, Steven, 45, 180 (2F-C3)  
 Thompson, Mary E., 31, 109 (1F-C2)  
 Thomson, David J., 49, 196, 197 (3B-C2)  
 Torabi, Mahmoud, 26, 84 (1C-C1)  
 Townsend, Mary, 44, 170 (2F-I3)  
 Trejo, Esther Perez, 28 (1D-D), 31, 107 (1F-C1)  
 Troupé, Marylène, 36, 133 (2B-C3)  
 Trubnikov, Maxim, 24, 75 (1C-I2)  
 Tu, Dongsheng, 27, 91 (1C-C3)  
 Tufts, Jocelyn, 41 (2D-D)
- Vaillant, Jean, 36, 133 (2B-C3)  
 Valmy, Larissa, 36, 133 (2B-C3)  
 Variyath, Asokan, 51, 206 (3C-I3)  
 Vera, Francisco, 29, 100 (1F-I3)  
 Villandré, Luc, 44, 173 (2F-C1)  
 Vincent, Kyle, 45, 179 (2F-C3)  
 Viveros-Aguilera, Roman, 39 (2C-C3), 51 (3C-I3), 151 (2C-C3), 207 (3C-I3)  
 Vrbik, Irene, 42, 160 (2D-P)
- Waddell, Adrian, 44, 172 (2F-C1)  
 Waite, Tim, 37, 136 (2C-I1)  
 Walter, Stephen D, 48, 193 (3B-C1)  
 Wang, Qiong, 23, 70 (1B-C3)  
 Wang, Xikui, 48, 194 (3B-C1)  
 Wang, Xu, 44, 173 (2F-C1)  
 Wang, Yuan, 35, 126 (2B-C1)  
 Wang, Zilin, 28, 92 (1D-P)  
 Warriyar, Vineetha K.V., 49, 199 (3B-C3)  
 Waterhouse, Tim, 37, 136 (2C-I1)
- Watier, Francois, 26, 87 (1C-C2)  
 Weerasinghe, Swarna, 50, 201 (3B-C3)  
 Wei, Rong, 29, 100 (1F-I3)  
 Weir, Alison, 38, 142 (2C-I4)  
 Wellner, Jon, 38, 145 (2C-C1)  
 Wells, Marty, 34, 120 (2B-I3)  
 Welsh, Alan, 30 (1F-I5), 46 (3B-I1), 104 (1F-I5), 183 (3B-I1)  
 White, Bethany, 29, 96 (1F-I1)  
 Wiens, Doug, 30, 104 (1F-I5)  
 Wilder, Kate, 45, 180 (2F-C3)  
 Willan, Andrew, 43, 165 (2F-A)  
 Williams, Brian, 38, 143 (2C-I5)  
 Wilmut, Michael, 30, 104 (1F-I5)  
 Wilson, Shanna, 42, 160 (2D-P)  
 Wolfson, David, 21 (1B-I3), 28 (1D-P), 58 (1B-I3), 94 (1D-P)  
 Wong, Hoi Suen, 30, 107 (1F-C1)  
 Wong, Monica, 42, 161 (2D-P)  
 Woodard, Dawn, 33, 117 (2B-I1)  
 Woods, Dave, 37, 136 (2C-I1)  
 Wright, Tommy, 31, 112 (1F-C3)  
 Wu, Changbao, 34, 122 (2B-I4)  
 Wu, Lang, 44, 175 (2F-C2)  
 Wu, Xiaojing, 27 (1D-D)  
 Wu, Ying, 27 (1D-D)  
 Wulff, Shaun S., 39, 151 (2C-C3)
- Xin, Lu, 28 (1D-P), 40 (2C-C3), 94 (1D-P), 152 (2C-C3)  
 Xu, Xiaojian, 26, 83 (1C-C1)  
 Xu, Ximing, 22, 65 (1B-C1)  
 Xu, Yuan, 42, 161 (2D-P)  
 Xun, Dong, 27 (1D-D)
- Yan, Guohua, 30, 103 (1F-I4)  
 Yang, Jie, 37, 135 (2C-I1)  
 Yang, Lin, 39, 150 (2C-C2)  
 Yang, Xiao Di, 28 (1D-D)  
 Yao, Fang, 20, 54 (1B-I1)  
 Yao, Xue, 27 (1D-D)  
 Yasui, Yutaka, 38, 145 (2C-C1)  
 Yi, Grace, 30, 106 (1F-C1)  
 Yi, Grace Y., 44, 175 (2F-C2)  
 Yi, Yanqing, 48, 194 (3B-C1)  
 Yu, Hao, 19, 28 (1D-P), 31 (1F-C2), 92 (1D-P), 110 (1F-C2)  
 Yu, Lisha, 41 (2D-D)  
 Yuan, Ao, 24, 73 (1C-I1)

Zaslavsky, Alan, 20, 56 (1B-I2)  
Zerpa, Carlos, 28 (1D-D)  
Zhang, Lin, 42, 161 (2D-P)  
Zhao, Yang, 21, 58 (1B-I3)  
Zhaozhi, Fan, 45, 175 (2F-C2)  
Zheng, Yanbing, 48, 191 (3B-I5)  
Zhou, Julie, 30, 104 (1F-I5)  
Zhou, Lutong, 41 (2D-D)  
Zhou, Xiaowen, 47, 189 (3B-I4)  
Zhou, Zhou, 43, 166 (2F-I1)  
Zhu, Jingqin, 41 (2D-D)  
Zhu, Jun, 48, 191 (3B-I5)  
Zidek, James, 48, 191, 192 (3B-I5)  
Zou, Fei, 43, 168 (2F-I2)  
Zou, Frank, 29, 100 (1F-I3)