

Société Statistique  
statistique Society  
du Canada of Canada

44<sup>th</sup> Annual Meeting  
of the  
Statistical Society of Canada

44<sup>e</sup> Congrès annuel  
de la  
Société statistique du Canada

May 29 - June 1, 2016  
29 mai au 1 juin 2016

Brock University

# Table of Contents • Table des matières

<b>Table of Contents • Table des matières</b>	<b>1</b>
<b>Welcome to • Bienvenue à St. Catharines</b>	<b>2</b>
<b>Sponsors • Commanditaires</b>	<b>3</b>
<b>Exhibitors • Exposants</b>	<b>3</b>
<b>Organizers • Organismes</b>	<b>4</b>
<b>General Information • Informations générales</b>	<b>5</b>
<b>The Conference • Le congrès</b>	<b>7</b>
<b>Social Events • Activités sociales</b>	<b>9</b>
<b>Committees and Meetings • Comités et réunions</b>	<b>11</b>
<b>Social and Information Events • Événements sociaux et informatifs</b>	<b>14</b>
<b>List of Sessions • Liste des sessions</b>	<b>16</b>
<b>Program • Programme</b>	<b>19</b>
<b>Abstracts • Résumés</b>	<b>67</b>
<b>Author List • Liste des auteurs</b>	<b>269</b>

## Welcome to • Bienvenue à St. Catharines

It is with great pleasure that the members of the Department of Mathematics and Statistics at Brock University welcome you to the 44<sup>th</sup> Annual Meeting of the Statistical Society of Canada. Founded in 1964, Brock University is located atop the Niagara Escarpment in Ontario's Niagara peninsula. With 18,000 students and nearly 600 faculty, Brock is buzzing with activity and welcoming visitors to the campus every day.

St. Catharines is a vibrant city of 132,000 people located in one of the most historic parts of Canada — the Niagara region. It is also known as the "The Garden City," which is renowned for its beautiful parks, gardens and trails. The city is home to the picturesque harbour front village of Port Dalhousie. In addition, the Welland Canal runs through the east side of St. Catharines and the locks are worth a visit. The city has a thriving arts and culture scene, and downtown St. Catharines is a good place to start. In particular, St. Paul Street and the surrounding area is an entertainment district with many small restaurants and cafés, often with art shows and live music. There are city bus routes that connect downtown and Brock/conference hotels directly. St. Catharines is also close to the tourist centres of Niagara Falls and Niagara on the Lake. There are many wonderful wineries and an enviable system of recreational trails and parks for everyone to explore. We encourage you to leave yourself some time to enjoy the features our city and region has to offer. We sincerely hope that you have a productive and fun time throughout the conference.

For more information to help you plan your trip, please visit <http://www.yourbus.com/>, <http://www.tourismstcatharines.ca> and <http://www.niagarafallstourism.com/>.

C'est avec un très grand plaisir que les membres du Département de mathématiques et de statistique de la Brock University vous accueillent au 44<sup>e</sup> congrès annuel de la Société statistique du Canada. Fondé en 1964, Brock est situé en haut de l'Escarpement du Niagara, dans la Péninsule du Niagara, en Ontario. Avec 18 000 étudiants et près de 600 enseignants, Brock grouille d'activité et accueille tous les jours de nouveaux visiteurs sur son campus.

St. Catharines est une ville dynamique de 132 000 habitants, située dans l'une des régions les plus historiques du Canada, la région de Niagara. Également connue sous le nom de « Ville-jardin », elle est réputée pour ses parcs, ses jardins et ses sentiers de promenade. La ville abrite aussi le pittoresque village de pêche de Port Dalhousie. Par ailleurs, le canal Welland traverse l'est de la ville : ses écluses méritent le détour. Le centre-ville de St. Catharines compte aussi une collectivité des arts et de la culture florissante. La rue St. Paul, notamment, et le quartier avoisinant abonde de petits restaurants et de cafés qui proposent régulièrement des manifestations artistiques et des concerts. Plusieurs lignes de bus relient directement le centre-ville à Brock et aux hôtels du congrès. St. Catharines est aussi tout proche des centres touristiques de Niagara Falls et de Niagara on the Lake. La région abonde de vignobles, de sentiers de promenade et de parcs à explorer. Nous vous encourageons vivement à prendre le temps de profiter des trésors que notre ville et notre région ont à offrir. Nous espérons sincèrement que vous passerez un séjour productif et agréable tout au long du congrès.

Pour vous aider à planifier votre séjour, veuillez consulter : <http://www.yourbus.com/>, <http://www.tourismstcatharines.ca> et <http://www.niagarafallstourism.com/>.

## Sponsors • Commanditaires

The Statistical Society of Canada would like to thank each of the sponsors, whose generous contributions have made this conference possible:

La Société statistique du Canada désire remercier chacun de ses commanditaires dont la généreuse contribution a rendu possible la tenue de ce congrès :

- President's Office, Brock University
- Faculty of Mathematics and Science, Brock University
- Department of Mathematics and Statistics, Brock University
- The Fields Institute
- Pacific Institute for the Mathematical Sciences
- Centre de Recherches Mathématiques
- Canadian Statistical Sciences Institute /  
Institut canadien des sciences statistiques

## Exhibitors • Exposants

The exhibitors, providing displays for examination and purchase, are located in the corridors of the Thistle Complex leading to rooms for the scientific sessions. The exhibitors will be available Monday, May 30 and Tuesday, May 31 from 9:00 am to 4:00pm.

Les kiosques des exposants, où vous pourrez bouquiner ou acheter, sont situés dans les couloirs du complexe Thistle menant aux salles des sessions scientifiques. Les exposants seront présents lundi 30 mai et mardi 31 mai de 9 h à 16 h.

- Maplesoft <http://www.maplesoft.com/>
- Pearson <http://www.pearsoncanada.ca/>
- Wiley <http://ca.wiley.com/WileyCDA>

# Organizers • Organisateurs

## Local Arrangements Committee • Comité des arrangements locaux

- Wai Kong (John) Yuen  
*Brock University*
- Ejaz Ahmed  
*Brock University*
- Mei Ling Huang  
*Brock University*
- Jan Vrbik  
*Brock University*
- Xiaojian Xu  
*Brock University*
- Peter Macdonald  
*McMaster University*

The local arrangements committee would like to thank all those who helped pull this event together. Among other responsibilities, our student volunteers have handled AV support and manned the registration desk. At Brock University, Sonya Forsey of Conference Services has been indispensable over the past two years. At the SSC Office, Larisa Valachko and Alan Kelm have provided crucial support for many aspects, including registration and exhibitors. Thanks to Edward Susko, Patrick Brown, Rhonda Rosychuk, Brian Allen and other previous local arrangements chairs, we did not have to start from scratch, and have appreciated their willingness to answer our (numerous!) questions. The SSC Meetings Coordinator Changbao Wu and other SSC executive members have offered insights and opinions that helped inform key decisions. We also greatly appreciate the accurate and timely translations provided by Catherine Cox and Pascale Giroux-Denis.

Le comité des arrangements locaux est très reconnaissant à tous ceux qui ont aidé à mettre sur pied cet événement. Nous tenons tout particulièrement à remercier les étudiants diplômés bénévoles qui ont contribué, entre autres, au soutien audiovisuel et au bureau des inscriptions. À Brock, Sonya Forsey des Services de conférence nous a été d'une aide inestimable ces deux dernières années. Au bureau de la SSC, Larisa Valachko et Alan Kelm ont fourni un appui essentiel pour divers aspects, dont l'inscription et les exposants. Merci à Edward Susko, Patrick Brown, Rhonda Rosychuk, Brian Allen et aux autres anciens présidents des comités des arrangements locaux qui ont généreusement partagé leurs expériences et volontiers répondu à nos nombreuses questions. Le coordonnateur des congrès Changbao Wu ainsi que d'autres membres de l'exécutif de la SSC ont offert leurs avis et opinions et contribué à éclairer la prise de décisions clés. Nous avons également beaucoup apprécié les traductions précises et rapides fournies par Catherine Cox et Pascale Giroux-Denis.

## Program Committee • Comité du programme

- Edward Susko (Chair • Président) *Dalhousie University*
- Karen Kopciuk *University of Calgary*, Biostatistics - Biostatistique
- Gail Ivanoff *University of Ottawa*, Probability - Probabilité
- Hugh Chipman *Acadia University*, Business and Industrial Statistics- Industrielle et Gestion
- Karla Fox *Statistics Canada*, Survey Methods - Enquête
- Peggy Ng *York University*, Statistical Education - Éducation en statistique
- Étienne Marceau *Laval University*, Actuarial Science - Science Actuarielle

## General Information • Informations générales

### Directions • Emplacement

Campus maps are available on the conference website, in the conference bag and on the back cover of this program. Those staying at Four Points by Sheraton St. Catharines Niagara Suites can walk to the campus (15 minutes to Thistle Complex) and the direction is available at <http://ssc.ca/en/meetings/2016>.

Des cartes du campus sont disponibles sur le site web du congrès et dans ce programme. Pour ceux qui logent au Four Points by Sheraton St. Catharines Niagara Suites, les itinéraires pour se rendre au campus (15 minutes du Thistle Complex) sont disponibles sur <http://ssc.ca/fr/congrès/2016>.

## Registration • Inscription

Registered participants can pick up their badges at the registration desk at the times indicated below. Walk-ins should register online and present their receipt at the registration desk. Registration will be available at:

- Sunday, May 29: 8:00 am to 5:30 pm  
Thistle Corridor across from Sean O'Sullivan Theatre
- Sunday, May 29: 5:45 pm to 8:00 pm  
Hallway outside of DeCew Dining Hall
- Monday, May 30: 7:30 am to 5:00 pm  
Thistle Corridor across from Sean O'Sullivan Theatre
- Tuesday, May 31: 8:15 am to 5:00 pm  
Thistle Corridor across from Sean O'Sullivan Theatre
- Wednesday, June 1: 8:15 am to 4:00 pm  
Thistle Corridor across from Sean O'Sullivan Theatre

Les participants déjà inscrits peuvent venir chercher leur matériel de conférence à l'endroit et aux heures indiqués ci-dessous. Ceux qui veulent s'inscrire au moment du congrès devraient le faire en ligne et présenter leur reçu au kiosque d'inscription.

- Dimanche 29 mai : 08 h à 17 h 30  
Corridor Thistle, en face du théâtre Sean O'Sullivan
- Dimanche 29 mai : 17 h 45 à 20 h  
Couloir d'entrée du réfectoire DeCew
- Lundi 30 mai : 7 h 30 à 17 h  
Corridor Thistle, en face du théâtre Sean O'Sullivan
- Mardi 31 mai : 8 h 15 à 17 h  
Corridor Thistle, en face du théâtre Sean O'Sullivan
- Mercredi 1er juin : 8 h 15 à 17 h  
Corridor Thistle, en face du théâtre Sean O'Sullivan

## Parking on Campus • Stationnement sur le campus

Participants staying in the Alan Earp Residence at Brock University receive complimentary parking in the residence parking lot. Participants staying in area hotels who wish to drive to Brock University may park in the Visitor Parking Lot D at cash per entry rate of \$7.00. A visitor parking location map is available at <http://ssc.ca/en/meetings/2016/transportation>.

Les participants logeant à la résidence Alan Earp sur le campus de Brock pourront stationner sans frais dans le stationnement de la résidence. Les participants logeant dans les hôtels de la région et qui souhaitent se rendre sur le campus en voiture peuvent utiliser le stationnement visiteurs D pour 7 \$, payable en liquide. Vous trouverez la carte des stationnements visiteurs à : <http://ssc.ca/fr/congrès/2016/transport>.

## Campus Security • Sécurité sur le campus

- Fire/Police/Ambulance: 911
- Campus Security Services:  
905-688-5550 ext. 4300 (day) / 3200 (emergency/night)
- Niagara Regional Police: 905-688-4111
- Pompiers/Police/Ambulance : 911
- Force policière du campus :  
905-688-5550 poste 4300 (jour) / 3200 (urgences/nuit)
- Niagara Regional Police : 905-688-4111

## Internet Access • Accès internet

Wireless access is available for all attendees. The wireless network name is BrockGuest. Once you are connected to the network, open a browser and you should be re-directed to a User Agreement Policy page. After

L'accès sans fil est disponible pour tous les participants. Le nom du réseau à utiliser est BrockGuest. Après vous être connecté sur ce réseau, ouvrez une fenêtre de navigateur : vous serez redirigé vers une page d'accord utili-

clicking on the *Accept* button located at the bottom of the page, you will be connected to the internet.

sateur. Cliquez sur le bouton « Accept » en bas de cette page pour vous connecter à l'Internet.

## Food on Campus and in the City • Nourriture sur le campus et en ville

Lunch Monday to Wednesday is provided with registration and is served at DeCew Dining Hall. You will receive three lunch tickets, one for each day, when you register. Lunch is served from 12:00pm to 1:30pm. Lunch-time committee meetings will occur in the scheduled rooms, served with boxed lunches to be picked up close to the meeting rooms. Other venues that provide food on a cash basis on campus include the Guernsey Market (Monday-Friday: 10:00 am to 3:00 pm), Tim Hortons – Tower (Monday-Friday: 7:30 am to 4:00 pm).

For dinner, there are many good restaurants in and around St Catharines, and you are encouraged to explore them. A list of restaurants recommended by our friends and colleagues is available at <http://ssc.ca/en/meetings/2016/restaurants>. A few restaurants and a convenience store, within walking distance of the campus, can be found on the opposite side of Glenridge Ave across the campus and in Four Points by Sheraton St. Catharines Niagara Suites.

Le repas de midi est fourni avec l'inscription du lundi au mercredi ; il sera servi au réfectoire DeCew. Vous recevrez trois coupons repas, un pour chaque jour, au moment de vous inscrire. Le lunch est servi de midi à 13h30. Les diners-réunions de comités se tiendront dans les salles prévues : vous trouverez une boîte à lunch près de votre salle de réunion. Vous pouvez également acheter à manger ailleurs sur le campus : au marché Guernsey (lundi-vendredi, de 10h00 à 15h00) et chez Tim Hortons – Tower (lundi-vendredi, de 7h30 à 16h00).

Le soir, nous vous encourageons à explorer les nombreux bons restaurants de St. Catharines et de la région. Vous trouverez une liste de restaurants recommandés par nos amis et collègues à : <http://ssc.ca/fr/congrès/2016/restaurants>. Il existe aussi quelques restaurants et un dépanneur à distance de marche du campus, de l'autre côté de l'avenue Glenridge et au Four Points by Sheraton St. Catharines Niagara Suites.

## Athletics Facilities • Installations sportives

Walker Complex is the main fitness facility at Brock (<https://brocku.ca/recreation-services/facilities-and-hours>). It has two gyms, a fitness centre (cardio room), a pool, an indoor running track and courts for squash and tennis. Full access day passes are available for \$7.

Walker Complex est le centre d'entraînement principal de Brock (<https://brocku.ca/recreation-services/facilities-and-hours>). Il comprend un gymnase, une salle de cardio, une piscine, une piste d'athlétisme intérieure et des courts de squash et de tennis. Des passes de jour sont disponibles au prix de 7 \$.





## The Conference • Le congrès

### Language • Langue





An important feature of our meetings is the presentation of the abstracts and of the plenary session visual aids in both official languages. This translation was very ably carried out under the supervision of the Bilingualism Committee (chaired by Geneviève Lefebvre) and the translators Caroline Gras, Catherine Cox, Caroline Pelletier, Michelle Blaquière and Caroline Petit-Turcotte.

Une caractéristique importante de nos congrès est la présentation des résumés et des supports visuels des sessions plénières dans les deux langues officielles. Cette traduction a été très habilement menée sous la supervision du Comité du bilinguisme (présidé par Geneviève Lefebvre) et des traductrices Caroline Gras, Catherine Cox, Caroline Pelletier, Michelle Blaquière et Caroline



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

Petit-Turcotte.

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

## Rooms • Salles

Coffee breaks, exhibits and posters will be in the corridors of the Thistle Complex leading to rooms for the scientific sessions. Plenary sessions on Monday and Tuesday mornings are in David S. Howes Theatre located in Welch Hall. Scientific sessions will be in one of three buildings: Thistle Complex, Welch Hall or Academic South.

Les pauses-café, le salon et les affiches se feront dans les couloirs du complexe Thistle qui mènent aux salles des sessions scientifiques. Les séances plénières du matin, lundi et mardi, se tiendront dans le théâtre David S. Howes, dans l'édifice Welch Hall. Les sessions scientifiques se tiendront dans les trois édifices suivants : Complexe Thistle, édifice Welch Hall ou édifice Academic South.

## Poster Sessions • Séances d'affichage

Contributed posters and Case Study posters will be in the corridors of the Thistle Complex leading to rooms for the scientific sessions, with Case study posters being displayed Monday and contributed research posters appearing on Tuesday. The authors will be with their posters until the end of coffee break at 3:30 pm.

Les séances d'affiches libres et d'études de cas seront dans les couloirs du complexe Thistle menant aux salles des sessions scientifiques, avec les affiches des études de cas exposées le lundi et les affiches libres exposées le mardi. Les affiches seront exposées de midi à 17h30. Les auteurs seront avec leur affiche de 13h30 jusqu'à la fin de la pause-café à 15h30.

## Workshops • Ateliers

**Sunday, May 29, 9:00 - 4:00 pm • Dimanche 29 mai, 9h00 - 16h00**

### **Thistle Complex**

Workshops organized by the sections will be held Sunday 9:00 am to 4:00 pm in the Thistle Complex. Coffee breaks will be held in the corridor at 10:15am and 2:15pm. Lunch break is at 12 noon.

Les ateliers organisés par les groupes auront lieu le dimanche de 9 heures à 16 heures dans le complexe Thistle. Les pauses-café auront lieu dans le hall du complexe à 10h15 et à 14h15. Le lunch est à midi.

## Other Meetings • Autres reunions

**SSC Annual General Meeting:** Monday 5-6pm  
Thistle 242

**Section AGM's:** Tuesday 5-6pm  
Various rooms in Thistle Complex

**Committee meetings:** Noontime business meetings will occur in the Thistle Complex or Welch Hall. Business meeting participants meeting in the Thistle Complex should pick up their box lunches in the Thistle Corridor across from Thistle 255; and those meeting in Welch Hall should pick up their lunch boxes in the hallway outside Welch 208.

**Information on 2016 NSERC Discovery Grant Competition:** Sunday 5-6  
Thistle 258

**NSERC Discovery Grant Application Workshop:**  
Tuesday 5-6  
Thistle 256  
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.

**Assemblée générale annuelle :** lundi 17h-18h  
Thistle 242

**AGM des groupes :** mardi 17h-18h  
plusieurs salles à Thistle Complex

**Réunions de comités :** Les réunions d'affaires de midi se tiendront dans le complexe Thistle ou à Welch Hall. Les participants aux réunions au complexe Thistle sont invités à prendre une boîte à lunch dans le corridor Thistle en face de la salle Thistle 255; ceux qui se réunissent à Welch Hall sont invités à prendre une boîte à lunch dans le corridor en face de la salle Welch 208.

**Informations sur le concours de subventions à la découverte 2016 du CRSNG :** dimanche 17h-18h  
Thistle 258

**Atelier subventions du CRSNG :** mardi 17h-18h  
Thistle 256.  
Cet atelier, présenté par le personnel des subventions à la découverte du CRSNG, 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 pour la préparation d'une demande de subvention à la découverte. À la fin de l'atelier, les participants pourront poser des questions.

## Social Events • Activités sociales

### Welcome Reception • Réception de bienvenue

**Sunday, May 29, 6:00 - 8:00 pm • Dimanche 29 mai, 18h00 - 22h00 ..... DeCew Dining Hall**

The welcome reception will be held in the DeCew Dining Hall.

La réception de bienvenue aura lieu dans le réfectoire DeCew.

## Barbeque • Barbecue

**Monday, May 30, 6:30 - 8:30 pm • Lundi 30 mai, 18h30 - 20h30 .....Pond Inlet**

The student BBQ is free for undergraduate and graduate students; registration is required. The BBQ will take place in the Pond Inlet.

Le barbecue étudiant est gratuit pour les étudiants du premier cycle et des cycles supérieurs ; l'inscription est requise. Il aura lieu au Pond Inlet.

## Banquet

**Tuesday, May 31, 6:15 pm • Mardi 31 mai, 18h15 .....Market Hall**

The banquet will be held at the Market Hall inside the Thistle Complex, with cocktails and a cash bar at 6:15 pm and dinner service beginning at 7:05 pm. All conference participants who have selected a meal for the banquet will find a banquet ticket in their registration envelope. You are asked to place this on the table in front of you. If you have a ticket but do not wish to attend the banquet, please return your ticket to the registration desk by noon on Tuesday, May 31. Buses are arranged to transport participants from the four conference hotels to Market Hall (Pick-up times: Holiday Inn - 5:30 pm; Best Western - 5:40 pm; Days Inn - 5:45 pm; Four Point - 6:00 pm and 6:10 pm) and return participants after the banquet (Return times: Four Point - 9:45 pm and 10:00 pm ; Holiday Inn, Best Western and Days Inn - 10:00 pm). We recommend that participants staying in the conference hotels use the bus service, and we ask that you show your banquet ticket prior to boarding the bus.

Le banquet se tiendra au Market Hall dans le complexe Thistle, avec cocktails et bar payant à partir de 18h15, suivi du souper servi à table à partir de 19h05. Les participants au congrès qui ont choisi de prendre part au banquet trouveront un ticket repas dans leur enveloppe d'inscription. Nous vous demanderons de le placer devant vous à la table. Si vous avez un ticket mais que vous ne souhaitez plus participer au banquet, veuillez remettre votre ticket au bureau des inscriptions, mardi 31 mai à midi au plus tard. Des bus ont été loués pour transporter les participants des quatre hôtels du congrès jusqu'au Market Hall (départs : Holiday Inn – 17h30 ; Best Western - 17h40 ; Days Inn - 17h45 ; Four Point – 18h00 et 18h10) et y accompagner les participants à la fin du banquet (départs : Four Point – 21h45 et 22h00 ; Holiday Inn, Best Western and Days Inn - 22h00). Nous recommandons aux participants qui logent dans les hôtels du congrès d'utiliser ce service et nous vous demandons de présenter votre ticket de banquet avant de monter à bord.

## Other Social Events • Autres événements sociaux

**Informal Gathering of New Investigators** Monday 7:30 pm  
Kelsey's Bar & Grill, 221 Glendale Ave., St. Catharines

**Repas des nouveaux chercheurs** lundi à 19h30  
Kelsey's Bar & Grill, 221 Glendale Ave., St. Catharines

## Committees and Meetings • Comités et réunions

### Saturday May 28

samedi 28 mai

18:00 - 23:00 Executive Committee 1 / Comité exécutif 1  
*Four Points by Sheraton St. Catharines Niagara Suites*

### Sunday May 29

dimanche 29 mai

9:30 - 12:00 Fundraising Committee / Comité de collecte de fonds  
*Sankey Chamber*

12:00 - 17:00 Board of Directors 1 / Conseil d'administration 1  
*Sankey Chamber*

14:30 - 17:00 Department Heads Meeting / Rencontre des directeurs de départements  
*Thistle 258*

### Monday May 30

lundi 30 mai

12:00 - 13:30 Business and Industrial Statistics Section Executive Committee 1 / Comité exécutif du groupe de statistique industrielle et de gestion 1  
*Thistle 255*

12:00 - 13:30 Actuarial Science Section Executive Committee 1 / Comité exécutif du groupe d'actuariat 1  
*Thistle 242*

12:00 - 13:30 Biostatistics Section Executive Committee 1 / Comité exécutif du groupe de biostatistique 1  
*Thistle 246*

12:00 - 13:30 Finance Committee / Comité des finances  
*Thistle 259*

12:00 - 13:30 Probability Section Executive Committee 1 / Comité exécutif du groupe de probabilité 1  
*Thistle 256*

12:00 - 13:30 Publications Committee / Comité des publications  
*Welch 207*

12:00 - 13:30 Research Committee / Comité de la recherche  
*Welch 209*

12:00 - 13:30 Statistical Education Section Executive Committee 1 / Comité exécutif du groupe d'éducation en statistique 1  
*Thistle 257*

12:00 - 13:30 Committee on Women in Statistics / Comité sur les femmes en statistique  
*Thistle 258*

12:00 - 13:30 Survey Methods Section Executive Committee Meeting 1 / Comité exécutif du groupe de méthodes d'enquête 1  
*Thistle 325*

12:00 - 18:00 Award for Case Studies in Data Analysis Committee / Comité du prix pour les études de cas et l'analyse de données  
*Thistle 254*

17:00 - 18:00 SSC Annual General Meeting / Assemblée générale annuelle de la SSC  
*Thistle 242*

### Tuesday May 31

mardi 31 mai

7:30 - 8:30 Statistics Education Committee / Comité d'éducation en statistique  
*Thistle 255*

9:00 - 13:30	Student Research Presentation Award Committee 1 / Comité du prix pour la présentation de recherche étudiante 1 <i>Thistle 254</i>
12:00 - 13:30	CJS Editorial Board / Comité de rédaction de la RCS <i>Thistle 242</i>
12:00 - 13:30	Census at School Committee / Comité de recensement à l'école <i>Thistle 255</i>
12:00 - 13:30	Office Committee / Comité du Bureau <i>Thistle 259</i>
12:00 - 13:30	Public Relations Committee / Comité des relations publiques <i>Thistle 258</i>
12:00 - 13:30	Committee on New Investigators / Comité des nouveaux chercheurs <i>Thistle 257</i>
12:00 - 13:30	Accreditation General Meeting / Assemblée générale des membres accrédités <i>Thistle 246</i>
12:00 - 13:30	Committee on Membership / Comité de recrutement <i>Thistle 256</i>
13:30 - 18:00	Student Research Presentation Award Committee 2 / Comité du prix pour la présentation de recherche étudiante 2 <i>Mackenzie Chown A323</i>
17:00 - 18:00	Business and Industrial Statistics Section General Meeting / Groupe de statistique industrielle et de gestion assemblée générale <i>Thistle 257</i>
17:00 - 18:00	Actuarial Science Section General Meeting / Groupe d'actuariat assemblée générale <i>Thistle 255</i>
17:00 - 18:00	Biostatistics Section General Meeting / Groupe de biostatistique assemblée générale <i>Thistle 242</i>
17:00 - 18:00	Probability Section General Meeting / Groupe de probabilité assemblée générale <i>Thistle 258</i>
17:00 - 18:00	Statistical Education Section General Meeting / Groupe d'éducation en statistique assemblée générale <i>Thistle 259</i>
17:00 - 18:00	Survey Methods Section General Meeting / Groupe de méthodes d'enquête assemblée générale <i>Thistle 246</i>

**Wednesday June 1****mercredi 1er juin**

12:00 - 13:30	Accreditation Committee / Comité d'accréditation <i>Thistle 259</i>
12:00 - 13:30	Actuarial Science Section Executive Committee 2 / Comité exécutif du groupe d'actuariat 2 <i>Thistle 254</i>
12:00 - 13:30	Biostatistics Section Executive Committee 2 / Comité exécutif du groupe de biostatistique 2 <i>Thistle 255</i>
12:00 - 13:30	Probability Section Executive Committee 2 / Comité exécutif du groupe de probabilité 2 <i>Thistle 256</i>
12:00 - 13:30	Program Committee / Comité du programme <i>Thistle 242</i>
12:00 - 13:30	Statistical Education Section Executive Committee 2 / Comité exécutif du groupe d'éducation en statistique 2 <i>Thistle 257</i>

12:00 - 13:30	Survey Methods Section Executive Committee Meeting 2 / Comité exécutif du groupe de méthodes d'enquête 2 <i>Thistle 258</i>
12:00 - 13:30	Student and Recent Graduate Committee / Comité des étudiants et diplômés récents <i>Thistle 246</i>
16:45 - 18:00	Board of Directors 2 / Conseil d'administration 2 <i>Thistle 253</i>
18:00 - 19:00	Executive Committee 2 / Comité exécutif 2 <i>Thistle 253</i>

## Social and Information Events • Événements sociaux et informatifs

### Sunday May 29

dimanche 29 mai

8:00 - 17:30	Registration / Inscription <i>Thistle Corridor across from Sean O'Sullivan Theatre</i>
15:15 - 17:15	Afternoon Excursion- Fee event, Hernder Estate Wines / Excursion en après-midi- activité payante, Hernder Estate Wines
17:00 - 18:00	Information on 2016 NSERC Discovery Grant Competition / Information sur le concours 2016 des subventions à la découverte du CRSNG <i>Thistle 258</i>
17:45 - 20:00	Registration / Inscription <i>Hallway outside of DeCew Dining Hall</i>
18:00 - 20:00	Welcome Reception / Réception de bienvenue <i>DeCew Dining Hall</i>

### Monday May 30

lundi 30 mai

7:30 - 17:00	Registration / Inscription <i>Thistle Corridor across from Sean O'Sullivan Theatre</i>
9:50 - 10:30	Coffee Break / Pause-café <i>Thistle Corridor across from Thistle 255</i>
14:45 - 15:30	Coffee Break / Pause-café <i>Thistle Corridor across from Thistle 255</i>
18:30 - 20:30	Student BBQ / Barbecue étudiant <i>Pond Inlet</i>
19:30	Informal Gathering of New Investigators / Rassemblement informel des nouveaux chercheurs <i>Kelsey's Bar &amp; Grill, 221 Glendale Ave., St. Catharines</i>

### Tuesday May 31

mardi 31 mai

8:15 - 17:00	Registration / Inscription <i>Thistle Corridor across from Sean O'Sullivan Theatre</i>
9:50 - 10:30	Coffee Break / Pause-café <i>Thistle Corridor across from Thistle 255</i>
12:00 - 13:00	Maplesoft Demo <i>Academic South 217</i>
14:45 - 15:30	Coffee Break / Pause-café <i>Thistle Corridor across from Thistle 255</i>
17:00 - 18:00	NSERC Discovery Grant Application Workshop / Atelier sur les demandes de subventions à la découverte du CRSNG <i>Thistle 256</i>
18:15 - 19:05	Banquet- Cocktails and Cash Bar / Banquet- Bar payant <i>Market Hall</i>
19:05 - 22:00	Banquet- Dinner Service / Banquet- Service du repas <i>Market Hall</i>

### Wednesday June 1

mercredi 1er juin

8:15 - 16:00	Registration / Inscription <i>Thistle Corridor across from Sean O'Sullivan Theatre</i>
--------------	---

---

9:50 - 10:30	Coffee Break / Pause-café <i>Thistle Corridor across from Thistle 255</i>
14:30 - 15:00	Coffee Break / Pause-café <i>Thistle Corridor across from Thistle 255</i>

---



## List of Sessions • Liste des sessions

	<b>Title</b>	<b>Page</b>	<b>Titre</b>
1A-A2	SSC Presidential Invited Address	22	L'allocution de l'invité du président de la SSC
1B-I1	CANSSI at the SSC: A Statistical Science Showcase	22	L'INCASS à la SSC : vitrine de la science statistique
1B-I2	Quantile regression and Extreme Value Analysis	23	Régression quantile et analyse des valeurs extrêmes
1B-I3	Statistical Methods in Public Health	23	Méthodes statistiques en santé des populations
1B-I4	Statistical Models in Actuarial Science	23	Modèles statistiques en actuariat
1B-I5	Stochastic Modelling for Energy Markets	24	Modélisation stochastique des marchés de l'énergie
1B-I6	What Do School Teachers and Students Need from the SSC?	24	Que peut faire la SSC pour les enseignants et les élèves ?
1B-C1	Biostatistics: Methodological Innovation 1	24	Biostatistique : innovation méthodologique 1
1B-C2	Biostatistics: Methods and Applications 1	25	Biostatistique : méthodes et applications 1
1B-C3	Statistical Methods Utilizing Copulas	26	Méthodes statistiques utilisant des copules
1B-C4	Statistical Methods and Applications 1	26	Méthodes statistiques et applications 1
1B-C5	Survey Methodology 1	27	Méthodologie d'enquête 1
1C-D1	Case Study 1: Can Google Flu Trends Predict the Frequency and Results of Tests for Influenza and Other Respiratory Illnesses?	28	Étude de cas 1: Le site Google Flu Trends peut-il prédire la fréquence et le résultat des tests pour la grippe et autres maladies respiratoires ?
1C-D2	Case Study 2: What Predicts Sustainability of Canadian Charities?	29	Étude de cas 2: Pouvons-nous prédire la viabilité des organismes de bienfaisance canadiens ?
1D-I1	Analysis of Big Data	29	Analyse des mégadonnées
1D-I2	New Perspectives in Causal Inference	30	Nouvelles perspectives en inférence causale
1D-I3	Small Area Estimation: New Developments	30	Estimation pour petits domaines : nouveaux développements
1D-I4	Statistical Modelling in Environmental and Health Studies	30	Modélisation statistique dans les études sur l'environnement et la santé
1D-I5	Imaging Genomics - A New Frontier for Statistical Methodology	31	Imagerie génomique – une nouvelle frontière pour la méthodologie statistique
1D-I6	Opportunities for statistical scientists in CIHR programs: a panel discussion	31	Opportunités pour les statisticiens dans les programmes IRSC : table ronde
1D-C1	Actuarial Science and Finance 1	31	Actuariat et finance 1
1D-C2	Biostatistics: Methods and Applications 2	32	Biostatistique : méthodes et applications 2
1D-C3	Statistical Methods and Applications 2	32	Méthodes statistiques et applications 2
1D-C4	Statistical Theory	33	Théorie statistique 1
1D-C5	Statistics Education	34	Éducation en statistique
1E-I1	Advances in Longitudinal Data Analysis	34	Progrès récents en analyse de données longitudinales
1E-I2	Advances in Regression Tree Modelling	35	Progrès récents en modélisation par les arbres de régression
1E-I3	An Introduction to the Joint Modelling of Longitudinal and Survival Data	35	Introduction à la modélisation conjointe des données longitudinales et de survie
1E-I4	Missing Data in Survey Sampling	35	Données manquantes en échantillonnage
1E-I5	Probability and Statistical Problems and Results Arising from Neural Models	36	Probabilités et problèmes statistiques et résultats découlant de modèles neuronaux
1E-I6	Statistical Education Section Presidential Invited Address	36	Allocution du l'invité du président du Groupe d'éducation en statistique
1E-C1	Biostatistics: Causal Inference and Measurement Error	36	Biostatistique : inférence causale et erreur de mesure
1E-C2	Biostatistics: Methodological Innovation 2	37	Biostatistique : innovation méthodologique 2

	<b>Title</b>	<b>Page</b>	<b>Titre</b>
1E-C3	Biostatistics: Survival and Current Status Data	37	Biostatistique : données de survie et d'état actuel
1E-C4	Statistical Methods and Applications 3	38	Méthodes statistiques et applications 3
1E-C5	Statistics in Action	39	La statistique en action
2A-A	Gold Medal Address	40	Allocution du récipiendaire de la Médaille d'or
2B-I1	Advanced Statistical and Machine Learning Techniques for High Dimensional Data Analysis	40	Techniques statistiques et apprentissage machine avancés pour l'analyse de données en haute dimension
2B-I2	Advancements to State-Space Models for Fisheries Science	41	Progrès récents dans les modèles à espace d'états pour les sciences halieutiques
2B-I3	Contributions of Professor Ian B. MacNeill to Statistical Theory, Methodology, Teaching and Practice	41	Contributions du professeur Ian B. MacNeill à la théorie statistique, Méthodologie, enseignement et la pratique
2B-I4	Graduate Students in Actuarial Science	41	Étudiants de cycles supérieurs en actuariat
2B-I5	Recent Developments in Designs of Statistical Experiments. In honour of the 65th birthday of Professor Douglas P. Wiens	42	Progrès récents en planification d'expériences. En l'honneur du 65e anniversaire du professeur Douglas P. Wiens
2B-C1	Biostatistics: Applications	42	Biostatistique : applications
2B-C2	Methodological Advances: A Smörgåsbord	43	Progrès méthodologiques : un buffet abondant
2B-C3	Mixture models	44	Modèles de mélange
2B-C4	Multivariate Data	44	Données multivariées
2B-C5	Survey Methodology 2	45	Méthodologie d'enquête 2
2C-P	Poster Session	46	Séance d'affichage
2D-I1	Developing Curriculum Guidelines for Canadian Statistics Undergraduate Programs	48	Comment mettre au point des directives pour les programmes de premier cycle en statistique au Canada
2D-I2	Genomic Data Analysis, Modelling and Testing	48	Analyse, modélisation et tests pour des données de séquençage génomique
2D-I3	Long Memory Time Series	48	Séries chronologiques à mémoire longue
2D-I4	Methods in Survival Analysis Motivated by challenges in clinical data	49	Méthodes d'analyse de survie motivées par les problèmes de données cliniques
2D-I5	On Some Applications of James-Stein Shrinkage Estimation	49	À propos de quelques applications des estimateurs de rétrécissement à la James Stein
2D-I6	Risk Measures in Actuarial Science	49	Mesures du risque en actuariat
2D-C1	Actuarial Science and Finance 2	50	Actuariat et finance 2
2D-C2	Biostatistics: Treatment Comparisons	50	Biostatistique : comparaisons de traitements
2D-C3	Statistical Computing and Visualization	51	Calcul statistique et visualisation
2D-C4	Statistical Genetics	52	Génétique statistique
2D-C5	Statistical Inference and Applications 1	52	Inférence statistique et applications 1
2E-I1	Adaption and Approximation in Markov Chain Monte Carlo	53	Adaptation et approximation pour la méthode de Monte-Carlo par chaînes de Markov
2E-I2	Building a Research Career: From New Investigator to Present	53	Comment bâtir une carrière en recherche : de nouveau chercheur à aujourd'hui
2E-I3	Design and Methodological Issues in Health Studies of Vulnerable Populations	53	Problèmes de conception et de méthodologie des études sur la santé des populations vulnérables
2E-I4	Inference and Prediction from Distributed Data	54	Inférence et prévisions à partir de données distribuées
2E-I5	Statistical Challenges in Large-Scale Inference	54	Défis statistiques en inférence à grande échelle
2E-I6	Survey Methodology in a Big Data World	54	Méthodes d'enquête dans un monde de mégadonnées
2E-C1	Biostatistics: New Inferential Strategies	55	Biostatistique : nouvelles stratégies d'inférence
2E-C2	Probability and Estimation	55	Probabilité et estimation
2E-C3	Regression Models	56	Modèles de régression

	<b>Title</b>	<b>Page</b>	<b>Titre</b>
2E-C4	Statistical Applications and Methods	56	Méthodes et applications statistiques
2E-C5	Statistical Inference and Applications 2	57	Inférence statistique et applications 2
3A-A1	CRM-SSC Prize in Statistics Address	59	Allocution du récipiendaire du Prix CRM-SSC en statistique
3A-A2	SSC Impact Award Session	59	Allocution du récipiendaire du Prix pour impact de la SSC
3B-I1	Analysis of Complex Survey Data	59	Analyse de données d'enquête complexes
3B-I2	Analysis of Genomic Sequencing Data in Families	60	Analyse des données de séquençage génomique dans les familles
3B-I3	High-dimensional Statistics: Challenges and Recent Developments	60	Statistiques en haute dimension : défis et progrès récents
3B-I4	Statistical Theory and Modeling for Functional Data Analysis	60	Théorie et modélisation statistiques pour l'analyse de données fonctionnelles
3B-I5	Staying Up-To-Date As a Practicing Biostatistician	61	Comment rester à la page en tant que biostatisticien praticien
3B-I6	The Theory of Dispersion Models with Applications	61	La théorie de la dispersion avec des applications
3B-C1	Biostatistics: Survival Analysis	61	Biostatistique : analyse de survie
3B-C2	Time Dependent Data	62	Données dépendantes du temps
3C-I1	Isobel Loutit Lecture	63	Allocution invitée Isobel Loutit
3C-I2	Survey Methods Section Presidential Invited Address	63	Allocution de l'invité du président du Groupe des méthodes d'enquête
3C-A1	CJS Award Address	63	Allocution du récipiendaire du Prix de la RCS
3C-A2	Pierre Robillard Award Address	64	Allocution du récipiendaire du Prix Pierre-Robillard
3D-I1	Financial and Actuarial Mathematics	64	Mathématique actuarielle et financière
3D-I2	High Frequency Data	64	Données de haute fréquence
3D-I3	Network Meta-Analysis	65	Méta-analyse de réseaux
3D-I4	Statistical Methods for Analysing Cancer Data	65	Méthodes statistiques pour analyse des données sur le cancer

Program • Programme

**Saturday May 28**

**samedi 28 mai**

**18:00 - 23:00**

**Four Points by Sheraton St. Catharines Niagara Suites**

**Executive Committee 1**

**Comité exécutif 1**

**Sunday May 29****dimanche 29 mai****8:00 - 17:30****Thistle Corridor across from Sean O'Sullivan Theatre****Registration  
Inscription****9:00 - 16:00****Thistle 245****Analysis of Categorical Data  
Analyse de données catégorielles**

9:00 - 16:00 Christopher R. Bilder (University of Nebraska)

9:00 - 16:00 Thomas M. Loughin (Simon Fraser University)

**9:00 - 16:00****Thistle 255****Bridging the Gap: Turning Classical Statistics Experience Directly into a Working Knowledge of Survey Data Analysis****Comblér le fossé : comment transformer une expérience de la statistique classique en une connaissance fonctionnelle de l'analyse des données d'enquête**

9:00 - 16:00 Claude Girard (Statistics Canada)

**9:00 - 16:00****Thistle 256****How to Speed Up Your R Computation with Vectorization and Parallel Programming  
Comment accélérer les calculs sous R grâce à la vectorisation et à la programmation parallèle**

9:00 - 16:00 Hao Yu (University of Western Ontario)

**9:00 - 16:00****Thistle 257****Teaching with Shiny Apps  
Comment enseigner avec des applis « Shiny »**

9:00 - 16:00 John Braun (University of British Columbia)

9:00 - 16:00 Alison Gibbs (University of Toronto)

9:00 - 16:00 Victor Veitch (University of Toronto)

**9:00 - 16:00****Thistle 246****Uncertainty Quantification and Optimization for Complex Models with Gaussian Processes  
Quantification de l'incertitude et optimisation de modèles complexes avec processus gaussiens**

9:00 - 16:00 Jason Loeppky (University of British Columbia)

**9:30 - 12:00****Sankey Chamber****Fundraising Committee  
Comité de collecte de fonds****12:00 - 17:00****Sankey Chamber****Board of Directors 1  
Conseil d'administration 1****14:30 - 17:00****Thistle 258****Department Heads Meeting  
Rencontre des directeurs de départements**

---

**15:15 - 17:15**

**Afternoon Excursion- Fee event, Hernder Estate Wines**  
**Excursion en après-midi- activité payante, Hernder Estate Wines**

---

**17:00 - 18:00**

**Thistle 258**

**Information on 2016 NSERC Discovery Grant Competition**  
**Information sur le concours 2016 des subventions à la découverte du CRSNG**

---

**17:45 - 20:00**

**Hallway outside of DeCew Dining Hall**

**Registration**  
**Inscription**

---

**18:00 - 20:00**

**DeCew Dining Hall**

**Welcome Reception**  
**Réception de bienvenue**

**Monday May 30****lundi 30 mai**

7:30 - 17:00

**Thistle Corridor across from Sean O'Sullivan Theatre****Registration  
Inscription**8:30 - 8:45 **Session 1A-A1****Welch - David Howes****Inaugural Session  
Séance inaugurale**

Special Session / Conférence spéciale



8:30 - 8:40 Brian Allen (SSC President)

8:40 - 8:45 Wai Kong Yuen (Local Arrangements Chair)

8:45 - 9:50 **Session 1A-A2** (abstract/résumé - page 68)**Welch - David Howes****SSC Presidential Invited Address****L'allocation de l'invité du président de la SSC**

Special Session / Conférence spéciale

8:45 - 9:50 John D. Kalbfleisch (University of Michigan)

Some Statistical Topics of Personal Interest / Quelques sujets statistiques d'intérêt personnel  



9:50 - 10:30

**Thistle Corridor across from Thistle 255****Coffee Break  
Pause-café**10:20 - 11:50 **Session 1B-I1** (abstracts/résumés - page 69)**Welch - David Howes****CANSSI at the SSC: A Statistical Science Showcase****L'INCASS à la SSC : vitrine de la science statistique**



Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: John Braun (University of British Columbia)



10:20 - 10:50 Alisha Albert-Green (University of Western Ontario), John Braun (University of British Columbia Okanagan), Charmaine Dean (University of Western Ontario)

A Spatio-Temporal Cluster Process for Modelling Storm Cells / Un processus de regroupement spatio-temporel pour la modélisation des cellules orageuses  

10:50 - 11:20 Derek Bingham (Simon Fraser University)

Design of Experiments in the CANSSI CRT "Statistical Modeling of the World: Computer and Physical Models in Earth, Ocean, and Atmospheric Sciences" / Conception d'expériences dans le PRC « Modélisation statistique du monde : modèles informatiques et physiques en sciences de la Terre, des océans et de l'atmosphère » de l'INCASS  

11:20 - 11:50 Radu V. Craiu (University of Toronto)

Modern Research Topics in Copula Methods: A Report from the CANSSI Collaborative Team Project / Sujets de recherche modernes sur les méthodes de copules : un rapport de l'équipe de projet collaborative INCASS  







---

**10:20 - 11:50    Session 1B-I2** (abstracts/résumés - page 71) **Welch 209**

**Quantile regression and Extreme Value Analysis**  
**Régression quantile et analyse des valeurs extrêmes**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Mei Ling Huang (Brock University)

- 10:20 - 10:50    Vincenzo Coia (University of British Columbia), Harry Joe & Natalia Nolde (The University of British Columbia)  
 Forecasting Extremes for Flooding / Prévion des extrêmes pour les inondations  
- 10:50 - 11:20    Keith Knight (University of Toronto)  
 Non-parametric Estimation of Extreme Conditional Quantiles and Lawson's Algorithm / L'estimation non paramétrique des quantiles conditionnels extrêmes et l'algorithme de Lawson  
- 11:20 - 11:50    Christine Nguyen, Mei Ling Huang & Wai Kong Yuen (Brock University)  
 On Weighted Quantile Regression / Régression quantile pondérée  

---







**10:20 - 11:50    Session 1B-I3** (abstracts/résumés - page 73) **Thistle 325**

**Statistical Methods in Public Health**  
**Méthodes statistiques en santé des populations**

Invited Paper Session / Présentations sur invitation

Chair/Président: Patrick Brown (University of Toronto)

Organizer/Responsable: Lennon Li (Public Health Ontario)

- 10:20 - 10:50    Rob Deardon (University of Calgary), Waleed Almutiry (University of Guelph)  
 Infectious Disease Modelling in the Presence of Underlying Contact Network Uncertainty / Modélisation de maladies infectieuses en présence d'incertitude du réseau de contact sous-jacent  
- 10:50 - 11:20    Andrew Lawson & Rachel Carroll (Medical University of South Carolina)  
 Spatio-temporal Latent Model Choice: With Application in Multiple Disease Small Area Analysis / Choix de modèles latents spatio-temporels : application dans plusieurs analyses de maladies de petits domaines  
- 11:20 - 11:50    Rhonda Rosychuk (University of Alberta)  
 Cluster Detection Tests for the Surveillance of Emergency Department Presentation Data / Tests de détection de grappes pour la surveillance des données de présentation des services des urgences  





---

**10:20 - 11:50    Session 1B-I4** (abstracts/résumés - page 75) **Thistle 242**



**Statistical Models in Actuarial Science**  
**Modèles statistiques en actuariat**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Jean-Philippe Boucher (UQAM)

- 10:20 - 10:50    Mathieu Pigeon (UQAM), Arthur Charpentier (Université du Québec à Montréal et Université de Rennes 1)  
 Macro and Micro Methods in Non-Life Claims Reserving / Modèles individuels et collectifs pour les réserves en assurance IARD  
- 10:50 - 11:20    Peng Shi (University of Wisconsin), Lu Yang (University of Wisconsin-Madison)  
 A Pair Copula Construction Model for Insurance Experience Rating / Modèle de construction de copule par paires pour la tarification d'antécédents en assurance  





11:20 - 11:50 Cary Chi-Liang Tsai (Simon Fraser University), Tzuling Lin (National Chung Cheng University)  
Applications of the Bühlmann Credibility Model to Mortality Forecasting / Applications du modèle de crédibilité de Bühlmann à la prévision de la mortalité  



---



**10:20 - 11:50 Session 1B-I5** (abstracts/résumés - page 77) **Thistle 246**

**Stochastic Modelling for Energy Markets**  
**Modélisation stochastique des marchés de l'énergie**

Invited Paper Session / Présentations sur invitation  
Chair/Président: Reg Kulperger (University of Western Ontario)  
Organizer/Responsable: Matt Davison (University of Western Ontario)

10:20 - 10:50 Matt Davison (University of Western Ontario), Nicolas Merener (Universidad Torcuato di Tella)  
Corn/Ethanol Price Spreads Emerging from an Inhomogeneous Set of Price-Influencing Ethanol Producers / Écarts de prix entre le maïs et l'éthanol dans un ensemble non homogène de producteurs d'éthanol ayant une influence sur les prix  

10:50 - 11:20 Anatoliy Swishchuk (University of Calgary)  
Stochastic Modelling and Pricing of Energy Markets' Contracts with Local Stochastic Delayed and Jumped Volatilities / Modélisation stochastique et tarification des contrats sur les marchés de l'énergie avec volatilités stochastiques locales à délais et à sauts  

11:20 - 11:50 Hans Tuentler (University of Toronto)  
Firming Wind Farm Revenues / Stabilisation des revenus des parcs éoliens  

---

**10:20 - 11:50 Session 1B-I6** (description - page 79) **Thistle 258**

**What Do School Teachers and Students Need from the SSC?**  
**Que peut faire la SSC pour les enseignants et les élèves ?**

Invited Paper Session / Présentations sur invitation  
Chair/Président: Lorna Deeth (University of Guelph)  
Organizer/Responsable: Alison Gibbs (University of Toronto)

10:20 - 11:50 Anne-Sophie Charest (Université Laval)

10:20 - 11:50 Steven Khan (Brock University)

10:20 - 11:50 Angelo Lillo (Sir Winston Churchill Secondary School)



10:20 - 11:50 Leonard Rak (Sheridan College)



---



**10:20 - 11:50 Session 1B-C1** (abstracts/résumés - page 80) **Thistle 259**

**Biostatistics: Methodological Innovation 1**  
**Biostatistique : innovation méthodologique 1**

Contributed Paper Session / Séance de communications libres  
Chair/Président: Caitlin Daly (McMaster University)

10:20 - 10:35 Ying Wu & Richard J. Cook (University of Waterloo)  
A Two-Phase Model For Chronic Disease Processes Under Intermittent Inspection / Modèle à deux phases pour les processus de la maladie chronique sous inspection intermittente  

10:35 - 10:50 Wenyan Zhong, Jingjing Wu & Xuewen Lu (University of Calgary)  
Group Selection in Proportional Odds Model for Right-Censored Data / Sélection de groupe dans le modèle des cotes proportionnelles pour les données censurées à droite  

10:50 - 11:05 Steve Ferreira Guerra, Mireille Schnitzer & Lucie Blais (Université de Montréal)  
A Self-Selecting Procedure for the Optimal Discretization of a Continuous Timeline for Longitudinal Causal Inference Methods / Une procédure de sélection automatique pour la discrétisation optimale d'une ligne de temps continue pour l'application de méthodes longitudinales en inférence causale  

- 11:05 - 11:20 Zelalem Negeri, Mateen Shaikh & Joseph Beyene (McMaster University)  
Bivariate Random Effects Meta-Analysis Models for Diagnostic Test Accuracy Studies Using Arcsine-Based Transformations / Modèles de meta-analyse bivariés à effets aléatoires pour les études sur l'exactitude des tests diagnostiques utilisant les transformations arc-sinus  
- 11:20 - 11:35 Maxime Turgeon (McGill University), Karim Oualkacha (Université du Québec à Montréal), Antonio Ciampi, Golsa Dehghan, Celia Greenwood & Aurélie Labbe (McGill University)  
Principal Component of Explained Variance: An Efficient and Optimal Data Dimension Reduction Framework for Association Studies / Composante principale de variance expliquée : une méthode de réduction dimensionnelle efficace et optimale pour les études d'association  
- 11:35 - 11:50 Marie-Hélène Toupin (Université Laval), Jean-François Quessy (Université du Québec à Trois-Rivières), Louis-Paul Rivest (Université Laval)  
On the Family of Chi-Square Copulas / À propos de la famille de copules Khi-carré  

---

**10:20 - 11:50**    **Session 1B-C2**    (abstracts/résumés - page 83)    **Thistle 255**

**Biostatistics: Methods and Applications 1**

**Biostatistique : méthodes et applications 1**

Contributed Paper Session / Séance de communications libres

Chair/Président: Ehsan Karim (McGill University)

- 10:20 - 10:35 Jill Ainsworth, Erica Moodie & Abbas Khalili (McGill University)  
Estimating Parameters from a Finite Mixture of Generalised Linear Mixed-Effect Models Using the Maximum Likelihood Method / Estimer des paramètres d'un mélange fini de modèles linéaires généralisés à effets mixtes en utilisant la méthode de maximum de vraisemblance  
- 10:35 - 10:50 Joycelyne Efua Ewusie (McMaster University), Monika Kastner (Li Ka Shing Knowledge Institute, St Michaels Hospital, Toronto), George Thomplinson (University of Toronto, Toronto), Sharon Straus & Jemila S. Hamid (Li Ka Shing Knowledge Institute, St Michaels Hospital, Toronto)  
Methods for Interrupted Time Series Analysis of Count Data / Méthodes pour l'analyse des séries chronologiques interrompues de données de dénombrement  
- 10:50 - 11:05 Tania Alarcon Falconi (Civil and Environmental Engineering, Tufts University), Bertha Estrella (Corporación Ecuatoriana de Biotecnología, Ecuador), Fernando Sempértegui (Central University, Ecuador), Magaly Koch (Center for Remote Sensing, Boston University), Elena Naimova (Friedman School of Nutrition Science and Policy, Tufts University)  
Characterization of Seasonality in Longitudinal Clinical Trials / Caractérisation des saisonnalités dans les essais cliniques longitudinaux  
- 11:05 - 11:20 Khan Mohammad Kaviul, Rafal Kustra & Zhengfei Chen (Dalla Lana School of Public Health, University of Toronto)  
Investigation of Change in Methylation in Prostate Cancer Patients after Receiving Treatment / Étude du changement dans la méthylation chez les patients atteints de cancer de la prostate après avoir reçu un traitement  
- 11:20 - 11:35 Victoria Borg-Debono, Philip J. Devereaux, Lawrence Mbuagbaw & Lehana Thabane (McMaster University)  
Interim Data Sharing: What Interim Results/Information Should Be Shared During the Interim of a Randomized Controlled Trial with Those Responsible for the Conduct of the Trial? / Partage de données intérimaires : quels résultats/information intérimaires devraient être partagés avec ceux qui mènent l'essai lors d'un essai contrôlé randomisé ?  
- 11:35 - 11:50 Shahriar Shams (University of Toronto), Eleanor Pullenayegum (CHES, The Hospital for Sick Children)  
Evaluate the Feasibility of Bayesian Inference as a Means of Reducing Scoring Uncertainty for the EQ-5D Health Utilities / Évaluer la faisabilité de l'inférence bayésienne pour réduire la cotation de l'incertitude pour l'utilité de santé EQ-5D  

---







**10:20 - 11:50    Session 1B-C3** (abstracts/résumés - page 87) **Thistle 256**

**Statistical Methods Utilizing Copulas**

**Méthodes statistiques utilisant des copules**

Contributed Paper Session / Séance de communications libres

Chair/Président: Zhenhua Lin (University of Toronto)

- 10:20 - 10:35    Elif Acar (University of Manitoba), Candida Geerdens & Paul Janssen (Hasselt University)  
Conditional Copula Models for Right-Censored Clustered Event Time Data / Modèles de copules conditionnelles pour données historiques d'événements groupés censurés à droite  
- 10:35 - 10:50    Tarik Bahraoui (Université de Sherbrooke)  
A Family of Goodness-of-fit Tests for Copulas Based on Characteristic Functions / Une famille de tests d'adéquation pour des copules fondée sur les fonctions caractéristiques  
- 10:50 - 11:05    Manoj Bahuguna (Oakland University, Rochester, MI), Ravindra Khattree (Oakland University, Rochester, MI)  
Copula Transformation, Prediction and Exploration / Transformation avec copule, prédiction et exploration  
- 11:05 - 11:20    Caren Hasler & Radu V. Craiu (University of Toronto), Louis-Paul Rivest (Université Laval)  
Vine Copulas for Imputation of Monotone Non-Response / Copules en vignes pour l'imputation de la non-réponse monotone  
- 11:20 - 11:35    David Lee & Harry Joe (University of British Columbia)  
Multivariate Extreme Value Copulas with Factor and Tree Dependence Structures / Copules de valeurs extrêmes multivariées avec structures factorielles et arborescentes de dépendance  
- 11:35 - 11:50    Yassir Rabhi & Taoufik Bouezmarni (Université de Sherbrooke)  
Copula Function Under Biased Sampling and Informative Censoring / Fonction de la copule dans le cadre d'un échantillonnage biaisé et d'une censure informative  

---









**10:20 - 11:50    Session 1B-C4** (abstracts/résumés - page 90) **Welch 8G**





**Statistical Methods and Applications 1**

**Méthodes statistiques et applications 1**

Contributed Paper Session / Séance de communications libres

Chair/Président: David Spiegelhalter (Cambridge University)

- 10:20 - 10:35    Jingjia Chu, Reg Kulperger & Hao Yu (University of Western Ontario)  
A Multivariate Time Series Model with an Additive GARCH Type Structure / Un modèle de séries temporelles multivarié avec une structure de type GARCH additif  
- 10:35 - 10:50    Reyhaneh Hosseini & Mahmoud Zarepour (University of Ottawa)  
A Bayesian Nonparametric Chi-squared Goodness-of-fit Test / Un test d'ajustement du chi-deux bayésien non paramétrique  
- 10:50 - 11:05    Aurélien Guetsop Nangue & Martin Bilodeau (Université de Montréal)  
Approximations to Permutation Tests of Independence Between Two Random Vectors / Approximation du test de permutation d'indépendance entre deux vecteurs aléatoires  
- 11:05 - 11:20    Aurélien Nicosia, Thierry Duchesne, Louis-Paul Rivest & Daniel Fortin (Université Laval)  
Equivalence Between the Random Walk Model in the Plane and Conditional Logistic Regression in a Multi-State Framework, with Application to the Analysis of Animal Movement / Équivalence entre un modèle de marche aléatoire dans le plan et une régression logistique conditionnelle dans un cadre multi-états, application à l'analyse du mouvement animalier  

- 11:20 - 11:35 Hossein Zareamoghaddam & Serge Provost (Western University), Syed Ejaz Ahmed (Brock University)  
A Moment-Based Bivariate Density Estimation Methodology for Large Data Sets / Une méthodologie basée sur les moments pour l'estimation de densités bivariées pour grands ensembles de données  
- 11:35 - 11:50 Bangxin Zhao & Wenqing He (Western University)  
Distance Techniques for High-dimensional Influence Measure / Techniques de distance pour mesure d'influence à dimensions élevées  

---

**10:20 - 11:50 Session 1B-C5** (abstracts/résumés - page 93) **Thistle 257**

**Survey Methodology 1**

**Méthodologie d'enquête 1**

Contributed Paper Session / Séance de communications libres

Chair/Président: John L. Eltinge (U.S. Bureau of Labor Statistics)

- 10:20 - 10:35 Jessica Andrews, Pierre Daoust & Matei Mireuta (Statistics Canada)  
Small Domains and Sparse Variables; Challenges with Creating an Active Collection Strategy / Petits domaines et variables rares; défis lors de la création d'une stratégie de collection active  
- 10:35 - 10:50 Susan Demedash (Statistics Canada), Matthew Buckley (Wilfred Laurier University)  
Job Vacancy and Wage Survey Paradata Log Files: A New Realm of Respondent Behavioural Analysis / Les fichiers de paradonnées de l'Enquête sur les postes vacants et les salaires : un nouveau monde d'analyse comportementale des répondants  
- 10:50 - 11:05 Michael Freiman & Amy Lauger (U.S. Census Bureau), Jerome Reiter (Duke University)  
Two Approaches to Disclosure Avoidance for a Microdata Analysis System at the U.S. Census Bureau / Deux approches pour prévenir la divulgation dans un système d'analyse des microdonnées au U.S. Census Bureau  
- 11:05 - 11:20 Zixin Nie, Kyle Vincent & Andee Cooper-Parks (International Justice Mission)  
Methodological Challenges Faced in the 2015-2016 Kolkata/Mumbai Commercial Sex Worker Study / Problèmes méthodologiques rencontrés dans l'étude 2015-2016 sur les travailleurs de l'industrie du sexe à Calcutta et à Mumbai  
- 11:20 - 11:35 Javier Oyarzun (Statistics Canada)  
Needle in a Haystack – How to Detect Break in Series: The Capital Expenditures Survey Integrated Business Statistics Program Experience / Une aiguille dans une botte de foin – La façon de détecter un bris dans une série : l'expérience de l'Enquête sur les dépenses en immobilisations dans le Programme intégré de la statistique des entreprises  
- 11:35 - 11:50 Qian Wei & J.N.K. Rao (Carleton University)  
Statistical Inference Based on Survey Weighted Composite Likelihood / Inférence statistique fondée sur la vraisemblance composite pondérée d'un sondage  

---

**12:00 - 13:30** **Thistle 242**

**Actuarial Science Section Executive Committee 1**

**Comité exécutif du groupe d'actuariat 1**

---

**12:00 - 18:00** **Thistle 254**

**Award for Case Studies in Data Analysis Committee**

**Comité du prix pour les études de cas et l'analyse de données**

<b>12:00 - 13:30</b> <b>Biostatistics Section Executive Committee 1</b> <b>Comité exécutif du groupe de biostatistique 1</b>	<b>Thistle 246</b>
<b>12:00 - 13:30</b> <b>Business and Industrial Statistics Section Executive Committee 1</b> <b>Comité exécutif du groupe de statistique industrielle et de gestion 1</b>	<b>Thistle 255</b>
<b>12:00 - 13:30</b> <b>Committee on Women in Statistics</b> <b>Comité sur les femmes en statistique</b>	<b>Thistle 258</b>
<b>12:00 - 13:30</b> <b>Finance Committee</b> <b>Comité des finances</b>	<b>Thistle 259</b>
<b>12:00 - 13:30</b> <b>Probability Section Executive Committee 1</b> <b>Comité exécutif du groupe de probabilité 1</b>	<b>Thistle 256</b>
<b>12:00 - 13:30</b> <b>Publications Committee</b> <b>Comité des publications</b>	<b>Welch 207</b>
<b>12:00 - 13:30</b> <b>Research Committee</b> <b>Comité de la recherche</b>	<b>Welch 209</b>
<b>12:00 - 13:30</b> <b>Statistical Education Section Executive Committee 1</b> <b>Comité exécutif du groupe d'éducation en statistique 1</b>	<b>Thistle 257</b>
<b>12:00 - 13:30</b> <b>Survey Methods Section Executive Committee Meeting 1</b> <b>Comité exécutif du groupe de méthodes d'enquête 1</b>	<b>Thistle 325</b>
<b>12:00 - 17:30</b> <b>Session 1C-D1</b> (description - page 96)	<b>Thistle Corridor</b>
<b>Case Study 1: Can Google Flu Trends Predict the Frequency and Results of Tests for Influenza and Other Respiratory Illnesses?</b>	
<b>Étude de cas 1 : Le site Google Flu Trends peut-il prédire la fréquence et le résultat des tests pour la grippe et autres maladies respiratoires ?</b>	
Case Study / Étude de cas	
Organizer and Chair / Responsable et président: Lisa Lix (University of Manitoba)	
(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)	
(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)	

12:00 - 17:30	Shomoita Alam, Shouao Wang & Zayd Omar (McGill University)
12:00 - 17:30	Shabnam Balamchi, Md. Erfanul Hoque, Faisal Atakora & Olawale Fatai Ayilara (University of Manitoba)
12:00 - 17:30	Zhoushanyue He & Danqiao Guo (University of Waterloo)
12:00 - 17:30	Yuanhao Lai, Tianpei Jiang & Samira Soleymani (University of Western Ontario)
12:00 - 17:30	Wei Lin, Hongyang Hu, Faizan Moshin, Ruoqi Yu & Yuxin Zou (University of Toronto)
12:00 - 17:30	Kuan Liu, Michael Moon, Kaviul Mohammad Khan & Sangook Kim (University of Toronto)
12:00 - 17:30	Lin Lu, Xu Zhao, Yishan Guo & Gemini Lam (University of Toronto)
12:00 - 17:30	Danny Tshitumbu, Hang Lai, Bisma Farooqi & Suj Sriskandarajah (York University)
12:00 - 17:30	Fahmida Yeasmin & Wenyan Zhong (University of Calgary)

---

**12:00 - 17:30    Session 1C-D2** (description - page 97) **Thistle Corridor**

**Case Study 2: What Predicts Sustainability of Canadian Charities?**

**Étude de cas 2 : Pouvons-nous prédire la viabilité des organismes de bienfaisance canadiens ?**

Case Study / Étude de cas

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

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)

(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

12:00 - 17:30	Isabelle Bernard, Samuel Labelle, Maikel Geagea & Jorge-Enrique Ruiz-Trujillo (HEC Montréal)
12:00 - 17:30	Bo Chen, Haosui Duanmu & Peiyang He (University of Toronto)
12:00 - 17:30	Huang Huang, Xiaoxiao Wang, Jin Ma, He Bian, Jingchun Pei, Yisen Lin & Banoo Madhanagopal (University of Toronto)
12:00 - 17:30	Wei Lin, Hongyang Hu, Faizan Moshin, Ruoqi Yu & Yuxin Zou (University of Toronto)
12:00 - 17:30	Tingbin Tan & Kexin Luo (University of Western Ontario)
12:00 - 17:30	Yanbo Tang, Yang Guan Jian Guo, Mufan Li, Gong Zhang & Harris Quach (University of Toronto)
12:00 - 17:30	Guanbo Wang, Alexander Levis & Mengian Pang (McGill University)
12:00 - 17:30	Fei Zuo, Xiaotian Wang, Jiajun Yan, Yao Xi & Jingjing Cao (University of Toronto)

---







**13:30 - 15:00    Session 1D-I1** (abstracts/résumés - page 98) **Welch - David Howes**

**Analysis of Big Data**

**Analyse des mégadonnées**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Francois Bellavance (HEC Montreal)

13:30 - 14:00	Luke Bornn (Simon Fraser University) From Pixels to Points: Using Tracking Data to Measure Performance in Professional Sports / Des pixels aux points : comment utiliser les données de poursuite pour mesurer la performance dans le sport professionnel  
14:00 - 14:30	Ershad Banijamali & Ali Ghodsi (University of Waterloo) Generative Mixture of Networks / Mélange de réseaux générateur  
14:30 - 15:00	Paul D. McNicholas (McMaster) Clustering Ultra High-Dimensional Data / Regroupement de données de très grande dimension  

---

---





**13:30 - 15:00**    **Session 1D-I2**    (abstracts/résumés - page 100)    **Thistle 325**

**New Perspectives in Causal Inference**

**Nouvelles perspectives en inférence causale**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Yeying Zhu (University of Waterloo)

- 13:30 - 14:00    Peisong Han (University of Waterloo), Lu Wang (University of Michigan)  
Multiply Robust Estimation in Causal Inference / Estimation multiplement robuste en inférence causale     
- 14:00 - 14:30    Wei Luo (Baruch College), Yeying Zhu (University of Waterloo), Debashis Ghosh (Colorado School of Public Health)  
On Estimating Regression-based Causal Effects Using Sufficient Dimension Reduction / Estimer les effets de causalité basés sur la régression en utilisant une réduction suffisante de la dimension     
- 14:30 - 15:00    Mireille Schnitzer (University of Montreal), Judith Lok & Susan Gruber (Harvard University), Mark van der Laan (University of California, Berkeley)  
Collaborative Double Robustness and the Connection to Data-Adaptive Nuisance Model Selection in Causal Inference / La double robustesse collaborative et le lien à la sélection de modèles de nuisance par algorithme d'apprentissage adaptatif en inférence causale     

---

**13:30 - 15:00**    **Session 1D-I3**    (abstracts/résumés - page 102)    **Thistle 258**







**Small Area Estimation: New Developments**

**Estimation pour petits domaines : nouveaux développements**

Invited Paper Session / Présentations sur invitation

Chair/Présidente: Karla Fox (Statistics Canada)

Organizer/Responsable: Mahmoud Torabi (University of Manitoba)

- 13:30 - 14:00    M. Giovanna Ranalli (University of Perugia, Italy), Gaia Bertarelli & Francesco Bartolucci (University of Perugia), Michele D'Alò & Fabrizio Solari (ISTAT)  
Time Series Small Area Estimation for Unemployment Rates using Latent Markov Models / Estimation pour petites régions de séries chronologiques relatives au taux de chômage à l'aide de modèles de Markov latents     
- 14:00 - 14:30    J.N.K. Rao (Carleton University)  
On Measuring Uncertainty of Small Area Estimators Under Area Level and Unit Level Models / Mesure de l'incertitude des estimateurs pour petites régions dans les modèles au niveau de la région et de l'unité     
- 14:30 - 15:00    Mahmoud Torabi (University of Manitoba)  
Prediction in Small-Area Spatially Correlated Data / Prédiction dans les données de petites régions spatialement corrélées     

---

**13:30 - 15:00**    **Session 1D-I4**    (abstracts/résumés - page 104)    **Welch 209**







**Statistical Modelling in Environmental and Health Studies**

**Modélisation statistique dans les études sur l'environnement et la santé**

Invited Paper Session / Présentations sur invitation

Chair/Président: Tim Swartz (Simon Fraser University)

Organizer/Responsable: Paramjit Gill (British Columbia-Okanogan)





- 13:30 - 14:00 Cindy Feng & Mehdi Rostamiforooshani (University of Saskatchewan)  
Modeling Spatially Correlated Survival Data: Impact of Misspecification of Correlation Structure on the Parameter Estimates / Modélisation de données de survie spatialement corrélées : impact des erreurs de spécification d'une structure de corrélation sur les estimations de paramètres  
- 14:00 - 14:30 Renjun Ma (University of New Brunswick)  
Mixed Models for Correlated Environmental/Health Data with Left Censoring and Right-Skewness / Modèles mixtes pour données environnementales/de santé corrélées avec censure à gauche et asymétrie à droite  
- 14:30 - 15:00 Gyanendra Pokharel & Rob Deardon (University of Calgary)  
Gaussian Process Emulator-Based Inference for Spatial Models Infectious Disease Systems / Inférence basée sur un émulateur à processus gaussien pour les modèles spatiaux de systèmes de maladies infectieuses  

---

**13:30 - 14:45 Session 1D-I5** (abstracts/résumés - page 106) **Thistle 246**

**Imaging Genomics - A New Frontier for Statistical Methodology**  
**Imagerie génomique – une nouvelle frontière pour la méthodologie statistique**

Organizer and Chair / Responsable et président: Linglong Kong (University of Alberta)

- 13:30 - 14:00 Michele Guindani (University of Texas)  
Bayesian Predictive Modeling for Imaging Genetics with Application to Schizophrenia / Modélisation prédictive bayésienne pour la génétique d'imagerie avec application à la schizophrénie  
- 14:00 - 14:45 Farouk Nathoo, Keelin Greenlaw & Mary Lesperance (University of Victoria), Elena Szefer & Jinko Graham (Simon Fraser University)  
A Bayesian Group Sparse Multi-Task Regression Model for Imaging Genomics / Modèle de régression multitâche bayésienne à groupes clairsemés en génomique de l'imagerie  

---

**13:30 - 14:56 Session 1D-I6** (description - page 108) **Thistle 242**

**Opportunities for statistical scientists in CIHR programs: a panel discussion**  
**Opportunités pour les statisticiens dans les programmes IRSC : table ronde**

Chair/Présidente: Mary Thompson (University of Waterloo)

- 13:30 - 13:50 Richard J. Cook (University of Waterloo)
- 13:50 - 14:12 Lisa Lix (University of Manitoba)
- 14:12 - 14:34 Robert W. Platt (McGill University)
- 14:34 - 14:56 Rhonda Rosychuk (University of Alberta)





---

**13:30 - 15:00 Session 1D-C1** (abstracts/résumés - page 109) **Thistle 257**

**Actuarial Science and Finance 1**  
**Actuariat et finance 1**

Contributed Paper Session / Séance de communications libres

Chair/Président: Étienne Marceau (Laval University)

- 13:30 - 13:45 Nicholas Beck (McGill University), Mélina Mailhot (Concordia University)  
A Consistent Estimator to the Orthant-Based Tail Value-at-Risk / Un estimateur convergent de l'aile de la valeur à risque en fonction de l'orthant  
- 13:45 - 14:00 Jean-François Bégin (HEC Montréal), Mathieu Boudreault (Université du Québec à Montréal), Delia Alexandra Doljanu & Geneviève Gauthier (HEC Montréal)  
Credit and Systemic Risks in the Financial Services Sector / Risques de crédit et systémique dans le secteur financier  



- 14:00 - 14:15 Lichen Chen, Adam Kolkiewicz & Tony Wirjanto (University of Waterloo)  
Bayesian Estimation of GARCH Models with General Parameter Constraints / Estimation bayésienne des modèles GARCH avec contraintes sur le paramètre général  
- 14:15 - 14:30 Wenjun Jiang & Jiandong Ren (Western University)  
Optimal Reinsurance Minimizing the Risk of Joint Party / Réassurance optimale minimisant le risque de la partie commune  
- 14:30 - 14:45 Luyao Lin, Li Chen, Yi Lu & Gary Parker (Simon Fraser University)  
Analysis of Survivorship Life Insurance Portfolio with Stochastic Interest Rates / Analyse d'un portefeuille d'assurance de survie avec taux d'intérêt stochastique  
- 14:45 - 15:00 Ying Wang & Jun Cai (University of Waterloo)  
Fair Premium for Reinsurance Contracts / Primes équitables des contrats de réassurance  

---










**13:30 - 15:00**    **Session 1D-C2**    (abstracts/résumés - page 112)    **Thistle 259**

**Biostatistics: Methods and Applications 2**

**Biostatistique : méthodes et applications 2**

Contributed Paper Session / Séance de communications libres

Chair/Président: Joseph Beyene (McMaster University)

- 13:30 - 13:45 Yang Jian, Wu Jingjing & Li Haocheng (University of Calgary)  
Minimum Hellinger Distance Estimation for Linear Regression Model / Méthode d'estimation de la distance minimale de Hellinger pour un modèle de régression linéaire  
- 13:45 - 14:00 Erin Lundy & Charmaine Dean (University of Western Ontario)  
Analyzing Heaped Counts and Longitudinal Presence/Absence Data in Joint Zero-inflated Poisson Regression Models / Analyse de données de dénombrement arrondies et de données longitudinales de présence et d'absence dans des modèles conjoints de régression de Poisson à surreprésentation de zéros  
- 14:00 - 14:15 Rachid Bentoumi & Mayer Alvo (Université d'Ottawa), Mhamed Mesfioui (Université du Québec à Trois-Rivières)  
Information Gain under Length-biased Sampling / Gain en information sous l'échantillonnage à biais de longueur  
- 14:15 - 14:30 Elnaz Ghadimi & Arusharka Sen (Concordia University)  
Multivariate Cure Rate Estimation under Random Censoring / Estimation multivariée du taux de guérison sous censure aléatoire  
- 14:30 - 14:45 Regina S. Kampo, Ashley Bonner & Joseph Beyene (McMaster University)  
Assessing the Influence of Non-adherence on Fixed-Effect Meta-Analysis for a Continuous Outcome: A Simulation Study / Évaluer l'influence de la non-adhérence sur la méta-analyse à effets fixes pour un résultat continu : une étude de simulation  
- 14:45 - 15:00 Yuying Xie, Yeying Zhu & Cecilia A. Cotton (University of Waterloo), Pan Wu (Value Institute, Christiana Care Health System)  
A Model Averaging Approach for Estimating Propensity Scores by Optimizing Balance / Approche par modèle moyen pour l'estimation de scores de propension en optimisant l'équilibre  

---











**13:30 - 15:00**    **Session 1D-C3**    (abstracts/résumés - page 115)    **Thistle 255**

**Statistical Methods and Applications 2**

**Méthodes statistiques et applications 2**

Contributed Paper Session / Séance de communications libres

Chair/Président: William Aeberhard (Dalhousie University)

- 13:30 - 13:45 Latifa Ben Hadj Slimene & Éric Marchand (Université de Sherbrooke), Theodoros Nicolieris (University of the Aegean, Greece)  
Inference of a Constrained Parameter in Presence of an Uncertain Constraint / Inférence d'une contrainte paramétrique en présence d'une contrainte incertaine  
- 14:00 - 14:15 Adam Rahman & Wayne Oldford (University of Waterloo)  
What Makes a Scatterplot Interesting? / Qu'est-ce qui rend un nuage de points pertinent?  
- 14:15 - 14:30 Yilei Wu, Yingli Qin & Mu Zhu (University of Waterloo)  
Large Covariance Matrix Estimation - A Factor Model Approach / Estimation d'une grande matrice de covariance – une approche par modèle factoriel  
- 14:30 - 14:45 Xin Liu & Wenqing He (Western University)  
Improve Performance of Support Vector Machine Classifiers with Data Adaptive Kernel / Amélioration de l'efficacité du classificateur d'une machine à vecteurs de support avec noyau adaptatif aux données  
- 14:45 - 15:00 Nathaniel Stevens (University of San Francisco), Christine M. Anderson-Cook (Los Alamos National Laboratory)  
Comparing the Reliability of Related Populations with the Probability of Agreement / Comparaison de la fiabilité des populations apparentées et de la probabilité d'accord  













**13:30 - 15:00 Session 1D-C4**

(abstracts/résumés - page 118)

**Welch 8G****Statistical Theory****Théorie statistique 1**

Contributed Paper Session / Séance de communications libres

Chair/Président: Jianfeng Yao (University of Hong Kong)

- 13:30 - 13:45 Haosui Duanmu & Daniel Roy (University of Toronto)  
Nonstandard Complete Class Theorem / Théorème de classe complète non standard  
- 13:45 - 14:00 Romain Kadje Kenmogne & François Perron (Université de Montréal)  
Density of the Ratio of Two Normal Random Variables / Densité du rapport de deux variables normalement distribuées  
- 14:00 - 14:15 Félix Camirand Lemyre & Taoufik Bouezmarni (Université de Sherbrooke), Jean-François Quesy (Université du Québec à Trois-Rivières)  
Nonparametric Measures of Local Causality and Tests of Non-Causality in Time Series / Mesures non paramétriques de la causalité locale et tests de non causalité pour des séries temporelles  
- 14:15 - 14:30 Victor Veitch & Daniel Roy (University of Toronto)  
Models and Inference for Sparse Random Graphs Using Exchangeable Random Measures / Modèles et inférence pour des graphes aléatoires épars à l'aide de mesures aléatoires échangeables  
- 14:30 - 14:45 Jun Yang & Daniel Roy (University of Toronto)  
Meta-Bayesian Analysis / Méta-analyse bayésienne  
- 14:45 - 15:00 Zhiyang Zhou (Simon Fraser University), Runchu Zhang (Nankai University and Northeast Normal University)  
A Generalized General Minimum Lower Order Confounding Criterion for Non-regular Designs / Critère généralisé de l'amalgame général minimal d'ordre inférieur pour des plans non réguliers  

---

**13:30 - 15:00**     **Session 1D-C5**     (abstracts/résumés - page 121)     **Thistle 256**

**Statistics Education****Éducation en statistique**

Contributed Paper Session / Séance de communications libres

Chair/Président: Alison Gibbs (University of Toronto)

- 13:30 - 13:45     Jeremy Balka (University of Guelph)  
A Traditional Introductory Statistics Course with Extensive Video Support / Un cours d'introduction aux statistiques traditionnelles avec un vaste support vidéo      
- 13:45 - 14:00     Dunham Bruce, Melissa Lee & Gaitri Yapa (University of British Columbia)  
Investigating How Students Interact with Simulation-based Applets / Étude de la manière dont les étudiants interagissent avec les applets basés sur la simulation      
- 14:00 - 14:15     Sohee Kang (University of Toronto Scarborough)  
Project-Based Statistics Learning Through Wikispace / Apprentissage des statistiques fondé sur le projet à l'aide de Wikispace      
- 14:15 - 14:30     Anne Michele Millar (Mount Saint Vincent University)  
P-values: Love them - Hate them - Leave them. Implications for Statistical Education. / Valeurs de p : aimez-les - détestez-les - laissez-les. Implications en éducation en statistique      
- 14:30 - 14:45     Lori Murray (University of Western Ontario)  
Five Years of Blending / Cinq années de mélange      
- 14:45 - 15:00     Nathan Taback & Alison Gibbs (University of Toronto)  
Students' Attitudes Towards Statistics in Online versus Flipped Classrooms / L'attitude des étudiants à l'égard des statistiques dans les cours en ligne par rapport aux classes inversées      

**14:45 - 15:30****Thistle Corridor across from Thistle 255****Coffee Break****Pause-café**


---







**15:30 - 17:00**     **Session 1E-I1**     (abstracts/résumés - page 124)     **Welch 209**

**Advances in Longitudinal Data Analysis****Progrès récents en analyse de données longitudinales**

Invited Paper Session / Présentations sur invitation

Chair/Président: Yogendra Chaubey (Concordia University)

Organizer/Responsable: Sanjoy Sinha (Carleton University)

- 15:30 - 16:00     Karelyn Davis (Health Canada)  
Longitudinal Analysis in Health Studies with Censored Observations / Analyse longitudinale d'études sur la santé avec observations censurées      
- 16:00 - 16:30     Abdus Sattar (Case Western Reserve University), Sanjoy Sinha (Carleton University)  
Joint Modeling of Longitudinal and Survival Data With a Covariate Subject to Limit of Detection / Modélisation conjointe de données longitudinales de survie avec une covariable soumise à une limite de détection      
- 16:30 - 17:00     Sanjoy Sinha (Carleton University)  
Methods for Longitudinal Data With Nonignorable Missing Responses and Outliers / Méthodes pour données longitudinales avec réponses manquantes et valeurs aberrantes ne pouvant être ignorées      
-

---







**15:30 - 17:00**    **Session 1E-I2**    (abstracts/résumés - page 126)    **Thistle 242**

**Advances in Regression Tree Modelling**

**Progrès récents en modélisation par les arbres de régression**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Matthew Pratola (Ohio State University)

- 15:30 - 16:00    Hugh Chipman (Acadia University), Robert McCulloch (University of Chicago), Matthew Pratola (Ohio State University), Edward George (University of Pennsylvania)  
Dispersion Modelling with an Ensemble of Trees / Modèle de dispersion avec ensemble d'arbres     
- 16:00 - 16:30    Tom Loughin & Andrew Henrey (Simon Fraser University)  
Robust Adaptively Pruned Random Forests Using Likelihood-Based Trees / Forêts aléatoires adaptatives élaguées et robustes utilisant des arbres fondés sur la vraisemblance     
- 16:30 - 17:00    Daniel Roy (University of Toronto)  
Mondrian Processes and their Statistical Applications / Les processus Mondrian et leurs applications statistiques     

---



**15:30 - 17:00**    **Session 1E-I3**    (abstract/résumé - page 128)    **Welch - David Howes**

**An Introduction to the Joint Modelling of Longitudinal and Survival Data**

**Introduction à la modélisation conjointe des données longitudinales et de survie**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Fernando Camacho (Damos Inc.)

- 15:30 - 17:00    Grace Y. Yi (University of Waterloo)  
Introduction of Joint Modeling of Longitudinal and Survival Data / Présentation d'une modélisation conjointe de données longitudinales et de survie     

---

**15:30 - 17:00**    **Session 1E-I4**    (abstracts/résumés - page 129)    **Thistle 258**







**Missing Data in Survey Sampling**

**Données manquantes en échantillonnage**

Invited Paper Session / Présentations sur invitation

Chair/Président: Michael Hidioglou (Statistics Canada)

Organizer/Responsable: Wes Yung (Statistics Canada)

- 15:30 - 16:00    Song Cai & J.N.K. Rao (Carleton University)  
Empirical Likelihood Inference Based on Estimation Equations for Complex Survey with Data Missing Completely At Random / Inférence de vraisemblance empirique fondée sur les équations d'estimation pour sondage complexe avec données manquantes aléatoirement     
- 16:00 - 16:30    David Haziza (Université de Montréal), Éric Lesage (INSEE)  
Properties of the Instrument Vector Calibration Estimator in the Presence of Unit Nonresponse / Propriétés de l'estimateur par calage instrumental en présence de non-réponse totale     
- 16:30 - 17:00    Xichen She & Changbao Wu (University of Waterloo)  
Fully Efficient Joint Fractional Imputation for Incomplete Bivariate Ordinal Responses / Imputation fractionnaire conjointe pleinement efficace pour des réponses bivariées ordinales incomplètes     
-







---

**15:30 - 17:00**     **Session 1E-I5**     (abstracts/résumés - page 131)     **Thistle 246**

**Probability and Statistical Problems and Results Arising from Neural Models**  
**Probabilités et problèmes statistiques et résultats découlant de modèles neuronaux**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Priscilla Greenwood (University of British Columbia)

- 15:30 - 16:00     Andre Longtin, Stephen Clarke & Len Maler (University of Ottawa)  
 Fisher Information Analysis of Neural Focussing Point Processes / Analyse d'information de  
 Fisher de processus de point focal neuronal      
- 16:00 - 16:30     Peter Rowat (University of California at San Diego), Priscilla (Cindy) Greenwood (University of  
 British Columbia)  
 Firing Patterns of a Stochastic Neural Model / Modes de décharge d'un modèle neuronal  
 stochastique      
- 16:30 - 17:00     Lawrence Ward & Priscilla (Cindy) Greenwood (University of British Columbia)  
 Quasicycles and Quasipatterns in the Brain? / Quasicycles et quasimotifs dans le cer-  
 veau?      



---

**15:30 - 17:00**     **Session 1E-I6**     (abstract/résumé - page 133)     **Thistle 325**

**Statistical Education Section Presidential Invited Address**  
**Allocution du l'invité du président du Groupe d'éducation en statistique**

Invited Paper Session / Présentations sur invitation

Chairs/Présidentes: Peggy Ng (York University) and/et Bethany White (University of Western Ontario)

- 15:30 - 17:00     David Spiegelhalter (Cambridge University)  
 The Ups and (Many) Downs of Trying to Be a 'Public Statistician' / Les hauts et les (nombreux)  
 bas de la vie d'un « statisticien public »      









---





**15:30 - 17:00**     **Session 1E-C1**     (abstracts/résumés - page 134)     **Thistle 255**

**Biostatistics: Causal Inference and Measurement Error**  
**Biostatistique : inférence causale et erreur de mesure**

Contributed Paper Session / Séance de communications libres

Chair/Président: Mireille Schnitzer (Université de Montréal)

- 15:30 - 15:45     Asma Bahamyirou, Mireille Schnitzer & Lucie Blais (Université de Montréal)  
 Comparison of Causal Inference Approaches for Estimating the Effect of Exposure to Inhaled  
 Corticosteroids on Mean Birth Weight and Gestational Age / Comparaison des approches  
 d'inférence causale pour l'estimation de l'effet de l'exposition aux corticostéroïdes inhalés sur  
 le poids moyen à la naissance et sur l'âge gestationnel      
- 15:45 - 16:00     Leila Golparvar, David A. Stephens & Robert W. Platt (McGill University)  
 Causal Structure Learning and Propensity Score Adjustment / Apprentissage de structure  
 causale et ajustement par score de propension      
- 16:00 - 16:15     Kamal Rai, Martin Lysy & Joel A. Dubin (University of Waterloo)  
 Bayesian Inference for Stochastic PK/PD Models / Inférence bayésienne pour les modèles  
 stochastiques pharmacocinétiques et pharmacodynamiques      
- 16:15 - 16:30     Di Shu & Grace Y. Yi (University of Waterloo)  
 Estimation of Causal Parameters in the Presence of Measurement Error / Estimation de pa-  
 ramètres causaux en présence d'erreur de mesure      

- 16:30 - 16:45 Denis Talbot (Université Laval), Amanda M. Rossi (McGill University), Simon L. Bacon (Concordia University), Juli Atherton & Geneviève Lefebvre (Université du Québec à Montréal)  
A Graphical Perspective of Marginal Structural Models when Estimating the Causal Relationships Between Physical Activity, Blood Pressure, and Mortality / Une perspective graphique aux modèles structureaux marginaux pour l'estimation des relations causales entre l'activité physique, la pression artérielle et la mortalité  
- 16:45 - 17:00 Yeying Zhu (University of Waterloo), Jennifer Savage (Pennsylvania State University), Debashis Ghosh (Colorado School of Public Health)  
A Kernel-Based Approach to Covariate Adjustment for Causal Inference / Méthode par noyaux pour l'ajustement de covariables à des fins d'inférence causale  

---

**15:30 - 16:45**      **Session 1E-C2**      (abstracts/résumés - page 137)      **Thistle 256**

**Biostatistics: Methodological Innovation 2**

**Biostatistique : innovation méthodologique 2**

Contributed Paper Session / Séance de communications libres

Chair/Président: Andrea Benedetti (McGill University)

- 15:30 - 15:45 Ashley Bonner & Joseph Beyene (McMaster University)  
Inference for Model Parameters in Sparse Canonical Correlation using Resampling Techniques / Inférence des paramètres des modèles pour la corrélation canonique rare à l'aide de techniques de rééchantillonnage  
- 15:45 - 16:00 Osvaldo Espin-Garcia (University of Toronto and Lunenfeld-Tanenbaum Research Institute), Radu V. Craiu (University of Toronto), Shelley B. Bull (University of Toronto and Lunenfeld-Tanenbaum Research Institute)  
Two-Phase Designs for Joint Quantitative-Trait-Dependent and GWAS-SNP-Dependent Sampling in Post-GWAS Regional Sequencing / Plans à deux phases pour trait quantitatif dépendant et échantillonnage GWAS-SNP conjoints dans le séquençage régional post-GWAS  
- 16:00 - 16:15 Jingxiong Xu (University of Toronto), Laurent Briollais (Lunenfeld-Tanenbaum Research Institute; University of Toronto), Wei Xu (Princess Margaret Cancer Centre: University of Toronto)  
Bayesian Statistical Approaches for Genetic Association with Next Generation Sequencing (NGS) Data / Approches statistiques bayésiennes pour l'association génétique avec des données de séquençage de prochaine génération (SPG)  
- 16:15 - 16:30 Qing Yu, Guohua Yan & Renjun Ma (University of New Brunswick)  
A Bayesian Poisson Mixed Modelling Approach to Zero-Inflated Cox Proportional Hazard Models / Approche bayésienne Poisson de modélisation mixte pour les modèles à risques proportionnels de Cox à surreprésentation de zéros  
- 16:30 - 16:45 Tian Feng, Joseph Beyene & Jemila S. Hamid (McMaster University)  
A Bayesian Method for Handling Missing Data in the Growth Curve Model / Une méthode bayésienne pour traiter les données manquantes dans le modèle de courbe de croissance  

---



**15:30 - 17:00**      **Session 1E-C3**      (abstracts/résumés - page 140)      **Thistle 259**











**Biostatistics: Survival and Current Status Data**

**Biostatistique : données de survie et d'état actuel**

Contributed Paper Session / Séance de communications libres

Chair/Président: Cindy Feng (University of Saskatchewan)

- 15:30 - 15:45 Longlong Huang & Xuewen Lu (University of Calgary), Karen Kopciuk (Alberta Health Services, University of Calgary)  
A Group Bridge Approach for Component Selection in Nonparametric Accelerated Failure Time Additive Regression Model / Une approche de pont groupé pour la sélection des composantes dans un modèle de régression additif non paramétrique du temps de défaillance accéléré  

- 15:45 - 16:00 Saima Khosa & Shahedul Khan (University of Saskatchewan)  
A Generalization of Log-logistic Distribution with Application in Survival Analysis / Une généralisation de la distribution log-logistique avec application à l'analyse de survie  
- 16:00 - 16:15 Shanshan Lu, Xuewen Lu & Jingjing Wu (University of Calgary)  
Exploring the Varying Covariate Effects in Partially Linear Proportional Odds Models with Current Status Data / Exploration des effets variables de covariables dans des modèles à cotes proportionnelles partiellement linéaires avec des données sur le statut actuel  
- 16:15 - 16:30 Yuan Dong & Xuewen Lu (University of Calgary), Radhey Singh (University of Waterloo)  
Efficient Estimation of Varying-Coefficient Partially Linear Proportional Hazards Models with Current Status Data / Estimation efficace des modèles de risques proportionnels à coefficients linéaires partiellement variables à l'aide de données sur l'état actuel  
- 16:30 - 16:45 Omidali Aghababaei Jazi, Masoud Asgharian & Abbas Khalili (McGill University)  
A Comparison of the Penalized Estimation Methods in the Cox Model for Right-censored Length-biased Data / Une comparaison des méthodes d'estimation pénalisées dans le modèle de Cox pour les données de longueur biaisée censurées à droite  
- 16:45 - 17:00 Dongdong Li & Joan Hu (Simon Fraser University)  
Modelling and Analysis of Bivariate Event-Times with Informative Censoring / Modélisation et analyse des temps d'événements bivariés avec censure de l'information  

**15:30 - 17:00 Session 1E-C4**

(abstracts/résumés - page 143)

**Welch 8G****Statistical Methods and Applications 3****Méthodes statistiques et applications 3**

Contributed Paper Session / Séance de communications libres

Chair/Président: Anand N. Vidyashankar (George Mason University)

- 15:30 - 15:45 Rojjar Haddadian, Liqun Wang & Yuliya Martsynyuk (University of Manitoba)  
Simulation-based Estimation in Regression Models with Ordinal Response Variable and Mis-measured Covariates / Estimation fondée sur la simulation dans les modèles de régression avec variable de réponse ordinale et covariables mesurées incorrectement  
- 15:45 - 16:00 Ren Mingchen, Alex de Leon & Yan Ying (University of Calgary)  
Likelihood Estimation of the Gaussian Copula Distribution for Mixed Data via PX-MCEM / Estimation de vraisemblance de la distribution de copules gaussiennes pour des données mixtes au moyen de l'algorithme EM de Monte-Carlo par expansion de paramètres  
- 16:00 - 16:15 Bouchra Nasri (INRS-ETE), Taoufik Bouezmarni (Université de Sherbrooke (UdeS)), André St-Hilaire (Institut national de recherche scientifique (INRS-ETE))  
Copula-Based Quantile Regression and Inference / Inférence pour un modèle de régression sur les quantiles basé sur des copules  
- 16:15 - 16:30 Monsur Chowdhury & Saumen Mandal (University of Manitoba)  
Optimal Designs for Maximum Likelihood Estimation / Conceptions optimales pour l'estimation du maximum de vraisemblance  
- 16:30 - 16:45 Maysum Panju (University of Waterloo)  
Topic Modelling in Natural Language Processing / Modélisation de sujet dans le traitement du langage naturel  
- 16:45 - 17:00 Mamadou Yauck & Louis-Paul Rivest (Laval University)  
Prediction of Population Sizes in Open-Population Capture-Recapture Models / Prédiction des tailles de populations dans les modèles de capture-recapture pour population ouverte  

---

**15:30 - 17:00**      **Session 1E-C5**      (abstracts/résumés - page 146)      **Thistle 257**

**Statistics in Action****La statistique en action**

Contributed Paper Session / Séance de communications libres

Chair/Président: Yi Yang (McGill University)

- 15:30 - 15:45      Devan Becker (University of Western Ontario), W. John Braun (University of British Columbia)  
Aerial Firefighting Fatigue Project / Projet sur la fatigue dans les services aériens de lutte  
contre les incendies       
- 15:45 - 16:00      Pierre-Jérôme Bergeron (PJB Consulting), Samuel Archibald (Université du Québec à Montréal)  
Valar Morghulis: A Survival Analysis of 'A Song of Ice and Fire' / Valar morghulis : une analyse  
de survie des livres du Trône de fer       
- 16:00 - 16:15      Debbie Dupuis (HEC Montreal), Marco Bee (University of Trento), Luca Trapin (IMT Institute for  
Advanced Studies Lucca)  
Realized Peaks over Threshold / «Realized Peaks over Threshold»       
- 16:15 - 16:30      Nicholas Mitsakakis (University of Toronto), Karen Bremner (University Health Network), Murray  
Krahn (THETA Collaborative)  
Investigating Structural Independence in Health Utility Instruments Using Graphical Models /  
Étude de l'indépendance structurelle des instruments d'utilité de santé au moyen de modèles  
graphiques       
- 16:30 - 16:45      David Riegert, David Thomson & Aaron Springford (Queen's University)  
Statistical Estimation of Transfer Functions to Examine the Relationship Between Geomag-  
netism and Power Systems / Estimation statistique des fonctions de transfert pour examiner  
les relations entre le géomagnétisme et les réseaux électriques       
- 16:45 - 17:00      Pavel Slavchev & Russell Steele (McGill University), Ian Shrier (Lady Davis Institute, Jewish Gen-  
eral Hospital and McGill University), Ashley Naimi (University of Pittsburgh), Nathalie Auger  
(Institut National de la Santé Public du Québec and University of Montreal Research Centre)  
The Analysis of Live Births as a Competing Event to Stillbirths / L'analyse des naissances  
vivantes comme événement concurrent aux accouchements de mort-nés       

---

**17:00 - 18:00**      **Thistle 242**

**SSC Annual General Meeting****Assemblée générale annuelle de la SSC**


---

**18:30 - 20:30**      **Pond Inlet**

**Student BBQ****Barbecue étudiant**


---

**19:30**      **Kelsey's Bar & Grill, 221 Glendale Ave., St. Catharines**

**Informal Gathering of New Investigators****Rassemblement informel des nouveaux chercheurs**



**Tuesday May 31****mardi 31 mai****7:30 - 8:30****Thistle 255**

**Statistics Education Committee**  
**Comité d'éducation en statistique**

**8:15 - 17:00****Thistle Corridor across from Sean O'Sullivan Theatre**

**Registration**  
**Inscription**

**8:45 - 9:50****Session 2A-A**



(abstract/résumé - page 149)

**Welch - David Howes****Gold Medal Address****Allocution du récipiendaire de la Médaille d'or**

Special Session / Conférence spéciale

8:45 - 9:50

Richard Lockhart (Simon Fraser University)

Big Data, High Dimensions, Goodness-of-fit: Something of This Nature / Données volumi-  
 neuses, grandes dimensions et adéquation : quelque chose de cette nature  

**9:00 - 13:30****Thistle 254****Student Research Presentation Award Committee 1****Comité du prix pour la présentation de recherche étudiante 1****9:50 - 10:30****Thistle Corridor across from Thistle 255**

**Coffee Break**  
**Pause-café**

**10:20 - 11:50****Session 2B-I1**

(abstracts/résumés - page 150)



**Welch - David Howes****Advanced Statistical and Machine Learning Techniques for High Dimensional Data Analysis****Techniques statistiques et apprentissage machine avancés pour l'analyse de données en haute dimension**

Invited Paper Session / Présentations sur invitation

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



10:20 - 10:50

Ryan Tibshirani (Carnegie Mellon University), Richard Lockhart (Simon Fraser University),  
 Jonathan Taylor & Robert Tibshirani (Stanford University), Will Fithian (University of California,  
 Berkeley), Jing Lei & Larry Wasserman (Carnegie Mellon University)

Recent Advances in Selective Inference / Avancées récentes en inférence sélective  



10:50 - 11:20

Anand N. Vidyshankar (George Mason University)

Efficiency Considerations in Post Model Selection Inference / Considérations d'efficacité dans  
 l'inférence post-sélection de modèle  

11:20 - 11:50

Yi Yang (McGill University), Hui Zou (University of Minnesota), Wei Qian (Rochester Institute of  
 Technology)

Variable Selection Deviation for High-dimensional Classification Methods / Déviation de la  
 sélection des variables pour les méthodes de classification à grande dimension  

---

**10:20 - 11:50**      **Session 2B-I2**      (abstracts/résumés - page 152)      **Thistle 242**

**Advancements to State-Space Models for Fisheries Science**

**Progrès récents dans les modèles à espace d'états pour les sciences halieutiques**

Invited Paper Session / Présentations sur invitation

Chair/Président: Rick Routledge (Simon Fraser University)

Organizer/Responsable: Joanna Mills-Flemming (Dalhousie University)

- 10:20 - 10:50      William Aeberhard & Joanna Mills Flemming (Dalhousie University), Eva Cantoni (University of Geneva), Chris Field (Dalhousie University), Ximing Xu (Nankai University)  
Robust and Consistent Estimation of Fixed Parameters in General State-Space Models / Estimation robuste et convergente des paramètres fixes de modèles espace-états       
- 10:50 - 11:20      David Campbell (Simon Fraser University)  
Nested Laplace Approximation Diagnostics and Fixes / Diagnostic et correction de l'approximation de Laplace emboîtée       
- 11:20 - 11:50      Joanna Mills Flemming (Dalhousie University)  
Addressing Statistical Challenges in Fisheries Science / Défis statistiques en sciences halieutiques       

---

**10:20 - 11:46**      **Session 2B-I3**      (abstracts/résumés - page 154)      **Academic South 217**









**Contributions of Professor Ian B. MacNeill to Statistical Theory, Methodology, Teaching and Practice**

**Contributions du professeur Ian B. MacNeill à la théorie statistique, Méthodologie, enseignement et la pratique**

Invited Paper Session / Présentations sur invitation

Chair/Président: Gary Umphrey (University of Guelph)

Organizer/Responsable: Krishna Jandhyala (Washington State University)

- 10:20 - 10:42      Abdel El-Shaarawi (The American University in Cairo), Sylvia R. Esterby (University of British Columbia Okanagan)  
Ian B. MacNeill and Environmetrics / Ian B. MacNeill et l'environnementrie       
- 10:40 - 11:02      Elena Naumova (Tufts University)  
Seasonality Analysis in Public Health Applications / Analyse de saisonnalité dans la santé publique       
- 11:02 - 11:24      Serge Provost (University of Western Ontario)  
On the q-Analogues of the Logistic and Generalized Extreme Value Distributions / Les analogues q des lois de valeurs extrêmes généralisées et logistiques       
- 11:24 - 11:46      Venkata Krishna Jandhyala (Washington State University), Shelemyahu Zacks (State University of New York at Binghamton)  
Contributions of Dr. MacNeill to Change-Point Methodology and its Environmental Applications / Les contributions du Dr MacNeill à la méthodologie des points de rupture et ses applications environnementales       

---

**10:20 - 12:05**      **Session 2B-I4**      (abstracts/résumés - page 156)      **Thistle 246**

**Graduate Students in Actuarial Science**

**Étudiants de cycles supérieurs en actuariat**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Andrei Badescu (University of Toronto)

- 10:20 - 10:35      Nicolas Essis-Breton (University of Concordia), Patrice Gaillardetz (Concordia University)  
Multi Trial Monte Carlo / Essais multiples de Monte-Carlo       







- 10:35 - 10:50 Dongchen Li (University of Waterloo), Xinfu Chen (University of Pittsburgh), David Landriault & Bin Li (University of Waterloo)  
On Minimizing Drawdown Risks of Lifetime Investments / Minimisation des risques de prélèvement sur les investissements à vie  
- 10:50 - 11:05 Haiyan Liu, Jun Cai & Ruodu Wang (University of Waterloo)  
Asymptotic Equivalence of Risk Measures under Dependence Uncertainty / Équivalence asymptotique des mesures de risque en cas d'incertitude de la dépendance  
- 11:20 - 11:35 Jingong Zhang, Ken Seng Tan & Chengguo Weng (University of Waterloo)  
Optimal Hedging with Basis Risk under Mean-Variance Criteria / Couverture optimale avec risque de base sous critères de moyenne-variance  
- 11:35 - 11:50 Ruixi Zhang (Western University), Kristina Sendova (The University of Western Ontario)  
On a Renewal Risk Process under a Dividend Barrier Strategy / Processus de risque de renouvellement sous une stratégie de barrière de dividendes  
- 11:50 - 12:05 Mohamed Amine Lkabous (Université du Québec à Montréal), Irmina Czarna (University of Wrocław), Jean-François Renaud (Université du Québec à Montréal)  
Parisian ruin for a refracted Lévy process / Ruine Parisienne pour un processus de Lévy réfracté  

---

**10:20 - 11:50 Session 2B-I5** (abstracts/résumés - page 159) **Thistle 243**  
**Recent Developments in Designs of Statistical Experiments. In honour of the 65th birthday of Professor Douglas P. Wiens**  
**Progrès récents en planification d'expériences. En l'honneur du 65e anniversaire du professeur Douglas P. Wiens**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Xiaojian Xu (Brock University)



- 10:20 - 10:50 Holger Dette (Ruhr-Universitaet Bochum), Kirsten Schorning & Kathrin Möllenhoff (Ruhr-Universität Bochum), Stanislav Volgushev (Cornell University), Frank Bretz (Novartis, Basel)  
Statistical Methodology for Comparing Curves / Méthodologie statistique pour la comparaison de courbes  
- 10:50 - 11:20 Linglong Kong & Doug Wiens (University of Alberta)  
Robust Designs for Nonlinear Quantile Regression / Plans robustes pour la régression quantile non linéaire  
- 11:20 - 11:50 Julie Zhou (University of Victoria)  
Computing Optimal Designs via Convex Optimization / Calculer des plans optimaux par l'optimisation convexe  

---

**10:20 - 11:50 Session 2B-C1** (abstracts/résumés - page 161) **Thistle 255**  
**Biostatistics: Applications**  
**Biostatistique : applications**

Contributed Paper Session / Séance de communications libres

Chair/Président: Lehana Thabane (McMaster University)

- 10:20 - 10:35 Mohammad Ehsanul Karim & Robert W. Platt (McGill University), Helen Tremlett & BeAMS Study group (University of British Columbia)  
Estimating Inverse Probability Weights using Super Learner when Weight-Model Specification is Unknown in a Marginal Structural Cox Model Context: An Application to Multiple Sclerosis / Estimation de poids de probabilité inverse à l'aide du super apprentissage lorsque la spécification des poids du modèle est inconnue dans un modèle de Cox structurel marginal Contexte : Une application à la sclérose en plaques  

- 10:35 - 10:50 Alexander Levis (McGill University), Daphna Harel (New York University), Linda Kwakkenbos (Radboud University), Brett Thombs (McGill University)  
Using Optimal Test Assembly Methods for Shortening Patient-Reported Outcome Measures: Development and Validation of the Cochin Hand Function Scale-6 / Utilisation de méthodes optimales d'ensembles de tests pour réduire le nombre de mesures de résultats rapportées par les patients : mise au point et validation de l'échelle 6 de l'indice fonctionnel de la main de Cochin  
- 10:50 - 11:05 Lin Liu (University of California, San Diego), Baron John (University of North Carolina), Lieberman David (Veteran Affairs Medical Center and Oregon Health and Science University), Jacobs Elizabeth (University of Arizona), Cross Amanda & Murphy Gwen (National Cancer Institute, NIH), Samir Gupta (University of California, San Diego)  
Statistical Issues in a Prognostic Model for Advanced Colorectal Neoplasia Recurrence / Problèmes statistiques dans un modèle de pronostique pour la récurrence de la néoplasie colorectale avancée  
- 11:05 - 11:20 Alexandre Piché & Russell Steele (McGill University), Ian Shrier (Center for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital), Niels Wedderkopp (Institute of Regional Health Research, University of Southern Denmark)  
Nested Dirichlet Process: an Application to Censored Data / Processus emboîté de Dirichlet : application à des données censurées  
- 11:20 - 11:35 Razvan Romanescu (University of Guelph), Rob Deardon (University of Calgary)  
Optimal Surveillance of Epidemics over Power Law Networks / Surveillance optimale des épidémies sur les réseaux de la loi de puissance  
- 11:35 - 11:50 Krishna Saha (Central CT State University)  
Estimating the Common Intra-class Correlation Coefficient for Clustered Binary Responses / Estimation du coefficient de corrélation intra-classe commun pour des réponses binaires groupées  

---

10:20 - 11:50 **Session 2B-C2** (abstracts/résumés - page 165)











**Thistle 256**



**Methodological Advances: A Smörgåsbord**

**Progrès méthodologiques : un buffet abondant**

Contributed Paper Session / Séance de communications libres

Chair/Président: Yves Berger (University of Southampton)

- 10:20 - 10:35 Harlan Campbell (University of British Columbia)  
A Non-parametric Rank Based Test for Semicontinuous Longitudinal Data / Un test non paramétrique basé sur les rangs pour les données longitudinales semi-continues  
- 10:35 - 10:50 Julan Al-Yassin, Abdulkadir Hussein & Tinglei Chen (University of Windsor)  
The GLARMA Model for Zero-Inflated Poisson Counts / Le modèle GLARMA pour les dénombrements Poisson à surreprésentation de zéros  
- 10:50 - 11:05 Fang He (University of Western Ontario)  
Carbon Dioxide Concentration Estimation / Estimation de la concentration de dioxyde de carbone  
- 11:05 - 11:20 Yuhong Wei & Paul D. McNicholas (McMaster University)  
Robust Mixture Modelling Based on Generalized Hyperbolic Distribution and its Limiting Case Multivariate Skew-t Distribution with Incomplete Data / Modélisation robuste par mélange pour données incomplètes fondée sur une distribution hyperbolique généralisée et son cas limite en distribution t-asymétrique multivariée  
- 11:20 - 11:35 Yoshihiro Taniguchi & Christiane Lemieux (University of Waterloo)  
Importance Sampling Techniques for Archimedean Copula Models / Techniques d'échantillonnage d'importance pour les modèles de copules archimédiennes  

- 11:35 - 11:50 Patricia Judd & Yildiz Yilmaz (Memorial University of Newfoundland)  
Two-Phase Response-Dependant Sampling Designs for Time-to-Event Analysis / Plans d'échantillonnage à deux phases dépendant de la réponse dans l'analyse de temps d'événement  











**10:20 - 12:05 Session 2B-C3** (abstracts/résumés - page 168) **Thistle 259**

**Mixture models**

**Modèles de mélange**

Contributed Paper Session / Séance de communications libres

Chair/Président: Hugh Chipman (Acadia University)

- 10:20 - 10:50 Martin Blostein (McMaster University), Antonio Punzo (University of Catania), Paul D. McNicholas (McMaster University)  
Robust High-Dimensional Modeling with the Contaminated Gaussian Distribution / Modélisation robuste de grande dimension avec une distribution gaussienne contaminée  
- 10:50 - 11:05 Michael Patrick Brian Gallagher & Paul D. McNicholas (McMaster University)  
Extending Fractionally Supervised Classification to Non-Gaussian Mixture Models / Étendre la classification fractionnaire supervisée aux modèles de mélanges non-gaussiens  
- 11:05 - 11:20 Yang Tang (McMaster University), Antonio Punzo (University of Catania), Paul D. McNicholas (McMaster University)  
Model-Based Clustering of Categorical Data with Extreme Patterns / Mise en grappes fondée sur un modèle de données catégoriques avec motifs extrêmes  
- 11:35 - 11:50 Utkarsh J. Dang (McMaster University), Ryan Browne (University of Waterloo), Paul D. McNicholas (McMaster University)  
Power Exponential Mixtures and Skewed Extensions / Mélanges de puissance exponentielle et extensions biaisées  
- 11:50 - 12:05 Anjali Silva, Steven Rothstein & Sanjeena Dang (University of Guelph)  
Mixture Model Selection for Cluster Analysis of RNA Sequencing Data / Sélection de modèles de mélange pour l'analyse du regroupement des données de séquençage ARN  









**10:20 - 11:50 Session 2B-C4** (abstracts/résumés - page 171) **Thistle 257**





**Multivariate Data**

**Données multivariées**

Contributed Paper Session / Séance de communications libres

Chair/Président: Lisa Lix (University of Manitoba)

- 10:20 - 10:35 Brian Franczak & Paul D. McNicholas (McMaster University)  
Discriminant Analysis via Parsimonious Multiple Scaled Mixtures / Analyse discriminante à l'aide de mélanges parcimonieux à multiples échelles  
- 10:35 - 10:50 Haocheng Li & Yukun Zhang (University of Calgary), Sarah Kozey-Keadle, Joshua Sampson & Charles Matthews (National Cancer Institute), Raymond Carroll (Texas A&M University)  
Multivariate Longitudinal Data with Mixed Types of Measurements / Données longitudinales multivariées avec types de mesures mixtes  
- 10:50 - 11:05 Yubin (Éric) Li (University of Windsor), Sévérien Nkurunziza (University of Windsor/Université de Sherbrooke)  
Estimation in Multivariate Regression with Measurement Error and Change-Points / Estimation en régression multivariée avec erreur de mesure et points de rupture  
- 11:05 - 11:20 Makosso-Kallyth Sun (McMaster University)  
Dimension Reduction of Histogram Data / Réduction de la dimension de variables de type histogramme  

- 11:20 - 11:35 Irene Vrbik (McGill University)  
Cluster Analysis of Genetic Sequence Data via the Gap Procedure / Analyse de mise en grappes de données de séquence génétique en utilisant le procédé des écarts  
- 11:35 - 11:50 Yang Zhao (University of Regina)  
Classification and Factor Analysis for Big Data Inference: A Study of Gambling Behaviours and the Associate Factors / Classification et analyse factorielle pour l'inférence de données massives : une étude des comportements de joueurs et des facteurs associés  

---











**10:20 - 11:35 Session 2B-C5** (abstracts/résumés - page 174) **Thistle 258**

**Survey Methodology 2**

**Méthodologie d'enquête 2**

Contributed Paper Session / Séance de communications libres

Chair/Président: Mahmoud Torabi (University of Manitoba)

- 10:20 - 10:35 Hyukjun (Jay) Gweon & Matthias Schonlau (University of Waterloo), Lars Kaczmirek & Michael Blohm (GESIS)  
New Approaches to Automated Occupation Coding Using Statistical Learning / Nouvelles approches pour le codage automatisé de la profession à l'aide de l'apprentissage statistique  
- 10:35 - 10:50 Ibrahima Ousmane Ida & Louis-Paul Rivest (Laval University)  
Balanced Sampling by Using the Cube Method and the Rejective Algorithm / L'échantillonnage équilibré par la méthode du cube et la méthode réjective  
- 10:50 - 11:05 Isabelle Lefebvre & David Haziza (Université de Montréal), Jean-François Beaumont (Statistique Canada)  
Simplified Variance Estimation for Complex Designs / Estimation simplifiée de la variance pour des plans complexes  
- 11:05 - 11:20 Thuva Vanniyasingam, Chuck E Cunningham, Gary Foster & Lehana Thabane (McMaster University)  
Determining the Impact of Different Design Features on Relative Design Efficiency in Discrete Choice Experiments / Déterminer l'effet de différents aspects d'un plan sur l'efficacité relative du plan dans les expériences avec choix discrets  
- 11:20 - 11:35 Wei Lin & Nancy Reid (University of Toronto), Karla Fox (Statistics Canada)  
Analysis of an Embedded Experiment in a Survey / Analyse d'une expérience intégrée dans le cadre d'une enquête  

---

**12:00 - 13:00**  
**Maplesoft Demo**

**Academic South 217**

---

**12:00 - 13:30**  
**Accreditation General Meeting**  
**Assemblée générale des membres accrédités**

**Thistle 246**

---

**12:00 - 13:30**  
**CJS Editorial Board**  
**Comité de rédaction de la RCS**

**Thistle 242**

---

<b>12:00 - 13:30</b>			<b>Thistle 255</b>
<b>Census at School Committee</b> <b>Comité de recensement à l'école</b>			
<b>12:00 - 13:30</b>			<b>Thistle 256</b>
<b>Committee on Membership</b> <b>Comité de recrutement</b>			
<b>12:00 - 13:30</b>			<b>Thistle 257</b>
<b>Committee on New Investigators</b> <b>Comité des nouveaux chercheurs</b>			
<b>12:00 - 13:30</b>			<b>Thistle 259</b>
<b>Office Committee</b> <b>Comité du Bureau</b>			
<b>12:00 - 13:30</b>			<b>Thistle 258</b>
<b>Public Relations Committee</b> <b>Comité des relations publiques</b>			
<b>12:00 - 17:30</b>	<b>Session 2C-P</b>	(abstracts/résumés - page 177)	<b>Thistle Corridor</b>
<b>Poster Session</b> <b>Séance d'affichage</b>			
Contributed Poster Session / Séance d'affichage libre (Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30) (Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)			
12:00 - 17:30	Huda Al-Wahsh, Abdulkadir Hussein & Weiwei Liu (University of Windsor) Scan Statistics in Cox's Model with Frailty / Statistiques de balayage dans le modèle de Cox avec fragilité		
12:00 - 17:30	Justin Angevaare & Zeny Feng (University of Guelph), Rob Deardon (University of Calgary) Phylodynamic Individual Level Models: Strategies for Simulation and Inference / Modèles phylodynamiques individuels : des stratégies de simulation et d'inférence		
12:00 - 17:30	Faisal Atakora & Po Yang (University of Manitoba) Construction of Optimal Foldover Designs with the General Minimum Lower-Order Confounding / Élaboration de plans de repliement optimal avec le critère de l'amalgame général minimal d'ordre inférieur		
12:00 - 17:30	Andrea Benedetti, Muhammad Mullah, Russell Steele & James Hanley (McGill University) Evaluation of Estimation Techniques for Smoothing via Generalized Linear Mixed Models / Évaluation des techniques d'estimation pour le lissage à l'aide de modèles mixtes linéaires généralisés		
12:00 - 17:30	Shelley B. Bull (University of Toronto), Jihyung Shin & Ruiyang Yi (Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital) Bootstrap Resampling in Genetic Association Analysis of Low-Frequency Variants / Rééchantillonnage bootstrap pour une analyse d'association génétique avec variantes peu fréquentes		

- 12:00 - 17:30 Bo Chen & Andrea Benedetti (McGill University)  
Quantifying Heterogeneity in Individual Participant Data Meta-Analysis with Binary Outcomes / Quantification de l'hétérogénéité dans la méta-analyse de données relatives au participant avec réponses binaires
- 12:00 - 17:30 Jonathan Earl & Paul D. McNicholas (McMaster University)  
Ranking and Tanking: GLMs for Football / Classement et sous-performance délibérée : une approche MLG au football
- 12:00 - 17:30 Yuanhao Lai & McLeod A. Ian (Western University)  
Computing Numerical Distribution Functions for Periodicity Tests in Microarray Time Series / Calcul numérique de fonctions de répartition pour des tests de périodicité dans des séries chronologiques en biopuce
- 12:00 - 17:30 Sean McGrath (McGill University), Xiaofei Zhao (RI MUHC), Andrea Benedetti (McGill University)  
Meta Analysis of Medians / Méta-analyse des valeurs médianes
- 12:00 - 17:30 Paul Nguyen & Lennon Li (Public Health Ontario), Patrick Brown (Cancer Care Ontario), Steven Johnson & John McLaughlin (Public Health Ontario)  
Spatial Cluster Detection for Ontario Public Health Research / Détection de grappes spatiales aux fins d'une recherche de Santé publique Ontario
- 12:00 - 17:30 Nidhi Patel, Utkarsh J. Dang & Paul D. McNicholas (McMaster University)  
Longitudinal Clustering via Mixtures of Multivariate Power Exponential Distributions / Mise en grappes longitudinale par mélanges de lois de puissance exponentielles multivariées
- 12:00 - 17:30 Reuben Pereira (University of Toronto), Patrick Brown (Cancer Care Ontario), Paul Nguyen (Public Health Ontario)  
Comparing Spatial Models for Mapping Larynx Cancer: A Simulation Study / Comparaison de modèles spatiaux pour cartographier le cancer du larynx : une étude de simulation
- 12:00 - 17:30 Gregory Pond (McMaster University), Cecilia A. Cotton & Richard J. Cook (University of Waterloo)  
A Biostatistics Training Initiative for Cancer Research / Projet de formation en biostatistique pour la recherche sur le cancer
- 12:00 - 17:30 Anna Theresa Santiago, Malcolm Binns & Wendy Lou (University of Toronto)  
Application of Dimension Reduction to Acquired Brain Injury Assessment / Application de la réduction dimensionnelle à l'évaluation de la lésion cérébrale acquise
- 12:00 - 17:30 Samira Soleymani & Ian McLeod (Western University)  
Simulating with Box-Cox transformations / Étude de simulation avec transformations de Box-Cox
- 12:00 - 17:30 Mohsen Soltanifar, Armend Lokku, Shahriar Shams & Paul Corey (University of Toronto)  
Factors Associated to Brain Reaction Time: A Higher Ordered Complete Random Block Design Study / Facteurs associés au temps de réaction du cerveau : étude d'un plan en blocs aléatoires complet d'ordre supérieur
- 12:00 - 17:30 Aaron Springford & David Thomson (Queen's University)  
Predicting Long-Term Missing Ground-Level Ozone Using Precursor, Solar and Climate Data / Prévoir l'absence à long terme de l'ozone au sol à l'aide de données de précurseurs, solaires et climatiques
- 12:00 - 17:30 Ian Waudby-Smith (Cancer Care Ontario; University of Waterloo), Zhihui (Amy) Liu (Cancer Care Ontario), Olli Saarela (University of Toronto)  
Multi-state Models for Chronic Kidney Disease Prevalence Projections in Ontario / Modèles multi-états pour des prévisions sur la prévalence de la néphropathie chronique en Ontario
- 12:00 - 17:30 Da Zhong (Dexen) Xi (Western University), Steve Taylor (Pacific Forestry Centre)  
Joint Modeling of Duration and Burn Area of Lightning Caused Forest Fires / Modélisation conjointe de la durée et de l'étendue des feux de forêt causés par les éclairs
- 12:00 - 17:30 Xu Xuan Zhang & Paul D. McNicholas (McMaster University)  
Non-Gaussian Mixture Model Averaging for Clustering / Approche par modèle moyen de mélange non gaussien pour la répartition en grappes
-



**13:30 - 18:00****Mackenzie Chown A323****Student Research Presentation Award Committee 2****Comité du prix pour la présentation de recherche étudiante 2****13:30 - 15:00 Session 2D-I1**

(description - page 186)

**Academic South 217****Developing Curriculum Guidelines for Canadian Statistics Undergraduate Programs****Comment mettre au point des directives pour les programmes de premier cycle en statistique au Canada**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Bethany White (University of Western Ontario)

13:30 - 14:15 Bethany White (The University of Western Ontario)

14:15 - 15:00 Jim Stallard (University of Calgary)

**13:30 - 15:00 Session 2D-I2**



(abstracts/résumés - page 187)

**Welch - David Howes****Genomic Data Analysis, Modelling and Testing****Analyse, modélisation et tests pour des données de séquençage génomique**



Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Brad McNeney (Simon Fraser University)



13:30 - 14:00 Ali Karimnezhad &amp; David R. Bickel (University of Ottawa)

Robust Bayesian Multiple Hypothesis Testing with Application to Genetic Association Data / Test d'hypothèses multiples bayésien robuste avec applications aux données d'association génétiques  

14:00 - 14:30 David A. Stephens, Farhad Shokoohi &amp; Aurélie Labbe (McGill University)

Hidden Markov Models for Identifying Differentially Methylated Regions / Modèles de Markov cachés pour l'identification des régions différenciellement méthylées  

14:30 - 15:00 Elena Szefer (Simon Fraser University &amp; The EMMES Corporation), Jinko Graham, Faisal Beg &amp; Donghuan Lu (Simon Fraser University), Farouk Nathoo (University of Victoria)

A Multivariate Imaging Genomics Application in Alzheimer's Disease / Une application d'imagerie génomique multivariée dans les cas de la maladie d'Alzheimer  **13:30 - 15:00 Session 2D-I3**



(abstracts/résumés - page 189)

**Thistle 242****Long Memory Time Series****Séries chronologiques à mémoire longue**



Invited Paper Session / Présentations sur invitation

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



13:30 - 14:00 Clifford M. Hurvich, Rohit Deo &amp; Jun Liu (New York University)

The Slow Convergence of OLS Estimators of Alpha, Beta, and Portfolio Weights Under Long-Memory Stochastic Volatility / La lente convergence des estimateurs OLS d'alpha, de beta et des poids de portfolio selon la volatilité stochastique à mémoire longue  

14:00 - 14:30 Valderio Reisen (University of Espirito Santo), Celine Levy-Leduc (AgroParisTech), Fabio Fajardo (UFES), Murad Taqqu (Boston University)

Long-memory Processes. Robust Spectral Estimators Based on Frequency Domain / Processus à mémoire longue. Estimateurs spectraux robustes fondés sur un domaine de fréquence  

14:30 - 15:00 Aera Thavaneswaren (University of Manitoba)

Nonlinear and Long Memory Count Time Series Models and Inference Using Estimating Functions / Modèles de séries chronologiques de dénombrement non linéaires et à mémoire longue et inférence utilisant des fonctions d'estimation  







---

**13:30 - 15:00**     **Session 2D-I4**     (abstracts/résumés - page 191)     **Thistle 247**

**Methods in Survival Analysis Motivated by challenges in clinical data**  
**Méthodes d'analyse de survie motivées par les problèmes de données cliniques**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Lisa Le (UHN Research)

- 13:30 - 14:00     Audrey Boruvka (University of Michigan), Yidan Shi, Leilei Zeng & Richard J. Cook (University of Waterloo)  
 Analysis of Clinical Trial Endpoints under Dual Censoring / Analyse de résultats d'essais cliniques sous double censure      
- 14:00 - 14:30     Melania Pintilie (Princess Margaret Cancer Centre), Harmke Groot & Michael Schaapveld (The Netherlands Cancer Institute)  
 Power for a Case-cohort Study in the Presence of Competing Risks / Puissance d'une étude de cas-cohorte en la présence de risques concurrents      
- 14:30 - 15:00     Eleanor Pullenayegum (The Hospital for Sick Children)  
 Using Recurrent Event Models to Handle Longitudinal Data Subject to Irregular Follow-Up / Analyse de données longitudinales à suivi irrégulier par modèles d'événements récurrents      





---

**13:30 - 15:00**     **Session 2D-I5**     (abstracts/résumés - page 193)     **Thistle 246**

**On Some Applications of James-Stein Shrinkage Estimation**  
**À propos de quelques applications des estimateurs de rétrécissement à la James Stein**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Melody Ghahramani (University of Winnipeg)

- 13:30 - 14:00     Syed Ejaz Ahmed (Brock University)  
 Shrinkage Strategy: From the Margin to the Centre of Universe / Stratégie de rétrécissement : de la bordure au centre de l'univers      
- 14:00 - 14:30     Shakhawat Hossain & Trevor Thomson (University of Winnipeg)  
 Shrinkage Estimation for Generalized Linear Mixed Models / Estimation de rétrécissement pour les modèles linéaires mixtes généralisés      
- 14:30 - 15:00     Bahadır Yüzbaşı (Inonu University), Syed Ejaz Ahmed (Brock University)  
 Pretest, Shrinkage, Positive Shrinkage Ridge Estimators in Linear Regression Models / Estimateurs de prétest, de rétrécissement et de rétrécissement positif pseudo-orthogonal dans les modèles de régression linéaire      

---



**13:30 - 15:00**     **Session 2D-I6**     (abstracts/résumés - page 195)     **Thistle 256**

**Risk Measures in Actuarial Science**  
**Mesures du risque en actuariat**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Jun Cai (University of Waterloo)

- 13:30 - 14:00     Edward Furman (York University), Ricardas Zitikis (Western University)  
 A New Risk Measure for Heavy Tailed Risks / Nouvelle mesure de risque pour les risques à queues lourdes      
- 14:00 - 14:30     Mélina Mailhot (Concordia University), Mhamed Mesfioui (Université du Québec à Trois-Rivières), Manuel Morales (Université de Montréal), Nicholas Beck (Concordia University)  
 Multivariate TVaR Risk Decomposition Techniques / Allocation de capital basée sur la TVaR multivariée      

14:30 - 15:00 Ruodu Wang (University of Waterloo)  
Risk Aversion in Monetary Risk Measures / Aversion au risque dans les mesures du risque monétaire  



**13:30 - 15:00 Session 2D-C1** (abstracts/résumés - page 197) **Thistle 255**



**Actuarial Science and Finance 2**



**Actuariat et finance 2**



Contributed Paper Session / Séance de communications libres



Chair/Président: Cary Tsai (Simon Fraser University)



13:30 - 13:45 Peter Adamic (Laurentian University), Jenna Guse (University of Waterloo)  
LOESS Density Estimation for Multivariate Survival Data Subject to Censoring and Masking / Estimation de densité LOESS pour données de survie multivariées soumises à de la censure et à du masquage  

13:45 - 14:00 Taehan Bae (University of Regina), Alex Kreinin (IBM)  
A Backward Construction and Simulation of Correlated Poisson Processes / Une construction en arrière et simulation de processus Poisson corrélés  

14:00 - 14:15 Yizhou Fang & Martin Lysy (University of Waterloo)  
Stochastic Volatility Modeling With Non-linear Filtering / Modélisation de la volatilité stochastique avec filtrage non-linéaire  

14:15 - 14:30 Rahim M Khorasany (Western University), R. Mark Reesor (SHARCNET Research Chair in Financial Mathematics Department of Statistical and Actuarial Sciences Richard Ivey School of Busin)  
Monte Carlo Valuation of Multiple Exercise Options: A Structured Regression Approach / Évaluation de Monte-Carlo de multiples levées d'options : une approche de régression structurée  

14:30 - 14:45 Etienne Marceau & Hélène Cossette (Université Laval)  
A Note on Extreme Negative Dependence / Une note sur la dépendance négative extrême  

14:45 - 15:00 Natalia Nolde (University of British Columbia), Johanna Ziegel (University of Bern)  
Elicitability and Backtesting / Élicibilité et contrôle ex-post  



**13:30 - 15:00 Session 2D-C2** (abstracts/résumés - page 200) **Thistle 254**



**Biostatistics: Treatment Comparisons**

**Biostatistique : comparaisons de traitements**

Contributed Paper Session / Séance de communications libres

Chair/Président: Gyanendra Pokharel (University of Calgary)

13:30 - 13:45 Siu Hung Cheung (The Chinese University of Hong Kong), Wenfu Xu (Renmin University), Feifang Hu (George Washington University)  
Adaptive Treatment Allocation for Non-inferiority Trials with Multiple Experimental Treatments / Allocation de traitement adaptatif pour les essais de non-infériorité avec traitements expérimentaux multiples  

13:45 - 14:00 James Hanley (McGill University), Zihui (Amy) Liu (Cancer Care Ontario), Olli Saarela (University of Toronto)  
Fitting a Model of the Mortality Reductions Produced by One/Several Rounds of Cancer Screening: Time and Sample Size Considerations / Ajuster un modèle de réduction de la mortalité produit pour un ou plusieurs cycles de dépistage du cancer : considérations de temps et de taille d'échantillon  

- 14:00 - 14:15 Taddele Cherinet Kibret, Tian Feng & Regina S. Kampo (McMaster University), Jemila S. Hamid (St. Michael's Hospital, Toronto), Joseph Beyene (McMaster University)  
Bayesian Network Meta-analysis (NMA) using MCMC and Integrated Nested Laplace Approximation (INLA): A Comparative Study / Méta-analyse bayésienne par réseau à l'aide de chaînes de Markov et approximation de Laplace emboîtée intégrée (ALEI) : une étude comparative  
- 14:15 - 14:30 Shun Fu Lee (Population Health Research Institute), Allan Donner & Neil Klar (University of Western Ontario)  
Adjusted  $I^2$  Statistic and its Confidence Intervals / Statistique  $I^2$  corrigée et ses intervalles de confiance  
- 14:30 - 14:45 Tong-Yu Lu (China Jiliang University), Wai-Yin Poon, Siu Hung Cheung & Ping Yang (The Chinese University of Hong Kong)  
Covariates-Adjusted Comparison of Treatments with Ordinal Responses / Comparaison de traitements par covariables ajustées, avec des réponses ordinales  
- 14:45 - 15:00 Binod Neupane & Joseph Beyene (McMaster University)  
Performance of Bayesian Network Meta-analysis for Sparse Data / Efficacité de la méta-analyse en réseau bayésienne pour des données rares  

**13:30 - 15:00**    **Session 2D-C3**    (abstracts/résumés - page 203)    **Thistle 257**

**Statistical Computing and Visualization**

**Calcul statistique et visualisation**

Contributed Paper Session / Séance de communications libres

Chair/Président: Martin Lysy (University of Waterloo)

- 13:30 - 13:45 Simon Bonner (University of Western Ontario), Matthew Schofield (University of Otago), Steven Price (University of Kentucky)  
Mark-recapture, Misidentification, and Markov Bases / Capture-recapture, identification erronée et bases de Markov  
- 13:45 - 14:00 Philippe Gagnon & Mylène Bédard (Université de Montréal), Alain Desgagné (Université du Québec à Montréal)  
Weak Convergence and Automation of the Reversible Jump Algorithm / Convergence faible et automatisation de l'algorithme à sauts réversibles  
- 14:00 - 14:15 Yun Ling & Martin Lysy (University of Waterloo)  
Superfast Gaussian Likelihood Solver and Application / Solveur très rapide du maximum de vraisemblance gaussien et application  
- 14:15 - 14:30 Wayne Oldford (University of Waterloo), Adrian Waddell (Roche, Basel Switzerland)  
Loon – the Interactive R Visualization Toolkit for Exploration / Loon : une boîte à outils interactive pour la visualisation avec R à des fins exploratoires  
- 14:30 - 14:45 Yukun Zhang (Yukun Zhang), Haocheng Li (University of Calgary), Sarah Kozey-Keadle & Charles Matthews (National Institutes of Health/National Cancer Institute), Raymond Carroll (Texas A&M University)  
PAactivPAL: An R Package for Statistical Analysis on Physical Activity Data / « PAactivPAL » : un paquet R pour l'analyse statistique de données sur l'activité physique  
- 14:45 - 15:00 Andrej Blejec (National Institute of Biology, Ljubljana, Slovenia)  
Presentation of Statistical Concepts with Animated Graphics and Simulations in R / Présentation des concepts statistiques à l'aide de graphiques animés et de simulations en R  

---

**13:30 - 14:45**    **Session 2D-C4**    (abstracts/résumés - page 206)    **Thistle 259**

**Statistical Genetics**  
**Génétique statistique**

Contributed Paper Session / Séance de communications libres  
Chair/Président: David Campbell (Simon Fraser University)

- 13:30 - 13:45    Bo Chen, Radu V. Craiu & Lei Sun (University of Toronto)  
Bayesian Model Averaging Method for X Chromosome Inactivation Problem / Méthode de la moyenne bayésienne des modèles pour l'inactivation du chromosome X     
- 13:45 - 14:00    Lu Cheng & Mu Zhu (University of Waterloo)  
Epistasis Detection by Reducing Disease Model Space using Clustering / Détection de l'épistasie en réduisant l'espace du modèle de maladie à l'aide de la classification     
- 14:00 - 14:15    Gun Ho Jang, Lincoln Stein & John Bartlett (Ontario Institute for Cancer Research)  
Locally Constant Signal Segmentation using a Relative Belief Inference / Segmentation de signal localement constant utilisant une inférence de croyance relative     
- 14:15 - 14:30    Mateen Shaikh, Jen Stearns, Michael Surette, Sonia Anand, Russell de Souza & Joseph Beyene (McMaster University)  
The Beta Negative Binomial Distribution for the Analysis of Microbiome Data / La distribution bêta-binomiale négative pour l'analyse des données sur le microbiome     
- 14:30 - 14:45    Jihyung Shin (Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto), Shelley B. Bull (Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Dalla Lana School of Public Health, University of Toronto)  
On the Analysis of Low-Frequency Genetic Variants in the Presence of Quantitative Covariates Using Prospective and Retrospective Penalized Logistic Regression Models / À propos de l'analyse de variants génétiques rares en présence de covariables quantitatives utilisant des modèles prospectifs et rétrospectifs par régression logistique pénalisée     

---

**13:30 - 15:00**    **Session 2D-C5**    (abstracts/résumés - page 209)    **Thistle 258**

**Statistical Inference and Applications 1**  
**Inférence statistique et applications 1**

Contributed Paper Session / Séance de communications libres  
Chair/Président: Jeffrey Rosenthal (University of Toronto)

- 13:30 - 13:45    Mikelis Bickis (University of Saskatchewan)  
Statistical Relativity / La relativité en statistique     
- 13:45 - 14:00    Edit Gombay (University of Alberta)  
Change Detection Algorithms / Algorithmes de détection du changement     
- 14:00 - 14:15    Kevin Mongeon (Brock University)  
The Process of Winning and Losing Sporting Contests / Le processus de gagner ou perdre des événements sportifs     
- 14:15 - 14:30    Shaun Zheng Sun (University of the Fraser Valley)  
Improved Tests for Independence in Ordinal Contingency Tables / Tests améliorés pour l'indépendance des tableaux de contingence ordinaux     
- 14:30 - 14:45    Gary J. Umphrey (University of Guelph)  
Aesthetics and Measurement: The Measurement of Aesthetics and the Aesthetics of Measurement / Esthétique et mesure : la mesure de l'esthétique et l'esthétique de la mesure     
- 14:45 - 15:00    Xikui Wang, You Liang & Lysa Porth (University of Manitoba)  
Dynamic Risk Measures and Agricultural Insurance / Mesures de risque dynamiques et assurance agricole     
-

**14:45 - 15:30****Thistle Corridor across from Thistle 255****Coffee Break  
Pause-café****15:30 - 17:00****Session 2E-I1**

(abstracts/résumés - page 212)



**Thistle 242****Adaption and Approximation in Markov Chain Monte Carlo****Adaptation et approximation pour la méthode de Monte-Carlo par chaînes de Markov**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Aaron Smith (University of Ottawa)

15:30 - 16:00

Mylène Bédard (Université de Montréal)

Hierarchical Models: Local Proposal Variances for the RWM-within-Gibbs / Modèles hiérarchiques : variances instrumentales locales pour le RWM-dans-Gibbs  



16:00 - 16:30

Daniel Jerison (Cornell University)

Honest MCMC Convergence Guarantees / Garanties honnêtes de convergence MCMC  

16:30 - 17:00

Matti Vihola (University of Jyväskylä)

Unbiased Estimators and Multilevel Monte Carlo / Estimateurs non biaisés et Monte-Carlo multiniveaux  **15:30 - 17:00****Session 2E-I2**

(abstracts/résumés - page 214)



**Welch - David Howes****Building a Research Career: From New Investigator to Present****Comment bâtir une carrière en recherche : de nouveau chercheur à aujourd'hui**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Martin Lysy (University of Waterloo)

15:30 - 16:00

James Ramsay (McGill University)

The Art of the Awkward Question (and Some Psychology, Too) / De l'art de la question délicate (et de la psychologie)  



16:00 - 16:30

Jeffrey Rosenthal (University of Toronto)

Lessons From a Twisted Career Path / Leçons d'un chemin de carrière tordu  

16:30 - 17:00

Mary Thompson (University of Waterloo)

Early Career Research Back Then and Now / Recherches de début de carrière : hier et aujourd'hui  **15:30 - 17:00****Session 2E-I3**

(abstracts/résumés - page 216)

**Thistle 247****Design and Methodological Issues in Health Studies of Vulnerable Populations****Problèmes de conception et de méthodologie des études sur la santé des populations vulnérables**



Invited Paper Session / Présentations sur invitation

Chair/Président: John Petkau (University of British Columbia)

Organizer/Responsable: Lehana Thabane (McMaster University)

15:30 - 16:00



Lisa Lix (Ottawa University of Manitoba)

Methodological Challenges in Aboriginal Health Research / Problèmes méthodologiques en recherche en santé autochtone  

16:00 - 16:30

Robert W. Platt (McGill University)

Randomized Trials in Obstetric Populations / Essais randomisés en milieu obstétrique  







- 16:30 - 17:00 Lehana Thabane (McMaster University)  
 Planning for the Twilight Years: Why We Need More Statisticians Involved in Trials in the Elderly Living in Long-Term Care Settings / Planification du crépuscule de la vie : pourquoi il nous faut plus de statisticiens dans les essais menés auprès des personnes âgées vivant en centres de soins de longue durée  

**15:30 - 17:00 Session 2E-I4** (abstracts/résumés - page 218) **Thistle 258**

**Inference and Prediction from Distributed Data**  
**Inférence et prévisions à partir de données distribuées**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Jean-Francois Plante (HEC Montreal)







- 15:30 - 16:00 Doina Caragea, Li HongMin, Sopova Oleksandra & Herndon Nic (Kansas State University)  
 Learning Domain Adaptation Classifiers from Multiple Distributed Sources / Classificateurs d'adaptation de domaines d'apprentissage à partir de sources multiples distribuées  
- 16:00 - 16:30 George Ostrouchov (Oak Ridge National Laboratory & University of Tennessee)  
 Statistical Computing with R on Distributed and Multicore Platforms / Calcul statistique avec R sur des plateformes distribuées et multicœur  
- 16:30 - 17:00 Ruibin Xi (Peking University), Nan Lin (Washington University in St. Louis)  
 Statistical Aggregation for Big Data Analysis / Agrégation statistique pour l'analyse de données volumineuses  

**15:30 - 17:00 Session 2E-I5** (abstracts/résumés - page 220) **Thistle 246**

**Statistical Challenges in Large-Scale Inference**  
**Défis statistiques en inférence à grande échelle**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Kun Liang (University of Waterloo)



- 15:30 - 16:00 Chunming Zhang (University of Wisconsin-Madison)  
 Single-Index Modulated Multiple Testing / Tests multiples modulés à indice unique  
- 16:00 - 16:30 Wenguang Sun (University of Southern California), Tony Cai (University of Pennsylvania), Weinan Wang (University of Southern California)  
 CARS: Covariate Assisted Ranking and Screening for Large-Scale Two-Sample Inference / CARS : Classement et triage assistés de covariables pour inférence à deux échantillons à grande échelle  
- 16:30 - 17:00 Ruth Heller (Tel-Aviv University), Marina Bogomolov (Technion), Daniel Yekutieli (Tel-Aviv University)  
 Assessing Replicability Across High-Dimensional Studies / Évaluer la reproductibilité entre les études à grande dimension  



**15:30 - 17:00 Session 2E-I6** (abstracts/résumés - page 222) **Academic South 217**

**Survey Methodology in a Big Data World**  
**Méthodes d'enquête dans un monde de mégadonnées**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Nancy Reid (University of Toronto)









- 15:30 - 16:15 Anne de Moliner (Université de Montréal), Hervé Cardot & Camelia Goga (Université de Bourgogne)  
 Estimating a Mean Electricity Consumption Curve in the Presence of Missing Data Using Survey Sampling Techniques / Estimation de courbe moyenne de consommation électrique par sondage en présence de valeurs manquantes  

- 16:15 - 17:00 John Eltinge (U.S. Bureau of Labor Statistics)  
 Characterization, Modeling and Management of Data Quality, Risk and Cost in the Integration of Surveys with Alternative Data Sources / Caractérisation, modélisation et gestion de la qualité, du risque et du coût de données dans l'intégration des enquêtes avec sources de données alternatives  

**15:30 - 16:30 Session 2E-C1** (abstracts/résumés - page 224) **Thistle 255**

**Biostatistics: New Inferential Strategies**  
**Biostatistique : nouvelles stratégies d'inférence**











Contributed Paper Session / Séance de communications libres  
 Chair/Président: Karen Kopciuk (University of Calgary)

- 15:30 - 15:45 Christopher Gravel & Robert W. Platt (McGill University)  
 Inverse Probability Weighted Estimation in Confounded Binary Data with Outcome Misclassification / Estimation pondérée selon la probabilité inverse pour les données binaires confondues avec classification erronée du résultat  
- 15:45 - 16:00 Qiuguang Sang, Guohua Yan & Renjun Ma (University of New Brunswick)  
 Bayesian Compound Poisson Mixed Models for Longitudinal Semi-Continuous Data with Non-Ignorable Missingness / Modèles composés mixtes bayésiens de Poisson pour des données longitudinales semi-continues manquantes de façon non-ignorable  
- 16:00 - 16:15 Farhad Shokoohi, Abbas Khalili & Masoud Asgharian (McGill University), Shili Lin (The Ohio State University)  
 Feature Selection in High-Dimensional Heterogeneous Time-to-Event Data; A Study on Ovarian Cancer / Sélection de traits pour des données de temps d'événement hétérogènes et de grande dimension ; étude sur le cancer de l'ovaire  
- 16:15 - 16:30 Meng Zhao & Russell Steele (McGill University), Ian Shrier (Lady Davis Institute, Jewish General Hospital and McGill University)  
 Application of the Trend-Renewal Process to Recurrent Sports Injury / Application du processus de tendance-renouvellement (PTR) aux blessures sportives récurrentes  



**15:30 - 17:00 Session 2E-C2** (abstracts/résumés - page 226) **Thistle 257**

**Probability and Estimation**  
**Probabilité et estimation**

Contributed Paper Session / Séance de communications libres  
 Chair/Président: Fernando Camacho (Damos Inc.)

- 15:30 - 15:45 Shanoja Naik & Taehan Bae (University of Regina)  
 On Convolution of Inverse Gaussian and Gamma Distribution and Auto Regressive Model / A propos de la convolution des lois inverse gaussienne et Gamma pour le modèle autorégressif  
- 15:45 - 16:00 Steve Puglia & Xu Xiaojian (Brock University)  
 D-optimal Design of Linear Measurement Error Models / Système optimal D de modèles linéaires de mesure d'erreur  
- 16:00 - 16:15 Jean Vaillancourt (École Nationale d'Administration Publique), Bruno Rémillard (HEC Montréal)  
 Loser Takes All! / Qui perd gagne !  
- 16:15 - 16:30 Myron Hlynka & Zishan Huang (University of Windsor)  
 Probabilistic Laplace Transform Inversion / Transformée inverse probabiliste de Laplace  
- 16:30 - 16:45 Chel Hee Lee & Mikelis Bickis (University of Saskatchewan)  
 Geometrical View of Imprecise Prior for Imprecise Inference / Vision géométrique des lois a priori imprécises pour l'inférence imprécise  













- 16:45 - 17:00 Deli Li (Lakehead University), Han-Ying Liang (Tongji University, China), Andrew Rosalsky (University of Florida, USA)  
A Note on the Symmetrization Procedures for the Laws of Large Numbers / Remarque sur les procédures de symétrisation des lois des grands nombres  

**15:30 - 17:00 Session 2E-C3** (abstracts/résumés - page 229) **Thistle 254**

**Regression Models**  
**Modèles de régression**





Contributed Paper Session / Séance de communications libres  
Chair/Président: Juli Atherton (UQAM)









- 15:30 - 15:45 Mohamed Amezziane (Central Michigan University), Syed Ejaz Ahmed (Brock University)  
Variable Selection via Individual Shrinkage of Regression Coefficients / Sélection de variable par le rétrécissement individuel des coefficients de régression  
- 15:45 - 16:00 Mai Ghannam (University of Windsor), Sévérin Nkurunziza (University of Windsor/Université de Sherbrooke)  
Some Stein Rules Methods in Tensor Regression Models with High and Ultrahigh Dimensional Data / Certaines règles de Stein dans les modèles de régression tensorielle avec des données de dimensions supérieures  
- 16:00 - 16:15 Bashir Khan (Saint Mary's University), A.K.Md. Ehsanes Saleh (Carleton University)  
Estimation in a Regression Model with Autocorrelated Errors / Estimation dans un modèle de régression avec erreurs autocorrélées  
- 16:15 - 16:30 Shahedul Khan (University of Saskatchewan)  
Exponentiated Weibull Regression Model for Time-to-Event Data / Modèle de régression de Weibull exponentié pour des données de temps d'événement  
- 16:45 - 17:00 Ka Ho Wu (The Chinese University of Hong Kong)  
Error Model Fitting in Benchmarking / Ajustement du modèle d'erreur à l'étalonnage  

**15:30 - 17:00 Session 2E-C4** (abstracts/résumés - page 232) **Thistle 256**

**Statistical Applications and Methods**  
**Méthodes et applications statistiques**

Contributed Paper Session / Séance de communications libres  
Chair/Président: Duncan Murdoch (University of Western Ontario)

- 15:30 - 15:45 Baktiar Hasan (European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium), Cong Chen (Merck & Co., USA), Laurence Collette (European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium), Xanthi Pedeli (Frontier Science Foundation - Hellas, Athens, Greece), Yue Shentu (Merck & Co., USA), Jessica Menis (European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium), Urania Dafni (Frontier Science Foundation - Hellas, Athens, Greece.)  
Adaptive Study Design: Dealing with Differential Treatment Effect across Subgroups, a Case Study from an EORTC-ETOP Randomized, Phase III Trial in Early Stage Non-Small Cell Lung Cancer (NSCLC) / Plan d'étude adapté à un effet différentiel de traitement selon une classification en sous-groupes : une phase III de EORTC-ETOP en exemple  
- 15:45 - 16:00 Michael McIsaac (Queen's University)  
Overcoming Missing Data in Physical Activity Studies Involving Accelerometers / Surmonter le problème des données manquantes dans les études d'activité physique impliquant des accéléromètres  

- 16:00 - 16:15 Khokan C. Sikdar (University of Calgary), Georgiy Bobashev (RTI International), Hude Quan, Danielle Southern & Matthew Thomas James (University of Calgary)  
Shifting Epidemiologic Paradigm from Population-level to Individual-level: Cardiovascular Mortality and Morbidity in Hypertensive Patients / Déplacer le paradigme épidémiologique du niveau de la population au niveau individuel : mortalité et morbidité cardiovasculaires chez des patients hypertendus  
- 16:15 - 16:30 Jenna Sykes (St. Michael's Hospital), Sanja Stanojevic (The Hospital for Sick Children), Christopher H. Goss (University of Washington), Bradley S. Quon (University of British Columbia), Anne L. Stephenson (St. Michael's Hospital)  
A Standardized Approach to Estimating Survival Statistics for Population-Based Registry Cohorts / Approche standardisée pour estimer les statistiques de survie de cohortes de registres basées sur des populations  
- 16:30 - 16:45 Ernest Dankwa, Taraneh Abarin & Hong Wang (Memorial University)  
Bayesian Approach to a Linear Regression model with Measurement Error in Covariate / Approche bayésienne pour un modèle de régression linéaire avec erreur de mesure dans la covariable  
- 16:45 - 17:00 Yisub Kye & Edward Furman (York University)  
Economic Capital Analysis within a Portfolio of Dependent and Heavy-tailed Risks / Analyse du capital économique d'un portefeuille de risques dépendants et à ailes lourdes  

**15:30 - 17:00**    **Session 2E-C5**    (abstracts/résumés - page 236)













**Thistle 259**

**Statistical Inference and Applications 2**

**Inférence statistique et applications 2**

Contributed Paper Session / Séance de communications libres

Chair/Président: Ali Karimnezhad (University of Ottawa)

- 15:30 - 15:45 Jules de Tibeiro (Université de Moncton), Luigi D'Ambra (Dipartimento di Economia, Management Istituzioni, University of Naples, Italy.), Antonello D'Ambra (Dipartimento di Economia, Second University of Naples, Italy)  
Different Constrained Ordination Methods for Contingency Table with External Information / Différentes méthodes d'ordination sous contraintes pour des tableaux de contingence à variables instrumentales  
- 15:45 - 16:00 Othmane Kortbi (UAE University), Aziz L'Moudden & Éric Marchand (Université de Sherbrooke), William E. Strawderman (Rutgers University)  
On Predictive Density Estimation with Constraints under Kullback-Leibler Loss / Estimation de densité prédictive avec des contraintes sous la perte Kullback-Leibler  
- 16:00 - 16:15 Chantal Larose (State University of New York at New Paltz), Ofer Harel (University of Connecticut), Katarzyna Kordas (University of Bristol), Dipak Dey (University of Connecticut)  
An Entropy-Based Method for Choosing a Number of Classes for Incomplete Data LCA / Méthode fondée sur l'entropie pour choisir le nombre de structures dans l'analyse de structures latentes  
- 16:15 - 16:30 Tiago Magalhães, Denise Botter & Monica Sandoval (University of São Paulo)  
A Wald-type Test in Proper Dispersion Models / Test de type Wald dans les modèles de dispersion propre  
- 16:30 - 16:45 Fahimeh Moradi & Ali Karimnezhad (University of Ottawa)  
Bayesian Parameter Learning in Diagnostic Networks / Apprentissage de paramètres bayésiens dans les réseaux de diagnostics  
- 16:45 - 17:00 Chen Xu (University of Ottawa), Shaobo Lin & Jian Fang (Xi'an Jiaotong University), Runze Li (The Pennsylvania State University)  
Prediction-Based Termination Rule for Greedy Learning with Massive Data / Règle de terminaison fondée sur la prévision pour l'apprentissage glouton avec des données volumineuses  



---

<b>17:00 - 18:00</b> <b>NSERC Discovery Grant Application Workshop</b> <b>Atelier sur les demandes de subventions à la découverte du CRSNG</b>	<b>Thistle 256</b>
<b>17:00 - 18:00</b> <b>Actuarial Science Section General Meeting</b> <b>Groupe d'actuariat assemblée générale</b>	<b>Thistle 255</b>
<b>17:00 - 18:00</b> <b>Biostatistics Section General Meeting</b> <b>Groupe de biostatistique assemblée générale</b>	<b>Thistle 242</b>
<b>17:00 - 18:00</b> <b>Business and Industrial Statistics Section General Meeting</b> <b>Groupe de statistique industrielle et de gestion assemblée générale</b>	<b>Thistle 257</b>
<b>17:00 - 18:00</b> <b>Probability Section General Meeting</b> <b>Groupe de probabilité assemblée générale</b>	<b>Thistle 258</b>
<b>17:00 - 18:00</b> <b>Statistical Education Section General Meeting</b> <b>Groupe d'éducation en statistique assemblée générale</b>	<b>Thistle 259</b>
<b>17:00 - 18:00</b> <b>Survey Methods Section General Meeting</b> <b>Groupe de méthodes d'enquête assemblée générale</b>	<b>Thistle 246</b>
<b>18:15 - 19:05</b> <b>Banquet- Cocktails and Cash Bar</b> <b>Banquet- Bar payant</b>	<b>Market Hall</b>
<b>19:05 - 22:00</b> <b>Banquet- Dinner Service</b> <b>Banquet- Service du repas</b>	<b>Market Hall</b>

**Wednesday June 1****mercredi 1er juin****8:15 - 16:00****Thistle Corridor across from Sean O'Sullivan Theatre****Registration  
Inscription****8:45 - 9:45      Session 3A-A1      (abstract/résumé - page 239)      Thistle 325****CRM-SSC Prize in Statistics Address****Allocution du récipiendaire du Prix CRM-SSC en statistique**

Special Session / Conférence spéciale



Chair/Président: Jose Garrido (Concordia University)

8:45 - 9:45      Radu V. Craiu (University of Toronto)  
 Design Strategies for Adaptive MCMC / Stratégies de conception pour le MCMC adaptatif  

**8:45 - 9:45      Session 3A-A2      (abstract/résumé - page 240)      Welch - David Howes****SSC Impact Award Session****Allocution du récipiendaire du Prix pour impact de la SSC**

Special Session / Conférence spéciale

Chair/Président: Rick Routledge (Simon Fraser University)



8:45 - 9:45      Shelley B. Bull (University of Toronto)  
 Statistical Issues and Opportunities in Molecular and Genetic Epidemiology / Problèmes et possibilités statistiques en épidémiologie moléculaire et génétique  



**9:50 - 10:30****Thistle Corridor across from Thistle 255****Coffee Break  
Pause-café****10:20 - 11:50      Session 3B-I1      (abstracts/résumés - page 241)      Thistle 258****Analysis of Complex Survey Data****Analyse de données d'enquête complexes**



Invited Paper Session / Présentations sur invitation

Chair/Président: Karla Fox (Statistics Canada)

Organizer/Responsable: Lenka Mach (Statistics Canada)

10:20 - 10:50      Yves Berger (University of Southampton), Melike Oguz-Alper (Statistics Norway)  
 An Empirical Likelihood Approach for Complex Sampling / Une approche par vraisemblance empirique pour l'échantillonnage complexe  

10:50 - 11:20      Elisabeth Neusy & Harold Mantel (Statistics Canada)  
 Confidence Intervals for Proportions Estimated from Complex Survey Data / Intervalles de confiance pour des proportions estimées à partir de données d'enquêtes complexes  

11:20 - 11:50      Changbao Wu (University of Waterloo), J.N.K. Rao (Carleton University)  
 Empirical Likelihood for Public-Use Survey Data / Vraisemblance empirique pour données de sondage d'utilité publique  







---

**10:20 - 11:50**     **Session 3B-I2**     (abstracts/résumés - page 243)     **Welch - David Howes**

**Analysis of Genomic Sequencing Data in Families**  
**Analyse des données de séquençage génomique dans les familles**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Jinko Graham (Simon Fraser University)

- 10:20 - 10:50     M'Hamed Lajmi Lakhel Chaieb, Martin Leclerc & Jacques Simard (Université Laval)  
 SNP-Set Association Testing for Survival Outcomes in the Presence of Intrafamilial Correlation / Tests d'association d'ensembles de SNPs pour des réponses de survie en présence de corrélation intra-familiale      
- 10:50 - 11:20     Marie-Hélène Roy-Gagnon, Kelly Burkett, Jean-François Lefebvre, Cheng Wang, Bénédicte Fontaine-Bisson & Lise Dubois (University of Ottawa)  
 A Comparison of Statistical Methods for the Discovery of Genetic Risk Factors Using Longitudinal Family Study Designs / Comparaison de méthodes statistiques pour l'identification de facteurs de risque génétiques dans des devis d'études familiales longitudinales      
- 11:20 - 11:50     Jianping Sun (McGill), Karim Oualkacha (Université du Québec à Montréal), Celia Greenwood (Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal)  
 Multivariate Association Test for Rare Variants Controlling for Cryptic and Family Relatedness / Test d'association à plusieurs variables pour variantes rares contrôlées pour parenté familiale et cryptique      






---

**10:20 - 11:50**     **Session 3B-I3**     (abstracts/résumés - page 245)     **Thistle 242**

**High-dimensional Statistics: Challenges and Recent Developments**  
**Statistiques en haute dimension : défis et progrès récents**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Yingli Qin (University of Waterloo)

- 10:20 - 10:50     Shoja'eddin Chenouri, Jiayi Liang & Christopher G. Small (University of Waterloo)  
 Nonlinear Dimension Reduction: Performance Analysis and Robustness / Réduction de la dimension non linéaire : analyse de l'efficacité et robustesse      
- 10:50 - 11:20     Jeff Yao (University of Hong Kong), Zeng Li (The University of Hong Kong), Wang Qinwen (University of Pennsylvania)  
 Identifying the Number of Factors from Singular Values of a Large Sample Auto-Covariance Matrix / Identification du nombre de facteurs à l'aide des valeurs singulières d'une grande matrice d'autocovariance empirique      
- 11:20 - 11:50     Ping-Shou Zhong (Michigan State University), Runze Li (Pennsylvania State University)  
 Homogeneity Test of Covariance Matrices and Change-Points Identification with High-Dimensional Longitudinal Data / Test d'homogénéité de matrices de covariance et l'identification de points de rupture avec des données longitudinales de grande dimension      



---





**10:20 - 11:50**     **Session 3B-I4**     (abstracts/résumés - page 247)     **Thistle 246**

**Statistical Theory and Modeling for Functional Data Analysis**  
**Théorie et modélisation statistiques pour l'analyse de données fonctionnelles**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Haocheng Li (University of Calgary)

- 10:20 - 10:50     Kexin Ji & Joel A. Dubin (University of Waterloo)  
 Semiparametric Stochastic Mixed Models for Bivariate Periodic Longitudinal Data / Modèles stochastiques mixtes semi-paramétriques pour des données longitudinales périodiques bidimensionnelles      

- 10:50 - 11:20 Bei Jiang (University of Alberta), Eva Petkova (New York University), Thaddeus Tarpey (Wright State University), R. Todd Ogden (Columbia University)  
Latent Class Modeling Using Matrix-valued Covariates with Application to Identifying Early Placebo Responders Based on EEG Signals / Modélisation de classes latentes utilisant des covariables à valeurs matricielles pour identifier les répondeurs hâtifs au placebo basée sur les signaux EEG  
- 11:20 - 11:50 Zhenhua Lin & Fang Yao (University of Toronto), Hans G. Müller (University of California at Davis)  
Mixture Models and Densities for Functional Data / Modèles de mélange et de densités pour données fonctionnelles  

---

**10:20 - 11:50 Session 3B-I5** (description - page 249) **Thistle 325**

**Staying Up-To-Date As a Practicing Biostatistician**  
**Comment rester à la page en tant que biostatisticien praticien**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et présidente: Eleanor Pullenayegum (Hospital for Sick Children)

- 10:20 - 10:50 Mohamed Abdollel (Dalhousie University)
- 10:50 - 11:20 Rhonda Rosychuk (University of Alberta)
- 11:20 - 11:50 Rinku Sutradhar (Institute for Clinical Evaluative Sciences)





---

**10:20 - 11:50 Session 3B-I6** (abstracts/résumés - page 250) **Academic South 217**

**The Theory of Dispersion Models with Applications**  
**La théorie de la dispersion avec des applications**

Invited Paper Session / Présentations sur invitation

Organizer and Chair / Responsable et président: Renjun Ma (University of New Brunswick)

- 10:20 - 10:30 Renjun Ma, Discussant (University of New Brunswick)
- 10:40 - 11:20 Rodrigo Labouriau (Aarhus University)  
Multivariate Generalized Linear Mixed Models with High Complexity / Modèles linéaires généralisés mixtes multivariés avec complexité élevée  
- 11:20 - 11:50 Peter XK Song (University of Michigan)  
Dispersion Models and Beyond / Les modèles de dispersion et au-delà  







---







**10:20 - 11:50 Session 3B-C1** (abstracts/résumés - page 252) **Thistle 256**

**Biostatistics: Survival Analysis**  
**Biostatistique : analyse de survie**

Contributed Paper Session / Séance de communications libres

Chair/Président: John Koval (University of Western Ontario)

- 10:20 - 10:35 Samantha-Jo Caetano, Gregory Pond, Peter Ellis & David Dawe (McMaster University)  
Proposing and Analysing Methods which Improve the Kaplan-Meier Estimates of Altered Time-to-Event Data / Proposer et analyser des méthodes qui améliorent les estimations Kaplan-Meier de données altérées de temps d'événement  
- 10:35 - 10:50 Noori Akhtar-Danesh (McMaster University), Lina Jansen & Hermann Brenner (German Cancer Research Centre)  
Estimating Complete Life Tables from Abridged Life Tables in Relative Survival Analysis / Estimer des tables de mortalité complètes dans une analyse de survie relative à partir de tables de mortalité abrégées  
- 10:50 - 11:05 Coraline Danieli & Michal Abrahamowicz (McGill University)  
Competing Risk Modelling of Time-Varying Exposures / Modélisation du risque concurrent d'expositions variant dans le temps  

- 11:05 - 11:20 Hua Shen (University of Calgary)  
Methods for Incomplete Covariates in Clustered Survival Data with Left Truncation / Méthodes pour les covariables incomplètes dans des données de survie regroupées avec troncation à gauche  
- 11:20 - 11:35 Yingying Xu, Joon Lee & Joel A. Dubin (University of Waterloo)  
Personalized Time-to-Event Prediction in Intensive Care / Préviation personnalisée des occurrences dans le temps pour les soins intensifs  
- 11:35 - 11:50 Yayuan Zhu, Jerald F. Lawless & Cecilia A. Cotton (University of Waterloo)  
Non- and Semiparametric Analysis of Dependently Interval-Censored Failure Time Data Due to Intermittent Visits at Disease Clinics / Analyse non paramétrique et semi-paramétrique de données de temps de défaillance avec censure par intervalle dépendante en raison de visites intermittentes à une clinique de traitement de maladies  

---









**10:20 - 11:20**    **Session 3B-C2**    (abstracts/résumés - page 255)    **Thistle 255**

**Time Dependent Data**

**Données dépendantes du temps**

Contributed Paper Session / Séance de communications libres

Chair/Président: Edward Chen (Statistics Canada)

- 10:20 - 10:35 Fuqi Chen & Rogemar Mamon (Western University), Sévérin Nkurunziza (University of Windsor)  
Estimation and Inference for a Change-point Problem under a Generalised Ornstein-Uhlenbeck Setting / Estimation et inférence pour un problème de point de changement sous une situation généralisée Ornstein-Uhlenbeck  
- 10:35 - 10:50 François Marshall (Queen's University)  
Robust Spectrum Estimation: Application to Relative Ionospheric Opacity Meters / Estimation spectrale robuste : application à des instruments de mesure de l'opacité ionosphérique relative  
- 10:50 - 11:05 Frédéric Picard & Steve Matthews (Statistics Canada)  
The Addition of Trend-Cycle Estimates to Selected Publications at Statistics Canada / L'ajout de la tendance-cycle à quelques publications de Statistique Canada  
- 11:05 - 11:20 Reza Ramezan (California State University, Fullerton), Paul Marriott & Shoja'eddin Chenouri (University of Waterloo)  
Time Scale in Neuronal Data: A Multiscale Model / Échelle de temps dans les données neuronales : un modèle à multi-échelles  

---

**12:00 - 13:30**    **Thistle 259**

**Accreditation Committee**

**Comité d'accréditation**

---

**12:00 - 13:30**    **Thistle 254**

**Actuarial Science Section Executive Committee 2**

**Comité exécutif du groupe d'actuariat 2**

---

**12:00 - 13:30**    **Thistle 255**

**Biostatistics Section Executive Committee 2**

**Comité exécutif du groupe de biostatistique 2**

---

**12:00 - 13:30** **Thistle 256**  
**Probability Section Executive Committee 2**  
**Comité exécutif du groupe de probabilité 2**

---

**12:00 - 13:30** **Thistle 242**  
**Program Committee**  
**Comité du programme**

---

**12:00 - 13:30** **Thistle 257**  
**Statistical Education Section Executive Committee 2**  
**Comité exécutif du groupe d'éducation en statistique 2**

---

**12:00 - 13:30** **Thistle 246**  
**Student and Recent Graduate Committee**  
**Comité des étudiants et diplômés récents**

---

**12:00 - 13:30** **Thistle 258**  
**Survey Methods Section Executive Committee Meeting 2**  
**Comité exécutif du groupe de méthodes d'enquête 2**

---

**13:30 - 14:30**    **Session 3C-I1**    (abstract/résumé - page 257)    **Thistle 242**  
**Isobel Loutit Lecture**  
**Allocution invitée Isobel Loutit**

Invited Paper Session / Présentations sur invitation  
Organizer and Chair / Responsable et président: Hugh Chipman (Acadia)

13:30 - 14:30    Christine M. Anderson-Cook (Los Alamos National Laboratory)  
Realistic Optimal Designs in a Resource-Constrained World / Plans optimaux réalistes dans  
un monde à ressources limitées E E

---

**13:30 - 14:30**    **Session 3C-I2**    (abstract/résumé - page 258)    **Academic South 204**  
**Survey Methods Section Presidential Invited Address**  
**Allocution de l'invité du président du Groupe des méthodes d'enquête**

Invited Paper Session / Présentations sur invitation  
Organizer and Chair / Responsable et présidente: Karla Fox (Statistics Canada)



13:30 - 14:30    Michael Hidiroglou (Statistics Canada)  
Data Collection / Collecte de données E E

---

**13:30 - 14:30**    **Session 3C-A1**    (abstract/résumé - page 259)    **Welch - David Howes**  
**CJS Award Address**  
**Allocution du récipiendaire du Prix de la RCS**

Special Session / Conférence spéciale  
Chair/Président: Richard Lockhart (Simon Fraser University)



13:30 - 14:30 Grace Y. Yi (University of Waterloo)  
 Variable Selection and Inference Procedures for Marginal Analysis of Longitudinal Data with Missing Observations and Covariate Measurement Error / Sélection de variables et procédures d'inférence pour l'analyse marginale de données longitudinales avec observations manquantes et erreur de mesure des covariables  



---

**13:30 - 14:30 Session 3C-A2** (abstract/résumé - page 260) **Thistle 325**

**Pierre Robillard Award Address**

**Allocution du récipiendaire du Prix Pierre-Robillard**

Special Session / Conférence spéciale  
 Chair/Président: Juli Atherton (UQAM)  
 Organizer/Responsable: Gordon (Fick)

13:30 - 14:30 Yujie Zhong (Cambridge Institute of Public Health)  
 Inference About Within-Family Associations in Disease Onset Times Under Biased Sampling / Inférence sur les associations familiales dans les moments d'apparition de la maladie en cas d'échantillonnage biaisé  

---

**14:30 - 15:00 Thistle Corridor across from Thistle 255**

**Coffee Break**

**Pause-café**



---



**15:00 - 16:30 Session 3D-I1** (abstracts/résumés - page 261) **Thistle 242**



**Financial and Actuarial Mathematics**

**Mathématique actuarielle et financière**

Invited Paper Session / Présentations sur invitation  
 Organizer and Chair / Responsable et président: Cody Hyndman (Concordia University)

15:00 - 15:30 Alex Badescu (University of Calgary), Juan-Pablo Ortega (CNRS, Université de Franche-Comté), Zhenyu Cui (Stevens Institute of Technology)  
 Non-Affine GARCH Option Pricing Models, Variance Dependent Kernels, and Diffusion Limits / Modèles d'évaluation d'options GARCH non affine, noyaux dépendant de la variance et limites de diffusion  

15:30 - 16:00 Cody Hyndman & Samira Shirgir (Concordia University)  
 Issuing a Convertible Bond with Call-Spread Overlay: Incorporating the Effects of Convertible Arbitrage / Émission d'une obligation convertible avec superposition d'opérations mixtes sur option d'achat : comment inclure les effets de l'arbitrage convertible  

16:00 - 16:30 Adam Metzler (Wilfred Laurier University), Alexandre Scott (BMO)  
 Importance Sampling for Portfolio Loss Probabilities under Conditional Independence / Échantillonnage préférentiel pour probabilités de perte sur portefeuille avec indépendance conditionnelle  







---

**15:00 - 16:30 Session 3D-I2** (abstracts/résumés - page 263) **Thistle 325**

**High Frequency Data**

**Données de haute fréquence**

Invited Paper Session / Présentations sur invitation  
 Organizer and Chair / Responsable et président: Rafal Kulik (University of Ottawa)

- 15:00 - 15:30 Sebastian Jaimungal (University of Toronto), Alvaro Cartea (Oxford University), Luhui Gan (University of Toronto)  
Option Pricing and Hedging with Limit and Market Orders / Fixation des prix en fonction de la valeur d'option et couverture avec ordres à cours limité et ordres au cours du marché  
- 15:30 - 16:00 Frederi Viens & Baron Law (Purdue University)  
A Pure-Jump Market-Making Model for High-Frequency Trading using Constrained Forward-Backward Stochastic Differential Equations / Un modèle de sauts pur pour teneur de marché faisant usage des équations différentielles stochastiques avant-arrière avec contrainte  
- 16:00 - 16:30 Ibrahim Abdelrazeq (Rhodes College)  
Model Checking for the Stochastic Volatility Model: Lévy Driven CARMA(2,1) / Vérification de modèle pour modèle de volatilité stochastique : processus CARMA(2,1) régi par un processus de Lévy  

---







**15:00 - 16:30**    **Session 3D-I3**    (abstracts/résumés - page 265)    **Thistle 246**

**Network Meta-Analysis**

**Méta-analyse de réseaux**

Invited Paper Session / Présentations sur invitation

Organizers/Responsables: Jemila Hamid (St. Michael's Hospital) and/et Lehana Thabane (McMaster University)

- 15:00 - 15:30 Caitlin Daly (McMaster University) (dalych@mcmaster.ca), Sharon Straus (Li Ka Shing Knowledge Institute, St. Michael's Hospital), Lehana Thabane (McMaster University), Jemila S. Hamid (Li Ka Shing Knowledge Institute, St. Michael's Hospital)  
Ranking Probabilities and Surface Under Cumulative Ranking Probabilities (SUCRA) in Network Meta-Analysis (NMA) / Probabilités de classement et surface selon le classement cumulatif de probabilités (SUCRA) dans la méta-analyse en réseau (NMA)  
- 15:30 - 16:00 Jemila S. Hamid (St. Michael's Hospital) (jhamid@mcmaster.ca), Caitlin Daly (McMaster University), Sharon Straus (St. Michael's Hospital)  
Integrating Ranking Probabilities in Network Meta-Analysis / Intégrer les probabilités de classement dans la méta-analyse en réseau  
- 16:00 - 16:30 Argie Veroniki (St. Michael's Hospital) (veronikiA@smh.ca)  
Addressing Consistency in Networks of Randomized Trials / Traiter de la cohérence dans les réseaux des essais aléatoires  

---

**15:00 - 16:30**    **Session 3D-I4**    (abstracts/résumés - page 267)    **Welch - David Howes**





**Statistical Methods for Analysing Cancer Data**



**Méthodes statistiques pour analyse des données sur le cancer**

Invited Paper Session / Présentations sur invitation

Chair/Président: Gregory Pond (McMaster University)

Organizer/Responsable: Amy Liu (Cancer Care Ontario)

- 15:00 - 15:30 Patrick Brown & Jamie Stafford (University of Toronto), Hedy Jiang (Cancer Care Ontario), Paul Nguyen & Lennon Li (Public Health Ontario), Reuben Pereira & Laura Feldman (University of Toronto)  
Spatial Statistics with Cancer Registry Data / Les statistiques spatiales avec les données des registres du cancer  
- 15:30 - 16:00 Karen Kopciuk (University of Calgary), Laurent Briollais (Lunenfeld-Tanenbaum Research Institute), Yun-Hee Choi (The University of Western Ontario), Patrick Parfrey & Jane Green (Memorial University)  
Risk Estimation in Family Data / Estimation du risque dans les données familiales  

16:00 - 16:30 Olli Saarela (University of Toronto), Zhihui (Amy) Liu (Cancer Care Ontario)  
A New Weighted Partial Likelihood Method for Estimating Marginal Structural Hazard Models  
/ Nouvelle méthode de vraisemblance partielle pondérée pour l'estimation des modèles de  
risques structurels marginaux  

---

**16:45 - 18:00**

**Thistle 253**

**Board of Directors 2**

**Conseil d'administration 2**

---

**18:00 - 19:00**

**Thistle 253**

**Executive Committee 2**

**Comité exécutif 2**

---

Abstracts • Résumés

**1A-A2: SSC Presidential Invited Address**  
**L'allocution de l'invité du président de la SSC**

---

**Room/Salle: Welch - David Howes**

**Abstract/Résumé**

---

**JOHN D. KALBFLEISCH** (University of Michigan)

[Monday May 30 / lundi 30 mai, 8:45]

*Some Statistical Topics of Personal Interest*

*Quelques sujets statistiques d'intérêt personnel*

I will discuss some statistical ideas and issues that have been recurring themes in my recent methodological research and work in applications. The topics covered will be somewhat idiosyncratic and will include issues of randomization and experimental design, fixed and random effects, empirical Bayes, and profiling of medical providers.

Je discuterai d'idées et de sujets statistiques qui constituent des thèmes récurrents dans mes récentes recherches en méthodologie et dans leurs applications. Les sujets couverts seront quelque peu idiosyncratiques et incluront la randomisation et la planification d'expériences, les effets fixes et aléatoires, les méthodes bayésiennes empiriques et le profilage des fournisseurs médicaux.

**1B-11: CANSSI at the SSC: A Statistical Science Showcase**  
**L'INCASS à la SSC : vitrine de la science statistique**

---

Organizer and Chair / Responsable et président: John Braun (University of British Columbia)

**Room/Salle: Welch - David Howes**

**Abstracts/Résumés**

---

**ALISHA ALBERT-GREEN** (University of Western Ontario), **JOHN BRAUN** (University of British Columbia Okanagan), **CHARMAINE DEAN** (University of Western Ontario)

[Monday May 30 / lundi 30 mai, 10:20]

*A Spatio-Temporal Cluster Process for Modelling Storm Cells*

*Un processus de regroupement spatio-temporel pour la modélisation des cellules orageuses*

Storm cells are the smallest component of a storm-producing system. A cluster of such cells is referred to as a storm and a storm system consists of a cluster of storms. This research develops a model for these storm cells over space and time. Specifically, we extend the modified Thomas process, which is commonly employed for the analysis of clustered point processes, to account for the hierarchical clustering present in our data. We do this by allowing the parents to follow a doubly stochastic process, namely a log-Gaussian Cox process. This model is applied to storm cell data from the Bismarck radar station in North Dakota, USA and parameter estimation is done using minimum contrast estimation.

Les cellules orageuses sont les plus petites composantes d'un système produisant des tempêtes. Un regroupement de ces cellules est appelé tempête et un système de tempêtes consiste en un regroupement de tempêtes. Cette recherche développe un modèle pour ces cellules orageuses à travers le temps et l'espace. Spécifiquement, nous élargissons le processus modifié de Thomas, qui est couramment employé dans l'analyse des processus de regroupement de points, pour prendre en compte le regroupement hiérarchique présent dans nos données. Pour ce faire, nous permettons aux parents de suivre un processus doublement stochastique, c'est-à-dire un processus Cox log-gaussien. Ce modèle est appliqué aux données de cellules orageuses provenant de la station radar Bismarck dans le Dakota du Nord, USA, et l'estimation de paramètres est effectuée en utilisant une estimation contraste minimum.

---

**DEREK BINGHAM** (Simon Fraser University)

[Monday May 30 / lundi 30 mai, 10:50]

*Design of Experiments in the CANSSI CRT ``Statistical Modeling of the World: Computer and Physical Models in Earth, Ocean, and Atmospheric Sciences''*

*Conception d'expériences dans le PRC « Modélisation statistique du monde : modèles informatiques et physiques en sciences de la Terre, des océans et de l'atmosphère » de l'INCASS*

Design and analysis of experiments continue to make important and far-reaching contributions to scientific investigation. Indeed, it is hard to imagine any field where it does not play a role. In the CANSSI Collaborative Research Team project ``Statistical Modeling of the World: Computer and Physical Models in Earth, Ocean, and Atmospheric Sciences'', experimental design is used for both physical and computer experiments. This talk will

La conception et l'analyse d'expériences continuent d'apporter des contributions importantes et de grande ampleur à la recherche scientifique. Il est en effet difficile d'imaginer un domaine où ces concepts ne jouent aucun rôle. Dans le Projet de recherche en collaboration « Modélisation statistique du monde : modèles informatiques et physiques en sciences de la Terre, des océans et de l'atmosphère » de l'INCASS, le plan expérimental sert à la fois pour les expériences physiques et informatiques. Cette présentation précisera le rôle

## 1B-11: CANSSI at the SSC: A Statistical Science Showcase L'INCASS à la SSC : vitrine de la science statistique

---

outline the role of experimental design for both settings, and a new design approach for exploring non-convex regions that are common in geophysical application will be presented.

du plan d'expériences dans les deux contextes et proposera une nouvelle approche de conception permettant d'explorer les régions non-convexes si communes dans les applications géophysiques.

---

**RADU V. CRAIU** (University of Toronto)

[Monday May 30 / lundi 30 mai, 11:20]

*Modern Research Topics in Copula Methods: A Report from the CANSSI Collaborative Team Project*

*Sujets de recherche modernes sur les méthodes de copules: un rapport de l'équipe de projet collaborative INCASS*

We will summarize some of the research activities undertaken by the members of the Copula CANSSI Collaborative Research Team: Christian Genest, Louis-Paul Rivest, Elif Acar, Harry Joe, Johanna Neslehova, Jean-Francois Quessy, Bruno Remillard and Radu Craiu.

Nous résumerons quelques activités de recherche entreprises par les membres de l'équipe de projet collaborative sur les copules INCASS : Christian Genest, Louis-Paul Rivest, Elif Acar, Harry Joe, Johanna Neslehova, Jean-Francois Quessy, Bruno Remillard et Radu Craiu.

Focus will be placed on our own research efforts on inference for conditional copula models in which the calibration function depends on multiple covariates. We will discuss a novel approach that relies on a Gaussian process prior used in conjunction with a single index model formulation. The methodology will be illustrated using simulations and real data.

Nous nous intéresserons principalement à notre effort de recherche sur l'inférence pour les modèles de copules conditionnels pour lesquels la fonction de redressement dépend de plusieurs covariables. Nous discuterons d'une approche novatrice qui s'appuie sur un processus a priori gaussien utilisé conjointement avec une formulation de modèle à indice unique. Cette méthodologie sera illustrée à l'aide de simulations et de données réelles.

## 1B-I2: Quantile regression and Extreme Value Analysis Régression quantile et analyse des valeurs extrêmes

---

Organizer and Chair / Responsable et présidente: Mei Ling Huang (Brock University)

Room/Salle: Welch 209

### Abstracts/Résumés

---

**VINCENZO COIA** (University of British Columbia), **HARRY JOE & NATALIA NOLDE** (The University of British Columbia)

[Monday May 30 / lundi 30 mai, 10:20]

*Forecasting Extremes for Flooding*

*Prévision des extrêmes pour les inondations*

When forecasting using past data, methods such as regression and maximum likelihood are available for reporting a predictive distribution that reflects what a typical outcome might be. However, these techniques are inadequate at structuring and fitting the predictive distribution's tail -- a shortfall when the end-user is only interested in what an extreme outcome might be. To get around these shortcomings, a copula-based nonlinear modelling technique and a composite quantile regression model-fitting technique are proposed. The result is an approach for building the tail of a predictive distribution that addresses questions like "how bad could it get?", and is applied to flood forecasting of the Bow River in Alberta.

Pour les prévisions utilisant des données historiques, les méthodes telles que la régression et le maximum de vraisemblance sont disponibles pour fournir une loi prédictive qui reflète ce qu'un résultat typique pourrait être. Cependant, ces techniques sont inadéquates pour structurer et faire correspondre les ailes de la loi prédictive -- un manque lorsque l'utilisateur final s'intéresse uniquement aux résultats extrêmes possibles. Pour contourner ces lacunes, une technique de modélisation non linéaire fondée sur les copules et une technique de régression quantile composite ajustée sur le modèle sont proposées. Le résultat est une approche pour la construction d'une aile de la loi prédictive qui aborde des questions telles que «jusqu'où cela va dégénérer?» et s'applique à la prévision d'inondations de la rivière Bow en Alberta.

---

**KEITH KNIGHT** (University of Toronto)

[Monday May 30 / lundi 30 mai, 10:50]

*Non-parametric Estimation of Extreme Conditional Quantiles and Lawson's Algorithm*

*L'estimation non paramétrique des quantiles conditionnels extrêmes et l'algorithme de Lawson*

Lawson's algorithm, developed in the early 1960s by C.L. Lawson, is an algorithm for computing  $L_\infty$  estimates in a linear regression model; it is perhaps the original iteratively reweighted least squares (IRLS) algorithm. We will explore the use of a similar IRLS algorithm to compute non-parametric estimates of extreme (and near-extreme) conditional quantiles.

L'algorithme de Lawson, développé au début des années 1960 par C.L. Lawson, est un algorithme pour calculer les estimateurs  $L_\infty$  dans un modèle de régression linéaire; il est peut-être le premier algorithme utilisant les moindres carrés itérativement repondérés (IRLS). Nous examinons l'utilisation d'un algorithme IRLS similaire afin de calculer des estimateurs non paramétriques des quantiles conditionnels extrêmes (et quasi-extrêmes).

---

**CHRISTINE NGUYEN, MEI LING HUANG & WAI KONG YUEN** (Brock University)

[Monday May 30 / lundi 30 mai, 11:20]

*On Weighted Quantile Regression*

*Régression quantile pondérée*



## 1B-I2: Quantile regression and Extreme Value Analysis Régression quantile et analyse des valeurs extrêmes

---

In recent years, studies of heavy tailed distributions have rapidly developed. For multivariate heavy tailed distributions, estimation of conditional quantiles at very high or low tails is of interest in numerous applications. Quantile regression uses an L1- loss function, and the optimal solution of linear programming for estimating coefficients of regression. This paper proposes a weighted quantile regression method on high quantile regression for certain extreme value sets. The Monte Carlo simulations show good results of the proposed weighted method. Comparisons of the proposed method and existing methods are given. The paper also investigates real-world examples by using the proposed weighted method.

Ces dernières années, les études sur les lois à ailes lourdes se sont rapidement développées. Pour les lois à ailes lourdes multivariées, l'estimation de quantiles conditionnels à ailes très hautes ou très basses est d'intérêt dans de nombreuses applications. La régression quantile utilise une fonction de perte L1, ainsi que la solution optimale de la programmation linéaire pour l'estimation des coefficients de régression. Cet article propose une méthode de régression quantile pondérée sur régression quantile élevée pour certains ensembles de valeurs extrêmes. Les simulations Monte-Carlo affichent de bons résultats pour la méthode pondérée proposée. Des comparaisons entre la méthode proposée et des méthodes existantes sont présentées. Cet article examine aussi des exemples concrets en utilisant la méthode pondérée proposée.

**1B-I3: Statistical Methods in Public Health**  
**Méthodes statistiques en santé des populations**

---

Chair/Président: Patrick Brown (University of Toronto)  
Organizer/Responsable: Lennon Li (Public Health Ontario)

**Room/Salle: Thistle 325**

**Abstracts/Résumés**

---

**ROB DEARDON** (University of Calgary), **WALEED ALMUTIRY** (University of Guelph)

[Monday May 30 / lundi 30 mai, 10:20]

*Infectious Disease Modelling in the Presence of Underlying Contact Network Uncertainty*

*Modélisation de maladies infectieuses en présence d'incertitude du réseau de contact sous-jacent*

Infectious disease transmission is often best modelled through contact networks, especially in human populations. These contact networks can represent sexual or social relationships, or shared employment, education or domestic status.

La transmission de maladies infectieuses est mieux modélisée par des réseaux de contact, particulièrement dans les populations humaines. Ces réseaux de contact peuvent représenter des relations sociales ou sexuelles, ou un emploi, une éducation ou un statut domestique partagés.

However, contact network data are often difficult and/or costly to collect; there may be privacy issues, or issues of incomplete or incorrect recall in surveys. As a result, contact network data desirable for the modelling may be missing entirely, only partially recorded, or recorded with uncertainty.

Cependant, les données des réseaux de contact sont souvent difficiles et/ou coûteuses à recueillir ; il peut y avoir des problèmes de protection de la vie privée, ou des problèmes dus à des rappels incomplets ou erronés dans les sondages. Par conséquent, des données de réseaux de contact désirées pour la modélisation peuvent être entièrement man-

Such missing/uncertain data can be incorporated within a data-augmented Bayesian framework using methods such as Markov chain Monte Carlo

quantes, partiellement enregistrées, ou enregistrées avec incertitude.

(MCMC). Here, we consider disease transmission modelling in the presence of uncertain contact networks, touching upon issues of computational difficulty.

Des telles données manquantes/incertaines peuvent être incorporées à l'intérieur d'un cadre bayésien augmenté en données en utilisant des méthodes Monte-Carlo par chaînes de Markov. Nous considérons la modélisation de la transmission de maladies en présence de réseaux de contact incertains, touchant aux questions de difficulté computationnelles.

---

**ANDREW LAWSON & RACHEL CARROLL** (Medical University of South Carolina)

[Monday May 30 / lundi 30 mai, 10:50]

*Spatio-temporal Latent Model Choice: With Application in Multiple Disease Small Area Analysis*

*Choix de modèles latents spatio-temporels : application dans plusieurs analyses de maladies de petits domaines*

Space-time variation in disease is the natural focus of many health studies. Conventionally the variation of the space-time risk fields is modeled via random effects. However, it is possible to consider spatial and temporal risk components se-

Les variations spatiotemporelles des maladies sont le centre d'intérêt naturel de nombreuses études. En général, la variation des zones de risques spatiotemporelles est modélisée au moyen d'effets aléatoires. Cependant, il est possible de considérer que les composantes spatiales et temporelles du

## 1B-I3: Statistical Methods in Public Health Méthodes statistiques en santé des populations

---

lected from predefined sets of terms and linked via mixing parameters. In addition there can be important information gained from the analysis of multiple diseases within this context, where shared components could inform about shared environmental risks. In this presentation I will demonstrate the development of multivariate spatio-temporal (MVST) linked models for analysis of shared risk. Some simulation results and a data example based on lung and bronchus cancer and melanoma in South Carolina (1996-2009) will be provided

risque sont sélectionnées à partir d'un ensemble de termes prédéfinis et reliés au moyen de paramètres de mélange. De plus, il est possible de gagner beaucoup d'information importante à partir de l'analyse de plusieurs maladies, dans un contexte où les composantes partagées pourraient informer sur les risques environnementaux partagés. Dans cet exposé, je présenterai le développement de modèles multivariés spatiotemporels pour l'analyse du risque partagé. Des résultats de simulations ainsi qu'un exemple de données basé sur le cancer du poumon et des bronches, ainsi que sur les mélanomes en Caroline du Sud (1996-2009) seront aussi présentés.

---

**RHONDA ROSYCHUK** (University of Alberta)

[Monday May 30 / lundi 30 mai, 11:20]

*Cluster Detection Tests for the Surveillance of Emergency Department Presentation Data*

*Tests de détection de grappes pour la surveillance des données de présentation des services des urgences*

Alberta has been one of the few provinces in Canada to have an extensive ambulatory care database that includes both hospital-based and community-based services. Presentations to Emergency Departments are part of the database and when linked with other provincial databases, excellent population-based analyses can result. The provincial health authority sometimes uses this data to provide analyses based on geography. While cluster detection tests have been popular in the statistical literature, they have been underutilized in this jurisdiction. This talk will focus on data issues, cluster detection tests, and knowledge translation activities in the context of Alberta Emergency Department presentation data.

L'Alberta est l'une des seules provinces canadiennes à avoir une base de données détaillée de ses soins ambulatoires qui comprend les services en milieu hospitalier et communautaire. Les présentations aux services des urgences font parties de la base de données et lorsque connectée avec d'autres base données provinciales, d'excellentes analyses basées sur la population peuvent en résulter. L'autorité sanitaire provinciale utilise parfois ces données pour produire des analyses basées sur la géographie. Alors que les tests de détection ont été populaires dans la littérature statistique, ils ont été sous-utilisés dans cette juridiction. Cet exposé sera axé sur les questions des données, les tests de détection de grappes, ainsi que les activités de translation des connaissances dans le contexte de données de présentations des services des urgences en Alberta.

## 1B-I4: Statistical Models in Actuarial Science Modèles statistiques en actuariat

---

Organizer and Chair / Responsable et président: Jean-Philippe Boucher (UQAM)

Room/Salle: Thistle 242

### Abstracts/Résumés

---

**MATHIEU PIGEON** (UQAM), **ARTHUR CHARPENTIER** (Université du Québec à Montréal et Université de Rennes 1)

[Monday May 30 / lundi 30 mai, 10:20]

*Macro and Micro Methods in Non-Life Claims Reserving*

*Modèles individuels et collectifs pour les réserves en assurance IARD*

Traditional methods for loss reserving in non-life insurance (Chain-Ladder, Bornhuetter Ferguson, London Chain, etc.) are constructed for aggregated data. Progress over the past decades in computing resources and greater availability of individual statistical data have resulted in the development of new models for individual dataset. In this talk, we will analyze the impacts of both approaches (macro and micro, respectively) on the evaluation of the reserve risk based on some parametric models (Poisson regression, quasi-Poisson regression, etc.) and generalized method of moments estimators. Results will be illustrated with simulations and a dataset from the industry.

La plupart des techniques traditionnelles permettant l'évaluation des réserves en assurances IARD (Chain-Ladder, Bornhuetter-Ferguson, London Chain, etc.) sont construites pour des données dites agrégées. Le progrès des outils informatiques et la disponibilité accrue de données détaillées ont permis récemment l'éclosion de techniques adaptées à des données individuelles (ou non agrégées). Dans cette présentation, nous analyserons les impacts de l'utilisation de données agrégées ou non agrégées sur les risques liés à l'évaluation des réserves. Nous ferons cette analyse à partir de modèles construits autour de la distribution de Poisson (régression Poisson, régression quasi-Poisson, etc.) et d'estimateurs basés sur la méthode des moments généralisés. Les résultats seront illustrés à l'aide de simulations et de données réelles provenant de l'industrie.

---

**PENG SHI** (University of Wisconsin), **LU YANG** (University of Wisconsin-Madison)

[Monday May 30 / lundi 30 mai, 10:50]

*A Pair Copula Construction Model for Insurance Experience Rating*

*Modèle de construction de copule par paires pour la tarification d'antécédents en assurance*

In non-life insurance, insurers use experience rating to adjust premium to reflect the policyholder's previous claim experience. Performing prospective experience rating can be challenging when the claim distribution is complex. In this article, we introduce a mixed vine pair copula construction framework for modeling semicontinuous longitudinal claims. Specifically, a two-component mixture regression is employed to accommodate the zero inflation and thick tails in claim distribution. The temporal dependence among repeated observations is modeled using a sequence of bivariate

Dans le cadre d'une assurance non-vie, les assureurs utilisent la tarification personnalisée pour ajuster les primes afin de refléter l'expérience précédente en matière de sinistres du titulaire de police. Les évaluations de l'expérience prospective peuvent être difficiles lorsque la distribution des demandes d'indemnisation est complexe. Dans cet exposé, nous présentons une construction de copules en vignes mixtes par paires pour la modélisation de demandes d'indemnisations semi-continues longitudinales. Notamment, une régression par mélange à deux composantes est employée pour prendre en compte des ailes épaisses à sur-représentation de zéros dans la distribution des demandes

## 1B-14: Statistical Models in Actuarial Science Modèles statistiques en actuariat

---

conditional copulas based on a mixed D-vine. In the application of government property insurance from the state of Wisconsin, we show that the proposed approach offers substantial opportunities for separating risks and identifying profitable business when compared with alternative experience rating methods.

d'indemnisation. La dépendance temporelle parmi les observations répétées est modélisée au moyen d'une séquence de copules conditionnelles bivariées en fonction d'une vigne D mixte. Dans le cadre de l'application de l'assurance des biens du gouvernement à partir de l'État du Wisconsin, nous démontrons que l'approche proposée offre des occasions substantielles de séparer les risques et identifier des affaires rentables lorsqu'on les compare à d'autres méthodes d'évaluation d'expérience.

---

**CARY CHI-LIANG TSAI** (Simon Fraser University), **TZULING LIN** (National Chung Cheng University)  
[Monday May 30 / lundi 30 mai, 11:20]

*Applications of the Bühlmann Credibility Model to Mortality Forecasting*

*Applications du modèle de crédibilité de Bühlmann à la prévision de la mortalité*

In this presentation, we first propose the Bühlmann credibility model to forecast mortality rates, and then compare forecasting performances between the proposed Bühlmann model and some leading mortality models. Empirical results based on mortality data for both genders of some country, a wide range of fitting year spans, and three forecasting periods show that the Bühlmann credibility model contributes to much better forecasting performances measured by the MAPE (mean absolute percentage error). Moreover, we give meaningful credibility interpretations regarding the decrement rates of an individual time trend for age  $x$  and a group time trend for all ages, and discuss the effects of the slope and intercept of the linear functions for the forecasted mortality rates under the underlying models.

Dans cette présentation, nous commençons par proposer que le modèle de crédibilité de Bühlmann peut servir à la prévision des taux de mortalité, puis nous en comparons la performance en matière de prévision à d'autres modèles de mortalité bien connus. Des résultats empiriques basés sur les données de mortalité pour les deux sexes dans un pays donné, diverses fourchettes d'années d'ajustement et trois périodes de prévision montrent que le modèle de crédibilité de Bühlmann contribue à des performances de prévision bien meilleures telles que mesurées par le MAPE (pourcentage d'erreur absolue moyen). Par ailleurs, nous donnons des interprétations valables de la crédibilité en ce qui concerne les taux de décrémentation d'une tendance temporelle individuelle pour un âge  $x$  et d'une tendance temporelle de groupe pour tous âges, avant de discuter des effets de la pente et de l'ordonnée des fonctions linéaires pour les taux de mortalité prévus par les modèles sous-jacents.

## 1B-15: Stochastic Modelling for Energy Markets Modélisation stochastique des marchés de l'énergie

---

Chair/Président: Reg Kulperger (University of Western Ontario)  
Organizer/Responsable: Matt Davison (University of Western Ontario)

**Room/Salle: Thistle 246**

### Abstracts/Résumés

---

**MATT DAVISON** (University of Western Ontario), **NICOLAS MERENER** (Universidad Torcuato di Tella)

[Monday May 30 / lundi 30 mai, 10:20]

*Corn/Ethanol Price Spreads Emerging from an Inhomogeneous Set of Price-Influencing Ethanol Producers*

*Écart de prix entre le maïs et l'éthanol dans un ensemble non homogène de producteurs d'éthanol ayant une influence sur les prix*

There is some evidence that the recent trend towards producing ethanol from corn has induced a correlation between corn and ethanol prices that was not previously seen. Corn ethanol producers may be considered to hold a real exchange option between the two commodities. We examine a simple model of a set of corn ethanol producers, with different efficiencies, whose activities in the corn and ethanol markets influence prices in those markets. We assume the facilities start (or are idled) in "merit order" of efficiency and examine the prediction of such a model on corn/ethanol spreads. We close by presenting our progress toward discerning related patterns in actual market data.

Il semble bien que la tendance récente à produire de l'éthanol à partir du maïs a conduit à une corrélation entre le prix du maïs et celui de l'éthanol qui n'existait pas auparavant. On peut considérer que les producteurs d'éthanol de maïs détiennent une option d'échange entre ces deux produits. Nous examinons un modèle simple d'un ensemble de producteurs d'éthanol de maïs, plus ou moins efficaces, dont les activités sur les marchés du maïs et de l'éthanol en influencent les prix. Nous supposons que les installations démarrent (ou sont mises hors service) par « ordre de mérite » de leur efficacité et examinons la prévision par ces modèles des écarts de prix entre le maïs et l'éthanol. Pour conclure, nous présentons nos progrès vers l'identification de tendances liées dans les données de marché réelles.

---

**ANATOLIY SWISHCHUK** (University of Calgary)

[Monday May 30 / lundi 30 mai, 10:50]

*Stochastic Modelling and Pricing of Energy Markets' Contracts with Local Stochastic Delayed and Jumped Volatilities*

*Modélisation stochastique et tarification des contrats sur les marchés de l'énergie avec volatilités stochastiques locales à délais et à sauts*

In this talk we study stochastic modelling and pricing of electricity, gas and temperature markets' contracts with delay and jumps, modelled by independent increments processes. The spot price models are based on a sum of non-Gaussian Ornstein-Uhlenbeck processes (including geometric and arithmetic models), describing the long- and short-term fluctuations of the spot dynamics. The models include jumps not only in the spot dynamics but also in the stochastic volatility to describe typical features like spikes of energy spot prices and

Dans cette présentation, nous étudions la modélisation stochastique et la tarification des contrats sur les marchés de l'électricité, du gaz et de la température avec délais et sauts, telles que modélisées par des processus d'incrémentés indépendants. Les modèles de prix au comptant (spot) sont basés sur une somme de processus d'Ornstein-Uhlenbeck non gaussiens (notamment des modèles géométriques et arithmétiques), qui décrivent les fluctuations à court et long terme de la dynamique du spot. Les modèles incluent des sauts non seulement dans la dynamique du spot, mais aussi dans la volatilité stochastique pour décrire des caractéristiques ty-

## 1B-15: Stochastic Modelling for Energy Markets Modélisation stochastique des marchés de l'énergie

---

jumps in the volatility. The basic products in these markets are spot, futures and forward contracts, swaps and options written on these, which will be investigated in our talk.

piques comme les pics de prix de l'énergie au comptant et les sauts de volatilité. Les produits de base sur ces marchés sont les contrats spot, à terme et à livrer, les swaps et les options sur ces contrats, que nous étudierons ici.

---

**HANS TUENTER** (University of Toronto)

[Monday May 30 / lundi 30 mai, 11:20]

*Firming Wind Farm Revenues*

*Stabilisation des revenus des parcs éoliens*

Wind is an intermittent energy source and introduces randomness in a wind farm's power output and revenues. We show how tailored financial products, such as production puts or swaps, can increase the economic viability of a wind farm. To price these products, we use historical simulation of wind speed in combination with a power curve. A case study will be presented to illustrate the approach.

Le vent est une source d'énergie intermittente, ce qui introduit de la variabilité dans la production d'énergie et de revenus des parcs éoliens. Nous démontrons comment des produits financiers adaptés, comme des «puts» ou des «swaps» de production, peuvent améliorer la viabilité économique d'un parc éolien. Pour calculer le prix de ces produits, nous utilisons des simulations historiques de vitesse du vent ainsi qu'une courbe de puissance. Une étude de cas permettra d'illustrer cette approche.

## **1B-16: What Do School Teachers and Students Need from the SSC? Que peut faire la SSC pour les enseignants et les élèves ?**

---

Chair/Président: Lorna Deeth (University of Guelph)  
Organizer/Responsable: Alison Gibbs (University of Toronto)

**Room/Salle: Thistle 258**

### **Description**

The purpose of this panel session is to explore the role of the SSC, both current and potential, in statistics education in Canadian schools. First, we will discuss the current pedagogy of training future teachers in the areas of mathematics and statistics, at both the elementary/junior and secondary levels, with emphasis on the required critical thinking, verbal, and evaluation skills. This will be followed by a discussion on the current state of statistics education at the secondary level, and the challenges faced by mathematics educators. Finally, a discussion on the statistics education curriculum in Ontario, and the potential role of the SSC in future developments, will be presented. The panelists include educators with strong connections to schools and teachers, who will share their experiences and open a dialogue about what the Statistical Society of Canada and its members can and should do to support statistics education, and educators, in Canada.

Cette réunion de spécialistes a pour but d'explorer le rôle de la Société statistique du Canada (SSC), à la fois actuel et éventuel, dans l'enseignement de la statistique dans les écoles canadiennes. Nous aborderons tout d'abord la pédagogie en vigueur pour la formation des futurs enseignants dans les domaines des mathématiques et des statistiques, à la fois au primaire et au secondaire, en insistant sur les aptitudes requises sur le plan de la pensée critique, de l'expression verbale et de l'évaluation. Nous traiterons ensuite de l'état actuel de l'enseignement de la statistique au secondaire et des défis auxquels font face les professeurs de mathématiques. Enfin, nous ferons le point sur le curriculum d'éducation statistique en Ontario et sur le rôle potentiel de la SSC dans leur élaboration future. Parmi les panélistes se trouvent des éducatrices en relation étroite avec les établissements scolaires et les enseignants qui viendront faire part de leur expérience, tout en amorçant un échange sur ce que peuvent et doivent faire la SSC et ses membres pour appuyer l'enseignement de la statistique, et ses éducatrices, au Canada.



# 1B-C1: Biostatistics: Methodological Innovation 1

## Biostatistique : innovation méthodologique 1

---

Chair/Président: Caitlin Daly (McMaster University)

Room/Salle: Thistle 259

### Abstracts/Résumés

---

**YING WU & RICHARD J. COOK** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 10:20]

*A Two-Phase Model For Chronic Disease Processes Under Intermittent Inspection*

*Modèle à deux phases pour les processus de la maladie chronique sous inspection intermittente*

A two-phase model is developed for chronic diseases with an indolent phase which is followed by a phase with more active disease resulting in progression and damage. Weakly parametric models with piecewise constant baseline hazard and rate functions are specified and an expectation-maximization algorithm is described for model fitting. Simulation studies examining the performance of the proposed model show good performance under maximum likelihood and two-stage estimation. An application to data from the motivating study of disease progression in psoriatic arthritis illustrates the procedure and identifies new human leukocyte antigens associated with the duration of the indolent phase.

Un modèle à deux phases est mis au point pour les maladies chroniques avec une phase indolente suivie d'une phase plus active de la maladie qui entraîne son évolution et des dommages. Des modèles paramétriques faibles avec des fonctions de risque de base et de taux constantes par morceaux sont précisés et un algorithme d'espérance-maximisation est décrit pour l'ajustement du modèle. Des études de simulation portant sur le modèle proposé montrent la bonne performance sous une estimation par vraisemblance maximale et en deux étapes. Une application aux données tirées de l'étude de l'évolution de la polyarthrite psoriasique illustre la procédure et identifie de nouveaux antigènes des leucocytes humains associés à la durée de la phase indolente.

---

**WENYAN ZHONG, JINGJING WU & XUEWEN LU** (University of Calgary)

[Monday May 30 / lundi 30 mai, 10:35]

*Group Selection in Proportional Odds Model for Right-Censored Data*

*Sélection de groupe dans le modèle des cotes proportionnelles pour les données censurées à droite*

We consider the problem of variable selection for proportional odds model with right-censored data. Proportional odds model is an important alternative to Cox model when proportional hazards assumption is not satisfied. To fit the proportional odds model, we proposed an estimation by minimizing a negatively weighted partial log-likelihood subject to a group bridge penalty. This penalty encourages sparse solutions and selects significant groups and, simultaneously, individual variables within significant groups. To implement this newly proposed method, we developed a computational algorithm that has been shown efficient through

Nous soupons le problème de la sélection des variables pour le modèle proportionnel de cotes pour les données censurées à droite. Ce dernier modèle est une importante solution de rechange au modèle de Cox lorsque l'hypothèse des risques proportionnels n'est pas satisfaite. Pour ajuster le modèle des cotes proportionnelles, nous proposons une estimation en minimisant un logarithme de vraisemblance partielle pondéré négativement et assujetti à une pénalité liée au groupe. Cette pénalité encourage les solutions à faible densité et sélectionne des groupes significatifs et, simultanément, des variables individuelles dans les groupes significatifs. Pour mettre en œuvre cette nouvelle méthode proposée, nous avons formulé un algorithme computationnel dont

## 1B-C1: Biostatistics: Methodological Innovation 1

### Biostatistique : innovation méthodologique 1

---

simulation studies. Some theoretical properties of the proposed method were also presented.

des études en simulation ont montré l'efficacité. Nous avons aussi présenté certaines propriétés théoriques de cette méthode.

---

**STEVE FERREIRA GUERRA, MIREILLE SCHNITZER & LUCIE BLAIS** (Université de Montréal)

[Monday May 30 / lundi 30 mai, 10:50]

*A Self-Selecting Procedure for the Optimal Discretization of a Continuous Timeline for Longitudinal Causal Inference Methods*

*Une procédure de sélection automatique pour la discrétisation optimale d'une ligne de temps continue pour l'application de méthodes longitudinales en inférence causale*

In health care research, administrative databases have become abundantly used to conduct causal inference for drug effects. In longitudinal settings, we generally rely on a discretization of the patient timeline and bias may result when coarsening is arbitrarily chosen by the researcher. This is partially due to discarded information about time-dependent confounders. Using a Longitudinal Targeted Maximum Likelihood Estimation loss based function, we developed a cross-validation self-selecting procedure of the optimal discretization of a continuous timeline. We use a simulation study to evaluate the bias-variance tradeoff of such a method.

En santé, les bases de données administratives sont abondamment utilisées afin d'appliquer l'inférence causale aux effets de médicaments. Dans un contexte longitudinal, nous discrétisons généralement la ligne du temps et un biais peut survenir lorsque le raffinement du pas est arbitrairement choisi par le chercheur, biais partiellement attribuable à la perte d'information sur les variables de confusion dépendantes du temps. Utilisant une fonction de perte basée sur l'estimation par maximum de vraisemblance ciblée longitudinale, nous avons développé une méthode de validation croisée auto-sélective pour la discrétisation optimale d'une ligne de temps continue. Nous évaluons le compromis biais-variance de cette méthode par simulation.

---

**ZELALEM NEGERI, MATEEN SHAIKH & JOSEPH BEYENE** (McMaster University)

[Monday May 30 / lundi 30 mai, 11:05]

*Bivariate Random Effects Meta-Analysis Models for Diagnostic Test Accuracy Studies Using Arcsine-Based Transformations*

*Modèles de meta-analyse bivariés à effets aléatoires pour les études sur l'exactitude des tests diagnostiques utilisant les transformations arc-sinus*

The bivariate random-effects model using the logit transformation is commonly employed to synthesize the sensitivity and specificity of diagnostic test accuracy studies. We propose the arcsine square root and the Freeman-Tukey double arcsine transformation to overcome some shortcomings of the logit transformation. We evaluated the performance of the three transformations using extensive simulations in terms of bias, root mean square error and coverage probability. We varied several parameters including number of studies and sample size. The proposed transformations outperformed the logit transformation in terms of all of the performance criteria. The methods have also been illustrated using real data sets.

Le modèle bivarié à effets aléatoires utilisant la transformation logit est souvent employé afin de synthétiser la sensibilité et la spécificité des études sur l'exactitude des tests diagnostiques. Nous proposons la racine carrée de l'arc-sinus et la transformation de Freeman-Tukey à double arc-sinus afin de surmonter certaines lacunes de la transformation logit. Nous avons évalué l'efficacité des trois transformations au moyen de simulations permettant d'évaluer le biais, l'erreur quadratique moyenne et la probabilité de couverture. Nous avons varié plusieurs paramètres, dont le nombre d'études et la taille de l'échantillon. Les transformations proposées ont surpassé la transformation logit concernant tous les critères d'efficacité. Nous avons également illustré les méthodes proposées à l'aide de jeux de données réelles.

## 1B-C1: Biostatistics: Methodological Innovation 1 Biostatistique : innovation méthodologique 1

---

**MAXIME TURGEON** (McGill University), **KARIM OUALKACHA** (Université du Québec à Montréal), **ANTONIO CIAMPI, GOLSA DEGHAN, CELIA GREENWOOD & AURÉLIE LABBE** (McGill University)

[Monday May 30 / lundi 30 mai, 11:20]

*Principal Component of Explained Variance: An Efficient and Optimal Data Dimension Reduction Framework for Association Studies*

*Composante principale de variance expliquée: une méthode de réduction dimensionnelle efficace et optimale pour les études d'association*

The genomics era has led to an increase in the dimensionality of the data collected. In this context, dimension-reduction techniques can be used to summarize high-dimensional signals, to further test for association with the covariates of interest. We revisit one such approach, renamed here as Principal Component of Explained Variance (PCEV). This method seeks a linear combination of outcomes by maximising the proportion of variance explained by the covariates of interest. We propose a general analytical framework that is conceptually simple and free of tuning parameters, and we provide a simple computational strategy for high-dimensional outcomes, along with testing procedures.

L'ère génomique a entraîné une augmentation de la dimension des données. Dans ce contexte, les techniques de réduction dimensionnelle peuvent être utilisées pour réduire un signal multidimensionnel, et ensuite tester l'association avec des covariables d'intérêt. Nous revisitons une de ces méthodes, rebaptisée Composante Principale de Variance Expliquée. Cette méthode cherche la combinaison linéaire des variables réponses qui maximise la proportion de variance expliquée par les covariables. Nous proposons un cadre analytique général qui est conceptuellement simple et dépourvu de paramètres de réglage. Nous proposons aussi une stratégie computationnelle simple pour des variables réponses de grande dimension, accompagnée de procédures pour tests d'hypothèse.

**MARIE-HÉLÈNE TOUPIN** (Université Laval), **JEAN-FRANÇOIS QUESSY** (Université du Québec à Trois-Rivières), **LOUIS-PAUL RIVEST** (Université Laval)

[Monday May 30 / lundi 30 mai, 11:35]

*On the Family of Chi-Square Copulas*

*À propos de la famille de copules Khi-carré*

The family of Chi-square copulas that has recently appeared in the literature is very attractive because it generalizes the Gaussian copula and allows for flexible modeling for high dimensional random vectors. This presentation will explore the theoretical properties and the practical usefulness of this class of dependence structures. An application of the Chi-square copulas will also be developed, namely parameter estimation.

La famille de copules Khi-carré qui a récemment fait son apparition dans la littérature est très attirante, car elle généralise la copule normale et permet la modélisation flexible de vecteurs aléatoires de grande dimension. Dans cette présentation, on explorera les propriétés théoriques et l'utilité pratique de cette classe de structures de dépendance. Une application de la copule Khi-carré sera également développée, à savoir l'estimation de paramètre.

## 1B-C2: Biostatistics: Methods and Applications 1 Biostatistique : méthodes et applications 1

---

Chair/Président: Ehsan Karim (McGill University)

Room/Salle: Thistle 255

### Abstracts/Résumés

---

**JILL AINSWORTH, ERICA MOODIE & ABBAS KHALILI** (McGill University)

[Monday May 30 / lundi 30 mai, 10:20]

*Estimating Parameters from a Finite Mixture of Generalised Linear Mixed-Effect Models Using the Maximum Likelihood Method*

*Estimer des paramètres d'un mélange fini de modèles linéaires généralisés à effets mixtes en utilisant la méthode de maximum de vraisemblance*

Consider a dataset with the following features that the analysis should take into account: (i) discrete outcomes, (ii) the longitudinal measurements, and (iii) possibility of subpopulations across which covariate effects differ. In this paper, I propose the use of a finite mixture of generalised linear mixed-effect models (FinMix GLMM) as an appropriate analytic approach. The performance of maximum likelihood estimator of the model parameters are explored by a simulation study as well as an analysis of data from the Scottish Early Rheumatoid Arthritis (SERA) cohort.

Considérez un ensemble de données avec les caractéristiques suivantes, qu'une analyse doit prendre en considération : (i) résultats discrets, (ii) les mesures longitudinales, et (iii) la possibilité de sous-populations à travers lesquelles les effets des covariables diffèrent. Dans cet article, je propose l'utilisation d'un mélange fini de modèles généralisés à effets mixtes (FinMix GLMM) comme approche analytique appropriée. La performance de l'estimateur de maximum de vraisemblance des paramètres du modèle est étudiée à l'aide d'une étude de simulation et d'une analyse des données de la cohorte écossaise de l'arthrite rhumatoïde précoce (SERA).

---

**JOYCELYNE EFUA EWUSIE** (McMaster University), **MONIKA KASTNER** (Li Ka Shing Knowledge Institute, St Michaels Hospital, Toronto), **GEORGE THOMPLINSON** (University of Toronto, Toronto), **SHARON STRAUS & JEMILA S. HAMID** (Li Ka Shing Knowledge Institute, St Michaels Hospital, Toronto)

[Monday May 30 / lundi 30 mai, 10:35]

*Methods for Interrupted Time Series Analysis of Count Data*

*Méthodes pour l'analyse des séries chronologiques interrompues de données de dénombrement*

This study evaluates different methods of analyzing count data in interrupted time series (ITS) designs. Fixed, random and zero-inflated Poisson models were fitted to an empirical data obtained from a previous ITS study evaluating the effect of an implementation tool to improve osteoporosis disease management. Extensive simulations with different scenarios were also performed to investigate performance. Preliminary results show that the fixed effects and zero-inflated Poisson models were best-fitting for estimating the effect of intervention on number of patients given BMD testing

Cette étude évalue les différentes méthodes d'analyse de données de dénombrement dans les modèles de séries chronologiques interrompues (SCI). Les modèles de Poisson fixes, aléatoires et à surreprésentation de zéros étaient ajustés à des données empiriques d'une étude SCI précédente évaluant l'effet d'un outil de mise en œuvre pour améliorer la gestion de la maladie de l'ostéoporose. Des études de simulations approfondies ont également été réalisées sous différents scénarios pour étudier la performance. Les résultats préliminaires montrent que les modèles de Poisson à effets fixes et à surreprésentation de zéros conviennent le mieux pour estimer l'effet de l'intervention sur le nombre de

## 1B-C2: Biostatistics: Methods and Applications 1

### Biostatistique : méthodes et applications 1

---

and osteoporosis medication, respectively.

patients ayant reçu respectivement un test de DMO et des médicaments contre l'ostéoporose.

---

**TANIA ALARCON FALCONI** (Civil and Environmental Engineering, Tufts University), **BERTHA ESTRELLA** (Corporación Ecuatoriana de Biotecnología, Ecuador), **FERNANDO SEMPÉRTEGUI** (Central University, Ecuador), **MAGALY KOCH** (Center for Remote Sensing, Boston University), **ELENA NAUMOVA** (Friedman School of Nutrition Science and Policy, Tufts University)

[Monday May 30 / lundi 30 mai, 10:50]

*Characterization of Seasonality in Longitudinal Clinical Trials*

*Caractérisation des saisonnalités dans les essais cliniques longitudinaux*

The impact of disease prevention interventions can be affected by characteristic temporal patterns in probability and dose of exposures. Lack of proper accounting for such seasonal dynamics can affect observed disease incidence and lead to misinterpretation of intervention results. We characterized seasonality of respiratory and diarrheal infections in a longitudinal clinical trial of supplementation in children in Ecuador using mixed effects models. We accounted for cohort age effects and associated cross-immunity by considering different temporal components and children's age. We also explored temporal effects of climatic and meteorological parameters (temperature, precipitation, UV radiation, photoperiod, cloud cover) on cohort seasonal behavior.

L'impact des interventions de prévention des maladies peut être affecté par des motifs temporels caractéristiques au niveau des probabilités et des doses d'exposition. Le manque de bonne comptabilisation de ces dynamiques saisonnières peut affecter l'incidence de la maladie observée et conduire à une mauvaise interprétation des résultats d'intervention. Nous avons caractérisé la saisonnalité des infections respiratoires et diarrhéiques dans un essai clinique longitudinal de la supplémentation chez les enfants en Équateur en utilisant des modèles à effets mixtes. Nous avons tenu compte des effets de l'âge de la cohorte et de l'immunité croisée associée en considérant différentes composantes temporelles et l'âge des enfants. Nous avons également exploré les effets temporels des paramètres climatiques et météorologiques (température, précipitations, rayonnement UV, photopériode, couverture nuageuse) sur le comportement saisonnier de la cohorte.

---

**KHAN MOHAMMAD KAVIUL, RAFAL KUSTRA & ZHENGFEI CHEN** (Dalla Lana School of Public Health, University of Toronto)

[Monday May 30 / lundi 30 mai, 11:05]

*Investigation of Change in Methylation in Prostate Cancer Patients after Receiving Treatment*

*Étude du changement dans la méthylation chez les patients atteints de cancer de la prostate après avoir reçu un traitement*

An experimental study was conducted on 33 prostate cancer patients undergoing chemotherapy. The patients were assigned to a treatment. Based on changes in their prostate-specific-antigen (PSA) responder or non-responder to the treatment was declared. Methylation status from blood sample was collected before treatment and weekly for seven weeks after treatment. We divided the whole genome into 1744 chromosome regions. We tested whether change in methylation status over time in a specific chromosome region varies in re-

Une étude expérimentale a été réalisée sur 33 patients atteints d'un cancer de la prostate subissant une chimiothérapie. Les patients ont été assignés à un traitement. En fonction des variations de leur antigène prostatique spécifique (PSA), ils ont été déclarés répondants ou non-répondants au traitement. L'état de méthylation de leur échantillon de sang a été prélevé avant le traitement puis chaque semaine pendant les sept semaines suivant le traitement. Nous avons divisé l'ensemble du génome en 1 744 régions chromosomiques. Nous avons vérifié si le changement dans leur statut de méthylation au fil du temps dans une région spéci-

## 1B-C2: Biostatistics: Methods and Applications 1

### Biostatistique : méthodes et applications 1

---

sponders and non-responders using linear mixed effect model. Since there were 1744 such regions adjustment was done for multiple comparisons.

fique du chromosome varie chez les répondants et les non-répondants à l'aide d'un modèle linéaire à effet mixte. Étant donné qu'il y avait 1 744 régions, un ajustement a été fait pour les comparaisons multiples.

---

**VICTORIA BORG-DEBONO, PHILIP J. DEVEREAUX, LAWRENCE MBUAGBAW & LEHANA THABANE** (McMaster University)

[Monday May 30 / lundi 30 mai, 11:20]

*Interim Data Sharing: What Interim Results/Information Should Be Shared During the Interim of a Randomized Controlled Trial with Those Responsible for the Conduct of the Trial?*

*Partage de données intérimaires : quels résultats/information intérimaires devraient être partagés avec ceux qui mènent l'essai lors d'un essai contrôlé randomisé ?*

It is unclear what interim trial information a Data Safety Monitoring Board (DSMB) should share with non-DSMB members during a trial's conduct. To understand the opinions of those involved in clinical trials, an online survey targeted at trialists and statisticians was conducted in 2015. The total response rate was 17.8%. Preliminary results suggest that the majority of respondents believe that the interim combined event rate should be shared (65% [95%CI: 60% to 71%]) with various parties responsible for a trial's conduct. Further analyses on the reasons for/against sharing interim information will be presented. Low response rate is a limitation.

On ne sait pas quelle information d'un essai intérimaire un comité indépendant de contrôle des données (CICD) devrait partager avec les membres non-CICD pendant l'essai. Afin de comprendre les opinions de ceux impliqués dans les essais cliniques, une enquête en ligne ciblant les spécialistes des essais et les statisticiens a été réalisée en 2015. Le taux de réponse total était de 17,8%. Les résultats préliminaires suggèrent que la majorité des répondants croient que le taux d'événements combinés intérimaire devraient être partagés (65% [IC à 95% : 60% à 71%]) avec les différentes parties responsables de la mise en œuvre de l'essai. Des analyses plus poussées sur les raisons pour/contre le partage de l'information intérimaire seront présentées. Le faible taux de réponse est une limite.

---

**SHAHRIAR SHAMS** (University of Toronto), **ELEANOR PULLENAYEGUM** (CHES, The Hospital for Sick Children)

[Monday May 30 / lundi 30 mai, 11:35]

*Evaluate the Feasibility of Bayesian Inference as a Means of Reducing Scoring Uncertainty for the EQ-5D Health Utilities*

*Évaluer la faisabilité de l'inférence bayésienne pour réduire la cotation de l'incertitude pour l'utilité de santé EQ-5D*

Health utilities are the quality weights used in estimating quality-adjusted life years, and can be measured using the EQ-5D-3L questionnaire which classifies a respondent into one of 243 health states. The health states are converted to utilities through a prediction equation, developed in a valuation study that directly valued 43 health states. Prediction error is large and is currently ignored. Model mis-specification ( $\delta$ ) explains 84% of the mean squared prediction error. Incorporating the posterior mean of  $\delta$  substantially increased the precision of the valued-states. Furthermore, modelling the correlation structure of  $\delta$ 's may increase

Les utilités de santé sont les pondérations des qualités, utilisées dans l'estimation des années de vie ajustée en fonction de la qualité. Elles peuvent être mesurées au moyen du questionnaire EQ-5D-3L qui classe un répondant dans l'un des 243 états de santé possibles. Les états sont convertis au moyen d'une équation de prédiction développée dans une étude d'évaluation qui a directement valorisé 43 états de santé. L'erreur de prédiction est grande et n'actuellement pas prise en compte. L'erreur de spécification du modèle ( $\delta$ ) explique 84 % de l'erreur quadratique moyenne de prédiction. L'intégration de la moyenne a posteriori de  $\delta$  permet d'augmenter substantiellement la précision des états évalués. De plus, la modélisation de la structure de corrélation

## 1B-C2: Biostatistics: Methods and Applications 1

### Biostatistique : méthodes et applications 1

---

the precision of the 200 unvalued-states.

des  $\delta$  pourrait augmenter la précision des 200 états non-évalués.

## 1B-C3: Statistical Methods Utilizing Copulas Méthodes statistiques utilisant des copules

---

Chair/Président: Zhenhua Lin (University of Toronto)

Room/Salle: Thistle 256

### Abstracts/Résumés

---

**ELIF ACAR** (University of Manitoba), **CANDIDA GEERDENS & PAUL JANSSEN** (Hasselt University)  
[Monday May 30 / lundi 30 mai, 10:20]

*Conditional Copula Models for Right-Censored Clustered Event Time Data*

*Modèles de copules conditionnelles pour données historiques d'événements groupés censurés à droite*

This work proposes a modelling strategy to infer the impact of a covariate on the dependence structure of right-censored clustered event time data. The joint survival function of the event times is modelled using a parametric conditional copula whose parameter depends on a cluster-level covariate. We use a local likelihood approach to estimate the form of the copula parameter and outline a generalized likelihood ratio-type test strategy to formally test its constancy. We apply the methods to data from the Diabetic Retinopathy Study to assess the impact of disease onset age on the loss of visual acuity.

Cette étude propose une stratégie de modélisation pour déterminer l'impact d'une covariable sur la structure de dépendance des données historiques d'événements groupés censurés à droite. Cette fonction de survie conjointe des temps d'événements est modélisée en utilisant une copule conditionnelle paramétrique dont le paramètre dépend d'une covariable au niveau de la grappe. Nous utilisons une approche de vraisemblance locale pour estimer la forme du paramètre de la copule et pour décrire une stratégie de test du rapport de vraisemblance généralisée pour tester formellement sa constance. Nous appliquons les méthodes à des données provenant de l'étude sur la rétinopathie diabétique pour évaluer l'impact de l'âge d'apparition de la maladie sur la perte d'acuité visuelle.

---

**TARIK BAHRAOUI** (Université de Sherbrooke)  
[Monday May 30 / lundi 30 mai, 10:35]

*A Family of Goodness-of-fit Tests for Copulas Based on Characteristic Functions*

*Une famille de tests d'adéquation pour des copules fondée sur les fonctions caractéristiques*

A general class of rank statistics based on the characteristic function is introduced for testing composite goodness-of-fit hypotheses about multivariate copula families. These statistics are defined as  $L_2$  weighted functional distances between a non-parametric estimator and a semi-parametric estimator of the characteristic function associated to a copula. It is shown that these statistics behave asymptotically as a degenerate  $V$ -statistic of order 4 and that the limit distribution is, up to a constant, a weighted sum of independent chi-square random variables. A parametric bootstrap is suggested. At the end, simulation study and illustrated on a multivariate dataset

Une classe générale des statistiques de rangs fondée sur la fonction caractéristique est présentée pour tester les hypothèses d'adéquation composées sur les familles de copules multivariées. Ces statistiques sont définies en tant que distances fonctionnelles pondérées  $L_2$  entre un estimateur non-paramétrique et un estimateur semiparamétrique de la fonction caractéristique associée à une copule. Il est démontré que ces statistiques se comportent asymptotiquement comme une statistique- $V$  dégénérée d'ordre 4 et que la loi limite est, jusqu'à une constante, une somme pondérée de variables aléatoires indépendantes khi carrée. Un bootstrap paramétrique est suggéré. En terminant, une étude de simulation est illustrée avec un ensemble de données multivariées.



## 1B-C3: Statistical Methods Utilizing Copulas Méthodes statistiques utilisant des copules

---

**MANOJ BAHUGUNA** (Oakland University, Rochester, MI), **RAVINDRA KHATTREE** (Oakland University, Rochester, MI)

[Monday May 30 / lundi 30 mai, 10:50]

*Copula Transformation, Prediction and Exploration*  
*Transformation avec copule, prédiction et exploration*

Copulas have been used in various applications in biomedical sciences and finance. In this talk, we suggest copula as a generic all-purpose transformation which can enable one to apply various standard procedures more efficiently and with better statistical properties and results. We evaluate and illustrate various applications including those in regression, principal component analysis and factor analysis, where analysis using the copula transformation results in improvement in implementation, interpretation as well as prediction. Specifically, emphasis is in introducing the multivariate normality using this all purpose transformation.

Les copules ont été utilisées dans plusieurs applications des sciences biomédicales et en finance. Dans cet exposé, nous recommandons la copule comme transformation générique tout usage qui peut rendre possible l'application plus efficace de différentes procédures standards avec de meilleurs propriétés statistiques et résultats. Nous évaluons et illustrons diverses applications y compris celles en régression, en analyse de composante principale et en analyse factorielle, où l'analyse utilisant la transformation avec copule a pour résultat l'amélioration de l'exécution, de l'interprétation ainsi que de la prédiction. Spécifiquement, l'accent est mis sur la présentation de la normalité en utilisant cette transformation tout usage.

**CAREN HASLER & RADU V. CRAIU** (University of Toronto), **LOUIS-PAUL RIVEST** (Université Laval)

[Monday May 30 / lundi 30 mai, 11:05]

*Vine Copulas for Imputation of Monotone Non-Response*  
*Copules en vignes pour l'imputation de la non-réponse monotone*

We investigate the use of vines to impute for monotone non-response. We use vine copulas and factorize the density of the observed variables into a cascade of bivariate copulas to flexibly model their joint distribution. The structure of the vine depends on the non-response pattern. We build on the work of Aas et al. (2009) and propose a method to estimate the parameters of the bivariate copulas entering into the model. Imputations are carried out by simulating the missing values using the constructed model. We discuss the generalization of our results to more global non-response patterns. Project supported by CANSSI.

Nous explorons l'application des copules en vignes pour l'imputation de la non-réponse monotone. Nous considérons des copules en vignes et factorisons la densité des variables observées en une cascade de copules par paire afin de modéliser leur distribution jointe avec une grande flexibilité. Nous nous basons sur les travaux de Aas et al. (2009) et proposons une méthode d'estimation des paramètres du modèle. L'imputation se fait par simulation des observations manquantes au moyen du modèle construit. Nous discutons l'extension de notre travail à des répartitions de données manquantes plus générales. Projet financé par l'INCASS.

**DAVID LEE & HARRY JOE** (University of British Columbia)

[Monday May 30 / lundi 30 mai, 11:20]

*Multivariate Extreme Value Copulas with Factor and Tree Dependence Structures*  
*Copules de valeurs extrêmes multivariées avec structures factorielles et arborescentes de dépendance*

In multivariate modelling, factor and tree dependence structures are parsimonious assumptions often reasonable in practice, for example in finance

Dans la modélisation multivariée, les structures factorielles et arborescentes de dépendance sont des hypothèses parcimonieuses souvent raisonnables dans la pratique comme

## 1B-C3: Statistical Methods Utilizing Copulas Méthodes statistiques utilisant des copules

---

where stock prices are driven by common latent factors, or in spatial applications where stations form a tree network. We propose copula extensions for extreme observations that incorporate such structures. These models allow intuitive interpretation of dependence relationships underlying the processes that generate the extremes; this is illustrated through a data example, which also suggests that our proposed models may perform better than a fully unconstrained one.

dans le domaine de la finance où les cours de l'action sont dirigés par des facteurs latents communs, ou bien dans des applications spatiales où les stations forment un réseau arborescent. Nous proposons des extensions de copules d'observations extrêmes qui incorporent de telles structures. Ces modèles permettent une interprétation intuitive de la relation de dépendance sous-jacente aux processus qui génèrent les extrêmes. Nous donnons un exemple de données qui suggère également que les modèles que nous proposons sont plus efficaces qu'un modèle sans contrainte.

---

**YASSIR RABHI & TAOUFIK BOUEZMARNI** (Université de Sherbrooke)

[Monday May 30 / lundi 30 mai, 11:35]

*Copula Function Under Biased Sampling and Informative Censoring*

*Fonction de la copule dans le cadre d'un échantillonnage biaisé et d'une censure informative*

In observational studies, incidence cohort sampling is ideally adopted to study individuals, who have not experienced a disease, from disease onset to a failure event. Logistic or other constraints may however preclude the possibility of recruiting incident cases. A feasible alternative in such circumstances is to sample subjects who have already experienced the disease onset, through cross-sectional sampling. In this presentation, we discuss the estimation of the copula function for right-censored length-biased data. Copula function is known for its use in modeling the dependence structure between two variables. We introduce a nonparametric estimation method for the copula and its functionals.

Dans le cadre des études d'observation, on adopte idéalement l'échantillonnage de l'incidence des cohortes pour étudier les individus qui n'ont pas connu de maladie, à partir du début de la maladie jusqu'à un événement négatif déterminé. Certaines contraintes, d'ordre logistiques ou autres, peuvent cependant empêcher de recruter des cas d'incident. Une alternative possible, dans de telles circonstances, est d'échantillonner les sujets qui montrent les premières manifestations de la maladie au moyen d'un échantillonnage transversal. Dans cet exposé, nous discutons de l'estimation de la fonction de la copule des données censurées à droite de longueur biaisée. La fonction de la copule est connue pour son utilisation dans la modélisation de la structure de dépendance entre deux variables. Nous présentons une méthode d'estimation non paramétrique pour la copule et ses fonctions.

## 1B-C4: Statistical Methods and Applications 1 Méthodes statistiques et applications 1

---

Chair/Président: David Spiegelhalter (Cambridge University)

Room/Salle: Welch 8G

### Abstracts/Résumés

---

**JINGJIA CHU, REG KULPERGER & HAO YU** (University of Western Ontario)

[Monday May 30 / lundi 30 mai, 10:20]

*A Multivariate Time Series Model with an Additive GARCH Type Structure*

*Un modèle de séries temporelles multivarié avec une structure de type GARCH additif*

A new class of multivariate time series model with an additive GARCH type structure was proposed, which could be used to capture the common risk among equities. The dynamic conditional covariances between series are aggregated by a common risk term, which has been the key to characterize the conditional correlation. The model ergodicity and stationarity were proved as well as the identifiability theorem in terms of the second moment. A Monte Carlo simulation example will be shown in the talk.

Une nouvelle classe de modèles de séries temporelles multivariés avec une structure de type GARCH additif, qui pourrait être utilisée pour capturer le risque commun des actions, a été proposée. Les covariances conditionnelles dynamiques entre les séries sont agrégées par un terme de risque commun, qui est la clé pour caractériser la corrélation conditionnelle. L'ergodicité et la stationnarité du modèle ont été prouvées ainsi que le théorème d'identification en termes du deuxième moment. Un exemple de simulation de Monte-Carlo sera montré dans la présentation.

---

**REYHANEH HOSSEINI & MAHMOUD ZAREPOUR** (University of Ottawa)

[Monday May 30 / lundi 30 mai, 10:35]

*A Bayesian Nonparametric Chi-squared Goodness-of-fit Test*

*Un test d'ajustement du khi-deux bayésien non paramétrique*

The Bayesian nonparametric inference and Dirichlet process are popular tools in statistical methodologies. In this work, we employ the Dirichlet process in hypothesis testing to propose a Bayesian nonparametric chi-squared goodness-of-fit test. In our Bayesian nonparametric approach, we consider the Dirichlet process as the prior for the distribution of data and carry out the test based on the Kullback-Leibler distance between the updated Dirichlet process and the hypothesized distribution. We prove that this distance asymptotically converges to the same chi-squared distribution as the classical test does. Similarly, a Bayesian nonparametric chi-squared test of independence for a contingency table is provided.

L'inférence bayésienne non paramétrique et le processus de Dirichlet sont des outils populaires dans les méthodes statistiques. Dans ce travail, nous employons le processus de Dirichlet dans un test d'hypothèse afin de proposer un test d'ajustement du khi-deux bayésien non paramétrique. Dans notre approche bayésienne non paramétrique, nous considérons le processus de Dirichlet en tant que préalable à la distribution des données et effectuons le test selon la distance de Kullback-Leibler entre le processus de Dirichlet mis à jour et la distribution hypothétique. Nous montrons que cette distance converge asymptotiquement vers la même distribution du khi-deux que le test classique. De même, un test du khi-deux bayésien non paramétrique de l'indépendance pour un tableau de contingence est fourni.

## 1B-C4: Statistical Methods and Applications 1 Méthodes statistiques et applications 1

---

**AURÉLIEN GUETSOP NANGUE & MARTIN BILODEAU** (Université de Montréal)

[Monday May 30 / lundi 30 mai, 10:50]

*Approximations to Permutation Tests of Independence Between Two Random Vectors*  
*Approximation du test de permutation d'indépendance entre deux vecteurs aléatoires*

The main result establishes the equivalence of power between the  $\alpha$ -distance covariance and the Hilbert-Schmidt independence criterion (*HSIC*) tests with the characteristic kernel of a stable probability distribution of index  $\alpha$  with a sufficiently small bandwidth. Large-scale simulations reveal the superiority of these tests over other tests using copula. The Pearson type III approximation to the exact permutation distribution yields tests with more accurate type I error rates than the gamma approximation used for *HSIC*. A new method for bandwidth selection in *HSIC* tests is proposed which improves power in three simulations, two of which are from machine learning.

Le travail établit une équivalence de puissance entre tests basés sur la alpha-distance de covariance et sur le critère d'indépendance d'Hilbert-Schmidt (*HSIC*) avec fonction caractéristique de distribution de probabilité stable d'indice alpha avec paramètre d'échelle suffisamment petit. Les simulations en grandes dimensions montrent la supériorité de ces tests en comparaison à certains tests utilisant les copules. La distribution de Pearson de type III approche la distribution exacte de permutation des tests et donne des erreurs de type I plus précises que l'approximation gamma pour *HSIC*. Des simulations empruntées de l'apprentissage machine montrent l'amélioration de la puissance par notre nouvelle méthode adaptative.

**AURÉLIEN NICOSIA, THIERRY DUCHESNE, LOUIS-PAUL RIVEST & DANIEL FORTIN** (Université Laval)

[Monday May 30 / lundi 30 mai, 11:05]

*Equivalence Between the Random Walk Model in the Plane and Conditional Logistic Regression in a Multi-State Framework, with Application to the Analysis of Animal Movement*  
*Équivalence entre un modèle de marche aléatoire dans le plan et une régression logistique conditionnelle dans un cadre multi-états, application à l'analyse du mouvement animalier*

In this talk, we introduce a multi-state version of conditional logistic regression that compares the locations that an animal has chosen with randomly sampled locations. The impacts of different sampling procedures are discussed. A hidden process structure enables the modeling of situations where the animal exhibits various choice behaviors. We prove its equivalence with a random walk model in the plane, in which the directional part of the process is an angular regression model. We illustrate the use of the methods by modeling the movement of an animal in Prince Albert National Park (Saskatchewan, Canada).

Dans cette présentation, nous introduisons une version multi-états d'un modèle de régression logistique conditionnelle. Le modèle de régression caractérise quelle localisation un animal choisit face à des localisations échantillonnées au hasard. L'impact des différentes procédures d'échantillonnage est discuté. Une structure de processus cachée permet de modéliser des situations où l'animal manifeste des comportements différents. Nous prouvons son équivalence avec un modèle de marche aléatoire dans le plan, où la partie directionnelle du processus est un modèle de régression angulaire. Nous illustrons son utilisation en modélisant le mouvement d'un animal dans le parc national de Prince Albert (Saskatchewan, Canada).

## 1B-C4: Statistical Methods and Applications 1

### Méthodes statistiques et applications 1

---

**HOSSEIN ZAREAMOGHADDAM & SERGE PROVOST** (Western University), **SYED EJAZ AHMED** (Brock University)

[Monday May 30 / lundi 30 mai, 11:20]

*A Moment-Based Bivariate Density Estimation Methodology for Large Data Sets*

*Une méthodologie basée sur les moments pour l'estimation de densités bivariées pour grands ensembles de données*

We propose a moment-based methodology for approximating the density function associated with a bivariate distribution, which consists in applying an adjustment involving orthogonal polynomials to an initial density approximation. The use of standard polynomials is then shown to be mathematically equivalent. As well, the proposed technique is extended to the estimation of bivariate density functions, in which case joint sample moments are being utilized. It will be explained that this approach is ideally suited to model 'big data', which will be illustrated by applying it to large data sets. Generalizations to multivariate distributions shall also be discussed.

Nous proposons une méthodologie basée sur les moments pour l'approximation de la fonction de densité associée à une distribution bivariée, consistant à appliquer un ajustement utilisant des polynômes orthogonaux à une approximation initiale de la densité. L'utilisation de polynômes standards est alors montrée comme un équivalent mathématique. De même, la technique proposée est étendue à l'estimation des fonctions de densité bivariées et dans ce cas, les moments conjoints échantillonnaires sont utilisés. Nous verrons que cette approche convient idéalement à la modélisation des « données massives », ce qui sera illustré par son application à de grands ensembles de données. Nous traiterons aussi de généralisations à des distributions multivariées.

**BANGXIN ZHAO & WENQING HE** (Western University)

[Monday May 30 / lundi 30 mai, 11:35]

*Distance Techniques for High-dimensional Influence Measure*

*Techniques de distance pour mesure d'influence à dimensions élevées*

In this talk, a new statistic based on Hellinger Distance is developed to evaluate the influence of individual observation on high-dimensional data. We first give the rigorous definition of this statistic which itself is a function of probability mass functions based on marginal correlations, and then explain how this statistic accurately evaluates the influence of individual observations. The asymptotic distribution of the proposed statistic can be found by setting the dimension of the explanatory variable to approach infinity, which shows inference can also be conducted. Simulation and real data analysis are given as illustration of the usefulness of the proposed method.

Dans cet exposé, une nouvelle statistique fondée sur la distance de Hellinger est élaborée pour évaluer l'influence de l'observation individuelle sur des données à dimensions élevées. Après avoir rigoureusement défini cette statistique qui est en soi une fonction des fonctions de masse basées sur les corrélations marginales, nous expliquerons ensuite de quelle façon elle évalue avec exactitude l'influence des observations individuelles. On peut dériver la distribution asymptotique de la statistique proposée en posant vers l'infini la dimension des variables explicatives, ce qui montre la possibilité d'une inférence. La simulation et l'analyse de données réelles illustrent l'utilité de la méthode proposée.

## 1B-C5: Survey Methodology 1 Méthodologie d'enquête 1

---

Chair/Président: John L. Eltinge (U.S. Bureau of Labor Statistics)

Room/Salle: Thistle 257

### Abstracts/Résumés

---

**JESSICA ANDREWS, PIERRE DAOUST & MATEI MIREUTA** (Statistics Canada)

[Monday May 30 / lundi 30 mai, 10:20]

*Small Domains and Sparse Variables; Challenges with Creating an Active Collection Strategy*

*Petits domaines et variables rares; défis lors de la création d'une stratégie de collection active*

Statistics Canada has recently put into place the Integrated Business Statistics Program (IBSP) as a common framework to process its business surveys. One aspect of this new system has been to introduce an active collection strategy to improve efficiency, data quality and lower costs. The Quality Indicator and Measure of Impact (QIMI) collection tool was built to facilitate this. There have, however, been a number of challenges in turning a theoretical tool into a practical system. In this paper we explore some of the challenges faced including the impact of misreported values, small domains and sparsely reported variables.

Statistique Canada a récemment mis en place le programme intégré de la statistique des entreprises (IBSP) comme cadre commun pour traiter ses enquêtes auprès des entreprises. Un aspect de ce nouveau système a été l'introduction d'une stratégie de collection active pour améliorer l'efficacité, la qualité des données et pour réduire les coûts. Pour faciliter ceci, l'outil de collection Indicateur de qualité et mesure de l'impact (QIMI) a été développé. Par contre, il y a eu plusieurs défis lors de la transformation d'un outil théorique en un système pratique. Dans cet article, nous explorons quelques-uns de ces défis dont l'impact des valeurs erronées, des petits domaines et des variables rarement rapportées.

---

**SUSAN DEMEDASH** (Statistics Canada), **MATTHEW BUCKLEY** (Wilfred Laurier University)

[Monday May 30 / lundi 30 mai, 10:35]

*Job Vacancy and Wage Survey Paradata Log Files: A New Realm of Respondent Behavioural Analysis*

*Les fichiers de paradonnées de l'Enquête sur les postes vacants et les salaires : un nouveau monde d'analyse comportementale des répondants*

The Job Vacancy and Wage Survey collects information from Canadian employers on a number of labour issues, such as job vacancies and hourly wages by occupation. Started in February 2015, this survey is Statistics Canada's largest business survey with 100,000 business locations surveyed quarterly. The survey is conducted via electronic questionnaire and underlying paradata information on the communication from the respondent's computer to an internal host server was deciphered and analysed. This paper presents paradata results of the early quarters of this survey, which can help improve our questionnaire and reduce response burden.

L'Enquête sur les postes vacants et les salaires recueille de l'information reliée au travail auprès des employeurs canadiens. L'enquête, qui a débuté en février 2015, est la plus vaste enquête de Statistique Canada auprès des entreprises, avec 100 000 emplacements échantillonnés trimestriellement. L'information est recueillie par questionnaire électronique et les paradonnées sous-jacentes résultant des communications entre les ordinateurs des répondants et nos serveurs internes ont été déchiffrées et étudiées. L'analyse des paradonnées permet d'améliorer le questionnaire et de réduire le fardeau des répondants. Cet article présente certains résultats de cette analyse pour les premiers trimestres de l'enquête.

## 1B-C5: Survey Methodology 1 Méthodologie d'enquête 1

---

**MICHAEL FREIMAN & AMY LAUGER** (U.S. Census Bureau), **JEROME REITER** (Duke University)

[Monday May 30 / lundi 30 mai, 10:50]

*Two Approaches to Disclosure Avoidance for a Microdata Analysis System at the U.S. Census Bureau*

*Deux approches pour prévenir la divulgation dans un système d'analyse des microdonnées au U.S. Census Bureau*

The U.S. Census Bureau is developing an online Microdata Analysis System where users request a table or other analysis of Census Bureau data and receive the results without seeing the underlying microdata. We compare two approaches to maintaining the confidentiality of the data: output suppression, where results are withheld if they are deemed too risky, and pre-tabulation perturbation, where the underlying microdata are modified before the analysis is performed. Our research shows severe limitations to a suppression approach, with adequate protection requiring an unreasonably large number of tables to be suppressed. Ongoing research on pre-tabulation methods is discussed.

Le U.S. Census Bureau développe un système d'analyse des microdonnées en ligne où les utilisateurs demandent un tableau ou une autre analyse des données du Census Bureau et reçoivent les résultats sans voir les microdonnées sous-jacentes. Nous comparons deux approches pour maintenir la confidentialité des données : suppression de sortie, où les résultats sont retenus s'ils sont jugés trop risqués, et perturbation des pré-tabulations, où les microdonnées sous-jacentes sont modifiées avant l'analyse. Notre recherche montre des limites importantes à une approche de suppression, avec une protection adéquate nécessitant un nombre déraisonnablement grand de tableaux à supprimer. Nous discutons les recherches en cours sur les méthodes de pré-tabulation.

**ZIXIN NIE, KYLE VINCENT & ANDEE COOPER-PARKS** (International Justice Mission)

[Monday May 30 / lundi 30 mai, 11:05]

*Methodological Challenges Faced in the 2015-2016 Kolkata/Mumbai Commercial Sex Worker Study*

*Problèmes méthodologiques rencontrés dans l'étude 2015-2016 sur les travailleurs de l'industrie du sexe à Calcutta et à Mumbai*

International Justice Mission (IJM) conducted a study in 2015-2016 to measure the prevalence of minors working in the commercial sex industry of Kolkata and Mumbai, India. To sell minors for sex is illegal, and therefore the population is hidden and transient, resulting in several methodological challenges. We present the methods used to address and account for such issues, along with the results based on the resulting inference procedure, and we discuss how to address these limitations for future studies.

La Mission Internationale De Justice a mené une étude en 2015-2016 afin de mesurer la prévalence des mineurs travaillant dans l'industrie du sexe à Calcutta et à Mumbai, en Inde. La vente de services sexuels par des mineurs est illégale. Ainsi la population est cachée et isolée, ce qui engendre plusieurs problèmes méthodologiques. Nous présentons les méthodes utilisées pour aborder et prendre en compte ces problèmes, et nous présentons les résultats basés sur la procédure d'inférence qui en découle. Nous abordons également la façon dont on peut remédier à ces limitations dans les futures études.

**JAVIER OYARZUN** (Statistics Canada)

[Monday May 30 / lundi 30 mai, 11:20]

*Needle in a Haystack – How to Detect Break in Series: The Capital Expenditures Survey Integrated Business Statistics Program Experience*

*Une aiguille dans une botte de foin – La façon de détecter un bris dans une série : l'expérience de l'Enquête sur les dépenses en immobilisations dans le Programme intégré de la statistique des entreprises*

## 1B-C5: Survey Methodology 1 Méthodologie d'enquête 1

---

Statistics Canada's Capital Expenditures Survey (CES), which collects data on capital and repair expenditures, was integrated into the Integrated Business Statistics Program (IBSP) for 2013. The transition from a stand-alone survey to the CES-IBSP survey introduced several methodological changes that could have potentially introduced a break in annual estimates between 2012 (CES-based) and 2013 (IBSP-based). Using SAS High-Performance Forecasting and along with the Time Series Research and Analysis Centre, a modeling approach was developed to determine whether a break in the series exists. An overview of the methodology used and results using the published 2013 estimates will be presented.

L'Enquête sur les dépenses en immobilisations (EDI) de Statistique Canada, qui recueille des données sur les dépenses en immobilisations et réparations, a été intégrée dans le Programme intégré de la statistique des entreprises (PISE) en 2013. La transition d'une enquête indépendante à l'enquête EDI-PISE a entraîné plusieurs changements méthodologiques qui ont pu introduire un bris dans les estimations annuelles entre 2012 (estimations de l'EDI) et 2013 (estimations du PISE). En utilisant «SAS High-Performance Forecasting» et en partenariat avec le Centre de recherche et d'analyse en séries chronologiques, une approche de modélisation a été élaborée afin de déterminer si un bris dans la série existe. Un aperçu de la méthodologie utilisée ainsi que les résultats obtenus en utilisant les estimations publiées de 2013 seront présentés.

---

**QIAN WEI & J.N.K. RAO** (Carleton University)

[Monday May 30 / lundi 30 mai, 11:35]

*Statistical Inference Based on Survey Weighted Composite Likelihood*

*Inférence statistique fondée sur la vraisemblance composite pondérée d'un sondage*

Survey weighted composite likelihood (WCL) method (Rao et al. 2013) has been proposed to estimate multilevel model parameters when data are from a complex survey with a multistage sample design. In this paper, we study inferential properties of the WCL method, such as the asymptotic normality of the maximum WCL estimators and likelihood ratio tests. Simulation studies are conducted to evaluate the performance of the proposed test statistics.

La méthode de la vraisemblance composite pondérée (VCP) d'un sondage (Rao et coll. 2013) a été proposée pour estimer les paramètres de modèles multiniveaux lorsque les données proviennent d'un sondage complexe avec un plan d'échantillonnage à plusieurs étapes. Dans cet article, nous étudions les propriétés inférentielles de la méthode VCP, comme la normalité asymptotique des estimateurs VCP maximaux et les tests du rapport des vraisemblances. Des études de simulation sont menées pour évaluer la performance des statistiques de test proposées.

Key Words: Composite Likelihood, Multilevel Models, Multi-stage Survey Data

Mots-clés : vraisemblance composite, modèles multiniveaux, données de sondage à plusieurs étapes



**1C-D1: Case Study 1: Can Google Flu Trends Predict the Frequency and Results of Tests for Influenza and Other Respiratory Illnesses?**

**Étude de cas 1: Le site Google Flu Trends peut-il prédire la fréquence et le résultat des tests pour la grippe et autres maladies respiratoires ?**

---

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

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)

(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

**Room/Salle: Thistle Corridor**

**Abstract/Résumé**

**SHOMOITA ALAM, SHOUAO WANG & ZAYD OMAR** (McGill University)

**SHABNAM BALAMCHI, MD. ERFANUL HOQUE, FAISAL ATAKORA & OLAWALE FATAI AYILARA** (University of Manitoba)

**ZHOUSHANYUE HE & DANQIAO GUO** (University of Waterloo)

**YUANHAO LAI, TIANPEI JIANG & SAMIRA SOLEYMANI** (University of Western Ontario)

**WEI LIN, HONGYANG HU, FAIZAN MOSHIN, RUOQI YU & YUXIN ZOU** (University of Toronto)

**KUAN LIU, MICHAEL MOON, KAVIUL MOHAMMAD KHAN & SANGOOK KIM** (University of Toronto)

**LIN LU, XU ZHAO, YISHAN GUO & GEMINI LAM** (University of Toronto)

**DANNY TSHITUMBU, HANG LAI, BISMA FAROOQI & SUJ SRISKANDARAJAH** (York University)

**FAHMIDA YEASMIN & WENYAN ZHONG** (University of Calgary)

**1C-D2: Case Study 2: What Predicts Sustainability of Canadian Charities?  
Étude de cas 2: Pouvons-nous prédire la viabilité des organismes de bienfaisance  
canadiens ?**

---

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

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)

(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

**Room/Salle: Thistle Corridor**

**Abstract/Résumé**

**ISABELLE BERNARD, SAMUEL LABELLE, MAIKEL GEAGEA & JORGE-ENRIQUE RUIZ-TRUJILLO** (HEC Montréal)

**BO CHEN, HAOSUI DUANMU & PEIYANG HE** (University of Toronto)

**HUANG HUANG, XIAOXIAO WANG, JIN MA, HE BIAN, JINGCHUN PEI, YISEN LIN & BANO MADHANAGOPAL** (University of Toronto)

**WEI LIN, HONGYANG HU, FAIZAN MOSHIN, RUOQI YU & YUXIN ZOU** (University of Toronto)

**TINGBIN TAN & KEXIN LUO** (University of Western Ontario)

**YANBO TANG, YANG GUAN JIAN GUO, MUFAN LI, GONG ZHANG & HARRIS QUACH** (University of Toronto)

**GUANBO WANG, ALEXANDER LEVIS & MENGIAN PANG** (McGill University)

**FEI ZUO, XIAOTIAN WANG, JIAJUN YAN, YAO XI & JINGJING CAO** (University of Toronto)

## 1D-I1: Analysis of Big Data Analyse des mégadonnées

---

Organizer and Chair / Responsable et président: Francois Bellavance (HEC Montreal)

Room/Salle: Welch - David Howes

### Abstracts/Résumés

---

**LUKE BORNN** (Simon Fraser University)

[Monday May 30 / lundi 30 mai, 13:30]

*From Pixels to Points: Using Tracking Data to Measure Performance in Professional Sports*

*Des pixels aux points : comment utiliser les données de poursuite pour mesurer la performance dans le sport professionnel*

In this talk I will explore how players perform, both individually and as a team, on a basketball court. By blending advanced spatio-temporal models with geography-inspired mapping tools, we are able to understand player skill far better than either individual tool allows. Using optical tracking data consisting of hundreds of millions of observations, I will demonstrate these ideas by characterizing defensive skill and decision making in NBA players.

Dans cette présentation, j'explorerai comment les joueurs réagissent, individuellement et en équipe, sur un terrain de basketball. La combinaison de modèles spatiotemporels avancés et d'outils cartographiques inspirés de la géographie nous permet une compréhension de l'habileté du joueur bien meilleure que celle que nous offre l'un ou l'autre de ces outils. Je présenterai ces idées à l'aide de données de poursuite optique consistant en des centaines de millions d'observations, qui permettent de caractériser l'habileté défensive et la prise de décision chez les joueurs de la NBA.

---

**ERSHAD BANIJAMALI & ALI GHODSI** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 14:00]

*Generative Mixture of Networks*

*Mélange de réseaux générateur*

In this talk I will introduce a generative model by training deep architecture. The model starts with dividing the input data into K clusters and feeding each of them into a separate network. So, there is a big network which consists of K sub-networks. The goal of training each sub-network is to generate samples which have close distribution to the underlying distribution of its assigned training set. Empirical estimation of Maximum Mean Discrepancy (MMD) is used as a measure of distance between these two distributions. Subsequently, an algorithm is employed whose goal is to further train the networks, jointly with updating the clusters of the training set by a non-parametric likelihood estimation. We call this algorithm, *Mixture of Networks*.

Dans cette présentation, j'introduirai un modèle générateur créé par entraînement d'une architecture profonde. Pour commencer, les données d'entrée sont divisées en K grappes et intégrées dans des réseaux distincts. Nous nous retrouvons avec un grand réseau composé de K sous-réseaux. L'entraînement de chaque sous-réseau vise à générer des échantillons dont la distribution est proche de la distribution sous-jacente de son ensemble d'entraînement attitré. L'estimation empirique de la divergence moyenne maximale (MMD) est utilisée comme mesure de la distance entre ces deux distributions. Par la suite, un algorithme est utilisé pour continuer l'entraînement des réseaux, tout en mettant à jour les grappes de l'ensemble d'entraînement par une estimation de la vraisemblance non paramétrique. Nous appelons cet algorithme *Mélange de réseaux*.

## 1D-I1: Analysis of Big Data Analyse des mégadonnées

---

---

**PAUL D. MCNICHOLAS** (McMaster)

[Monday May 30 / lundi 30 mai, 14:30]

*Clustering Ultra High-Dimensional Data*

*Regroupement de données de très grande dimension*

Some approaches to clustering ultra high-dimensional data are considered. Each approach is based on a mixture model, and each component thereof is taken as corresponding to a cluster. Some approaches depend on the assumption of an underlying low-dimensional (latent) space while others also draw on shrinkage. Real data are used for illustration.

Nous examinons diverses approches de regroupement de données de très grande dimension. Chaque approche est fondée sur un modèle de mélange, dont chaque composante est considérée comme correspondant à un groupe. Certaines approches reposent sur l'hypothèse d'un espace (latent) à faible dimension sous-jacent, tandis que d'autres ont recours au concept du rétrécissement. Nous illustrons ces approches à l'aide de données réelles.

## 1D-I2: New Perspectives in Causal Inference Nouvelles perspectives en inférence causale

---

Organizer and Chair / Responsable et présidente: Yeying Zhu (University of Waterloo)

Room/Salle: Thistle 325

### Abstracts/Résumés

---

**PEISONG HAN** (University of Waterloo), **LU WANG** (University of Michigan)

[Monday May 30 / lundi 30 mai, 13:30]

*Multiply Robust Estimation in Causal Inference*

*Estimation multiplement robuste en inférence causale*

Multiple robustness is a desirable property first established in the missing data literature. Estimators are multiply robust if they are consistent when any one of the multiple missingness probability models and multiple data distribution models is correctly specified. We will show how to construct multiply robust estimators for the average treatment effect for observational studies with binary treatments. The estimators are consistent if any one of the multiple propensity score models and multiple outcome regression models is correctly specified. Our estimation procedure can also achieve desired level of covariate balance between treatment and control groups by matching moments of the covariate distributions through reweighting the units in one or both groups.

La robustesse multiple est une propriété souhaitable d'abord établie dans la littérature des données manquantes. Des estimateurs sont multiplement robustes s'ils sont convergents lorsqu'un des modèles de probabilités de données manquantes multiples ou de loi de données multiples est correctement spécifié. Nous démontrerons comment construire des estimateurs multiplement robustes pour l'effet de traitement moyen lors d'études d'observations avec traitements binaires. Les estimateurs sont convergents si un des modèles de coefficients de propension multiples ou de régression des résultats multiples est correctement spécifié. Notre procédure d'estimation peut aussi atteindre le niveau souhaité de balance des covariables entre le traitement et les groupes témoins en faisant correspondre les moments des lois des covariables par la repondération des unités dans un groupe ou dans les deux groupes.

---

**WEI LUO** (Baruch College), **YEYING ZHU** (University of Waterloo), **DEBASHIS GHOSH** (Colorado School of Public Health)

[Monday May 30 / lundi 30 mai, 14:00]

*On Estimating Regression-based Causal Effects Using Sufficient Dimension Reduction*

*Estimer les effets de causalité basés sur la régression en utilisant une réduction suffisante de la dimension*

In many causal inference problems, the parameter of interest is often the regression causal effect, defined as the conditional mean difference in the potential outcomes given covariates. This paper discusses how sufficient dimension reduction can be used to assist in its estimation, and proposes a new estimator for the regression causal effect inspired by a sufficient dimension reduction method called the minimum average variance estimation. The estimator requires a weaker common support condition than the traditional propensity score-based ap-

Dans plusieurs problèmes d'inférence causale, le paramètre d'intérêt est souvent l'effet de causalité de la régression, défini comme étant la différence de la moyenne conditionnelle dans les résultats potentiels compte tenu des covariables. Cet article discute de la façon dont une réduction suffisante de la dimension peut être utilisée pour aider à son estimation et propose un nouvel estimateur pour l'effet causal de régression, inspiré d'une méthode de réduction suffisante de la dimension appelée estimation à variance moyenne minimale. L'estimateur nécessite une condition de support commun plus faible que dans les approches traditionnelles fon-

## 1D-I2: New Perspectives in Causal Inference Nouvelles perspectives en inférence causale

---

proaches. In addition, it can be easily converted to estimate the average causal effect, where it is shown to be asymptotically super efficient. The finite-sample properties of the proposed method are illustrated using simulation studies.

dées sur les coefficients de propension. De plus, il peut facilement être transformé pour estimer l'effet causal moyen, où il est asymptotiquement super efficace. Les propriétés d'échantillons finis de la méthode proposée sont illustrées par des études de simulation.

---

**MIREILLE SCHNITZER** (University of Montreal), **JUDITH LOK & SUSAN GRUBER** (Harvard University), **MARK VAN DER LAAN** (University of California, Berkeley)

[Monday May 30 / lundi 30 mai, 14:30]

*Collaborative Double Robustness and the Connection to Data-Adaptive Nuisance Model Selection in Causal Inference*

*La double robustesse collaborative et le lien à la sélection de modèles de nuisance par algorithme d'apprentissage adaptatif en inférence causale*

In causal inference and censored data methods, double robust estimators such as Targeted Minimum Loss-based estimation require the specification of two model components. Estimation will be consistent under the correct specification of either nuisance component, conditional on a sufficient set of confounding variables. However, this class of estimators also has a collaborative robustness property that allows for consistent effect estimation in a wider class of settings. We describe this class of estimators in single time point and longitudinal exposure settings, describe a cohesive approach to variable selection in causal inference, and illustrate the performance of data-adaptive learning algorithms for the nuisance models.

En inférence causale et méthodes pour données censurées, les estimateurs doublement robustes, telle l'estimation par maximum de vraisemblance ciblée, nécessitent la bonne spécification de deux composantes de la vraisemblance. L'estimation est convergente quand au moins une des deux composantes nuisibles est bien spécifiée, conditionnellement sur des facteurs de confusion suffisants. Cependant, cette classe d'estimateurs a aussi la propriété de robustesse collaborative permettant une estimation convergente dans certains cas, même si aucun des modèles n'est bien spécifié. Nous décrivons cette classe d'estimateurs pour des contextes d'exposition à un ou plusieurs points dans le temps, décrivons une approche cohérente pour la sélection de variables en inférence causale et illustrons la performance d'algorithmes d'apprentissage adaptatif pour l'estimation de modèles à composantes nuisibles.

## 1D-I3: Small Area Estimation: New Developments Estimation pour petits domaines : nouveaux développements

---

Chair/Présidente: Karla Fox (Statistics Canada)

Organizer/Responsable: Mahmoud Torabi (University of Manitoba)

**Room/Salle: Thistle 258**

### Abstracts/Résumés

---

**M. GIOVANNA RANALLI** (University of Perugia, Italy), **GAIA BERTARELLI & FRANCESCO BARTOLUCCI** (University of Perugia), **MICHELE D'ALÒ & FABRIZIO SOLARI** (ISTAT)

[Monday May 30 / lundi 30 mai, 13:30]

*Time Series Small Area Estimation for Unemployment Rates using Latent Markov Models*

*Estimation pour petites régions de séries chronologiques relatives au taux de chômage à l'aide de modèles de Markov latents*

In this work we develop a new area-level Small Area Estimation method using Latent Markov Models (LMMs) in a Bayesian setting. LMMs handle longitudinal data in which the characteristics of interest, and their evolution in time, are represented by a latent process that follows a Markov chain. Small areas are allowed to move between latent states over time. LMMs may be seen as an extension of Markov chain models to control for measurement errors and as a dynamic extension of Latent class models for longitudinal data. Estimation is conducted using a Gibbs sampler with data augmentation. The proposed model is applied to estimate quarterly unemployment rates for Italian Local Labour Market Areas using LFS data from 2004 to 2014.

Nous développons une nouvelle méthode d'estimation pour petites régions au niveau de la région, basée sur des modèles de Markov latents (LMM) dans un contexte bayésien. Les LMM permettent de traiter des données longitudinales dans lesquelles les caractéristiques d'intérêt et leur évolution dans le temps sont représentées par un processus latent qui suit une chaîne de Markov. Les petites régions peuvent passer d'un état latent à un autre au fil du temps. Les LMM peuvent être considérés comme une extension des modèles de chaînes de Markov qui tient compte des erreurs de mesure – et comme une extension dynamique des modèles de catégories latentes pour les données longitudinales. Nous effectuons l'estimation à l'aide d'un échantillonneur de Gibbs avec augmentation des données. Nous appliquons ce modèle à l'estimation du taux de chômage trimestriel pour diverses régions du marché du travail italien avec des données de l'EPA de 2004 à 2014.

---

**J.N.K. RAO** (Carleton University)

[Monday May 30 / lundi 30 mai, 14:00]

*On Measuring Uncertainty of Small Area Estimators Under Area Level and Unit Level Models*

*Mesure de l'incertitude des estimateurs pour petites régions dans les modèles au niveau de la région et de l'unité*

Model-based estimators of small area means that borrow strength from related areas through linking models are extensively used because direct area-specific estimators are imprecise due to small sample sizes within areas. We study different methods of estimating mean squared error (MSE) of the model-based estimators and appraise their relative properties. We also study methods of finding confi-

On a largement recours à des estimateurs modélisés de la moyenne régionale dont l'efficacité est renforcée au moyen de données empruntées des régions connexes grâce à des modèles de liaison, parce que les estimateurs directs de la moyenne régionale pour une région précise sont imprécis en raison de la petite taille des échantillons régionaux. Nous étudions différentes méthodes d'estimation de l'erreur quadratique moyenne (EQM) des estimateurs modélisés et en

## 1D-I3: Small Area Estimation: New Developments

### Estimation pour petits domaines : nouveaux développements

---

dence intervals on the small area means. We consider both area level and unit level models.

évaluons les propriétés respectives. Nous étudions également les méthodes de calcul des intervalles de confiance pour les moyennes régionales. Nous examinons des modèles au niveau de la région et de l'unité.

---

**MAHMOUD TORABI** (University of Manitoba)

[Monday May 30 / lundi 30 mai, 14:30]

*Prediction in Small-Area Spatially Correlated Data*

*Prédiction dans les données de petites régions spatialement corrélées*

In small area estimation, we need to predict characteristics of the sub-populations (areas) based on the coarse scale data. Small area predictors are improved by using (standard) mixed models. However, there are many situations where the characteristics are related to their locations. For example, it is an interest of policy makers (and public) to know the spatial pattern of a rare disease (e.g., chronic disease or cancer). We propose small area models in the class of spatial mixed models to be able to predict characteristics and also to obtain corresponding mean squared prediction error (MSPE) and MSPE estimate. The performance of our proposed approach is evaluated through simulations and by a real application.

Dans l'estimation pour petites régions, il faut prédire les caractéristiques des sous-populations (régions) en fonction de données à une échelle moins fine. Les prédicteurs sur petits domaines sont souvent améliorés à l'aide de modèles mixtes (standard). Cependant, dans de nombreuses situations, les caractéristiques sont liées à leur localisation. Par exemple, il est important pour les décideurs (et pour le grand public) de connaître la répartition géographique d'une maladie rare (maladie chronique ou cancer). Nous proposons des modèles pour petites régions de la classe des modèles spatiaux mixtes capables de prédire des caractéristiques et d'obtenir l'erreur quadratique moyenne de prédiction (EQMP) correspondante et une estimation de l'EQMP. Nous évaluons la performance de cette approche via des simulations et une application réelle.



## 1D-14: Statistical Modelling in Environmental and Health Studies Modélisation statistique dans les études sur l'environnement et la santé

---

Chair/Président: Tim Swartz (Simon Fraser University)  
Organizer/Responsable: Paramjit Gill (British Columbia-Okanogan)

Room/Salle: Welch 209

### Abstracts/Résumés

---

**CINDY FENG & MEHDI ROSTAMIFOROOSHANI** (University of Saskatchewan)

[Monday May 30 / lundi 30 mai, 13:30]

*Modeling Spatially Correlated Survival Data: Impact of Misspecification of Correlation Structure on the Parameter Estimates*

*Modélisation de données de survie spatialement corrélées: impact des erreurs de spécification d'une structure de corrélation sur les estimations de paramètres*

In epidemiological and environmental studies, time to event data are often grouped into clusters (e.g. families, clinical sites and geographical regions, etc.). Spatial survival models have been proposed in literature for modeling geographically correlated survival data, which includes a flexible spatially varying baseline hazard function to control for unmeasured spatial confounders and to borrow information across geographical units. The identified spatial pattern may highlight the regions requiring attention, which may assist public health professionals in their decision making. Few studies have been conducted to investigate the consequence of ignoring modeling the spatial correlation on the parameter estimates, so we conducted a simulation study to determine how the bias and efficiency in the parameter estimates change when misspecifying the correlation structure.

Dans les études épidémiologiques et environnementales, les données de durée avant événement sont souvent regroupées en grappes (par exemple les familles, les sites cliniques et les régions géographiques, etc.). Les modèles de survie spatiaux ont été présentés dans la littérature pour modéliser des données de survie corrélées géographiquement, ce qui inclut une fonction du risque de base flexible spatialement variable pour contrôler les facteurs de confusion spatiaux non mesurés et pour emprunter de l'information à travers les unités géographiques. Le schéma spatial identifié peut souligner les régions qui requièrent de l'attention, ce qui peut aider les professionnels de la santé dans le processus de prise de décision. Peu d'études ayant été menées pour étudier la conséquence d'ignorer la modélisation de la corrélation spatiale sur les estimations de paramètres, nous avons décidé de mener une étude de simulation pour déterminer le changement dans le biais et dans l'efficacité des estimations de paramètres lorsque la structure de corrélation est mal spécifiée.

---

**RENJUN MA** (University of New Brunswick)

[Monday May 30 / lundi 30 mai, 14:00]

*Mixed Models for Correlated Environmental/Health Data with Left Censoring and Right-Skewness*

*Modèles mixtes pour données environnementales/de santé corrélées avec censure à gauche et asymétrie à droite*

Environmental and health data are often subject to left censoring at detection limit. Such censored data are usually right-skewed continuous, but with a point mass at the detection limit. Examples of such data are water pollutants in fish, tumor size due to radioactivity treatments, lesion depth re-

Les données environnementales et de santé sont souvent sujettes à une censure à gauche à la limite de détection. Ces données censurées sont généralement asymétriques à droite et continues, mais avec une masse ponctuelle à la limite de détection. C'est le cas par exemple pour les polluants de l'eau dans les poissons, la taille d'une tumeur après trai-

## 1D-I4: Statistical Modelling in Environmental and Health Studies Modélisation statistique dans les études sur l'environnement et la santé

---

lated to ultrasound and precipitation. A special case of such data is so-called zero-inflated semi-continuous data where negligible amount is ignored. In this talk, we propose compound Poisson mixed models to characterize both the occurrence of the detection limit and size of the correlated data simultaneously. Our approach is illustrated with applications to real data.

tements radioactifs, la profondeur de la lésion en fonction des ultrasons et des précipitations. Cas spécial de ces données : les données dites semi-continues à surreprésentation de zéros où une quantité négligeable est ignorée. Dans cette présentation, nous proposons des modèles mixtes de Poisson composés pour caractériser simultanément l'occurrence de la limite de détection et l'importance des données corrélées. Nous illustrons notre approche par des applications sur des données réelles.

---

**GYANENDRA POKHAREL & ROB DEARDON** (University of Calgary)

[Monday May 30 / lundi 30 mai, 14:30]

*Gaussian Process Emulator-Based Inference for Spatial Models Infectious Disease Systems*

*Inférence basée sur un émulateur à processus gaussien pour les modèles spatiaux de systèmes de maladies infectieuses*

Statistical inference for mechanistic models of infectious disease spread is often very computationally expensive. Such models are generally fitted in a Bayesian Markov chain Monte Carlo (MCMC) framework that requires multiple calculation of likelihood function and is often computationally inefficient. This problem is more severe when incorporating large numbers of latent variables. Here, we propose a method of inference based on so-called emulation techniques. The method is again set in a Bayesian MCMC context, but avoids calculation of the computationally expensive likelihood function replacing it with a Gaussian process approximation. We show that such a method offers a significant gain in computation and can be used to infer the model parameters and underlying characteristics of spatial disease systems.

L'inférence statistique pour les modèles mécanistes de la propagation de maladies infectieuses est souvent très coûteuse en ressources informatiques. Ces modèles sont généralement ajustés dans un cadre de Monte-Carlo par chaîne de Markov (MCMC) bayésien qui exige des calculs multiples de la fonction de vraisemblance, ce qui est souvent inefficace d'un point de vue computationnel. Ce problème ne fait que s'aggraver en la présence de grands nombres de variables latentes. Ici, nous proposons une méthode d'inférence basée sur des techniques dites d'émulation. Cette méthode s'inscrit aussi dans un contexte de MCMC bayésien, mais elle évite le calcul de la fonction de vraisemblance, si coûteux en ressources, et le remplace par une approximation par processus gaussien. Nous montrons qu'une telle méthode offre un gain computationnel significatif et qu'elle permet d'inférer les paramètres du modèle et les caractéristiques sous-jacentes des systèmes spatiaux de la maladie.

## 1D-I5: Imaging Genomics - A New Frontier for Statistical Methodology Imagerie génomique -- une nouvelle frontière pour la méthodologie statistique

---

Organizer and Chair / Responsable et président: Linglong Kong (University of Alberta)

Room/Salle: Thistle 246

### Abstracts/Résumés

---

**MICHELE GUINDANI** (University of Texas)

[Monday May 30 / lundi 30 mai, 13:30]

*Bayesian Predictive Modeling for Imaging Genetics with Application to Schizophrenia*

*Modélisation prédictive bayésienne pour la génétique d'imagerie avec application à la schizophrénie*

By combining single-nucleotide polymorphism (SNP) arrays and functional magnetic resonance imaging (fMRI), we propose an integrative Bayesian risk prediction model that allows us to discriminate between individuals with schizophrenia and healthy controls, based on a sparse set of discriminatory regions of interest (ROIs) and SNPs. Inference on a regulatory network between SNPs and ROI intensities (ROI-SNP network) is used in a single modeling framework to inform the selection of the discriminatory ROIs and SNPs. We use simulation studies to assess the performance of our method and apply it to data collected from individuals with schizophrenia and healthy controls. We found our approach to outperform competing methods that do not link the ROI-SNP network to the selection of discriminatory markers.

En combinant des réseaux de polymorphismes nucléotidiques (SNP) et l'imagerie par résonance magnétique fonctionnelle (IRMf), nous proposons un modèle bayésien intégratif de prévision du risque qui nous permet de distinguer les individus schizophrènes des contrôles sains, basé sur un ensemble clairsemé de ROI (régions d'intérêt) et de SNP discriminants. Une inférence sur un réseau régulateur entre les ROI et les SNP est utilisée dans une modélisation unifiée pour informer la sélection de ROI et de SNP discriminants. Nous utilisons des études de simulation pour évaluer la performance de notre méthode, avant de l'appliquer à des données collectées auprès d'individus schizophrènes et de contrôles sains. Nous montrons que notre approche surpasse des méthodes qui ne relient pas le réseau ROI-SNP à la sélection de marqueurs discriminants.

---

**FAROUK NATHOO, KEELIN GREENLAW & MARY LESPERANCE** (University of Victoria), **ELENA SZEFER & JINKO GRAHAM** (Simon Fraser University)

[Monday May 30 / lundi 30 mai, 14:00]

*A Bayesian Group Sparse Multi-Task Regression Model for Imaging Genomics*

*Modèle de régression multitâche bayésienne à groupes clairsemés en génomique de l'imagerie*

Advances in technology for brain imaging and genotyping have motivated studies examining the relationships between genetic variation and brain structure. Wang et al. (Bioinformatics, 2012) developed an approach for simultaneous regression parameter estimation and SNP selection based on penalized regression with a group  $l_{2,1}$ -norm penalty. The group-norm penalty formulation incorporates the biological group structures among SNPs induced from their genetic arrangement and enforces

Les progrès technologiques en imagerie cérébrale et génotypage ont motivé plusieurs études qui examinent les relations qui existent entre la variation génétique et la structure du cerveau. Wang et al. (Bioinformatics, 2012) ont mis au point une approche pour l'estimation simultanée des paramètres de régression et la sélection de SNP basée sur la régression pénalisée avec une pénalité de groupe de norme  $l_{2,1}$ . La formulation de pénalité de groupe de norme intègre les structures de groupes biologiques des SNP induites par leur agencement génétique et impose la parci-

## **1D-I5: Imaging Genomics - A New Frontier for Statistical Methodology** **Imagerie génomique -- une nouvelle frontière pour la méthodologie statistique**

---

sparsity at the group level. In this paper, we propose a corresponding Bayesian model that allows for full posterior inference for the regression parameters using Gibbs sampling. Properties of our method are investigated using simulation studies and the methodology is applied to a large dataset collected as part of the Alzheimer's Disease Neuroimaging Initiative.

monie au niveau du groupe. Dans cet article, nous proposons une formulation bayésienne correspondante qui permet une inférence a posteriori complète par échantillonnage de Gibbs. Nous étudions les propriétés de notre méthode à l'aide d'études de simulation et appliquons notre méthodologie à un grand ensemble de données collectées dans le cadre de l'initiative en neuroimagerie de la maladie d'Alzheimer.

## **1D-16: Opportunities for statistical scientists in CIHR programs: a panel discussion Opportunit s pour les statisticiens dans les programmes IRSC : table ronde**

---

Chair/Pr sidente: Mary Thompson (University of Waterloo)

**Room/Salle: Thistle 242**

### **Description**

The Canadian Institutes of Health Research (CIHR) has recently introduced new open operating grant programs, such as Foundation Scheme and Project Scheme grants for individual researchers or teams working to solve problems, including methodological problems, about health and healthcare use. Other CIHR-funded programs of potential interest to statistical scientists include Collaborative Health Research Projects (joint with NSERC) and funding for training initiatives. The panelists will describe these opportunities from their own perspectives and share their advice on applying to CIHR programs.

Les Instituts de recherche en sant  du Canada (IRSC) ont pr sent  de nouveaux programmes ouverts de subventions de fonctionnement, comme les subventions – volet Fondation et volet Projet, destin s   des chercheurs ou  quipes de recherche s'appliquant   r soudre des probl mes, y compris d'ordre m thodologique, en mati re de sant  et d'utilisation des soins de sant . D'autres programmes financ s par les IRSC d'un int r t  ventuel pour les scientifiques de la statistique comprennent les Projets de recherche concert e sur la sant  (conjointement avec le CRSNG) et des subventions de projets en mati re de formation. Les pan listes d criront ces possibilit s de financement de leur point de vue respectif et expliqueront comment soumettre une demande aux programmes des IRSC.

## 1D-C1: Actuarial Science and Finance 1 Actuariat et finance 1

---

Chair/Président: Étienne Marceau (Laval University)

Room/Salle: Thistle 257

### Abstracts/Résumés

---

**NICHOLAS BECK** (McGill University), **MÉLINA MAILHOT** (Concordia University)

[Monday May 30 / lundi 30 mai, 13:30]

*A Consistent Estimator to the Orthant-Based Tail Value-at-Risk*

*Un estimateur convergent de l'aile de la valeur à risque en fonction de l'orthant*

In this paper we address the estimation of multivariate Value-at-Risk (VaR) and Tail Value-at-Risk (TVaR). We recall definitions for the bivariate lower and upper orthant VaR and bivariate lower and upper orthant TVaR, presented in Cossette et al. (2013, 2014). Here, we present estimators for both these measure extended to arbitrary dimension  $d > 2$  and establish the consistency of our estimator for the lower and upper orthant TVaR in any dimension. We then demonstrate these results by providing numerical examples that compare our estimator to theoretical results, from both real and simulated data.

Dans cet article nous abordons l'estimation de la valeur à risque multivariée (VaR) et de l'aile de la valeur à risque (TVaR). Nous nous rappelons certaines définitions pour la VaR bivariée de l'orthant inférieur et supérieur et pour la TVaR bivariée de l'orthant inférieur et supérieur, présentées dans Cossette et al. (2013, 2014). Nous présentons des estimateurs pour ces deux mesures élargies à une dimension arbitraire  $d > 2$  et nous établissons la convergence de notre estimateur pour la TVaR de l'orthant inférieur et supérieur dans n'importe quelle dimension. Nous démontrons ensuite ces résultats par des exemples numériques qui comparent notre estimateur à des résultats théoriques, pour des données réelles et simulées.

---

**JEAN-FRANÇOIS BÉGIN** (HEC Montréal), **MATHIEU BOUDREAU** (Université du Québec à Montréal), **DELIA ALEXANDRA DOLJANU** & **GENEVIÈVE GAUTHIER** (HEC Montréal)

[Monday May 30 / lundi 30 mai, 13:45]

*Credit and Systemic Risks in the Financial Services Sector*

*Risques de crédit et systémique dans le secteur financier*

The Great Recession has shaken the foundations of the financial industry and led to tighter solvency monitoring of both the banking and insurance industries. To this end, we develop a portfolio credit risk model that includes firm-specific Markov-switching regimes as well as individual stochastic and endogenous recovery rates. Using weekly credit default swap premiums for 35 financial firms, we analyze the credit risk of each of these companies and their statistical linkages, placing special emphasis on the 2005-2012 period. Moreover, we study the systemic risk affecting both the banking and insurance subsectors.

La Grande Récession a ébranlé les bases de l'industrie financière et a mené à un suivi plus serré de la solvabilité des firmes du secteur financier. À cet effet, nous développons un modèle de risque de crédit multivarié qui comprend des régimes propres à chaque firme ainsi qu'un taux de recouvrement stochastique. À l'aide de primes hebdomadaires pour 35 sociétés, nous analysons le risque de crédit de chacune de ces firmes et leurs liens statistiques, mettant un accent particulier sur 2005-2012. Nous étudions aussi le risque systémique affectant à la fois le sous-secteur bancaire et celui de l'assurance.

## 1D-C1: Actuarial Science and Finance 1 Actuariat et finance 1

---

**LICHEN CHEN, ADAM KOLKIEWICZ & TONY WIRJANTO** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 14:00]

*Bayesian Estimation of GARCH Models with General Parameter Constraints*

*Estimation bayésienne des modèles GARCH avec contraintes sur le paramètre général*

We discuss Bayesian estimation of GARCH models with general parameter constraints for non-negative conditional variance. The general parameter constraints are important for capturing long-memory type behaviour of financial asset return volatilities together with their short-term behaviour. However, the shapes of the parameter space and the likelihood surface under these constraints often pose computational problems to standard estimation algorithms. In our work we explain how Bayesian MCMC methods and model reparameterization can help us solve these difficulties.

Nous discutons l'estimation bayésienne des modèles GARCH avec des contraintes sur le paramètre général pour la variance conditionnelle non-négative. Les contraintes sur le paramètre général sont importantes pour saisir le comportement de type longue mémoire des volatilités du retour sur l'actif financier ainsi que leur comportement à court terme. Cependant, les formes de l'espace des paramètres et la surface de vraisemblance sous ces contraintes posent souvent des problèmes de calcul aux algorithmes standards d'estimation. Dans notre travail, nous expliquons comment les méthodes bayésiennes MCMC et le reparamétrage du modèle peuvent nous aider à résoudre ces difficultés.

**WENJUN JIANG & JIANDONG REN** (Western University)

[Monday May 30 / lundi 30 mai, 14:15]

*Optimal Reinsurance Minimizing the Risk of Joint Party*

*Réassurance optimale minimisant le risque de la partie commune*

Optimal reinsurance problem has been studied for a long time. By imposing restrictions on ceded function, analytical solutions can be obtained under certain risk measures. However, most studies are from the viewpoint of insurer or reinsurer. In this paper, we consider the bivariate Value-at-Risk of joint party, insurer and reinsurer, and take full use of geometric approach to obtain the optimal form of reinsurance with ceded function in  $C^1$  and  $C^2$  [Chi and Tan (2011)]. We further illustrate the solution is also Pareto-optimal and derive the optimal reinsurance parameters.

Le problème de réassurance optimale a été étudié depuis longtemps. En imposant des restrictions sur la fonction cédée, les solutions analytiques peuvent être obtenues sous certaines mesures du risque. Cependant, la plupart des études sont du point de vue de l'assureur ou du réassureur. Dans cet article, nous considérons la valeur à risque bvariée de la partie commune, assureur et réassureur, et utilisons pleinement l'approche géométrique pour obtenir la forme optimale de la réassurance avec fonction cédée en  $C^1$  et  $C^2$  [Chi et Tan (2011)]. De plus, nous illustrons que la solution est également Pareto optimale et dérivons les paramètres de réassurance optimale.

**LUYAO LIN, LI CHEN, YI LU & GARY PARKER** (Simon Fraser University)

[Monday May 30 / lundi 30 mai, 14:30]

*Analysis of Survivorship Life Insurance Portfolio with Stochastic Interest Rates*

*Analyse d'un portefeuille d'assurance de survie avec taux d'intérêt stochastique*

A general portfolio of joint survivorship life insurance contracts is studied in a stochastic interest rate environment with dependent mortality model. Two methods are used to derive the first two moments of the projected prospective loss random

Nous étudions un portefeuille général de contrats d'assurance de survie au moyen d'un modèle de mortalité dépendant dans un contexte de taux d'intérêt stochastique. Nous utilisons deux méthodes afin d'obtenir les deux premiers moments de la variable prospective de perte aléatoire. Nous

## 1D-C1: Actuarial Science and Finance 1

### Actuariat et finance 1

---

variable. The distribution function of the present value of future losses at given valuation time is derived and approximated for a homogeneous portfolio. The effects of the mortality dependence, the portfolio size and the policy type, as well as the impact of investment strategies on the riskiness of portfolios of survivorship life insurance policies are analyzed by means of moments and probability distributions.

dérivons la fonction de répartition de la valeur actuelle des futures pertes à un moment donné de l'évaluation et nous en faisons une approximation afin d'obtenir un portefeuille homogène. Nous analysons les effets de la dépendance de la mortalité, de la taille du portefeuille et du type de police, ainsi que l'incidence des stratégies d'investissement sur le niveau de risque des portefeuilles de polices d'assurance de survie au moyen de distributions des moments et de probabilités.

---

**YING WANG & JUN CAI** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 14:45]

*Fair Premium for Reinsurance Contracts*

*Primes équitables des contrats de réassurance*

Fair reinsurance premium principles considering risks of both insurer and reinsurer are introduced. Scholars usually investigate the optimal ceded function if the reinsurance premium is given. In this paper, quadratic and identity loss functions are applied to quantify the risks of both insurer and reinsurer. And two novel reinsurance premium principles are proposed. In addition, the properties for these two kinds of premium principles are studied.

Des principes de primes équitables en cas de réassurance, compte tenu des risques encourus à la fois par l'assureur et le réassureur sont présentés ici. En général, les chercheurs examinent la fonction de cession optimale, lorsque la prime de réassurance est connue. Dans cet article, les fonctions de perte quadratiques et d'identité sont appliquées pour quantifier les risques de l'assureur et du réassureur. Deux nouveaux principes de primes de réassurance sont proposés. De plus, les propriétés de ces deux types de principes de primes sont étudiées.



## 1D-C2: Biostatistics: Methods and Applications 2 Biostatistique : méthodes et applications 2

---

Chair/Président: Joseph Beyene (McMaster University)

Room/Salle: Thistle 259

### Abstracts/Résumés

---

**YANG JIAN, WU JINGJING & LI HAOCHENG** (University of Calgary)

[Monday May 30 / lundi 30 mai, 13:30]

*Minimum Hellinger Distance Estimation for Linear Regression Model*

*Méthode d'estimation de la distance minimale de Hellinger pour un modèle de régression linéaire*

Minimum Hellinger Distance estimation (MHDE) has been shown an appealing method of estimation for discrete data when the assumed model is suspected to be true. In this presentation, we first introduce Minimum Hellinger Distance (MHD) as well as the MHDE. Then, we will derive the MHDE for linear regression model. Some properties of the estimator are discussed and a comparison with MLE is carried out through Monte Carlo simulation studies. Lastly, we will apply this method to a breast cancer data set and demonstrate its implementation and efficiency in estimation.

L'estimation de la distance minimale de Hellinger (EDMH) se veut une méthode attrayante d'estimation pour les données discrètes lorsque le modèle supposé est soupçonné être le vrai. Dans cette présentation, nous présentons d'abord la distance minimale de Hellinger (DMH) et l'EDMH. Ensuite, nous allons dériver l'EDMH pour le modèle de régression linéaire. Certaines propriétés de l'estimateur sont discutées et une comparaison avec l'EMV est réalisée à l'aide d'études de simulation de Monte Carlo. Enfin, nous allons appliquer cette méthode à un jeu de données sur le cancer du sein et démontrer sa mise en œuvre et son efficacité dans l'estimation.

---

**ERIN LUNDY & CHARMAINE DEAN** (University of Western Ontario)

[Monday May 30 / lundi 30 mai, 13:45]

*Analyzing Heaped Counts and Longitudinal Presence/Absence Data in Joint Zero-inflated Poisson Regression Models*

*Analyse de données de dénombrement arrondies et de données longitudinales de présence et d'absence dans des modèles conjoints de régression de Poisson à surreprésentation de zéros*

Recurrent event data where a fraction of subjects are not at-risk for an event are frequently seen in longitudinal studies. In many settings, an aggregate count of the number of self-reported events over an observation period is recorded. Self-reported counts are often subject to heaping which yields a distorted distribution for the observed counts and therefore may bias estimation. Alternatively, the presence/absence of events between shorter periodic assessments may be recorded. Motivated by a major study of criminal behaviour, we compare the analysis of aggregate heaped count data and longitudinal presence/absence data using joint zero-inflated Poisson regression models.

Dans les études longitudinales, des données d'événements récurrents où une fraction des sujets n'est pas à risque pour un événement sont fréquemment observées. Dans de nombreuses situations, on enregistre des données de dénombrement agrégé du nombre d'événements autodéclarés sur une certaine période d'observation. Les données de dénombrement autodéclarées sont souvent arrondies, ce qui a une incidence sur la répartition des données observées et ce qui peut biaiser l'estimation. Par ailleurs, la présence ou l'absence d'événements entre des évaluations périodiques plus courtes peuvent être enregistrées. Motivés par une grande étude sur le comportement criminel, nous comparons l'analyse des données arrondies de dénombrement agrégé et les données de présence et d'absence au moyen de modèles de régression de Poisson à surreprésentation de zéros.

## 1D-C2: Biostatistics: Methods and Applications 2 Biostatistique : méthodes et applications 2

---

**RACHID BENTOUMI & MAYER ALVO** (Université d'Ottawa), **MHAMED MESFIOUI** (Université du Québec à Trois-Rivières)

[Monday May 30 / lundi 30 mai, 14:00]

*Information Gain under Length-biased Sampling*

*Gain en information sous l'échantillonnage à biais de longueur*

In epidemiological studies, subjects with disease (prevalent cases) differ from newly diseased (incident) cases. Methods for regression analyses have recently been proposed to measure the potential effects of covariates on survival. We propose to extend the measure of dependence based on information gain in the context of length-biased sampling. This will require development of an estimator for the covariate distribution. We will assess the asymptotic properties of the measure of dependence and estimated information gain and illustrate the methods using both simulation and application to the Canadian Study on Health.

Dans les études épidémiologiques, les sujets souffrant de la maladie (cas prévalents) diffèrent des cas nouvellement malades (incidents). Des méthodes pour les analyses de régression ont récemment été proposées pour mesurer les effets potentiels des covariables sur la survie. Nous proposons d'élargir la mesure de dépendance fondée sur le gain en information dans le contexte d'échantillonnage à biais de longueur. Ceci exigera le développement d'un estimateur pour la loi de la covariable. Nous évaluerons les propriétés asymptotiques de la mesure de dépendance et du gain en information estimé et nous illustrerons les méthodes en utilisant à la fois la simulation et l'application à l'étude canadienne sur la santé.

**ELNAZ GHADIMI & ARUSHARKA SEN** (Concordia University)

[Monday May 30 / lundi 30 mai, 14:15]

*Multivariate Cure Rate Estimation under Random Censoring*

*Estimation multivariée du taux de guérison sous censure aléatoire*

We considered a non-parametric multivariate cure rate estimator under random censoring. Individuals can be cured or fail. The event of interest is defined as death. Our data have censoring at the end of follow-up. Sen and Stute proposed a multivariate Kaplan-Meier estimator via mass-shifting method. We use the tail of the Sen-Stute estimator as an estimator of cure rate under random censoring. Taking a cue from Maller and Zhou without using martingale theory, the asymptotic normality of univariate and multivariate cure rate estimator is established and its variance estimator is obtained. The results are illustrated using simulation and real data.

Nous avons considéré un estimateur multivarié non-paramétrique du taux de guérison sous censure aléatoire. Les individus peuvent être guéris ou non. L'événement d'intérêt est défini comme étant le décès. Nos données sont censurées à la fin du suivi. Sen et Stute ont proposé un estimateur multivarié de Kaplan-Meier via la méthode de déplacement de la masse. Nous utilisons la queue de l'estimateur Sen-Stute comme estimateur du taux de guérison sous censure aléatoire. En s'inspirant de Maller et Zhou sans utiliser la théorie des martingales, la normalité asymptotique de l'estimateur univarié et multivarié du taux de guérison est établie et son estimateur de la variance est obtenu. Les résultats sont illustrés à l'aide de simulations et de données réelles.

**REGINA S. KAMPO, ASHLEY BONNER & JOSEPH BEYENE** (McMaster University)

[Monday May 30 / lundi 30 mai, 14:30]

*Assessing the Influence of Non-adherence on Fixed-Effect Meta-Analysis for a Continuous Outcome: A Simulation Study*

*Évaluer l'influence de la non-adhérence sur la méta-analyse à effets fixes pour un résultat continu : une étude de simulation*

## 1D-C2: Biostatistics: Methods and Applications 2

### Biostatistique : méthodes et applications 2

---

A traditional meta-analysis assumes complete adherence for an intervention. For some interventions such as exercise, change of diet etc., adherence might be a challenge that may compromise inferences. We investigated the influence of non-adherence on meta-analysis of a continuous outcome through simulations. We varied several key parameters and assessed the effect of non-adherence on estimation and hypothesis test properties. The findings from the simulation studies show that, the properties of estimation and hypothesis perform well under complete adherence but are not optimal as non-adherence is observed. Thus, researchers must treat non-adherence cautiously before using results in decision making.

Une méta-analyse traditionnelle suppose l'adhérence complète pour une intervention. Pour certaines interventions telles que l'exercice, le changement de régime alimentaire, etc., l'adhérence pourrait représenter un défi pouvant compromettre les inférences. Nous avons étudié l'influence de la non-adhérence sur la méta-analyse d'un résultat continu à l'aide de simulations. Nous avons fait varier plusieurs paramètres clés et avons évalué l'effet de la non-adhérence sur les propriétés de l'estimation et du test d'hypothèse. Les résultats des études de simulation montrent que les propriétés de l'estimation et de l'hypothèse fournissent de bons résultats sous adhérence complète, mais ne sont pas optimales lorsque la non-adhérence est observée. Ainsi, les chercheurs doivent traiter la non-adhérence avec prudence avant d'utiliser les résultats dans le processus décisionnel.

---

**YUYING XIE, YEYING ZHU & CECILIA A. COTTON** (University of Waterloo), **PAN WU** (Value Institute, Christiana Care Health System)

[Monday May 30 / lundi 30 mai, 14:45]

*A Model Averaging Approach for Estimating Propensity Scores by Optimizing Balance*

*Approche par modèle moyen pour l'estimation de scores de propension en optimisant l'équilibre*

Many approaches, including traditional parametric modelling and machine learning techniques, have been proposed to estimate propensity scores. We propose a new model averaging approach to propensity score estimation in which parametric and nonparametric estimates are combined to achieve covariate balance. Simulation studies are conducted across different scenarios varying in the degree of misspecification in the treatment model. It shows that the proposed method produces less bias and smaller standard error than existing approaches. This approach is applied to a real data set in evaluating the causal effect of formula or mixed feeding versus exclusive breastfeeding on a child's BMI Z-score.

De nombreuses approches, y compris la modélisation paramétrique traditionnelle et les techniques d'apprentissage automatique, ont été proposées pour estimer les scores de propension. Nous proposons une nouvelle approche par modèle moyen pour l'estimation des scores de propension, approche qui combine des estimations paramétriques et non paramétriques pour équilibrer les covariables. Des études de simulation sont menées selon divers scénarios dans lesquels le degré d'erreur de spécification du modèle de traitement varie. Nous voyons que contrairement aux approches existantes, il y a moins de biais et des erreurs types moins importantes avec la méthode proposée. Celle-ci est appliquée à un ensemble de données réelles afin d'évaluer l'effet causal de l'usage de formules ou d'une alimentation mixte pour les bébés comparativement au recours exclusif à l'allaitement sur le score-Z de l'IMC de l'enfant.

## 1D-C3: Statistical Methods and Applications 2 Méthodes statistiques et applications 2

---

Chair/Président: William Aeberhard (Dalhousie University)

Room/Salle: Thistle 255

### Abstracts/Résumés

---

**LATIFA BEN HADJ SLIMENE & ÉRIC MARCHAND** (Université de Sherbrooke), **THEODOROS NICOLERIS** (University of the Aegean, Greece)

[Monday May 30 / lundi 30 mai, 13:30]

*Inference of a Constrained Parameter in Presence of an Uncertain Constraint*

*Inférence d'une contrainte paramétrique en présence d'une contrainte incertaine*

We consider a statistical model  $X \sim f_\theta$  and the problem of estimating  $\theta$  under the parametric constraint  $\theta \in C \subset \mathbb{R}^p$ , with situations where uncertainty resides in the parametric constraint and where we take a simple hierarchical Bayes approach to describe the uncertainty relative to  $C$  and to  $\theta$ . We will focus on the case of a lower bound constraint. We provide various examples. For estimating a positive normal mean under squared error loss, we obtain hierarchical Bayes estimators which are also minimax. Finally, extensions to predictive density estimation are provided.

On considère le modèle statistique  $X \sim f_\theta$  et le problème de l'estimation de  $\theta$  sous la contrainte paramétrique  $\theta \in C \subset \mathbb{R}^p$ , où l'incertitude réside dans la contrainte paramétrique et où nous prenons une approche simple de Bayes hiérarchique pour décrire l'incertitude par rapport à  $C$  et  $\theta$ . On se concentre sur le cas d'une contrainte sur la borne inférieure. On fournit divers exemples. Pour estimer une moyenne normale positive sous la perte quadratique, on obtient les estimateurs de Bayes hiérarchiques qui sont aussi minimax. Enfin, les extensions à l'estimation par densité prédictive sont aussi prévues.

---

**ADAM RAHMAN & WAYNE OLDFORD** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 14:00]

*What Makes a Scatterplot Interesting?*

*Qu'est-ce qui rend un nuage de points pertinent?*

An important question that arises in modern data analytics is how to detect (and ultimately create) interesting structure in a set of points. Tools such as the minimum spanning tree and Delaunay triangulation can help to detect the presence of interesting structure. Using these concepts, Wilkinson and Wills introduced scagnostics, which uses the tools available in graph theory to numerically summarize the structure of a scatterplot in two dimensions. We will consider some of the limitations of scagnostics, propose new measures that capture interesting structure that escapes current scagnostics, and consider the possibilities of generalization to higher dimensions.

Une question importante qui surgit dans les analyses de données modernes concerne la façon de détecter (et finalement créer) une structure pertinente dans un ensemble de points. Des outils, tels que l'arbre recouvrant de longueur minimum et la triangulation de Delaunay, peuvent aider à détecter la présence d'une structure pertinente. Au moyen de ces concepts, Wilkinson et Wills ont introduit la librairie R scagnostic, qui utilise les outils disponibles dans la théorie des graphes pour résumer numériquement la structure d'un nuage de points en deux dimensions. Nous considérerons certaines limitations de scagnostic, et proposerons de nouvelles mesures qui permettent de capturer une structure pertinente qui échappe actuellement à scagnostic. Nous considérerons les possibilités de généralisation à des dimensions plus élevées.

## 1D-C3: Statistical Methods and Applications 2

### Méthodes statistiques et applications 2

---

**YILEI WU, YINGLI QIN & MU ZHU** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 14:15]

*Large Covariance Matrix Estimation - A Factor Model Approach*

*Estimation d'une grande matrice de covariance – une approche par modèle factoriel*

Estimating high-dimensional covariance matrix is a challenging and important problem. Motivated by lasso-type estimators in the literature, we reduce the complexity of the estimation problem by imposing a factor model structure for the population covariance matrix. We prove the consistency of our estimator under moderate conditions on the eigenvalues of the population covariance matrix. Our covariance matrix estimate can be used in various statistical procedures, such as discriminant analysis.

L'estimation d'une grande matrice de covariance est une préoccupation importante. Motivés par les estimateurs de type Lasso présentés dans la littérature, nous réduisons la complexité du problème que pose l'estimation en imposant une structure de modèle factoriel pour la matrice de covariance de la population. Nous prouvons également la convergence de notre estimateur sous des conditions modérées pour les valeurs propres de la matrice de covariance de la population. Notre estimation d'une matrice de covariance peut être utilisée pour diverses procédures statistiques, comme l'analyse discriminante.

**XIN LIU & WENQING HE** (Western University)

[Monday May 30 / lundi 30 mai, 14:30]

*Improve Performance of Support Vector Machine Classifiers with Data Adaptive Kernel*

*Amélioration de l'efficacité du classificateur d'une machine à vecteurs de support avec noyau adaptatif aux données*

In this talk, a new way to enhance the performance of an SVM classifier is presented. The initial kernel function is conformally re-scaled in an adaptive way so that the separation between two classes can be effectively enlarged, based on the prior knowledge obtained from the conventional SVM. The modified classifier takes into consideration the distribution of the support vectors in the feature space, and the correlation between voxels will be dealt with by selecting only limited numbers of parameters properly. Improvement of prediction accuracy from this data-dependent SVM is shown with numerical studies.

Nous présentons une nouvelle façon d'améliorer l'efficacité du classificateur d'une machine à vecteurs de support. La fonction initiale du noyau est pondérée de façon conforme et adaptative afin d'élargir efficacement la séparation entre deux classes, en fonction des connaissances préalables acquises à partir de la machine à vecteurs de support classique. Le classificateur modifié prend en considération la répartition des vecteurs de support dans l'espace des attributs, et la corrélation entre les voxels sera traitée en sélectionnant adéquatement seulement un certain nombre de paramètres. Au moyen d'études numériques, nous montrons l'amélioration de l'exactitude de la prédiction à partir de cette machine à vecteurs de support dépendante des données.

**NATHANIEL STEVENS** (University of San Francisco), **CHRISTINE M. ANDERSON-COOK** (Los Alamos National Laboratory)

[Monday May 30 / lundi 30 mai, 14:45]

*Comparing the Reliability of Related Populations with the Probability of Agreement*

*Comparaison de la fiabilité des populations apparentées et de la probabilité d'accord*

Combining information between different populations to improve precision, simplify future predictions or improve underlying understanding of relationships can be advantageous when considering

La combinaison de l'information entre différentes populations visant à améliorer la précision, à simplifier les prédictions futures ou à améliorer la compréhension sous-jacente des relations peut être un avantage lorsqu'on considère la

## 1D-C3: Statistical Methods and Applications 2

### Méthodes statistiques et applications 2

---

the reliability of several related sets of systems. Using the probability of agreement to help quantify the similarities of populations can help to give a realistic assessment of whether the systems have reliabilities that are sufficiently similar for practical purposes to be treated as a homogeneous combined population. The new method is described and illustrated with an example involving two generations of a complex system where the reliability is modeled using either a logistic or probit regression model.

fiabilité de plusieurs ensembles apparentés de systèmes. L'utilisation de la probabilité d'adéquation pour aider à quantifier les similitudes des populations peut contribuer à évaluer de façon réaliste si les systèmes ont des fiabilités qui sont suffisamment semblables, à des fins pratiques, pour être traitées comme une population homogène combinée. La méthode proposée est décrite et illustrée par un exemple impliquant deux générations d'un système complexe où la fiabilité est modélisée au moyen d'un modèle de régression logistique ou probit.

## 1D-C4: Statistical Theory Théorie statistique 1

---

Chair/Président: Jianfeng Yao (University of Hong Kong)

Room/Salle: Welch 8G

### Abstracts/Résumés

---

**HAOSUI DUANMU & DANIEL ROY** (University of Toronto)

[Monday May 30 / lundi 30 mai, 13:30]

*Nonstandard Complete Class Theorem*

*Théorème de classe complète non standard*

For finite parameter spaces under finite loss, there is a close link between optimal frequentist decision procedures and Bayesian procedures: every admissible procedure is Bayes. Using nonstandard analysis, we introduce the notion of a hyperfinite statistical decision problem and study the class of nonstandard Bayesian decision procedure—namely, those whose average risk with respect to some prior is within an infinitesimal of the optimal Bayes risk. We give some sufficient regularity conditions on standard statistical decision problems that imply every admissible procedure is nonstandard Bayes, and conditions such that nonstandard Bayes procedures are in fact Bayes ones.

Pour les espaces de paramètres finis sous la perte finie, il existe un lien étroit entre les procédures de décision fréquentistes optimales et les procédures bayésiennes : chaque procédure admissible est bayésienne. En utilisant une analyse non standard, nous introduisons la notion d'un problème statistique de décision hyperfini et nous étudions la classe de la procédure de décision bayésienne non standard, à savoir celle dont le risque moyen préalable est dans une infinitésimale du risque optimal de Bayes. Nous donnons des conditions suffisantes de régularité sur les problèmes de décision statistique standard qui impliquent que chaque procédure admissible est non standard Bayes et des conditions telles que les procédures de Bayes non standards sont en fait celles de Bayes.

---

**ROMAIN KADJE KENMOGNE & FRANÇOIS PERRON** (Université de Montréal)

[Monday May 30 / lundi 30 mai, 13:45]

*Density of the Ratio of Two Normal Random Variables*

*Densité du rapport de deux variables normalement distribuées*

The random vector  $(X, Y)$  has a multinormal distribution and we are looking for the distribution of the ratio  $X/Y$ . This issue was addressed in a long paper (T. PHAM GIA et al.) often cited on Google Scholar. The purpose of this work is to significantly reduce the length of the proofs and to explain how to numerically solve estimation problems in a Bayesian setting. We show that the density of  $X/Y$  is a mixture of new densities belonging to a family. We discuss the properties of this family. Some convergence results are given.

Le vecteur aléatoire  $(X, Y)$  est de loi multinormale et on cherche la loi du rapport  $X/Y$ . Ce sujet a été abordé dans un long article (T. PHAM-GIA et coll.) souvent cité sur Google Scholar. Le but de ce travail est de raccourcir considérablement la longueur des démonstrations et d'expliquer comment résoudre numériquement des problèmes d'estimation dans un cadre bayésien. Nous montrons que la densité de  $X/Y$  est un mélange de nouvelles densités appartenant à une certaine famille. Nous discutons des propriétés de cette famille. Des résultats de convergence sont donnés.

## 1D-C4: Statistical Theory Théorie statistique 1

---

**FÉLIX CAMIRAND LEMYRE & TAOUFIK BOUEZMARNI** (Université de Sherbrooke), **JEAN-FRANÇOIS QUESSY** (Université du Québec à Trois-Rivières)

[Monday May 30 / lundi 30 mai, 14:00]

*Nonparametric Measures of Local Causality and Tests of Non-Causality in Time Series*

*Mesures non paramétriques de la causalité locale et tests de non causalité pour des séries temporelles*

To study the causal relationships in a process  $(Y_t, Z_t)_{t \in \mathbb{Z}}$ , a widely-used approach is to consider the Granger causality. In the case of Markovian processes, the notion is based on the joint distribution of  $(Y_t, Z_{t-1})$  given  $Y_{t-1}$ . The Granger causality measures proposed so far are global, which means that if the relationship between  $Y_t$  and  $Z_{t-1}$  changes with the value of  $Y_{t-1}$ , this will not be captured. To circumvent this limitation, *local* Granger causality indices based on the conditional copula of  $(Y_t, Z_{t-1})$  given  $Y_{t-1} = x$  are here proposed, and the asymptotic behavior of nonparametric estimators is obtained for  $\alpha$ -mixing processes.

Pour étudier la dépendance entre  $(Y_t, Z_t)_{t \in \mathbb{Z}}$ , on s'intéresse souvent à la causalité de Granger. Dans le cas de processus Markoviens, cette notion se base sur la distribution de  $(Y_t, Z_{t-1})$  sachant  $Y_{t-1}$ . À ce jour, les mesures de la causalité de Granger sont globales, en ce sens où si la dépendance entre  $Y_t$  et  $Z_{t-1}$  change en fonction des valeurs de  $Y_{t-1}$ , alors ces changements ne pourront être détectés. Pour contourner ce problème, des mesures de causalité basées sur la copule conditionnelle de  $(Y_t, Z_{t-1})$  sachant  $Y_{t-1} = x$  seront ici proposées, et leur normalité asymptotique est établie pour des processus  $\alpha$ -mélangeants.

---

**VICTOR VEITCH & DANIEL ROY** (University of Toronto)

[Monday May 30 / lundi 30 mai, 14:15]

*Models and Inference for Sparse Random Graphs Using Exchangeable Random Measures*

*Modèles et inférence pour des graphes aléatoires épars à l'aide de mesures aléatoires échangeables*

We introduce a class of random graphs that meets many of the desiderata for a foundation for statistical analysis of real-world networks. The class of random graphs is defined by a probabilistic symmetry: invariance of the distribution of each graph to an arbitrary relabelings of its vertices. We interpret a symmetric simple point process on  $\mathbb{R}_+^2$  as the edge set of a random graph, and formalize the probabilistic symmetry as joint exchangeability of the point process. We give a representation theorem for the class of random graphs satisfying this symmetry.

Nous présentons une classe de graphes aléatoires qui répond à bon nombre des desideratas pour établir le fondement d'une analyse statistique des réseaux réels. La classe de graphes aléatoires se définit par une symétrie probabiliste : l'invariance de la distribution de chaque graphe au ré-étiquetage arbitraire de ses sommets. Nous interprétons un processus ponctuel symétrique sur  $\mathbb{R}_+^2$  comme l'ensemble des arêtes d'un graphe aléatoire et formalisons la symétrie probabiliste comme l'échangeabilité conjointe du processus ponctuel. Nous proposons un théorème de représentation pour la classe de graphes aléatoires qui répond à cette symétrie.

---

**JUN YANG & DANIEL ROY** (University of Toronto)

[Monday May 30 / lundi 30 mai, 14:30]

*Meta-Bayesian Analysis*

*Méta-analyse bayésienne*

The "optimality" of the Bayesian approach to inference does not hold when the model is misspecified. As essentially every statistical model is mis-

Le caractère optimal de l'approche bayésienne à l'inférence ne tient pas la route en cas d'erreur de spécification du modèle. Comme la spécification de tout modèle statistique est



## 1D-C4: Statistical Theory Théorie statistique 1

---

specified, this raises the question: what is a prior? We formalize the problem of choosing a (surrogate) prior as a Bayesian decision theory task, and develop theory and algorithms for choosing optimal priors. The resulting theory, which we call meta-Bayesian analysis, gives priors a pragmatic interpretation and has some surprising consequences. For example, the optimal prior may depend on the number of data points you plan to observe, and the number of predictions you expect to make.

essentiellement erronée, une question se pose : qu'est-ce qu'une loi a priori ? Nous formalisons le problème du choix d'une loi a priori (de substitution) comme une tâche de théorie de décision bayésienne et mettons au point une théorie et des algorithmes pour le choix de lois a priori optimales. Ladite théorie que nous appelons méta-analyse bayésienne donne une interprétation pragmatique aux lois a priori et mène à certaines conséquences étonnantes. Par exemple, une loi a priori optimale peut dépendre du nombre de données que vous prévoyez observer et des prévisions que vous comptez faire.

---

**ZHIYANG ZHOU** (Simon Fraser University), **RUNCHU ZHANG** (Nankai University and Northeast Normal University)  
[Monday May 30 / lundi 30 mai, 14:45]

*A Generalized General Minimum Lower Order Confounding Criterion for Non-regular Designs*  
*Critère généralisé de l'amalgame général minimal d'ordre inférieur pour des plans non réguliers*

We extend the work of Zhang et al. [Statistica Sinica 18, 1689--1705] for non-regular designs and propose two new concepts, i.e., the generalized aliasing effect-number pattern ( $G_2$ -AENP) and generalized general minimum lower order confounding ( $G_2$ -GMC). It proves that (i) isomorphic designs have the same  $G_2$ -AENP and (ii) the generalized minimum aberration (GMA) and minimum moment aberration (MMA) can both be treated as ones that optimize functions over the  $G_2$ -AENP. That is, the  $G_2$ -GMC criterion is more sensitive in the recognition of non-isomorphic designs.

Nous généralisons les travaux de Zhang et coll. [Statistica Sinica 18, 1689--1705] pour les plans non réguliers et proposons deux nouveaux concepts, c.-à-d. la suite numérique à effet de repliement de spectre généralisé ( $G_2$ -AENP) et l'amalgame général minimal d'ordre inférieur ( $G_2$ -GMC). Cela prouve (i) que les concepts isomorphiques ont le même  $G_2$ -AENP et (ii) que l'aberration minimale généralisée (AMG) et l'aberration minimale de moments (AMM) peuvent toutes deux être traitées comme celles qui optimisent les fonctions sur le  $G_2$ -AENP. En fait, le critère  $G_2$ -GMC est plus sensible dans la reconnaissance de plans non isomorphiques.

## 1D-C5: Statistics Education Éducation en statistique

---

Chair/Président: Alison Gibbs (University of Toronto)

Room/Salle: Thistle 256

### Abstracts/Résumés

---

**JEREMY BALKA** (University of Guelph)

[Monday May 30 / lundi 30 mai, 13:30]

*A Traditional Introductory Statistics Course with Extensive Video Support*

*Un cours d'introduction aux statistiques traditionnelles avec un vaste support vidéo*

In an introductory statistics course, I supply complete course notes, more than 100 accompanying videos, and exercise sets with full solutions. In a sense, the course has developed into an online course with lectures as a bonus. In this talk I will discuss my motivation for setting up the course in this way, some of the pros and cons of this approach, and some of my experiences teaching the course.

Dans un cours d'introduction aux statistiques, je donne des notes de cours complètes, plus de 100 vidéos d'accompagnement, et des ensembles d'exercices avec solutions complètes. En un sens, le cours est devenu un cours en ligne avec conférences en prime. Dans cette présentation, je discuterai de la raison pour laquelle j'ai créé le cours de cette façon, quelques pour et contre de cette approche, et un peu de mon expérience lors de l'enseignement de ce cours.

---

**DUNHAM BRUCE, MELISSA LEE & GAITRI YAPA** (University of British Columbia)

[Monday May 30 / lundi 30 mai, 13:45]

*Investigating How Students Interact with Simulation-based Applets*

*Étude de la manière dont les étudiants interagissent avec les applets basés sur la simulation*

The use of applets in helping students grasp statistical concepts has attracted much interest in the statistical education community. There is evidence such tools may improve student learning, but there is little research on how students engage with applets or how they should be designed. As part of the development of a set of new applets, we have interviewed students and observed how they interact with the applets. Suggestions for how to conduct such research will be discussed, along with some provisional findings.

L'utilisation d'applets pour aider les étudiants à saisir les concepts statistiques a suscité beaucoup d'intérêt dans le milieu de l'éducation en statistique. Il existe des preuves que de tels outils peuvent améliorer l'apprentissage des élèves, mais il existe peu de recherches sur la façon dont les élèves participent avec les applets ou comment ceux-ci doivent être conçus. Dans le cadre de l'élaboration d'un ensemble de nouveaux applets, nous avons interrogé les étudiants et observé comment ils interagissaient avec les applets. Des suggestions sur la façon de mener une telle recherche seront discutées, avec quelques conclusions provisoires.

---

**SOHEE KANG** (University of Toronto Scarborough)

[Monday May 30 / lundi 30 mai, 14:00]

*Project-Based Statistics Learning Through Wikispace*

*Apprentissage des statistiques fondé sur le projet à l'aide de Wikispace*

“Learn by doing” is the type of experience that educators strive to facilitate for students. Project based learning is an instructional strategy in which stu-

« Apprendre en faisant » est le type d'expérience que les éducateurs recherchent pour aider les étudiants. L'apprentissage fondé sur le projet est une stratégie d'enseignement

## 1D-C5: Statistics Education Éducation en statistique

---

dents work cooperatively to create a product. A final project in upper year statistics courses provides students with opportunities to build 21st century learning skills such as collaboration, communication, critical thinking, and the use of technology. In this talk, I share the classroom experience of using "Wikispace" to create a wiki-like page as a final product for my fourth year course, Multivariate Analysis. How to create and maintain projects will be demonstrated along with students' products.

dans laquelle les étudiants travaillent en collaboration pour créer un produit. Un projet final dans les cours de statistique de dernière année fournit aux élèves des occasions d'acquérir des compétences d'apprentissage du 21e siècle telles que la collaboration, la communication, la pensée critique et l'utilisation de la technologie. Dans cet exposé, je partage l'expérience en classe de l'utilisation de « Wikispace » pour créer une page wiki-like comme produit final de mon cours de quatrième année sur l'analyse multivariée. Nous montrerons comment créer et maintenir des projets ainsi que les produits des élèves.

---

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

[Monday May 30 / lundi 30 mai, 14:15]

*P-values: Love them - Hate them - Leave them. Implications for Statistical Education.*

*Valeurs de p : aimez-les - détestez-les - laissez-les. Implications en éducation en statistique*

The vast majority of our introductory statistics text books teach significance tests as the definitive method for statistical inference. In light of the ban on p-values by some journals, and the ASA's recent statement on p-values, should we be updating this approach? Great strides have been made in statistical education, stressing the teaching of concepts rather than calculations, introducing students to real world data from the beginning of the course. Many of our students still leave the first course with very fuzzy ideas about inference. We rarely see non-frequentist approaches taught in our introductory courses. Is it time for change?

La grande majorité de nos manuels d'introduction à la statistique enseignent les tests de signification en tant que méthode définitive pour l'inférence statistique. À la lumière de l'interdiction sur les valeurs p par certains journaux et de la récente déclaration de l'ASA sur les valeurs p, devrions-nous mettre à jour cette approche? De grands progrès ont été réalisés en éducation en statistique, en insistant sur l'enseignement des concepts plutôt que sur les calculs, en initiant les élèves aux données du monde réel au début du cours. Un grand nombre de nos étudiants quittent encore le premier cours avec des idées très floues au sujet de l'inférence. Nous voyons rarement des approches non-fréquentistes enseignées dans nos cours d'introduction. Le temps est-il venu pour un changement?

---

**LORI MURRAY** (University of Western Ontario)

[Monday May 30 / lundi 30 mai, 14:30]

*Five Years of Blending*

*Cinq années de mélange*

A blended course format is a combination of in-class and online learning. Although the unique format has been in use for several years, for many it is still a new concept. We have been using blended learning in an introductory statistics course at the University of Western Ontario since 2012. In this talk, I will give my experience as the instructor, highlighting some of the advantages and pitfalls that I have encountered along the way.

Un format de cours hybride est un mélange d'apprentissage en classe et en ligne. Pour beaucoup d'entre nous, c'est encore un nouveau concept, bien que ce format unique ait été utilisé depuis plusieurs années. Depuis 2012, nous avons donné un enseignement hybride dans un cours d'introduction à la statistique à l'Université de Western Ontario. Dans cet exposé, je vous ferai part de mon expérience à titre d'enseignant et je décrirai certains des avantages et des inconvénients de ce genre d'enseignement.

## 1D-C5: Statistics Education Éducation en statistique

---

**NATHAN TABACK & ALISON GIBBS** (University of Toronto)

[Monday May 30 / lundi 30 mai, 14:45]

*Students' Attitudes Towards Statistics in Online versus Flipped Classrooms*

*L'attitude des étudiants à l'égard des statistiques dans les cours en ligne par rapport aux classes inversées*

Technology has allowed instructors to experiment with different course delivery methods including flipped and fully online courses. In the Fall of 2015 at the University of Toronto, our large (N=1,413) introductory course was taught in both flipped and fully online formats, with some discipline-specific sections. To investigate if students' attitudes towards statistics differ depending on course delivery method, students were asked to complete pre and post course surveys (SATS-36) measuring attitudes towards statistics. We'll report on a comparison of attitude components between flipped and online sections, and discipline and non-discipline specific sections of the course, adjusted for baseline student covariates.

La technologie a permis aux enseignants d'essayer diverses méthodes de prestation de cours, y compris les cours en ligne et les classes inversées. À l'automne de 2015 à l'Université de Toronto, notre volumineux cours d'introduction (N=1413) a été donné dans l'un et l'autre format, avec toutefois des sections propres à la discipline. Pour savoir si l'attitude des étudiants à l'égard des statistiques différait en fonction de la méthode d'enseignement, des sondages préalables et subséquents au cours ont été menés auprès d'eux (SATS-36) afin de mesurer ce facteur. Nous ferons état d'une comparaison des aspects de leur attitude en fonction des sections enseignées en classe inversée et dans le cours en ligne ainsi qu'entre les sections propres ou non à la discipline, avec ajustement pour les covariables de base des étudiants.

## 1E-I1: Advances in Longitudinal Data Analysis Progrès récents en analyse de données longitudinales

---

Chair/Président: Yogendra Chaubey (Concordia University)  
Organizer/Responsable: Sanjoy Sinha (Carleton University)

**Room/Salle: Welch 209**

### Abstracts/Résumés

---

**KARELYN DAVIS** (Health Canada)

[Monday May 30 / lundi 30 mai, 15:30]

*Longitudinal Analysis in Health Studies with Censored Observations*

*Analyse longitudinale d'études sur la santé avec observations censurées*

In performing statistical analysis of national health studies, statisticians often encounter data that is correlated either due to a natural clustering or is collected longitudinally. In particular, when concerned with contaminant levels in humans (e.g. mercury, lead), laboratory data may be collected in such a fashion so as to determine significant patterns in vulnerable populations such as pregnant women and children. However, such contaminant levels may be so low as to be below the laboratory limit of detection and are thus classified as non-detects. In this presentation, analyses encountered from a Health Canada study are discussed which extend linear mixed models to incorporate parameter constraints and analyses of left-censored observations.

Lors d'analyses statistiques d'études nationales sur la santé, les statisticiens font souvent face à des données qui sont corrélées soit en raison d'une classification naturelle ou d'une collecte longitudinale. Notamment, lorsque préoccupés par les concentrations de contaminants chez les humains (par exemple, mercure, plomb), les données de laboratoire peuvent être recueillies de manière à déterminer des tendances significatives chez les groupes plus vulnérables tels que les femmes enceintes et les enfants. Par contre, ces concentrations de contaminants peuvent être assez faibles pour être sous la limite de détection du laboratoire et ainsi être classifiées en tant que concentration non détectée. Lors de cette présentation, des analyses provenant d'une étude de Santé Canada seront discutées. Ces analyses élargissent les modèles linéaires mixtes pour incorporer des contraintes de paramètres et des analyses d'observations censurées à gauche.

---

**ABDUS SATTAR** (Case Western Reserve University), **SANJOY SINHA** (Carleton University)

[Monday May 30 / lundi 30 mai, 16:00]

*Joint Modeling of Longitudinal and Survival Data With a Covariate Subject to Limit of Detection*

*Modélisation conjointe de données longitudinales de survie avec une covariable soumise à une limite de détection*

We develop and study innovative methods for jointly modeling longitudinal and time-to-event data with a covariate or biomarker subject to the limit of detection (LOD). We investigate the gain in efficiency from joint models over separate models when we attempt to estimate the effect of an endogenous time-dependent covariate on survival times. We also investigate the effects of misspecified random effects distributions on the likelihood inference. Our extensive simulation study indicates

Nous développons et étudions des méthodes novatrices pour modéliser conjointement des données longitudinales et de temps d'événement avec une covariable ou un biomarqueur soumis à une limite de détection (LOD). Nous examinons le gain en efficacité des modèles conjoints sur les modèles séparés lorsque nous tentons d'estimer l'effet d'une covariable endogène avec une dépendance temporelle sur les temps de survie. Nous examinons aussi les effets des lois des effets aléatoires mal spécifiés sur l'inférence de vraisemblance. Notre vaste étude de simulation indique que si la

## 1E-I1: Advances in Longitudinal Data Analysis

### Progrès récents en analyse de données longitudinales

---

that if the assumed random effects distributions deviate from the true distributions to a large extent, the maximum likelihood method can produce systematically biased estimators. We present an application of the proposed method using a large clinical trial data from the Genetic and Inflammatory Markers of Sepsis (GenIMS) study.

loi des effets aléatoires présumée dévie en grande partie de la vraie loi, la méthode de maximum de vraisemblance peut produire systématiquement des estimateurs biaisés. Nous présentons une application de la méthode proposée avec des données d'un vaste essai clinique de l'étude sur les marqueurs génétiques et inflammatoires du sepsis (GenIMS).

---

**SANJOY SINHA** (Carleton University)

[Monday May 30 / lundi 30 mai, 16:30]

*Methods for Longitudinal Data With Nonignorable Missing Responses and Outliers*

*Méthodes pour données longitudinales avec réponses manquantes et valeurs aberrantes ne pouvant être ignorées*

We encounter missing data in many longitudinal studies. When data are nonignorably missing, it is important to analyze the data by incorporating a missing data model into the observed data likelihood function. It is well-known that the maximum likelihood (ML) estimators are generally sensitive to potential outliers in the data. I propose and explore a robust method, which is developed in the framework of the ML estimation and is useful for down-weighting any influential observations when estimating the model parameters. The empirical properties of the robust estimators are studied in simulations. The proposed method is also illustrated in an example using actual longitudinal data on CD4 counts obtained from clinical trials of HIV-infected patients.

Plusieurs études longitudinales contiennent des données manquantes. Lorsque la non-réponse est non-ignorable, il est important d'analyser les données en incorporant un modèle de données manquantes dans la fonction de vraisemblance des données observées. Il est bien connu que les estimateurs de maximum de vraisemblance (ML) sont habituellement sensibles aux valeurs aberrantes potentielles dans les données. Je propose et explore une méthode robuste qui est développée selon le cadre de l'estimation ML et qui est utile pour réduire la pondération d'observations qui ont de l'influence lors de l'estimation des paramètres du modèle. Les propriétés empiriques des estimateurs robustes sont étudiées dans des simulations. La méthode proposée est aussi illustrée au moyen d'un exemple utilisant des données longitudinales réelles sur le CD4 provenant d'essais cliniques sur des patients infectés par le VIH.

## 1E-I2: Advances in Regression Tree Modelling Progrès récents en modélisation par les arbres de régression

---

Organizer and Chair / Responsable et président: Matthew Pratola (Ohio State University)

Room/Salle: Thistle 242

### Abstracts/Résumés

---

**HUGH CHIPMAN** (Acadia University), **ROBERT MCCULLOCH** (University of Chicago), **MATTHEW PRATOLA** (Ohio State University), **EDWARD GEORGE** (University of Pennsylvania)

[Monday May 30 / lundi 30 mai, 15:30]

*Dispersion Modelling with an Ensemble of Trees*  
*Modèle de dispersion avec ensemble d'arbres*

Bayesian additive regression trees (BART) is a flexible and scalable supervised learning model that offers accurate assessment of uncertainty via credible intervals. It makes the strong assumption of iid errors. Even when error variance is nonconstant, BART can still give accurate point predictions. However, credible intervals are unlikely to remain accurate or useful. We develop a novel heteroscedastic BART model to alleviate these concerns. This is achieved through the introduction of Bayesian Multiplicative Trees, which model the variance component of BART as a function of the predictors. We implement the approach and demonstrate it in several examples.

Les arbres additifs de régression bayésiennes (BART) forment un modèle d'apprentissage supervisé flexible et évolutif qui offre une évaluation précise de l'incertitude via des intervalles de crédibilité. Ce modèle se base sur la forte hypothèse des erreurs iid. Même lorsque l'erreur de la variance est non constante, BART peut donner des prédictions ponctuelles précises. Par contre, il est improbable que les intervalles de crédibilité demeurent précis ou utiles. Nous développons un nouveau modèle BART hétéroscédastique pour dissiper ces préoccupations. Cela est possible par l'introduction d'arbres multiplicatifs bayésiens, qui modélisent la composante de variance en tant que fonction des prédicteurs. Nous employons et démontrons cette approche à travers plusieurs exemples.

---

**TOM LOUGHIN & ANDREW HENREY** (Simon Fraser University)

[Monday May 30 / lundi 30 mai, 16:00]

*Robust Adaptively Pruned Random Forests Using Likelihood-Based Trees*  
*Forêts aléatoires adaptatives élaguées et robustes utilisant des arbres fondés sur la vraisemblance*

Random Forests for regression are typically constructed using standard regression trees. These trees make splits that minimize squared error, which implicitly assumes homoscedasticity. They are not robust against heteroscedasticity, and this can be passed on to the forests. As heteroscedasticity is prevalent in a substantial fraction of real datasets, random forests may underperform frequently. Furthermore, they are inefficient at fitting mean functions that are partially flat. We present a likelihood-based version of regression trees that explicitly model both the mean and variance, and use these trees as base learners in our random

Les forêts aléatoires de régression sont généralement construites à l'aide d'arbres de régression standards. Ces arbres font des divisions qui minimisent l'erreur carrée, ce qui suppose implicitement une homoscedasticité. Ils ne sont pas robustes en présence d'hétéroscédasticité, ce qui peut se transférer aux forêts. Comme l'hétéroscédasticité est courante dans une part substantielle des données réelles, les forêts aléatoires peuvent fréquemment être peu efficaces. De plus, ils sont inefficaces pour ajuster les fonctions moyennes qui sont partiellement planes. Nous présentons une version des arbres de régression basée sur la vraisemblance qui modélise explicitement à la fois la moyenne et la variance, et qui utilise ces arbres comme base d'apprentissage dans

## 1E-I2: Advances in Regression Tree Modelling

### Progrès récents en modélisation par les arbres de régression

---

forest. We also develop a fast pruning algorithm based on information criteria that improves the fit to partially flat mean functions.

notre forêt aléatoire. Nous développons aussi un algorithme rapide d'élagage fondé sur des critères d'information qui améliore l'ajustement avec les fonctions moyennes partiellement planes.

---

**DANIEL ROY** (University of Toronto)

[Monday May 30 / lundi 30 mai, 16:30]

*Mondrian Processes and their Statistical Applications*

*Les processus Mondrian et leurs applications statistiques*

Mondrian processes are a class of continuous time stochastic processes that induce a random hierarchical partition of a product space. In this talk, I will survey Mondrian processes and their application to a number of statistical problems, including network modeling and efficient online classification and regression via Mondrian forests.

Les processus Mondrian sont une classe de processus stochastiques en temps continu qui induisent une partition aléatoire hiérarchique d'un espace produit. Durant cet exposé, je passerai en revue les processus Mondrian et leurs applications à un certain nombre de problèmes statistiques, y compris les modèles de réseau et la classification efficace en ligne et la régression via les forêts Mondrian.



## 1E-I3: An Introduction to the Joint Modelling of Longitudinal and Survival Data Introduction à la modélisation conjointe des données longitudinales et de survie

---

Organizer and Chair / Responsable et président: Fernando Camacho (Damos Inc.)

Room/Salle: Welch - David Howes

### Abstract/Résumé

---

**GRACE Y. YI** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 15:30]

*Introduction of Joint Modeling of Longitudinal and Survival Data*

*Présentation d'une modélisation conjointe de données longitudinales et de survie*

Longitudinal data and survival data frequently arise together in biomedical studies, clinical trials and observational studies. These data are often associated. Separate analyses of longitudinal data and survival data may lead to inefficient or biased results. Joint models of longitudinal and survival data, on the other hand, incorporate all information simultaneously, and provide valid and efficient inferences. Over the last two decades, joint modeling of longitudinal and survival data has received increasing attention. A number of modeling strategies and inferential procedures have been developed in the literature. In this talk, I will give an introductory overview and discuss a broad range of issues on joint modeling.

Les données longitudinales et les données de survie apparaissent souvent ensemble dans les études biochimiques, les essais cliniques et les études d'observation, et sont souvent associées. Les analyses séparées de données longitudinales et de données de survie pourraient causer des estimations inefficaces ou biaisées. D'autre part, les modèles conjoints de données longitudinales et de données de survie intègrent toute l'information simultanément et fournissent des inférences valides et efficaces. Au cours des vingt dernières années, la modélisation conjointe de données longitudinales et de données de survie a suscité une attention croissante. Un certain nombre de stratégies et de procédures d'inférence ont été conçues dans la littérature. Dans cet exposé, je donnerai un aperçu et je discuterai des divers problèmes liés à la modélisation conjointe de ce type de données.

## 1E-I4: Missing Data in Survey Sampling Données manquantes en échantillonnage

---

Chair/Président: Michael Hidioglou (Statistics Canada)  
Organizer/Responsable: Wes Yung (Statistics Canada)

Room/Salle: Thistle 258

### Abstracts/Résumés

---

**SONG CAI & J.N.K. RAO** (Carleton University)

[Monday May 30 / lundi 30 mai, 15:30]

*Empirical Likelihood Inference Based on Estimation Equations for Complex Survey with Data Missing Completely At Random*

*Inférence de vraisemblance empirique fondée sur les équations d'estimation pour sondage complexe avec données manquantes aléatoirement*

We propose an empirical likelihood (EL) method for constructing confidence intervals for population parameters defined by estimation equations (EEs) with data from complex surveys that are missing completely at random. Instead of imputing the original data, we impute the EEs using fractional imputation with fixed number of draws. We then construct an EL ratio based on the imputed EEs and show that this ratio has a chi-bar-square limiting distribution in general. We also show that, when no missing data are present, the proposed EL ratio has a simple chi-square limiting distribution under PPS sampling with replacement. Moreover, we propose a proper bootstrap procedure to approximate the limiting distribution of the EL ratio for constructing confidence intervals for parameters of interest.

Nous proposons une méthode de vraisemblance empirique (EL) pour la construction d'intervalles de confiance pour des paramètres d'une population définis par des équations d'estimation (EEs) avec des données manquantes aléatoirement provenant de sondages complexes. Au lieu d'imputer les données originales, nous imputons les EEs en utilisant une imputation fractionnaire avec un nombre fixé de tirages. Nous construisons ensuite un rapport EL basé sur les EEs attribués et démontrons que ce rapport a généralement une loi limite khi carré. Nous démontrons aussi que, lorsqu'il n'y a aucune donnée manquante, le rapport EL proposé a une simple loi limite khi carré selon l'échantillonnage PPT avec remise. De plus, nous proposons une procédure bootstrap pour approximer la loi limite du rapport EL pour la construction d'intervalles de confiance pour des paramètres d'intérêt.

---

**DAVID HAZIZA** (Université de Montréal), **ÉRIC LESAGE** (INSEE)

[Monday May 30 / lundi 30 mai, 16:00]

*Properties of the Instrument Vector Calibration Estimator in the Presence of Unit Nonresponse*  
*Propriétés de l'estimateur par calage instrumental en présence de non-réponse totale*

In recent years, instrument vector calibration has received a lot of attention in the literature in the context of unit nonresponse in surveys. In this paper, we lay out the conditions under which the instrument vector calibration estimator is consistent. Even when the conditions are met, the resulting estimator may suffer from variance amplification. When the conditions are not met, the estimator may suffer from both bias and variance amplification. Results from a simulation study will be presented.

Ces dernières années, le calage instrumental a reçu beaucoup d'attention dans la littérature dans le contexte de la non-réponse totale dans les enquêtes. Dans cet article, nous exposons les conditions selon lesquelles l'estimateur par calage instrumental est convergent. Même lorsque les conditions sont remplies, l'estimateur résultant peut subir une amplification de la variance. Lorsque les conditions ne sont pas remplies, l'estimateur peut souffrir d'une amplification du biais et d'une amplification de la variance. Des résultats provenant d'une étude de simulation seront présentés.

## 1E-I4: Missing Data in Survey Sampling Données manquantes en échantillonnage

---

**XICHEN SHE & CHANGBAO WU** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 16:30]

*Fully Efficient Joint Fractional Imputation for Incomplete Bivariate Ordinal Responses*

*Imputation fractionnaire conjointe pleinement efficace pour des réponses bivariées ordinales incomplètes*

In medical studies and many fields of social sciences, pairwise ordinal responses are one of the common data formats collected for analysis. It is often that one or both of the responses are subject to missingness. In this paper, we propose a fully efficient joint fractional imputation procedure, which creates a single complete dataset with fractional weights. We show that this synthetic dataset not only leads to valid inference for simple parameters such as marginal proportions but also preserves the correlation structure, which further leads to valid association measure estimation. Theoretical results on inference procedures based on the fractionally imputed dataset are presented, and finite sample performances of the methods with comparisons to existing methods are investigated through simulation studies.

Dans les études médicales ainsi que dans plusieurs domaines des sciences sociales, les réponses ordinales par paires sont un des formats de données communs recueillis pour analyse. Fréquemment, une réponse ou même parfois les deux réponses sont manquantes. Dans cet article, nous proposons une procédure d'imputation fractionnaire conjointe pleinement efficace qui crée un ensemble de données complet unique avec des pondérations fractionnaires. Nous démontrons que cet ensemble de données synthétique ne conduit pas seulement à une inférence valide pour des paramètres simples tels que des proportions marginales mais préserve également la structure de la corrélation, ce qui conduit à une estimation de mesures d'association valide. Des résultats théoriques des procédures d'inférence basées sur les ensembles de données fractionnaires imputées sont présentés, et les performances de ces méthodes sur des échantillons finis avec comparaisons avec des méthodes existantes sont examinées au moyen d'études de simulation.

## 1E-15: Probability and Statistical Problems and Results Arising from Neural Models Probabilités et problèmes statistiques et résultats découlant de modèles neuronaux

---

Organizer and Chair / Responsable et présidente: Priscilla Greenwood (University of British Columbia)

Room/Salle: Thistle 246

### Abstracts/Résumés

---

**ANDRE LONGTIN, STEPHEN CLARKE & LEN MALER** (University of Ottawa)

[Monday May 30 / lundi 30 mai, 15:30]

*Fisher Information Analysis of Neural Focussing Point Processes*

*Analyse d'information de Fisher de processus de point focal neuronal*

The existence of a focus in a sensory system suggests that the neural activity acquires a specific feature around the focal point. Such a feature is currently unknown. This activity should be altered when the stimulus either looms towards or recedes away from the focal point. We first develop Fisher information theory for non-Poisson firing time point process data. We find that there is a distance at which this information is maximized during looming and receding object motions. This distance is in fact chosen by certain animals to view an object. Strikingly, this maximum occurs at a bifurcation between tonic firing and bursting patterns in the neuron's dynamics where the point process becomes clustered.

L'existence d'un point focal dans un système sensoriel suggère que l'activité neuronale acquiert une caractéristique spécifique autour de celui-ci. Cette caractéristique est actuellement inconnue. Cette activité devrait être modifiée lorsque le stimulus s'approche ou s'éloigne du point focal. Nous commençons par développer une théorie d'information de Fisher pour les données de points ponctuels de temps de décharge non Poisson. Nous trouvons qu'il existe une distance à laquelle cette information est maximisée pendant le rapprochement et l'éloignement d'un objet. Cette distance est en fait choisie par certains animaux pour visualiser un objet. Remarquablement, ce maximum se présente à une bifurcation entre les modes de décharge et de rafale tonique dans la dynamique du neurone où le processus ponctuel devient regroupé.

---

**PETER ROWAT** (University of California at San Diego), **PRISCILLA (CINDY) GREENWOOD** (University of British Columbia)

[Monday May 30 / lundi 30 mai, 16:00]

*Firing Patterns of a Stochastic Neural Model*

*Modes de décharge d'un modèle neuronal stochastique*

Neurons in the medial entorhinal cortex exhibit an hexagonal grid-like spatial pattern of spike rates that has been proposed to represent a neural code for the navigational ability of rats and other mammals. Such a spatial pattern can be simulated from a network of neurons with both excitatory and inhibitory connections. We use a simple stochastic model of a stellate cell, with input from a simulated spatial firing field, to explore the spiking patterns produced when an influential current,  $I_h$ , is varied.

Les neurones du cortex entorhinal médian présentent une configuration spatiale en grille hexagonale de taux d'impulsions qui pourrait constituer un code neuronal pour la capacité de navigation des rats et autres mammifères. Une telle configuration spatiale peut être simulée à partir d'un réseau de neurones avec des connexions excitatrices et inhibitrices. Nous utilisons un modèle stochastique simple d'une cellule stellaire, avec des données d'un champ d'impulsions spatiales simulé, pour explorer les configurations d'impulsions produites lorsqu'un courant influent,  $I_h$ , est modifié.

## 1E-15: Probability and Statistical Problems and Results Arising from Neural Models Probabilités et problèmes statistiques et résultats découlant de modèles neuronaux

---

**LAWRENCE WARD & PRISCILLA (CINDY) GREENWOOD** (University of British Columbia)

[Monday May 30 / lundi 30 mai, 16:30]

*Quasicycles and Quasipatterns in the Brain?*

*Quasicycles et quasimotifs dans le cerveau?*

Quasicycles are damped oscillations sustained by noise. Similarly, quasipatterns are damped spatial patterns sustained by noise. We will describe (1) one model that exhibits quasicycles, (2) an approximation that indicates that the model's behaviour comprises a stochastic rotation process and an Ornstein-Uhlenbeck amplitude process, (3) application of the approximation to gamma-band oscillations in living brains, and (4) production of quasipatterns in spatial arrays of quasicycle oscillators.

Les quasicycles sont des oscillations affaiblies mais soutenues par le bruit. Similairement, les quasimotifs sont des motifs spatiaux affaiblis également soutenus par le bruit. Nous discuterons (1) un modèle qui exhibe des quasicycles, (2) une approximation qui indique que le comportement du modèle comprend une rotation stochastique et un processus d'amplitude Ornstein-Uhlenbeck, (3) une application de l'approximation aux oscillations de bande-gamma dans les cerveaux vivants, et (4) la production de quasimotifs dans les rangs spatiaux des oscillateurs quasicycliques.

**1E-I6: Statistical Education Section Presidential Invited Address**  
**Allocution du l'invité du président du Groupe d'éducation en statistique**

---

Chairs/Présidentes: Peggy Ng (York University) and/et Bethany White (University of Western Ontario)

**Room/Salle: Thistle 325**

**Abstract/Résumé**

---

**DAVID SPIEGELHALTER** (Cambridge University)

[Monday May 30 / lundi 30 mai, 15:30]

*The Ups and (Many) Downs of Trying to Be a 'Public Statistician'*

*Les hauts et les (nombreux) bas de la vie d'un « statisticien public »*

Statisticians can have special insights into the numbers in the news, but don't tend to have much of a public role. I've got a unique job in trying to improve the way that stats and risk are discussed in society, although this is not an easy task. I will relate both positive and negative experiences from radio, TV, print and online, covering topics such as climate change, alcohol, polls, sky-diving, and sex. You can share the panic at being asked unanswerable questions live on radio, and the joys of doing well in Winter Wipeout due to careful study of the statistics.

Les statisticiens peuvent avoir des idées très utiles sur les chiffres qui font les nouvelles, mais on ne leur demande que rarement de jouer un rôle public. J'occupe un poste assez unique où je tente d'améliorer la discussion publique des statistiques et des risques – ce qui n'est pas toujours facile. J'aborderai des expériences positives et négatives tirées de la radio, de la télévision, de la presse écrite et en ligne sur des sujets aussi variés que le changement climatique, l'alcool, les sondages, le saut en parachute et la sexualité. Vous partagerez la panique que je ressens lorsqu'on me pose en direct à la radio des questions sans réponse possible et la joie que j'ai eue à réussir à Winter Wipeout après en avoir soigneusement étudié les statistiques.

## 1E-C1: Biostatistics: Causal Inference and Measurement Error Biostatistique : inférence causale et erreur de mesure

---

Chair/Président: Mireille Schnitzer (Université de Montréal)

Room/Salle: Thistle 255

### Abstracts/Résumés

---

**ASMA BAHAMYIROU, MIREILLE SCHNITZER & LUCIE BLAIS** (Université de Montréal)

[Monday May 30 / lundi 30 mai, 15:30]

*Comparison of Causal Inference Approaches for Estimating the Effect of Exposure to Inhaled Corticosteroids on Mean Birth Weight and Gestational Age*

*Comparaison des approches d'inférence causale pour l'estimation de l'effet de l'exposition aux corticostéroïdes inhalés sur le poids moyen à la naissance et sur l'âge gestationnel*

Regression and propensity score methods to deal with confounder adjustment for the estimation of a treatment effect can produce biased estimates. Semiparametric methods such as Targeted Minimum Loss-Based Estimation (TMLE) can be used to reduce dependence on parametric model specification. In a large covariate space, Collaborative TMLE (C-TMLE) can further improve upon mean squared error. This work aims to apply adaptive methods to estimate the causal effect of exposure to inhaled corticosteroids on mean birth weight and gestational age in pregnant women with mild asthma. We compare Inverse Probability of Treatment Weighting, TMLE and C-TMLE, all implemented with Super Learner

Les méthodes de score de propension et de régression pour traiter l'ajustement sur des facteurs de confusion pour l'estimation de l'effet d'un traitement peuvent produire des estimations biaisées. Des méthodes semiparamétriques telles que l'estimation ciblée de la perte minimale (TMLE) peuvent être utilisées pour réduire la dépendance à la spécification du modèle paramétrique. Dans un grand espace de covariables, la TMLE collaborative (C-TMLE) peut améliorer encore davantage l'erreur carrée moyenne. Cette étude a pour but l'application de méthodes adaptatives pour estimer l'effet causal de l'exposition aux corticostéroïdes inhalés sur le poids moyen à la naissance et sur l'âge gestationnel chez les femmes enceintes souffrant d'asthme léger. Nous comparons la pondération de la probabilité inverse du traitement, la TMLE et C-TMLE, toutes mises en oeuvre avec Super Learner.

---

**LEILA GOLPARVAR, DAVID A. STEPHENS & ROBERT W. PLATT** (McGill University)

[Monday May 30 / lundi 30 mai, 15:45]

*Causal Structure Learning and Propensity Score Adjustment*

*Apprentissage de structure causale et ajustement par score de propension*

Mathematical representation of causal dependencies among a set of variables via graphs has been in the statistical literature for more than a century. Predicting the effect of manipulations from non-experimental data involves two steps: first, discovery of causal structures, second identification and estimation of causal parameters. We adopted a frequentist constraint-based approach to discover the causal graph, and use the PC algorithm for causal discovery. To explore the use of the PC algorithm

La représentation mathématique des dépendances de causalité entre un ensemble de variables via des graphiques apparaît dans la littérature statistique depuis plus d'un siècle. Prédire l'effet des manipulations à partir de données non-expérimentales comporte deux étapes : premièrement, la découverte des structures causales, deuxièmement, l'identification et l'estimation des paramètres de causalité. Nous avons adopté une approche basée sur la contrainte fréquentiste pour découvrir le graphique de causalité et utilisons l'algorithme PC pour la découverte causale. Pour ex-

## 1E-C1: Biostatistics: Causal Inference and Measurement Error Biostatistique : inférence causale et erreur de mesure

---

in selection of confounders, we conduct a Monte Carlo simulation study. Results show that PC algorithm works very well when the sample size is moderate to large.

plorer l'utilisation de l'algorithme PC dans le choix des facteurs confondants, nous menons une étude de simulation de Monte-Carlo. Les résultats montrent que l'algorithme PC fonctionne très bien lorsque la taille de l'échantillon est modérée à grande.

---

**KAMAL RAI, MARTIN LYSY & JOEL A. DUBIN** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 16:00]

*Bayesian Inference for Stochastic PK/PD Models*

*Inférence bayésienne pour les modèles stochastiques pharmacocinétiques et pharmacodynamiques*

Differential equations (DEs) occupy a central role in the modeling of pharmacokinetic/pharmacodynamic (PK/PD) processes. To estimate the parameters of these equations, a statistical approach might augment the deterministic DE with a stochastic simulation model and attempt to solve the corresponding inverse problem. We present a Bayesian methodology and its software implementation for PK/PD parameter inference, accounting for three sources of stochastic variability: (1) measurement error, (2) within-subject process fluctuations, and (3) between-subject random effects. Through numerical experiments with a small-sample PK study, we explore the effect of increased model complexity on the bias-variance trade-off.

Les équations différentielles occupent un rôle central dans la modélisation de processus pharmacocinétiques et pharmacodynamiques. Afin d'estimer les paramètres de ces équations, une approche statistique peut augmenter les équations différentielles déterministes au moyen d'un modèle de simulation stochastique, et peut tenter de résoudre le problème inverse correspondant. Nous présentons une méthodologie bayésienne et la mise en œuvre du logiciel connexe pour l'inférence de paramètres pharmacocinétiques et pharmacodynamiques, en tenant compte de trois sources de variabilité stochastique : 1) l'erreur de mesure, 2) les fluctuations du processus intra-sujet et 3) les effets aléatoires intersujets. Nous explorons l'effet de la complexité accrue du modèle sur le compromis entre le biais et la variance au moyen d'expériences numériques avec une étude pharmacocinétique sur un petit échantillon.

---

**DI SHU & GRACE Y. YI** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 16:15]

*Estimation of Causal Parameters in the Presence of Measurement Error*

*Estimation de paramètres causaux en présence d'erreur de mesure*

Odds ratio, risk ratio and risk difference are widely used measures for comparing the performance of two treatments. In observational studies where confounders are inevitable, obtaining these measures with causal interpretation is of increasing interest. To reduce the confounding bias, many methods have been proposed, assuming that there is no measurement error. However, measurement error unavoidably occurs for various reasons. It is well known that ignoring measurement error effects can result in severely biased parameter estimates. In this talk, I'll discuss estimating causal parameters with measurement error. Numerical studies will be presented to assess the performance of pro-

Les rapports de cotes, les rapports de risques et les différences de risques sont des mesures largement utilisées pour comparer l'efficacité de deux traitements. Dans le cadre d'études d'observation où les facteurs de confusion sont inévitables, obtenir ces mesures avec une interprétation causale suscite un intérêt croissant. Afin de réduire le biais des facteurs de confusion, de nombreuses méthodes ont été proposées en supposant qu'il n'y ait pas d'erreur de mesure. Cependant, les erreurs de mesure se produisent inévitablement, pour diverses raisons. Il est bien connu que le fait de ne pas tenir compte des effets d'erreurs de mesure peut biaiser sérieusement les estimations des paramètres. Dans cet exposé, nous discuterons des paramètres causaux ainsi que de l'erreur de mesure. Nous présenterons des études numé-



## 1E-C1: Biostatistics: Causal Inference and Measurement Error Biostatistique : inférence causale et erreur de mesure

---

posed methods.

riques afin d'évaluer l'efficacité des méthodes proposées.

---

**DENIS TALBOT** (Université Laval), **AMANDA M. ROSSI** (McGill University), **SIMON L. BACON** (Concordia University), **JULI ATHERTON & GENEVIÈVE LEFEBVRE** (Université du Québec à Montréal)

[Monday May 30 / lundi 30 mai, 16:30]

*A Graphical Perspective of Marginal Structural Models when Estimating the Causal Relationships Between Physical Activity, Blood Pressure, and Mortality*

*Une perspective graphique aux modèles structuraux marginaux pour l'estimation des relations causales entre l'activité physique, la pression artérielle et la mortalité*

Estimating causal effects requires important prior subject-matter knowledge and, sometimes, sophisticated statistical tools. Marginal structural models (MSMs) are a relatively new class of causal models that effectively deals with the estimation of the effects of time-varying exposures. MSMs have traditionally been embedded in the counterfactual framework to causal inference. In this talk, we use the causal graph framework to enhance the implementation of MSMs. We illustrate our approach using data from a prospective cohort study, the Honolulu Heart Program.

L'estimation d'effets causaux requiert une importante connaissance *a priori* du domaine d'application et, parfois, l'utilisation d'outils statistiques sophistiqués. Les modèles structuraux marginaux (MSMs) constituent une classe de modèles causaux relativement nouvelle qui permet l'estimation des effets d'une exposition variant dans le temps. Ces modèles sont habituellement imbriqués dans le paradigme contrefactuel pour l'inférence causale. Afin de faciliter l'implantation des MSMs, nous proposons de considérer le paradigme graphique à l'inférence causale. Nous illustrons notre approche à l'aide des données d'une étude de cohorte prospective, le Honolulu Heart Program.

---

**YEYING ZHU** (University of Waterloo), **JENNIFER SAVAGE** (Pennsylvania State University), **DEBASHIS GHOSH** (Colorado School of Public Health)

[Monday May 30 / lundi 30 mai, 16:45]

*A Kernel-Based Approach to Covariate Adjustment for Causal Inference*

*Méthode par noyaux pour l'ajustement de covariables à des fins d'inférence causale*

In matching, an important goal is to achieve balance in the covariates among the treatment groups. In this article, we introduce the concept of distributionally balance preserving which requires the distribution of the covariates to be the same in different treatment groups. Meanwhile, we propose a new balance measure called kernel distance, which is the empirical estimate of the probability metric defined in the reproducing kernel Hilbert spaces. Compared to the traditional balance metrics, the kernel distance measures the difference in the two multivariate distributions instead of the difference in the finite moments of the distributions.

En appariement, équilibrer les covariables parmi les groupes de traitement est un objectif important. Cet article présente le concept de préservation distributionnelle de l'équilibre, ce qui exige une distribution identique des covariables dans différents groupes de traitement. Entre-temps, nous proposons une nouvelle mesure de l'équilibre appelée distance par noyaux, qui se veut une estimation empirique de la métrique de la probabilité définie dans les espaces de Hilbert à noyau reproduisant. Comparativement aux mesures traditionnelles de l'équilibre, la distance par noyaux mesure la différence dans les deux distributions multivariées plutôt que celle dans les moments finis des distributions.

## 1E-C2: Biostatistics: Methodological Innovation 2 Biostatistique : innovation méthodologique 2

---

Chair/Président: Andrea Benedetti (McGill University)

Room/Salle: Thistle 256

### Abstracts/Résumés

---

**ASHLEY BONNER & JOSEPH BEYENE** (McMaster University)

[Monday May 30 / lundi 30 mai, 15:30]

*Inference for Model Parameters in Sparse Canonical Correlation using Resampling Techniques*

*Inférence des paramètres des modèles pour la corrélation canonique rare à l'aide de techniques de rééchantillonnage*

Sparse canonical correlation analysis (sCCA) is a multivariate method that uses latent variables to explore sparse and complex associations between high-dimensional data types. Currently, there are limited inferential strategies to accompany the estimated loading values that researchers depend on to determine whether, and to what extent, each variable contributes to the latent relationship. Through simulation, we evaluate the performance of re-sampling techniques, including the non-parametric bootstrap, to explore distributional properties and create confidence intervals for sCCA loading values. Under certain data conditions, re-sampling may offer a means to infer some level of confidence in the complex relationships returned from sCCA.

L'analyse de corrélation canonique rare (sCCA) est une méthode multivariée qui utilise des variables latentes pour explorer les associations rares et complexes entre les types de données de grande dimension. Actuellement, il existe des limites dans les stratégies inférentielles pour accompagner les valeurs de chargement estimées dont dépendent les chercheurs pour déterminer si, et dans quelle mesure, chaque variable contribue à la relation latente. À l'aide de simulations, nous évaluons la performance des techniques de rééchantillonnage, y compris le bootstrap non-paramétrique, pour explorer les propriétés de distribution et bâtir des intervalles de confiance pour les valeurs de chargement sCCA. Sous certaines conditions des données, le rééchantillonnage peut offrir un moyen de déduire un certain niveau de confiance dans les relations complexes renvoyées par la sCCA.

---

**OSVALDO ESPIN-GARCIA** (University of Toronto and Lunenfeld-Tanenbaum Research Institute), **RADU V. CRAIU** (University of Toronto), **SHELLEY B. BULL** (University of Toronto and Lunenfeld-Tanenbaum Research Institute)

[Monday May 30 / lundi 30 mai, 15:45]

*Two-Phase Designs for Joint Quantitative-Trait-Dependent and GWAS-SNP-Dependent Sampling in Post-GWAS Regional Sequencing*

*Plans à deux phases pour trait quantitatif dépendant et échantillonnage GWAS-SNP conjoints dans le séquençage régional post-GWAS*

We evaluate two-phase designs to follow up findings from genomewide association study (GWAS) when regional sequencing in the entire cohort is too expensive. We develop a novel EM-estimation under a semiparametric maximum likelihood formulation tailored for post-GWAS inference. A GWAS-SNP serves as an auxiliary covariate in inferring association between a sequence variant and a non-

Nous évaluons les plans à deux phases pour assurer le suivi des résultats de l'étude d'association pangénomique (GWAS) lorsque le séquençage régional de l'ensemble de la cohorte est trop cher. Nous développons une nouvelle estimation EM sous une formulation du maximum de vraisemblance semiparamétrique sur mesure pour l'inférence post-GWAS. Un GWAS-SNP sert de covariable auxiliaire dans la déduction d'une association entre une variante de sé-

## 1E-C2: Biostatistics: Methodological Innovation 2 Biostatistique : innovation méthodologique 2

---

mally distributed quantitative trait (QT). We perform simulations to quantify efficiency and power of QT-SNP joint sampling under alternative sample allocations. A joint allocation balanced on GWAS-SNP genotype and extreme-QT strata yields power improvements compared to marginal QT or SNP-sampling counterparts.

quence et un trait quantitatif (CQ) normalement distribué. Nous effectuons des simulations pour quantifier l'efficacité et la puissance de l'échantillonnage conjoint CQ-SNP dans le cadre de répartitions d'échantillons alternatifs. Une répartition conjointe équilibrée sur le génotype GWAS-SNP et les strates extrêmes-QT donne des améliorations de puissance par rapport à CQ marginal ou à ses homologues SNP-échantillonnage.

---

**JINGXIONG XU** (University of Toronto), **LAURENT BRIOLLAIS** (Lunenfeld-Tanenbaum Research Institute; University of Toronto), **WEI XU** (Princess Margaret Cancer Centre: University of Toronto)

[Monday May 30 / lundi 30 mai, 16:00]

*Bayesian Statistical Approaches for Genetic Association with Next Generation Sequencing (NGS) Data*

*Approches statistiques bayésiennes pour l'association génétique avec des données de séquençage de prochaine génération (SPG)*

The discovery of rare variants is becoming a major challenge in genetic association studies and could help elucidating the genetic basis of common diseases. Because rare variants occur too infrequently in the general population, single-variant association tests lack power in NGS analyses. We developed several joint test statistics using a Bayes Factor approach to assess the evidence of association between a set of rare variants located on same chromosomal region and a cancer outcome. Our simulation studies show the advantages of the Bayesian approach compared to two popular approaches: the burden test and the sequence kernel association test (SKAT).

La découverte de variantes rares prend de l'importance pour les études d'associations génétiques et pourrait aider à élucider le fondement génétique de maladies communes. Comme les variantes rares sont peu fréquentes dans la population en général, les tests d'association à variante unique ont peu de puissance dans les analyses SPG. Nous avons élaboré plusieurs statistiques de tests conjointes à l'aide d'une approche fondée sur le facteur de Bayes pour évaluer les indices d'une association entre un ensemble de variantes rares situé dans une même zone chromosomique et l'apparition d'un cancer. Nos études de simulation montrent les avantages de l'approche bayésienne comparativement à deux approches populaires : le test de charge et le test d'association à noyau de séquences (SKAT).

---

**QING YU, GUOHUA YAN & RENJUN MA** (University of New Brunswick)

[Monday May 30 / lundi 30 mai, 16:15]

*A Bayesian Poisson Mixed Modelling Approach to Zero-Inflated Cox Proportional Hazard Models*

*Approche bayésienne Poisson de modélisation mixte pour les modèles à risques proportionnels de Cox à sur-représentation de zéros*

In Cox proportional hazard models, frailty can be introduced to incorporate individual susceptibility to a disease or an event of interest; however, some individuals may be immune or insusceptible to certain diseases. We use compound Poisson distributed frailties, which has a spike at zero, to characterize the heterogeneity in susceptibility among individuals. This approach allows us to accommodate both zero and positive susceptibilities in an integral way. We extend the Poisson modelling approach

Dans les modèles à risques proportionnels de Cox, un facteur de fragilité peut être introduit afin d'incorporer la susceptibilité individuelle à une maladie ou à un événement d'intérêt; toutefois, certaines personnes peuvent être immunisées contre une maladie ou y être insusceptibles. Nous utilisons des fragilités distribuées selon une Poisson composée, qui a un pic à zéro, pour caractériser l'hétérogénéité de la susceptibilité chez les individus. Cette approche nous permet de prendre en compte les susceptibilités nulles et positives. Nous étendons l'approche de modélisation de type Poisson

## 1E-C2: Biostatistics: Methodological Innovation 2

### Biostatistique : innovation méthodologique 2

---

proposed by Ma et al. (Biometrika, 2003) to estimation of our model. Our proposed approach is illustrated with a simulated dataset and a real life example.

proposée par Ma et coll. (Biometrika, 2003) à l'estimation de notre modèle. L'approche que nous proposons est illustrée par un ensemble de données simulées et un exemple réel.

---

**TIAN FENG, JOSEPH BEYENE & JEMILA S. HAMID** (McMaster University)

[Monday May 30 / lundi 30 mai, 16:30]

*A Bayesian Method for Handling Missing Data in the Growth Curve Model*

*Une méthode bayésienne pour traiter les données manquantes dans le modèle de courbe de croissance*

The growth curve model (GCM) is a generalized multivariate analysis of variance (GMANOVA) model. The standard GCM has several desirable properties when data is complete, but is problematic when there are missing observations. We propose a Bayesian method for handling missing data in GCM using the Gibbs sampling algorithm. An extensive simulation study is carried out to compare the proposed method with the expectation-maximization (EM) algorithm. The Bayesian method is found to be generally useful, and gives promising results compared with the EM algorithm. Classical datasets are used to illustrate the performance of the proposed method.

Le modèle de courbe de croissance (MCC) est un modèle d'analyse multivariée généralisée de la variance (GMANOVA). Le MCC standard possède plusieurs propriétés souhaitables lorsque des données sont complètes, mais est problématique en présence d'observations manquantes. Nous proposons une méthode bayésienne pour traiter les données manquantes dans le MCC en utilisant l'algorithme d'échantillonnage de Gibbs. Une vaste étude par simulation est réalisée pour comparer la méthode proposée à l'algorithme espérance-maximisation (EM). La méthode bayésienne se révèle être généralement utile et donne des résultats prometteurs par rapport à l'algorithme EM. Les jeux de données classiques sont utilisés pour illustrer la performance de la méthode proposée.

## 1E-C3: Biostatistics: Survival and Current Status Data Biostatistique : données de survie et d'état actuel

---

Chair/Président: Cindy Feng (University of Saskatchewan)

**Room/Salle: Thistle 259**

### Abstracts/Résumés

---

**LONGLONG HUANG & XUEWEN LU** (University of Calgary), **KAREN KOPCIUK** (Alberta Health Services, University of Calgary)

[Monday May 30 / lundi 30 mai, 15:30]

*A Group Bridge Approach for Component Selection in Nonparametric Accelerated Failure Time Additive Regression Model*

*Une approche de pont groupé pour la sélection des composantes dans un modèle de régression additif non paramétrique du temps de défaillance accéléré*

We propose a nonparametric accelerated failure time additive regression model whose covariates have nonparametric effects on the survival time. The proposed model can be fitted to high-dimensional censored data. B-splines are used to approximate the nonparametric components. A group bridge penalized variable selection approach based on the inverse probability-of-censoring weighted least-squares is developed to select nonparametric components. Simulation studies indicate that the proposed method is able to distinguish the nonzero components from the zero components and estimate the nonzero components simultaneously even with relatively high censoring rates. Real data analysis illustrates the application of the proposed method to survival data.

Nous proposons un modèle de régression additif non paramétrique du temps de défaillance accéléré dont les covariables ont des effets non paramétriques sur le temps de survie. Le modèle proposé peut être ajusté à des données censurées de grande dimension. Les B-splines sont utilisées pour approximer les composantes non paramétriques. Une approche de sélection des variables de pont groupé pénalisée basée sur l'inverse de la probabilité de censures des moindres carrés pondérés est développée pour sélectionner les composantes non paramétriques. Des études de simulation montrent que la méthode proposée peut distinguer les composantes non nulles à partir des composantes nulles et estimer les composantes non nulles simultanément même avec des taux de censure relativement élevés. L'analyse de données réelles illustre l'application de la méthode proposée aux données de survie.

---

**SAIMA KHOSA & SHAHEDUL KHAN** (University of Saskatchewan)

[Monday May 30 / lundi 30 mai, 15:45]

*A Generalization of Log-logistic Distribution with Application in Survival Analysis*

*Une généralisation de la distribution log-logistique avec application à l'analyse de survie*

The log-logistic distribution has wide applications in analyzing survival data. The model is closed under both multiplication of failure time and proportionality of odds. However, it is not a proportional hazard (PH) model. In survival analysis, PH models play a pivotal role in many applications. So in this study, we propose a generalization of the log-logistic distribution that falls into the PH family. A comparison between the generalized log-logistic and the Cox

La distribution log-logistique a de larges applications dans l'analyse des données de survie. Le modèle est fermé à la fois sous la multiplication du temps de défaillance et la proportionnalité des cotes. Cependant, ce n'est pas un modèle de risque proportionnel (RP). Dans l'analyse de survie, les modèles de RP jouent un rôle central dans de nombreuses applications. Ainsi dans cette étude, nous proposons une généralisation de la distribution log-logistique qui tombe dans la famille de RP. Une comparaison entre le log-logistique géné-

## 1E-C3: Biostatistics: Survival and Current Status Data Biostatistique : données de survie et d'état actuel

---

PH model reveals that the generalized log-logistic PH model performs reasonably well in analyzing different types of time-to-event data.

ralisé et le modèle de Cox RP révèle que le modèle généralisé log-logistique de RP performe assez bien dans l'analyse des différents types de données de temps d'événement.

---

**SHANSHAN LU, XUEWEN LU & JINGJING WU** (University of Calgary)

[Monday May 30 / lundi 30 mai, 16:00]

*Exploring the Varying Covariate Effects in Partially Linear Proportional Odds Models with Current Status Data*  
*Exploration des effets variables de covariables dans des modèles à cotes proportionnelles partiellement linéaires avec des données sur le statut actuel*

We consider a partially linear proportional odds model with current status data. This model enables one to examine the extent to which some covariates interact nonlinearly with an exposure variable, while other covariates present linear effects. B-spline approach and sieve maximum likelihood estimation method are used to get an integrated estimate for the linear coefficients, the varying-coefficient functions and the baseline function. The proposed parameter estimators are proved to be consistent and asymptotically normal, and the estimators for the nonparametric functions achieve the optimal rate of convergence. Simulation studies and a real data analysis are used for assessment and illustration.

Nous nous intéressons à un modèle à cotes proportionnelles partiellement linéaire avec des données sur le statut actuel. Ce modèle permet d'examiner dans quelle mesure certaines covariables interagissent de façon non linéaire avec une variable d'exposition alors que d'autres covariables présentent des effets linéaires. Nous utilisons l'approche de la courbe B-Spline et la méthode par tamisage du maximum de vraisemblance afin d'obtenir une estimation intégrée pour les coefficients linéaires, les fonctions de coefficient variable et la fonction de base. Nous démontrons que les estimateurs des paramètres sont convergents et asymptotiquement normaux et que les estimateurs des fonctions non-paramétriques atteignent le taux optimal de convergence. Nous utilisons des études de simulation et une analyse de données réelles pour l'évaluation et l'illustration.

---

**YUAN DONG & XUEWEN LU** (University of Calgary), **RADHEY SINGH** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 16:15]

*Efficient Estimation of Varying-Coefficient Partially Linear Proportional Hazards Models with Current Status Data*  
*Estimation efficace des modèles de risques proportionnels à coefficients linéaires partiellement variables à l'aide de données sur l'état actuel*

We consider a semiparametric varying-coefficient proportional hazards model with current status data. This mode enables one to assess possibly nonlinear effect of a certain covariate on the hazard rate. We apply B-splines to approximate both the unknown baseline hazard function and the varying-coefficient function. Sieve maximum likelihood estimation method is used for estimation. The rate of convergence of the estimators for the two unknown smooth functions is obtained and the estimator for the unknown parameter is shown to be asymptotically efficient and normally distributed. Simulation studies and data analysis are conducted to examine the finite-sample properties of the proposed estimators.

Nous considérons un modèle semi-paramétrique de risques proportionnels à coefficients variables avec des données sur l'état actuel. Ce mode permet d'évaluer l'effet éventuellement non-linéaire d'une certaine covariable sur le taux de risque. Nous appliquons B-splines pour approximer à la fois la fonction inconnue de risque de base et la fonction à coefficients variables. La méthode d'estimation du maximum de vraisemblance de Sieve est utilisée pour l'estimation. Le taux de convergence des estimateurs pour les deux fonctions lisses inconnues est obtenu et il est montré que l'estimateur pour le paramètre inconnu est asymptotiquement efficace et distribué normalement. Des études de simulation et une analyse des données sont menées pour étudier les propriétés des échantillons de taille finie des estimateurs proposés.

## 1E-C3: Biostatistics: Survival and Current Status Data Biostatistique : données de survie et d'état actuel

---

**OMIDALI AGHABABAEI JAZI, MASOUD ASGHARIAN & ABBAS KHALILI** (McGill University)

[Monday May 30 / lundi 30 mai, 16:30]

*A Comparison of the Penalized Estimation Methods in the Cox Model for Right-censored Length-biased Data*  
*Une comparaison des méthodes d'estimation pénalisées dans le modèle de Cox pour les données de longueur biaisée censurées à droite*

Length-biased data commonly arises in epidemiological studies when prevalent sampling is used for recruiting cohort subjects. Several estimation methods have been proposed for parameters in different survival models in the presence of length-bias in the last decade. In this talk, we will focus on variable selection and discuss how to penalize those methods with different penalty functions. We will compare the resulting estimators for the parameters in the Cox model in terms of their efficiency and employ the selected method to analyze a real right-censored length-biased data.

Les données de longueur biaisée surviennent généralement dans les études épidémiologiques lorsque l'échantillonnage prévalent est utilisé pour le recrutement des sujets de la cohorte. Plusieurs méthodes d'estimation ont été proposées pour les paramètres des différents modèles de survie en présence de longueur biaisée au cours de la dernière décennie. Dans cet exposé, nous allons nous concentrer sur la sélection des variables et discuter de la façon de sanctionner ces méthodes à l'aide de différentes fonctions de pénalité. Nous allons comparer les estimateurs résultants pour les paramètres du modèle de Cox selon leur efficacité et employer la méthode choisie pour analyser de vraies données de longueur biaisée censurées à droite.

**DONGDONG LI & JOAN HU** (Simon Fraser University)

[Monday May 30 / lundi 30 mai, 16:45]

*Modelling and Analysis of Bivariate Event-Times with Informative Censoring*  
*Modélisation et analyse des temps d'événements bivariés avec censure de l'information*

In attempt to evaluate risk of cancer patient to cardiovascular disease, we consider analysis of bivariate event-times with informative censoring. Our particular attention is to the inherent challenges to modelling association between the two event-times and their dependence with the censoring time. We explore various models and develop inference procedures under proposed models. The research is motivated and will be illustrated using records of cardiovascular disease related hospitalization from a cohort of breast cancer patients together with their cancer registry information. This is joint work with Professor Joan Hu.

Pour tenter d'évaluer le risque de cancer chez des patients atteints de maladie cardiovasculaire, nous considérons l'analyse des temps d'événements bivariés avec censure d'information. Nous portons une attention particulière aux défis liés à la modélisation de l'association entre deux temps d'événements et leur dépendance avec la censure de temps. Nous explorons divers modèles et élaborons des procédures d'inférence dans le cadre des modèles proposés. Cette recherche est motivée et illustrée par des données de dossiers d'hospitalisation sur les maladies cardiovasculaires provenant d'une cohorte de patients atteints de cancer du sein, ainsi que par les données contenues dans leur dossier d'information sur leur cancer. Ces travaux ont été réalisés conjointement avec la Professeure Joan Hu.

## 1E-C4: Statistical Methods and Applications 3 Méthodes statistiques et applications 3

---

Chair/Président: Anand N. Vidyashankar (George Mason University)

Room/Salle: Welch 8G

### Abstracts/Résumés

---

**ROJIAR HADDADIAN, LIQUN WANG & YULIYA MARTSYNYUK** (University of Manitoba)

[Monday May 30 / lundi 30 mai, 15:30]

*Simulation-based Estimation in Regression Models with Ordinal Response Variable and Mismeasured Covariates*  
*Estimation fondée sur la simulation dans les modèles de régression avec variable de réponse ordinale et covariables mesurées incorrectement*

A common problem in regression analysis is that some covariates are measured with errors. We present a simulation-based approach for the inference in the cumulative logit model for ordinal response variable and mismeasured covariates. We propose a simulation-based approach to overcome the computational difficulty of minimizing an objective function which involves multiple integrals and thus obtaining estimators for unknown parameters. This method does not require parametric assumptions for the distributions of the unobserved covariates and error components. Consistency and asymptotically normality for the proposed estimator are derived under some regularity conditions. Finally, the methodology is illustrated through simulation studies.

Un problème commun dans l'analyse de régression est l'erreur de mesure de certaines covariables. Nous présentons une approche fondée sur la simulation pour l'inférence dans le modèle logit cumulatif pour des variables explicatives ordinales et des covariables mesurées incorrectement. Nous proposons une approche fondée sur la simulation pour surmonter la difficulté de calcul de la minimisation d'une fonction objective qui inclut des intégrales multiples et ainsi obtenir des estimateurs pour les paramètres inconnus. Cette méthode ne nécessite pas d'hypothèses paramétriques sur les distributions des covariables non observées et sur les composantes d'erreur. La cohérence et la normalité asymptotique de l'estimateur proposé sont dérivées sous certaines conditions de régularité. Enfin, la méthodologie est illustrée à l'aide d'études de simulation.

---

**REN MINGCHEN, ALEX DE LEON & YAN YING** (University of Calgary)

[Monday May 30 / lundi 30 mai, 15:45]

*Likelihood Estimation of the Gaussian Copula Distribution for Mixed Data via PX-MCEM*  
*Estimation de vraisemblance de la distribution de copules gaussiennes pour des données mixtes au moyen de l'algorithme EM de Monte-Carlo par expansion de paramètres*

I employ the Gaussian copula distribution (GCD) to construct a joint distribution for mixed variables, where I adopt latent variable description of the binary variables. The model generalizes the grouped continuous model (GCM) or the multivariate probit model for binary data and its extension, the conditional GCM (CGCM), to mixed data. I used the parameter-expanded EM algorithm (PX-MCEM) to estimate the parameters. I also carried out a simulation study to investigate the finite-sample proper-

J'utilise la distribution des copules gaussiennes pour construire une distribution conjointe pour des variables mixtes, où j'adopte une description des variables binaires par des variables latentes. Le modèle généralise le modèle continu regroupé, le modèle probit multivarié pour les données binaires ainsi que son extension, le modèle continu regroupé conditionnel, au cas de données mixtes. J'ai utilisé l'algorithme EM avec expansion de paramètres pour estimer les paramètres et effectué une étude de simulation pour étudier les propriétés de l'échantillon de taille finie des estimations.



## 1E-C4: Statistical Methods and Applications 3

### Méthodes statistiques et applications 3

---

ties of the estimates. This is a joint work with A. R. de Leon and Y. Yan.

Ceci est un travail conjoint avec A. R. de Leon et Y. Yan.

---

**BOUCHRA NASRI** (INRS-ETE), **TAOUFIK BOUEZMARNI** (Université de Sherbrooke (UdeS)), **ANDRÉ ST-HILAIRE** (Institut national de recherche scientifique (INRS-ETE))

[Monday May 30 / lundi 30 mai, 16:00]

*Copula-Based Quantile Regression and Inference*

*Inférence pour un modèle de régression sur les quantiles basé sur des copules*

In this work, we introduce a new approach for non-linear quantile regression modelling based on the copula function. The main idea of this approach is to describe the conditional quantile by using the link between the conditional distribution of the variable of interest given the covariate with the copulas function and the marginal distribution. The quantile estimation will be done by estimating the parameters of the copula function and marginal distributions. Parametric and a semi-parametric estimators for the quantile functions are proposed and their asymptotic normality are established. Simulation results show the performance of these proposed estimators.

Nous présentons une nouvelle approche pour la modélisation par régression non-linéaire des quantiles, basée sur la fonction de copules. L'idée principale derrière cette approche est de décrire le quantile conditionnel grâce au lien existant entre la loi de la variable d'intérêt conditionnelle à la covariable, la fonction de copules et la distribution marginale. L'estimation du quantile utilise l'estimation des paramètres de la fonction de copules et des distributions marginales. Nous proposons des estimateurs paramétriques et semi-paramétriques des fonctions de quantiles et établissons leur normalité asymptotique. Les résultats de simulations montrent l'efficacité des estimateurs proposés.

---

**MONSUR CHOWDHURY & SAUMEN MANDAL** (University of Manitoba)

[Monday May 30 / lundi 30 mai, 16:15]

*Optimal Designs for Maximum Likelihood Estimation*

*Conceptions optimales pour l'estimation du maximum de vraisemblance*

There are many problems in statistics, which demand the calculation of one or more optimizing probability distributions. We consider one such problem in which we determine maximum likelihood estimates of cell probabilities under the hypothesis of marginal homogeneity in square contingency tables. This is a maximization problem subject to satisfying several constraints. We first formulate the Lagrangian function and then transform the problem to that of maximizing some functions of the cell probabilities simultaneously. Finally, we apply the methodologies in some real data sets. The methodologies are quite flexible and could be applied to a wide class of optimization problems.

Plusieurs problèmes en statistique exigent le calcul d'une ou plusieurs distributions de probabilité optimisante. Nous considérons un tel problème dans lequel nous déterminons les estimations du maximum de vraisemblance des probabilités de cellule sous l'hypothèse d'homogénéité marginale dans les tableaux de contingence carrés. Il s'agit d'un problème de maximisation sous réserve de satisfaire à plusieurs contraintes. Nous formulons d'abord la fonction lagrangienne puis transformons le problème en celui de la maximisation simultanée de certaines fonctions de probabilités de cellules. Enfin, nous appliquons les méthodologies à certains jeux de données réelles. Les méthodes sont assez flexibles et pourraient être appliquées à une grande classe de problèmes d'optimisation.

## 1E-C4: Statistical Methods and Applications 3

### Méthodes statistiques et applications 3

---

**MAYSUM PANJU** (University of Waterloo)

[Monday May 30 / lundi 30 mai, 16:30]

*Topic Modelling in Natural Language Processing*

*Modélisation de sujet dans le traitement du langage naturel*

Documents written in natural language typically are not just structured sequences of randomly selected words. The underlying distribution determining which words might be found in a given document typically depends on some latent factors describing the overall domain of the document, which we can refer to as "topics". Using statistical methods, these hidden topics can be discovered and identified based on the words that are observed across various documents. In this talk, we'll explore a high-level overview of some of these methods, which ultimately allow machines to infer the central topics and concepts that shape a body of text.

En général, les documents écrits en langage naturel ne sont pas seulement des séquences structurées de mots choisis au hasard. La distribution sous-jacente qui détermine quels mots peuvent être trouvés dans un document donné dépend de certains facteurs latents décrivant le domaine principal du document, que l'on peut appeler « sujets ». Grâce à des méthodes statistiques, ces sujets cachés peuvent être découverts et identifiés en fonction des mots qui sont observés dans divers documents. Dans cet exposé, nous explorerons une vue d'ensemble de haut niveau de certaines de ces méthodes, qui permettent à des machines d'inférer les sujets et concepts centraux formant le corps d'un texte.

**MAMADOU YAUCK & LOUIS-PAUL RIVEST** (Laval University)

[Monday May 30 / lundi 30 mai, 16:45]

*Prediction of Population Sizes in Open-Population Capture-Recapture Models*

*Prédiction des tailles de populations dans les modèles de capture-recapture pour population ouverte*

Capture-recapture models are used for the estimation of demographic parameters of a wildlife population. For the open-population experiment, we have models such as the Jolly-Seber approach (1963) and the robust-design of Kendall-Pollock-Brownie (1965). This presentation focuses on estimates for population sizes, seen as a prediction problem. We are mainly interested in the robust-design experiment and provide a better predictor than the Kendall-Pollock-Brownie and Rivest-Daigle (2004) estimators. We develop an empirical version of our new predictor, and provide an estimate for the prediction error.

Les modèles de capture-recapture sont utilisés pour estimer les paramètres démographiques de populations animales. Plusieurs approches sont disponibles pour les populations ouvertes, par exemple le modèle de Jolly et Seber (1963) et le design robuste de Kendall, Pollock et Brownie (1995). Cet exposé porte sur l'estimation des tailles de populations, considérée comme un problème de prédiction. Nous nous intéressons au cas du design robuste et proposons un prédicteur, meilleur que les estimateurs de Kendall et Pollock et de Rivest et Daigle (2004). Nous présentons ensuite une version empirique du nouveau prédicteur, et traitons enfin de l'estimation de l'erreur de prédiction.

## 1E-C5: Statistics in Action La statistique en action

---

Chair/Président: Yi Yang (McGill University)

Room/Salle: Thistle 257

### Abstracts/Résumés

---

**DEVAN BECKER** (University of Western Ontario), **W. JOHN BRAUN** (University of British Columbia)  
[Monday May 30 / lundi 30 mai, 15:30]

*Aerial Firefighting Fatigue Project*

*Projet sur la fatigue dans les services aériens de lutte contre les incendies*

In this ongoing project, we are attempting to find indicators of pilot fatigue in aerial firefighters using raw flight data (location, altitude, speed, yaw, pitch, roll, etc. recorded each second). We are applying Shewhart control charts to detect unusual flights based on summaries of the data. Each flight is different, but control charts are meant for homogeneous data. We must either find homogenous processes within the heterogeneous data or extend the methodology. This is still an open problem, but successful completion might impact fatigue regulations for non-typical flights and aid future applications of control charts with non-homogeneous data.

Dans ce projet en cours, nous tentons de trouver des indicateurs de la fatigue des pilotes chez les pompiers de l'air en utilisant des données de vol brutes (emplacement, altitude, vitesse, lacet, tangage, roulis, etc. enregistrés chaque seconde). Nous appliquons les graphiques de contrôle Shewhart pour détecter les vols inhabituels en se basant sur les résumés des données. Chaque vol est différent mais les graphiques de contrôles sont prévus pour des données homogènes. Nous devons donc soit trouver des processus homogènes à l'intérieur de ces données hétérogènes ou nous devons élargir la méthodologie. Ceci est toujours un problème ouvert, mais une conclusion réussie pourrait avoir une incidence sur les règlements portant sur la fatigue pour les vols non-typiques et pourrait aider l'application future des graphiques de contrôle à des données non-homogènes.

---

**PIERRE-JÉRÔME BERGERON** (PJB Consulting), **SAMUEL ARCHIBALD** (Université du Québec à Montréal)  
[Monday May 30 / lundi 30 mai, 15:45]

*Valar Morghulis: A Survival Analysis of 'A Song of Ice and Fire'*

*Valar morghulis: une analyse de survie des livres du Trône de fer*

George R. R. Martin's *A Song of Ice and Fire* is an ongoing series of fantasy novels with a large cast of characters, and numerous, sometimes shocking deaths. Readers do not know what to expect in terms of the eventual resolution of the story's many threads. We apply here the methods of survival analysis to see if we can discern patterns that could inform the literary analysis of the text. We discuss data selection, the choice of variables and units of measurements in defining event times in fiction. We then compare survival between different groups of characters.

La série des livres du *Trône de fer* de George R.R. Martin comporte un grand nombre de personnages qui souvent meurent, parfois de manière choquante. Les lecteurs ne savent pas à quoi s'attendre quant au dénouement des nombreuses intrigues de l'histoire. On applique ici des méthodes d'analyse de survie pour voir s'il existe des schémas permettant d'éclairer l'analyse littéraire du texte. On discute de la sélection des données, du choix des variables et des unités de mesure pour définir des temps d'événements dans la fiction. On compare ensuite la survie de différents groupes de personnages.

## 1E-C5: Statistics in Action La statistique en action

---

**DEBBIE DUPUIS** (HEC Montreal), **MARCO BEE** (University of Trento), **LUCA TRAPIN** (IMT Institute for Advanced Studies Lucca)

[Monday May 30 / lundi 30 mai, 16:00]

*Realized Peaks over Threshold*

«*Realized Peaks over Threshold*»

Recent contributions to the financial econometrics literature exploit high-frequency data to improve models for daily asset returns. In this talk, we present a new class of dynamic extreme value models that profit from high-frequency data when estimating the tails of daily asset returns. Our Realized Peaks-Over-Threshold approach provides estimates for the tails of the time-varying conditional return distribution, and remains applicable even when data are non-stationary. An in-sample fit to S&P 500 index returns shows that high-frequency data contain information on daily extreme returns beyond that included in low frequency data.

Les contributions récentes à la littérature de l'économétrie financière exploitent les données à haute fréquence afin d'améliorer les modèles pour les rendements quotidiens d'actifs. Dans cet exposé, nous présentons une nouvelle classe de modèles dynamiques pour valeurs extrêmes qui bénéficient des données à haute fréquence lors de l'estimation des queues des rendements quotidiens d'actifs. Notre approche «Realized Peaks-Over-Threshold» reste applicable lorsque les données sont non stationnaires. Un ajustement «in-sample» aux rendements de l'indice SP500 montre que les données à haute fréquence contiennent des informations sur les rendements quotidiens extrêmes au-delà de celles incluses dans les données à basse fréquence.

**NICHOLAS MITSAKAKIS** (University of Toronto), **KAREN BREMNER** (University Health Network), **MURRAY KRAHN** (THETA Collaborative)

[Monday May 30 / lundi 30 mai, 16:15]

*Investigating Structural Independence in Health Utility Instruments Using Graphical Models*

Étude de l'indépendance structurelle des instruments d'utilité de santé au moyen de modèles graphiques

In health economics cost effectiveness analyses rely on accurate estimation of health utility, a single global measure of health related quality of life (HRQoL). Health utilities are often measured with the use of questionnaire-type instruments, containing a number of items describing specific domains of HRQoL. The construction of these instruments relies on multi-attribute utility theory, assuming structural independence among the attributes. This property is rarely tested empirically. Here we applied discrete graphical models to multiple patient datasets to test the structural independence of a prostate cancer-specific instrument, Patient-Oriented Prostate Utility Scale (PORPUS-U). Results reveal undesired interdependence among items in PORPUS-U.

En économie de la santé, les analyses de rapports de coût-efficacité sont fondées sur une estimation précise de l'utilité de la santé, une mesure globale unique de la qualité de vie liée à la santé. Les utilités de santé sont souvent mesurées au moyen d'instruments de type questionnaires, contenant un certain nombre d'éléments décrivant des domaines précis de qualité de vie liés à la santé. La conception de ces instruments repose sur la théorie de l'utilité multi-attributs, qui suppose l'indépendance structurelle des attributs. Cette hypothèse est rarement testée manière empirique. Dans cette étude, nous avons appliqué des modèles graphiques discrets à plusieurs jeux de données de patients pour vérifier l'indépendance structurelle d'un instrument propre au cancer de la prostate, à savoir l'échelle d'utilité orientée sur les patients atteints du cancer de la prostate (PORPUS-U). Les résultats révèlent une interdépendance non souhaitée entre les éléments de l'échelle d'utilité orientée sur les patients atteints du cancer de la prostate.

## 1E-C5: Statistics in Action La statistique en action

---

**DAVID RIEGERT, DAVID THOMSON & AARON SPRINGFORD** (Queen's University)

[Monday May 30 / lundi 30 mai, 16:30]

*Statistical Estimation of Transfer Functions to Examine the Relationship Between Geomagnetism and Power Systems*

*Estimation statistique des fonctions de transfert pour examiner les relations entre le géomagnétisme et les réseaux électriques*

Geomagnetic disturbances disrupt ground-based systems in a profound manner; increasing the rate of corrosion in pipelines, damaging equipment in power transmission systems, and interfering with communications. We will focus specifically on the relationship between Earth's magnetic field and induced currents in power lines. In order to describe this relationship, transfer functions were estimated using current methods which are robust to outliers. We discuss the shortcomings of such current methods in the context of a Northwest US power system.

Les perturbations géomagnétiques interrompent considérablement les systèmes terrestres. Elles augmentent le taux de corrosion dans les pipelines, endommagent l'équipement dans les réseaux de transport d'énergie et interfèrent avec les réseaux de communication. Nous mettrons particulièrement l'accent sur la relation entre le champ magnétique de la terre et les courants induits dans les lignes électriques. Nous décrirons cette relation par l'estimation des fonctions de transfert au moyen des méthodes actuelles qui sont robustes aux valeurs aberrantes. Nous discuterons des faiblesses de ces méthodes actuelles dans le contexte d'un réseau électrique du nord-ouest des États-Unis.

**PAVEL SLAVCHEV & RUSSELL STEELE** (McGill University), **IAN SHRIER** (Lady Davis Institute, Jewish General Hospital and McGill University), **ASHLEY NAIMI** (University of Pittsburgh), **NATHALIE AUGER** (Institut National de la Santé Public du Québec and University of Montreal Research Centre)

[Monday May 30 / lundi 30 mai, 16:45]

*The Analysis of Live Births as a Competing Event to Stillbirths*

*L'analyse des naissances vivantes comme événement concurrent aux accouchements de mort-nés*

Understanding the causal risk factors for stillbirths should help develop more effective prevention programs. The Institut National de la Santé Public du Québec (INSPQ) birth registry data includes all births over the last 25 years in the province of Quebec. Classical time-to-event analyses (e.g. using Cox-proportional hazards models) for the INSPQ registry data on stillborn infants are complicated by two problems: clustering of births within mothers and live births that censor the time to stillbirth. In this presentation, we show how modelling live births as competing events to stillbirths can address both of these problems.

Comprendre les facteurs de risque causaux des accouchements d'un mort-né devrait contribuer à élaborer des programmes de prévention efficaces. Les données du registre des naissances de l'Institut National de la Santé Public du Québec (INSPQ) comprennent toutes les naissances au cours des 25 dernières années dans la province du Québec. Les analyses classiques de temps d'événement (p. ex., les modèles de régression des risques proportionnels de Cox) des données sur les accouchements d'un mort-né du registre de l'INSPQ sont compliquées par deux problèmes : le regroupement des naissances chez les mères et les naissances vivantes qui censurent le temps jusqu'à l'accouchement d'un mort-né. Dans cet exposé, nous démontrons comment la modélisation des naissances vivantes comme événement concurrent aux accouchements d'un mort-né peut régler ces deux problèmes.

**2A-A: Gold Medal Address**  
**Allocution du récipiendaire de la Médaille d'or**

---

**Room/Salle: Welch - David Howes**

**Abstract/Résumé**

---

**RICHARD LOCKHART** (Simon Fraser University)

[Tuesday May 31 / mardi 31 mai, 8:45]

*Big Data, High Dimensions, Goodness-of-fit: Something of This Nature*

*Données volumineuses, grandes dimensions et adéquation : quelque chose de cette nature*

Motivated by Zonker's Uncle Duke, I intend to try to trace a thread through big data, high dimensional inference, goodness-of-fit testing, transformations, invariance, contiguity and the role of theory. The plan is to give a high level overview of my attitude to the role of theory in high dimensional and big data problems by wandering through some of the areas of research in which I have been involved.

Motivé par l'oncle Duke de Zonker, je tenterai de tracer un fil conducteur qui relie les données volumineuses, l'inférence en grandes dimensions, les tests d'adéquation, les transformations, l'invariance, la contiguïté et le rôle de la théorie. L'objectif est de proposer un aperçu de mon attitude envers le rôle de la théorie dans les problèmes de grandes dimensions et de données volumineuses en parcourant certains domaines de recherche dans lesquels j'ai travaillé.

## 2B-I1: Advanced Statistical and Machine Learning Techniques for High Dimensional Data Analysis

### Techniques statistiques et apprentissage machine avancés pour l'analyse de données en haute dimension

---

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

Room/Salle: Welch - David Howes

#### Abstracts/Résumés

---

**RYAN TIBSHIRANI** (Carnegie Mellon University), **RICHARD LOCKHART** (Simon Fraser University), **JONATHAN TAYLOR & ROBERT TIBSHIRANI** (Stanford University), **WILL FITHIAN** (University of California, Berkeley), **JING LEI & LARRY WASSERMAN** (Carnegie Mellon University)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Recent Advances in Selective Inference*  
*Avancées récentes en inférence sélective*

I will talk about new sets of tools for inference after model selection. The highlight will be doing inference along steps of the forward stepwise, least angle regression, and lasso paths, though the framework for inference is applicable well beyond these cases, and other relevant problems will be briefly described as well. I will also discuss some asymptotic results about the robustness of the proposed tests to non-Gaussian noise distributions, and if time permits, a completely different framework based on conformal inference, that is essentially distribution-free. This all represents joint work with many different authors.

Je présenterai de nouveaux types d'outils pour l'inférence après sélection de modèle. Je traiterai notamment de l'inférence dans le cadre de la régression multiple ascendante, de la régression de least angle et LASSO, même si le cadre de l'inférence est applicable à bien d'autres cas, et je décrirai brièvement d'autres problèmes pertinents. Je discuterai également certains résultats asymptotiques relatifs à la robustesse des tests proposés pour les distributions de bruit non gaussien et, si le temps le permet, d'un cadre complètement différent basé sur l'inférence conforme, qui est essentiellement sans distribution. Il s'agit de travaux communs avec de nombreux auteurs différents.

---

**ANAND N. VIDYSHANKAR** (George Mason University)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Efficiency Considerations in Post Model Selection Inference*  
*Considérations d'efficacité dans l'inférence post-sélection de modèle*

It is common practice in high-dimensional data analysis that a model selection is first performed and then inference is carried out using the selected model presuming that the chosen model is the true model; that is, without accounting for model selection uncertainty. Recently, methods such as *clean and screen* are being used to account for model selection uncertainty. However, efficiency properties of the resulting statistical procedures are largely unknown. In this presentation, we provide a systematic account of efficiency properties of post-selection estimators. In the process we address some foundational questions concerning the role

Il est courant en analyse de données de grande dimension de commencer par la sélection de modèle puis de passer à l'inférence sur la base du modèle sélectionné en presumant que celui-ci est le vrai modèle, c'est-à-dire sans tenir compte de l'incertitude de sélection. Récemment, des méthodes dites de nettoyage et filtrage ont été développées pour tenir compte de cette incertitude. Cependant, l'efficacité des procédures statistiques résultantes demeure largement inconnue. Dans cette présentation, nous proposons une étude systématique des propriétés d'efficacité des estimateurs post-sélection. Nous répondrons notamment à des questions fondamentales concernant le rôle de la théorie de la déviation modérée dans l'étude de l'efficacité statistique.

## 2B-I1: Advanced Statistical and Machine Learning Techniques for High Dimensional Data Analysis

### Techniques statistiques et apprentissage machine avancés pour l'analyse de données en haute dimension

---

of moderate deviation theory in the study of statistical efficiency.

---

**YI YANG** (McGill University), **HUI ZOU** (University of Minnesota), **WEI QIAN** (Rochester Institute of Technology)  
[Tuesday May 31 / mardi 31 mai, 11:20]

*Variable Selection Deviation for High-dimensional Classification Methods*

*Déviation de la sélection des variables pour les méthodes de classification à grande dimension*

Despite wide applications of the regularization methods in many areas, sometimes we may question the reliability of the feature selection results, mainly for two reasons: (i) the outcome of selection depends heavily on the choice the tuning parameter, which could be very unstable if given small amount of data; (ii) although consistency in variable selection has been established for various regularization methods, it often depends on the assumption that the  $p$ -dimensional variable are sparse with many components being zero, which usually is unlike to hold for observational data. Thus we hope to develop the variable selection deviation (VSD) measures in the high-dimensional classification setting to evaluate the reliability of a set of selected variables.

Malgré les vastes applications des méthodes de régularisation dans de nombreux domaines, nous pouvons parfois remettre en question la fiabilité des résultats de la sélection caractéristique, principalement pour deux raisons : (i) le résultat de la sélection dépend fortement du choix du paramètre de réglage qui pourrait être très instable en présence d'une petite quantité de données ; (ii) bien que la convergence dans la sélection des variables ait été établie pour diverses méthodes de régularisation, elle dépend souvent de l'hypothèse que les variables à  $p$  dimensions sont rares avec de nombreuses composantes établies à zéro, ce qui ne tient généralement pas pour les données d'observation. Ainsi, nous espérons développer des mesures de la déviation de la sélection des variables (DSV) dans le contexte de la classification à grande dimension pour évaluer la fiabilité d'un ensemble choisi de variables.



## 2B-I2: Advancements to State-Space Models for Fisheries Science Progrès récents dans les modèles à espace d'états pour les sciences halieutiques

---

Chair/Président: Rick Routledge (Simon Fraser University)  
Organizer/Responsable: Joanna Mills-Flemming (Dalhousie University)

Room/Salle: Thistle 242

### Abstracts/Résumés

---

**WILLIAM AEBERHARD & JOANNA MILLS FLEMMING** (Dalhousie University), **EVA CANTONI** (University of Geneva), **CHRIS FIELD** (Dalhousie University), **XIMING XU** (Nankai University)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Robust and Consistent Estimation of Fixed Parameters in General State-Space Models*  
*Estimation robuste et convergente des paramètres fixes de modèles espace-états*

Fixed parameters in State-Space Models (SSMs) are traditionally estimated by maximum likelihood and typically include regression, auto-regression and scale parameters. The sensitivity of these estimates to deviations from the assumed model is problematic, all the more so since distributional assumptions about latent variables cannot be verified. Standard robust estimation techniques for generalized linear and time series models cannot be directly extended to SSMs, due mainly to the high-dimensional integrals that generally need to be approximated. We propose a robust and consistent estimating method by downweighting observations on the joint log-likelihood scale and by approximating the marginal log-likelihood by Laplace's method. Encouraging simulation results are presented with an application to a fish stock assessment.

Les paramètres fixes des modèles espace-états sont traditionnellement estimés par le maximum de vraisemblance et incluent typiquement des coefficients de régression, d'auto-régression et d'échelle. La sensibilité de ces estimations aux déviations par rapport au modèle postulé est problématique, d'autant plus que les postulats au niveau latent sont généralement invérifiables. Les estimateurs robustes pour modèles linéaires généralisés et séries temporelles ne peuvent être directement appliqués à cause d'intégrales de haute dimension à approximer. Nous proposons une méthode d'estimation robuste et convergente où des poids sont ajoutés à la vraisemblance jointe et la vraisemblance marginale est approximée par la méthode de Laplace. En sus de résultats de simulations encourageants, une évaluation de stock de poissons sera présentée.

---

**DAVID CAMPBELL** (Simon Fraser University)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Nested Laplace Approximation Diagnostics and Fixes*  
*Diagnostic et correction de l'approximation de Laplace emboîtée*

Laplace Approximations are used to marginalize out nuisance random effect variables from state space models. Under linear conditions with Gaussian perturbations this approximation is exact. In more complex models the approximation can be far from reasonable. Typically one must resort to importance sampling or MCMC as a marginalization tool to assess appropriateness of the approx-

Les approximations de Laplace servent à marginaliser les variables de bruit à effet aléatoire des modèles espace-état. Dans des conditions linéaires avec perturbations gaussiennes, cette approximation est exacte. Dans les modèles plus complexes, l'approximation peut au contraire être très imprécise. Il faut généralement recourir à l'échantillonnage par importance ou au MCMC comme outil de marginalisation pour évaluer la qualité de l'approximation. Nous présen-

## 2B-I2: Advancements to State-Space Models for Fisheries Science

### Progrès récents dans les modèles à espace d'états pour les sciences halieutiques

---

imation. In this work a quick and dirty estimate of the Kullback-Leibler (KL) divergence between the Laplace approximation and the target density to marginalize out parameters is presented. In the event of a large KL divergence we propose a probabilistic numerical integrator to marginalize over nuisance parameters. The proposed approach provides an intermediate step in the trade-off between computational complexity and level of approximation.

tons une estimation rapide de la divergence de Kullback-Leibler (KL) entre l'approximation de Laplace et la densité cible pour marginaliser les paramètres. Si la divergence de KL est élevée, nous proposons un intégrateur numérique probabiliste qui permet de marginaliser les paramètres de nuisance. Cette approche constitue une étape intermédiaire dans le compromis entre complexité computationnelle et niveau d'approximation.

---

**JOANNA MILLS FLEMMING** (Dalhousie University)

[Tuesday May 31 / mardi 31 mai, 11:20]

*Addressing Statistical Challenges in Fisheries Science*

*Défis statistiques en sciences halieutiques*

In a recent letter from Canada's Prime Minister to its Minister of Fisheries, top priorities included working to "protect the health of fish stocks, ..., use scientific evidence and the precautionary principle, and take into account climate change, when making decisions affecting fish stocks and ecosystem management". Fisheries science is faced with the challenge of providing management advice based on noisy survey and commercial fishery data that may only provide indirect information about fish stock size. Here I highlight our CANSII CRT's recent progress in building robust statistical methodologies for SSMs that are suitable for both fish stock assessments and for identifying drivers of animal movement in order to predict responses to climate change.

Une récente lettre du Premier ministre du Canada à son ministre des Pêches indique qu'une priorité absolue doit être accordée à « protéger la santé des stocks halieutiques, ..., s'appuyer sur les preuves scientifiques et le principe de la prudence et tenir compte des changements climatiques dans la prise de décisions ayant des répercussions sur les stocks halieutiques et la gestion des écosystèmes. » Les sciences halieutiques ont pour défi de donner des conseils en matière de gestion sur la base de données d'enquêtes bruitées, ou en provenance des pêches commerciales, qui n'offrent qu'indirectement une information sur la taille des stocks. Je discuterai des récents progrès de notre projet de recherche en collaboration de l'INCASS : nous avons développé des méthodologies statistiques rigoureuses pour les modèles espace-état (MEE) adaptées à la fois à l'évaluation des stocks halieutiques et à l'identification des moteurs de mouvements d'animaux, nous permettant ainsi de mieux prédire les réactions au changement climatique.

**2B-13: Contributions of Professor Ian B. MacNeill to Statistical Theory, Methodology,  
Teaching and Practice**  
**Contributions du professeur Ian B. MacNeill à la théorie statistique, Méthodologie,  
enseignement et la pratique**

---

Chair/Président: Gary Umphrey (University of Guelph)  
Organizer/Responsable: Krishna Jandhyala (Washington State University)

**Room/Salle: Academic South 217**

**Abstracts/Résumés**

---

**ABDEL EL-SHAARAWI** (The American University in Cairo), **SYLVIA R. ESTERBY** (University of British Columbia Okanagan)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Ian B. MacNeill and Environmetrics*

*Ian B. MacNeill et l'environnementrie*

Ian MacNeill made significant contributions to statistics for over 40 years at both the national and international levels. He made notable contributions to many areas of research, particularly to time series methods and change point analysis with applications to the environmental and health sciences. Ian also established the Department of Statistics at the University of Western Ontario on July 1, 1980. In this talk we will focus our discussion on his role in establishing The International Environmetrics Society (TIES) and its journal *Environmetrics*.

Ian MacNeill a fait d'importantes contributions aux statistiques pendant plus de 40 ans au niveau à la fois national et international. Il contribua de façon remarquable à plusieurs domaines de recherche, particulièrement les méthodes de séries chronologiques et l'analyse des points de changement avec applications aux sciences environnementales et de la santé. Ian a aussi créé le département de statistique à l'université Western Ontario le 1er juillet 1980. Dans cet exposé, je me concentrerai sur son rôle dans la création de l'International Environmetrics Society (TIES) et de son journal *Environmetrics*.

---

**ELENA NAUMOVA** (Tufts University)

[Tuesday May 31 / mardi 31 mai, 10:40]

*Seasonality Analysis in Public Health Applications*

*Analyse de saisonnalité dans la santé publique*

Several aspects of seasonality analysis are essential in public health applications. This includes an ability to formally compare timing, amplitude and duration of seasonal peaks across locations, populations, time, and causal agents. The contribution of Professor Ian B. MacNeill to the methodology development for seasonality assessment in health surveillance systems and specifically, the need for preservation of information relevant to health outcome of interest at the finest temporal resolution, the importance of considering social calendars in modeling seasonality, and the rationales why seasonal patterns have to be taken into account in designing cross-sectional and prospective epidemiological studies are discussed.

Plusieurs aspects de l'analyse de saisonnalité sont essentiels dans des applications du domaine de la santé publique. Cela inclut la capacité de comparer formellement le moment, l'amplitude et la durée des pics saisonniers à travers les emplacements, les populations et les agents causaux. La contribution du professeur Ian B. MacNeill au développement de la méthodologie pour l'évaluation de saisonnalité dans les systèmes de surveillance de la santé et particulièrement, le besoin de préserver l'information pertinente aux résultats en matière de santé avec la meilleure résolution temporelle possible, l'importance de considérer les calendriers sociaux lors de la modélisation de saisonnalité, et les raisons pour lesquelles les cycles saisonniers doivent être pris en considération lors de la conception d'études épidémiologiques prospectives et transversales sont abordés.

**2B-I3: Contributions of Professor Ian B. MacNeill to Statistical Theory, Methodology,  
Teaching and Practice**  
**Contributions du professeur Ian B. MacNeill à la théorie statistique, Méthodologie,  
enseignement et la pratique**

---

**SERGE PROVOST** (University of Western Ontario)

[Tuesday May 31 / mardi 31 mai, 11:02]

*On the q-Analogues of the Logistic and Generalized Extreme Value Distributions*

*Les analogues q des lois de valeurs extrêmes généralisées et logistiques*

The q-analogues of the logistic, generalized extreme value and related distributions are being introduced, the additional parameter allowing for greater flexibility for modeling purposes. It will be explained that the supports of these distributions can be infinite, bounded above or below, or even finite. Certain statistical functions such as their moments and quantile functions shall be derived. Potential fields of application include reliability theory, hydrology, biostatistics and survival analysis.

Les analogues q des lois de valeurs extrêmes généralisées et logistiques et autres lois connexes sont présentées, le paramètre additionnel permettant une plus grande flexibilité à des fins de modélisation. Nous expliquerons que les supports de ces lois peuvent être infinis, avoir une borne supérieure ou inférieure, ou même être finis. Certaines fonctions statistiques telles que leurs fonctions de quantiles ou de moments seront dérivées. Parmi les domaines d'application possibles, il y a la théorie de fiabilité, l'hydrologie, la biostatistique et l'analyse de survie.

**VENKATA KRISHNA JANDHYALA** (Washington State University), **SHELEMYAHU ZACKS** (State University of New York at Binghamton)

[Tuesday May 31 / mardi 31 mai, 11:24]

*Contributions of Dr. MacNeill to Change-Point Methodology and its Environmental Applications*

*Les contributions du Dr MacNeill à la méthodologie des points de rupture et ses applications environnementales*

Dr. MacNeill's contributions to change-point methodology and its applications are fundamental and are well known in the statistical literature. This talk reviews his main contributions to selected areas including: (i) sequences of independent random variables, (ii) linear regression models with possibly dependent errors, and (iii) spatial models. The review concludes with some examples from environmental time series.

Les contributions du Dr. MacNeill à la méthodologie des points de rupture et ses applications sont fondamentales et bien connues dans la littérature statistique. Cet exposé examine ses principales contributions à certains domaines, y compris : (i) séquences de variables aléatoires indépendantes, (ii) modèles de régression linéaire avec de possibles erreurs dépendantes, et (iii) modèles spatiaux. L'exposé se conclue avec quelques exemples de séries chronologiques environnementales.

## 2B-14: Graduate Students in Actuarial Science Étudiants de cycles supérieurs en actuariat

---

Organizer and Chair / Responsable et président: Andrei Badescu (University of Toronto)

Room/Salle: Thistle 246

### Abstracts/Résumés

---

**NICOLAS ESSIS-BRETON** (University of Concordia), **PATRICE GAILLARDETZ** (Concordia University)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Multi Trial Monte Carlo*

*Essais multiples de Monte-Carlo*

Multi Trial Monte Carlo allows to price arbitrarily complex path-dependent American option. The algorithm relies on the theory of approximate dynamic programming. In this talk, we review the available valuation techniques, explain why the pricing of certain options remains a challenge, and describe the solution that our algorithm offers.

L'essai multiple de Monte-Carlo permet d'établir le prix d'options américaines complexes et dépendantes du chemin parcouru. L'algorithme est fondé sur la théorie de programmation dynamique. Dans cet exposé, nous examinons les techniques d'évaluation existantes, nous expliquons pourquoi l'établissement des prix de certaines options reste un défi et nous décrivons la solution que notre algorithme apporte.

---

**DONGCHEN LI** (University of Waterloo), **XINFU CHEN** (University of Pittsburgh), **DAVID LANDRIAULT & BIN LI** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 10:35]

*On Minimizing Drawdown Risks of Lifetime Investments*

*Minimisation des risques de prélèvement sur les investissements à vie*

Drawdown measures the decline of portfolio value from its historic high-water mark. We study lifetime investment problems aiming at minimizing drawdown-related risk measures. Under the Black-Scholes framework, we examine two financial market models: a market with two risky assets, and a market with a risk-free asset and a risky asset. Closed-form optimal trading strategies are derived by solving the associated Hamilton-Jacobi-Bellman (HJB) equations. We show that it is optimal to minimize the portfolio variance when the fund value is at its historic high-water mark. Moreover, when the fund value drops, the proportion of wealth invested in the asset with a higher instantaneous rate of return should be increased.

Le prélèvement mesure le déclin de la valeur d'un portefeuille par rapport à son pic historique. Nous étudions des problèmes d'investissement à vie visant à minimiser les mesures de risque lié au prélèvement. En considérant le cadre de Black-Scholes, nous examinons deux modèles de marchés financiers : un marché avec deux actifs à risque et un autre avec un actif sans risque et un actif à risque. Nous dérivons des stratégies optimales d'exécution des opérations de formes fermées en résolvant les équations de Hamilton-Jacobi-Bellman (HJB) associées. Nous montrons qu'il est optimal de minimiser la variance du portefeuille lorsque la valeur du fonds est à son pic historique. De plus, lorsque la valeur du fonds chute, il faut augmenter la proportion de richesse investie dans l'actif dont le taux de rendement de retour instantané est le plus élevé.

---

**HAIYAN LIU, JUN CAI & RUODU WANG** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Asymptotic Equivalence of Risk Measures under Dependence Uncertainty*

*Équivalence asymptotique des mesures de risque en cas d'incertitude de la dépendance*

## 2B-I4: Graduate Students in Actuarial Science Étudiants de cycles supérieurs en actuariat

---

We study the aggregate risk of a large inhomogeneous portfolio with dependence uncertainty, evaluated by a generic risk measure. We establish general asymptotic equivalence results for the classes of distortion risk measures and convex risk measures under different mild conditions. The results implicitly suggest that it is only reasonable to implement a coherent risk measure for an aggregate risk of a large portfolio with uncertainty in the dependence structure, a relevant situation for risk management practice.

Nous étudions le risque global d'un gros portefeuille non homogène avec incertitude concernant la dépendance, tel qu'évalué par une mesure du risque générique. Nous établissons les résultats d'équivalence asymptotique généraux pour les classes de mesures de risque de distorsion et de risque convexe sous diverses hypothèses faibles. Les résultats suggèrent implicitement qu'il n'est que raisonnable d'utiliser une mesure de risque cohérente pour le risque global d'un gros portefeuille avec incertitude dans la structure de dépendance, situation pertinente en pratique pour la gestion des risques.

---

**JINGONG ZHANG, KEN SENG TAN & CHENGGUO WENG** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 11:20]

*Optimal Hedging with Basis Risk under Mean-Variance Criteria*

*Couverture optimale avec risque de base sous critères de moyenne-variance*

Optimal hedging for European options is studied under a mean-variance framework in the presence of basis risk, where the underlying asset is non-tradable and replaced by another closely related and tradable asset. The problem is formulated to determine the subgame Nash equilibrium hedging strategy. This problem differs from classic dynamic programming problems in that its objective function is not separable and therefore, the Bellman optimality principle does not apply. Optimal control process is derived by resorting to dynamic programming technique and an extended HJB equation. A closed-form optimal control process is obtained with the aid of a change-of-measure technique.

J'étudie la couverture optimale pour des options européennes dans un cadre de moyenne-variance en présence de risque de base, où l'actif sous-jacent n'est pas négociable et est remplacé par un autre actif étroitement lié négociable. Le problème est formulé de sorte qu'il permette de déterminer la stratégie de couverture d'équilibre de Nash par sous-jeux. Ce problème se différencie des problèmes de programmation dynamique classiques : sa fonction objectif n'est pas séparable si bien que le principe d'optimalité de Bellman ne s'applique pas. Je dérive le processus de contrôle optimal en utilisant une technique de programmation dynamique et une équation de HJB élargie. J'obtiens un processus de contrôle optimal de forme fermée avec une technique de changement de mesure.

---

**RUIXI ZHANG** (Western University), **KRISTINA SENDOVA** (The University of Western Ontario)

[Tuesday May 31 / mardi 31 mai, 11:35]

*On a Renewal Risk Process under a Dividend Barrier Strategy*

*Processus de risque de renouvellement sous une stratégie de barrière de dividendes*

We consider a renewal risk process in the presence of a constant dividend barrier. Using probabilistic arguments, the probability of ruin is shown to be 1, if either the claim sizes exceed the barrier with positive probability or 0 is a point of increase of the inter-claim time. For  $K_n$ -distributed inter-claim times, we offer a revised proof regarding the number of roots to the generalized Lundberg's equation. The density and moments of the time to

Nous considérons un processus de risque de renouvellement en présence d'une barrière de dividendes constante. En utilisant des arguments probabilistes, nous montrons que la probabilité de ruine est de 1, si les tailles de réclamations dépassent la barrière avec une probabilité positive ou si 0 est un point de hausse du temps entre réclamations. Pour des temps entre réclamations de distribution  $K_n$ , nous proposons une preuve révisée concernant le nombre de racines de l'équation de Lundberg généralisée. Nous effectuons une

## 2B-I4: Graduate Students in Actuarial Science Étudiants de cycles supérieurs en actuariat

---

ruin are approximated numerically for generalized Erlang( $n$ ) inter-claim times and rational-distributed claim sizes. However, if neither condition is satisfied, the probability of ruin may be reduced to 0. A dividend-reinsurance strategy inspired by the last observation is discussed. Finally, generalizations to certain dependent risk processes are included.

approximation numérique de la densité et des moments du temps avant la ruine pour des temps entre réclamations d'Erlang( $n$ ) généralisés et des tailles de réclamations à distribution rationnelle. Cependant, si aucune condition n'est remplie, la probabilité de ruine peut être réduite à 0. Nous discutons d'une stratégie de réassurance des dividendes inspirée par cette dernière observation. Enfin, nous incluons des généralisations à certains processus de risques dépendants.

---

**MOHAMED AMINE LKABOUS** (Université du Québec à Montréal), **IRMINA CZARNA** (University of Wrocław), **JEAN-FRANÇOIS RENAUD** (Université du Québec à Montréal)

[Tuesday May 31 / mardi 31 mai, 11:50]

*Parisian ruin for a refracted Lévy process*

*Ruine Parisienne pour un processus de Lévy réfracté*

Parisian ruin occurs if the time spent below zero by the risk process is longer than a fixed delay. In this talk, we investigate Parisian ruin for a refracted Lévy process. We generalize the result of Loeffen, Czarna and Palmowski (2013) for the probability of Parisian ruin of a standard Lévy insurance risk process. Other fluctuation identities with Parisian delay will be presented. Finally, we will give examples to illustrate the results.

Dans cet exposé, on s'intéresse à la ruine parisienne pour un processus de Lévy réfracté. Nous généralisons le résultat de Loeffen, Czarna and Palmowski (2013) pour la probabilité de ruine parisienne pour les processus Lévy spectralement négatifs. D'autres identités avec délai parisien seront aussi présentées. Enfin, nous présenterons quelques exemples pour illustrer les résultats.

**2B-15: Recent Developments in Designs of Statistical Experiments. In honour of the 65th birthday of Professor Douglas P. Wiens**  
**Progrès récents en planification d'expériences. En l'honneur du 65e anniversaire du professeur Douglas P. Wiens**

---

Organizer and Chair / Responsable et président: Xiaojian Xu (Brock University)

**Room/Salle: Thistle 243**

**Abstracts/Résumés**

---

**HOLGER DETTE** (Ruhr-Universität Bochum), **KIRSTEN SCHORNING & KATHRIN MÖLLENHOFF** (Ruhr-Universität Bochum), **STANISLAV VOLGUSHEV** (Cornell University), **FRANK BRETZ** (Novartis, Basel)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Statistical Methodology for Comparing Curves*

*Méthodologie statistique pour la comparaison de courbes*

An important problem in drug development is to establish the similarity between two dose response curves. We propose statistical methodology improving the current state of the art in two directions. On the one hand optimal designs are constructed minimizing the width of the confidence band for the difference between the regression functions, which is currently used for a test of similarity. The use of the optimal designs proposed yields a reduction of the width of the confidence band by more than 50 percent. On the other hand a new test for the hypothesis of "similarity" is developed, and it is demonstrated that this test has substantially more power than the test which is based on the confidence band.

Un problème important dans le développement de médicaments est la détermination des similitudes entre deux courbes dose-effet. Nous proposons une méthodologie statistique pour améliorer l'état actuel des choses dans deux directions. D'une part, des plans optimaux sont construits pour minimiser la largeur de la bande de confiance pour la différence entre les fonctions de régression, qui est actuellement utilisée pour un test de similarité. L'utilisation des plans optimaux proposés mène à une réduction de la largeur de la bande de confiance de plus de 50 pour cent. D'autre part, un nouveau test pour l'hypothèse de «similarité» est développé et il est démontré que ce test a considérablement plus de puissance que le test fondé sur la bande de confiance.

---

**LINGLONG KONG & DOUG WIENS** (University of Alberta)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Robust Designs for Nonlinear Quantile Regression*

*Plans robustes pour la régression quantile non linéaire*

We give methods for the construction of designs for regression models, when the purpose of the investigation is the estimation of the conditional quantile function, and the estimation method is nonlinear quantile regression. The designs are robust against misspecified response functions, and against unanticipated heteroscedasticity. The methods are illustrated by examples.

Nous offrons des méthodes pour la construction de plans pour les modèles de régression lorsque la recherche vise à estimer la fonction quantile conditionnelle et que la méthode d'estimation est une régression quantile non linéaire. Les plans sont robustes face à des fonctions de réponses mal spécifiées et face à une hétéroscédasticité imprévue. Les méthodes sont illustrées avec des exemples.

---

**JULIE ZHOU** (University of Victoria)

[Tuesday May 31 / mardi 31 mai, 11:20]

*Computing Optimal Designs via Convex Optimization*

*Calculer des plans optimaux par l'optimisation convexe*



**2B-15: Recent Developments in Designs of Statistical Experiments. In honour of the 65th birthday of Professor Douglas P. Wiens**  
**Progrès récents en planification d'expériences. En l'honneur du 65e anniversaire du professeur Douglas P. Wiens**

---

Optimal designs of experiments have applications in many research fields. When optimal designs are hard to construct analytically, we can use numerical algorithms to compute them. Various algorithms have been developed, including multiplicative, simulated annealing, coordinate exchange, semi-infinite programming, and genetic algorithms. In this talk, we will introduce another powerful algorithm for finding optimal designs for linear and nonlinear regression models with discrete design space. This algorithm is developed to solve semidefinite programming (SDP) problems. SDP problems form a special class of convex optimization problems, which have a linear objective function subject to a linear matrix being positive semidefinite. We will show that many optimal design problems can be transformed into SDP problems. Examples using the algorithm will be presented.

Les plans d'expériences optimaux ont des applications dans plusieurs domaines de recherche. Lorsque les plans optimaux sont difficiles à construire analytiquement, nous pouvons utiliser des algorithmes numériques pour les calculer. Différents algorithmes ont été développés, y compris le recuit simulé et multiplicatif, l'échange coordonné, la programmation semi-infinie, et les algorithmes génétiques. Dans cet exposé, nous présenterons un autre algorithme puissant pour trouver des plans optimaux pour les modèles de régression linéaire et non linéaire avec espace de design discret. Cet algorithme est développé pour résoudre les problèmes de programmation semi-définis (SDP). Les problèmes SDP constituent une classe spéciale de problèmes d'optimisation convexe qui ont une fonction objective soumise à une matrice linéaire positive semi-définie. Nous démontrerons que plusieurs problèmes de plans optimaux peuvent être transformés en problèmes SDP. Des exemples utilisant l'algorithme seront présentés.

## 2B-C1: Biostatistics: Applications Biostatistique : applications

---

Chair/Président: Lehana Thabane (McMaster University)

Room/Salle: Thistle 255

### Abstracts/Résumés

---

**MOHAMMAD EHSANUL KARIM & ROBERT W. PLATT** (McGill University), **HELEN TREMLETT & BEAMS STUDY GROUP** (University of British Columbia)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Estimating Inverse Probability Weights using Super Learner when Weight-Model Specification is Unknown in a Marginal Structural Cox Model Context: An Application to Multiple Sclerosis*

*Estimation de poids de probabilité inverse à l'aide du super apprentissage lorsque la spécification des poids du modèle est inconnue dans un modèle de Cox structurel marginal Contexte : Une application à la sclérose en plaques*

This study is motivated by the investigation of the impact of beta-interferon treatment in delaying disability progression in subjects with multiple sclerosis from British Columbia (1995-2008) using a marginal structural Cox model (MSCM). MSCM estimates are known to be highly sensitive to the weight-model misspecification. Through simulation, we show that, with a misspecified MSCM weight-model, Super Learner with a rich set of candidate algorithms gives lower MSE of the treatment effect than using logistic regression model or other statistical learning approaches. We apply the findings of the simulation studies in fitting the MSCM to the multiple sclerosis cohort data.

Cette étude est motivée par l'étude de l'impact du traitement par l'interféron bêta visant à retarder la progression du handicap chez les sujets atteints de sclérose en plaques en Colombie-Britannique (1995-2008) en utilisant un modèle de Cox structurel marginal (MCSM). Les estimations MCSM sont reconnues pour être très sensibles aux erreurs de spécification du poids du modèle. À l'aide de simulations, nous montrons qu'avec une erreur de spécification du poids du modèle, le super apprentissage avec un riche ensemble d'algorithmes candidats donne une EQM inférieure pour l'effet du traitement qu'un modèle de régression logistique ou d'autres approches d'apprentissage statistique. Nous appliquons les résultats des études de simulation à l'ajustement du MCSM aux données de la cohorte sur la sclérose en plaques.

---

**ALEXANDER LEVIS** (McGill University), **DAPHNA HAREL** (New York University), **LINDA KWAKKENBOS** (Radboud University), **BRETT THOMBS** (McGill University)

[Tuesday May 31 / mardi 31 mai, 10:35]

*Using Optimal Test Assembly Methods for Shortening Patient-Reported Outcome Measures: Development and Validation of the Cochin Hand Function Scale-6*

*Utilisation de méthodes optimales d'ensembles de tests pour réduire le nombre de mesures de résultats rapportées par les patients : mise au point et validation de l'échelle 6 de l'indice fonctionnel de la main de Cochin*

Patient-completed questionnaires measure important aspects of health, but may be burdensome to complete. Shortening these measures while maintaining measurement equivalence could reduce burden without compromising data quality. Currently, there are no standard methods for creating shortened forms. Optimal test assembly (OTA),

Les questionnaires remplis par les patients mesurent des aspects importants de la santé, mais le fait de les remplir peut représenter un fardeau. Réduire le nombre de ces mesures tout en conservant l'équivalence de mesure pourrait réduire ce fardeau sans compromettre la qualité des données. Actuellement, il n'y a pas de méthodes standards permettant de raccourcir les formulaires. L'ensemble de tests optimal est

## 2B-C1: Biostatistics: Applications Biostatistique : applications

---

used widely in high-stakes educational test design, can automate the selection of a subset of items through mixed-integer programming by generating short forms that maintain maximum Fisher information across the continuum of the latent trait while satisfying user-given constraints. This study explores the applicability of OTA through an example, the Cochin Hand Function Scale, commonly used in rheumatic diseases.

largement utilisé dans la conception de tests pédagogiques importants dans le domaine de l'éducation. Il peut automatiser, par une programmation linéaire mixte, la sélection d'un sous-ensemble de questions en générant de courts formulaires qui conservent le plus d'information de Fisher dans un continuum de trait latent tout en satisfaisant les contraintes données. Cette étude explore l'applicabilité de l'ensemble de tests optimal au moyen de l'exemple de l'échelle 6 de l'indice fonctionnel de la main de Cochin, souvent utilisé dans les maladies rhumatismales.

---

**LIN LIU** (University of California, San Diego), **BARON JOHN** (University of North Carolina), **LIEBERMAN DAVID** (Veteran Affairs Medical Center and Oregon Health and Science University), **JACOBS ELIZABETH** (University of Arizona), **CROSS AMANDA & MURPHY GWEN** (National Cancer Institute, NIH), **SAMIR GUPTA** (University of California, San Diego)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Statistical Issues in a Prognostic Model for Advanced Colorectal Neoplasia Recurrence*

*Problèmes statistiques dans un modèle de pronostique pour la récurrence de la néoplasie colorectale avancée*

We discuss statistical issues in a study examining whether model-based risk stratification could improve risk prediction over current US surveillance colonoscopy guidelines. Statistical challenges existed in model development, cut-point determination and validation. In a training set, bootstrap LASSO and Bayesian model averaging were used for model selection, cut-points were determined to achieve a priori gain in either sensitivity or specificity. In a validation set, sensitivity, specificity, over- and under-use rate of colonoscopy were estimated. Improvement in clinical benefit was assessed by bootstrapped confidence intervals and net reclassification improvement. This approach successfully identified an improved prognostic model and its potential benefit.

Nous abordons les problèmes statistiques dans une étude qui examine si la stratification du risque en fonction d'un modèle pourrait améliorer la prédiction du risque par rapport aux directives de surveillance actuelles des colonoscopies aux États-Unis. Des défis statistiques existaient dans l'élaboration du modèle, dans la détermination du seuil de risque et dans la validation. Dans un jeu de données d'apprentissage, nous avons utilisé les procédures bootstrap, LASSO et de moyenne de modèles bayésien pour la sélection du modèle et nous avons déterminé les seuils de risque pour gagner a priori soit de la sensibilité ou de la spécificité. Dans un jeu de données de validation, nous avons estimé la sensibilité, la spécificité ainsi que le taux de sur-utilisation et de sous-utilisation de la colonoscopie. Nous avons évalué l'amélioration de l'avantage clinique au moyen d'intervalles de confiance de Bootstrap et de l'amélioration nette de la reclassification. Cette approche a permis de définir avec succès un modèle de pronostique amélioré et d'évaluer son avantage potentiel.

---

**ALEXANDRE PICHÉ & RUSSELL STEELE** (McGill University), **IAN SHRIER** (Center for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital), **NIELS WEDDERKOPP** (Institute of Regional Health Research, University of Southern Denmark)

[Tuesday May 31 / mardi 31 mai, 11:05]

*Nested Dirichlet Process: an Application to Censored Data*

*Processus emboîté de Dirichlet : application à des données censurées*

## 2B-C1: Biostatistics: Applications Biostatistique : applications

---

The CHAMPS-study DK project has been collecting data on the physical well being and injuries of children in Denmark. Estimating the effect of a new sport school program on injury rate is difficult due to between-grade heterogeneity, the injury recurrence, and censoring due to vacation. We used a Bayesian non-parametric method that requires weak assumptions on the distribution of the underlying data. Specifically, we adapted the Nested Dirichlet Process model to handle right-censored data in order to cluster classes, and children within class. It also efficiently estimates the time to injury distribution by sharing statistical strength across classes.

Le projet d'étude du Danemark sur la santé, l'activité et la performance motrice des enfants a permis de recueillir des données sur le bien-être et les blessures physiques d'enfants au Danemark. Il est difficile d'estimer l'effet d'un nouveau programme de sport à l'école sur le taux de blessures à cause de l'hétérogénéité entre les classes, de la récurrence des blessures et de la censure en raison des vacances. Nous avons utilisé une méthode bayésienne non paramétrique qui nécessite des hypothèses faibles sur la distribution des données sous-jacentes. Notamment, nous avons adapté le modèle du processus emboîté de Dirichlet pour gérer les données censurées à droite afin de regrouper les classes et les enfants dans une classe. Ce modèle permet également d'estimer efficacement la distribution du temps écoulé jusqu'à la blessure en répartissant la puissance statistique au travers des classes.

---

**RAZVAN ROMANESCU** (University of Guelph), **ROB DEARDON** (University of Calgary)

[Tuesday May 31 / mardi 31 mai, 11:20]

*Optimal Surveillance of Epidemics over Power Law Networks*

*Surveillance optimale des épidémies sur les réseaux de la loi de puissance*

Properties of statistical alarms have been well studied for simple disease surveillance models, such as normally distributed incidence rates, with a sudden or gradual shift in mean at the start of an outbreak. We investigate the behavior of various alarms on simulated epidemics over non-homogeneous networks built by the Configuration Model, with a power law degree distribution. We develop a new regression-based alarm and show that it has the best performance over this network type, for diseases with relatively short infectious periods. Finally, we illustrate its use on observed rates of influenza-like illness from the French Sentinelles network.

Les propriétés des alarmes statistiques ont bien été étudiées pour des modèles simples de surveillance de maladies, tels que les taux d'incidence répartis normalement, avec un changement soudain ou graduel dans la moyenne au début d'une épidémie. Nous étudions le comportement des diverses alarmes sur des épidémies simulées sur des réseaux non homogènes conçus par le modèle de configuration, au moyen d'une loi de probabilité selon le degré de puissance. Nous mettons au point une nouvelle alarme fondée sur la régression et démontrons qu'elle est la plus efficace sur ce type de réseau pour des maladies dont les périodes d'infection sont relativement courtes. Enfin, nous illustrons son utilisation sur les taux observés de maladies telles que la grippe à partir du réseau français Sentinelles.

---

**KRISHNA SAHA** (Central CT State University)

[Tuesday May 31 / mardi 31 mai, 11:35]

*Estimating the Common Intra-class Correlation Coefficient for Clustered Binary Responses*

*Estimation du coefficient de corrélation intra-classe commun pour des réponses binaires groupées*

In the analysis of several treatment groups for binary responses arising from cluster randomized designs, it is often of interest in measuring the precision of the treatment effects in clinical trials. This inference problem can be addressed based on the

Dans l'analyse de plusieurs groupes de traitement et de réponses binaires provenant de plans d'échantillonnage en grappes aléatoires, il est souvent pertinent de mesurer la précision des effets de traitement dans des essais cliniques. Ce problème d'inférence peut être résolu à partir de l'inter-

## 2B-C1: Biostatistics: Applications Biostatistique : applications

---

confidence interval for a common intraclass correlation, and in many epidemiological applications it is preferable by practitioners. This article focuses on constructing the interval procedures for a common intraclass correlation of several treatment groups. We compare our proposed approaches with a number of large sample procedures through simulations, and illustrate the methodology with an example from biomedical fields.

valle de confiance de la corrélation intra-classe commune, que les praticiens préfèrent utiliser dans de nombreuses applications épidémiologiques. Cet exposé met l'accent la construction de tels intervalles pour une corrélation intra-classe commune dans le cas de plusieurs groupes de traitement. A l'aide de simulations, nous comparons les approches que nous proposons avec certaines procédures développées pour de grands échantillons, et nous illustrons la méthodologie avec un exemple provenant du domaine biomédical.

## 2B-C2: Methodological Advances: A Smörgåsbord Progrès méthodologiques : un buffet abondant

---

Chair/Président: Yves Berger (University of Southampton)

Room/Salle: Thistle 256

### Abstracts/Résumés

---

**HARLAN CAMPBELL** (University of British Columbia)

[Tuesday May 31 / mardi 31 mai, 10:20]

*A Non-parametric Rank Based Test for Semicontinuous Longitudinal Data*

*Un test non paramétrique basé sur les rangs pour les données longitudinales semi-continues*

Semicontinuous longitudinal data can prove challenging. Even under the simplest scenario, the `two-part model` (Olsen & Schafer, 2001), which uses correlated random-effects to account for both the semicontinuous and longitudinal characteristics of the data, requires prohibitively intensive numerical integration to estimate numerous parameters, these often difficult to interpret. Alternatively, reducing the data to binary and fitting a logistic mixed or GEE model, while simple and straightforward, leads to a substantial loss in power. A non-parametric rank-based approach, similar to that proposed by Lin, Li & Tan (2013), is introduced. Simulation studies show relatively high power across a range of scenarios.

Les données longitudinales semi-continues peuvent poser un défi. Même avec le scénario le plus simple, le « modèle à deux parties » (Olsen & Schafer, 2001), qui utilise des effets aléatoires corrélés pour tenir compte à la fois des caractéristiques semi-continues et longitudinales des données, nécessite une intégration numérique intensive prohibitive pour estimer de nombreux paramètres, ceux-ci étant souvent difficiles à interpréter. Par ailleurs, réduire les données à un caractère binaire et ajuster un modèle logistique mixte ou GEE, quoique simple et direct, conduit à une perte substantielle de puissance. Une approche non paramétrique basée sur le rang, similaire à celle proposée par Lin, Li & Tan (2013) est introduite. Des études de simulation montrent une puissance relativement élevée à travers une gamme de scénarios.

---

**JULAN AL-YASSIN, ABDULKADIR HUSSEIN & TINGLEI CHEN** (University of Windsor)

[Tuesday May 31 / mardi 31 mai, 10:35]

*The GLARMA Model for Zero-Inflated Poisson Counts*

*Le modèle GLARMA pour les dénombrements Poisson à surreprésentation de zéros*

We investigate the generalised linear autoregressive moving average (GLARMA) model with observations assumed to follow a Poisson distribution. Since the GLARMA model may not be appropriate for over-dispersed count data. We propose a new approach to the GLARMA model with a connection of the Zero-Inflated Poisson function. We develop likelihood estimation procedures, which allow the models to be estimated.

Nous étudions le modèle linéaire généralisé autorégressif à moyenne mobile (GLARMA) avec des observations qui sont présumées suivre une loi de Poisson. Puisque le modèle GLARMA peut ne pas être approprié pour des données de dénombrement sur-dispersées, nous proposons une nouvelle approche au modèle GLARMA avec une connexion de la fonction de Poisson à surreprésentation de zéros. Nous développons des procédures pour l'estimation de vraisemblance ce qui permet d'estimer les modèles.

---

**FANG HE** (University of Western Ontario)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Carbon Dioxide Concentration Estimation*

*Estimation de la concentration de dioxyde de carbone*

## 2B-C2: Methodological Advances: A Smörgåsbord Progrès méthodologiques : un buffet abondant

---

Carbon dioxide (CO<sub>2</sub>) is one of the necessary elements for plants to process photosynthesis. Changes in the CO<sub>2</sub> concentration have an effect on crop yields. The ultimate goal of this analysis is to predict CO<sub>2</sub> concentration based on the known information, such as previous years of flux tower data and corresponding MODIS data. In this way, sound decisions can be made concerning yields and the protection of environment in a more efficient way.

Le dioxyde de carbone (CO<sub>2</sub>) est l'un des éléments nécessaires aux plantes pour traiter la photosynthèse. Les variations de la concentration de CO<sub>2</sub> ont un impact sur les rendements des cultures. Le but ultime de cette analyse est de prédire la concentration de CO<sub>2</sub> sur la base d'information connue, comme les données de tours de flux des années précédentes et les données MODIS correspondantes. De cette façon, des décisions judicieuses peuvent être prises d'une manière plus efficace au sujet des rendements et de la protection de l'environnement.

---

**YUHONG WEI & PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 11:05]

*Robust Mixture Modelling Based on Generalized Hyperbolic Distribution and its Limiting Case Multivariate Skew-t Distribution with Incomplete Data*

*Modélisation robuste par mélange pour données incomplètes fondée sur une distribution hyperbolique généralisée et son cas limite en distribution t-asymétrique multivariée*

Clustering from incomplete data is a relevant topic because the real world datasets are fatter-tailed and asymmetric, and even worse, with arbitrary patterns of missing observations. Herein, we present flexible methods for mixture of the generalized hyperbolic and its limiting case multivariate skew-t models for learning from such datasets. Under MAR mechanisms, we formulate an analytically feasible EM algorithm for carrying out parameter estimation and imputation of missing values under mixture models and missing at random mechanisms. The proposed methodologies are investigated through a simulation study with varying proportions of synthetic missing values and illustrated through a real data set.

La mise en grappes à partir de données incomplètes est un sujet pertinent puisque les ensembles de données réelles sont à aile plus large et asymétrique, et pire encore, avec des patrons arbitraires d'observations manquantes. Nous présentons ici des méthodes flexibles de mélange de la distribution hyperbolique généralisée et son cas limite en distribution t-asymétrique multivariée pour l'apprentissage à partir de tels ensembles de données. Sous le mécanisme de données manquantes MAR, nous formulons un algorithme EM réalisable analytiquement pour effectuer l'estimation des paramètres et l'imputation des valeurs manquantes en vertu de modèles de mélange et de mécanismes manquant au hasard. Les méthodologies proposées sont examinées grâce à une étude de simulation avec des proportions variables de valeurs manquantes synthétiques et illustrées à l'aide d'un ensemble de données réelles.

---

**YOSHIHIRO TANIGUCHI & CHRISTIANE LEMIEUX** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 11:20]

*Importance Sampling Techniques for Archimedean Copula Models*

*Techniques d'échantillonnage d'importance pour les modèles de copules archimédiennes*

In risk management, the joint losses of assets in a portfolio are often modelled through a copula. An important problem in risk management is the accurate estimation of the extreme tail loss-distribution of a portfolio. As such extreme loss occurs with a very low probability, a plain Monte Carlo estimator is bound to be imprecise. Importance sampling is

En gestion du risque, les pertes conjointes de l'actif dans un portefeuille sont souvent modélisées à partir d'une copule. Un problème important en gestion du risque est l'estimation exacte de la distribution des pertes avec ailes extrêmes dans un portefeuille. Comme la probabilité de pertes extrêmes est très faible, une estimation simple selon la méthode de Monte-Carlo ne pourrait être qu'imprécise. Dans ce

## 2B-C2: Methodological Advances: A Smörgåsbord Progrès méthodologiques : un buffet abondant

---

a popular variance reduction technique in this context. In this talk, we present an efficient importance sampling procedure for Archimedean copula models. Our sampling techniques exploit the Marshall-Olkin representation of Archimedean copulas.

contexte, l'échantillonnage d'importance est une technique de réduction de la variance populaire. Dans cet exposé, nous présentons une procédure efficace d'échantillonnage d'importance pour les modèles de copules archimédiennes. Nos techniques d'échantillonnage font appel à la représentation Marshall-Olkin des copules archimédiennes.

---

**PATRICIA JUDD & YILDLIZ YILMAZ** (Memorial University of Newfoundland)

[Tuesday May 31 / mardi 31 mai, 11:35]

*Two-Phase Response-Dependant Sampling Designs for Time-to-Event Analysis*

*Plans d'échantillonnage à deux phases dépendant de la réponse dans l'analyse de temps d'événement*

Budgetary constraints may prevent measuring expensive covariates in the first phase for all individuals in a cohort. Expensive covariates are only measured for the selected individuals in the second phase. Case-cohort sampling is a cost-efficient design for the second phase to analyze time-to-event data. In a case-cohort design a random sample from the cases and the cohort is selected. We explore variations of case-cohort design which give more efficient association estimates for a given sample size. We stratify cases based on the observed time-to-event values and apply basic stratified sampling. Different strata sampling proportions change the efficiency of association estimates.

Des contraintes budgétaires peuvent empêcher la mesure de covariables coûteuses dans la première phase pour tous les individus d'une cohorte. Les covariables coûteuses ne sont mesurées que pour les individus sélectionnés dans la seconde phase. Lors de la deuxième phase, l'échantillonnage des cas-témoins est un modèle rentable pour analyser les données de temps d'événement. Dans un modèle de cas-témoins, un échantillon aléatoire est sélectionné à partir des cas et de la cohorte. Nous explorons les variations du modèle cas-témoins qui donnent des estimations d'association plus efficaces pour une taille d'échantillon donnée. Nous stratifions les cas selon les valeurs observées de temps d'événement et appliquons un échantillonnage stratifié de base. Différentes proportions pour l'échantillonnage à l'intérieur des strates changent l'efficacité des estimations d'association.



## 2B-C3: Mixture models Modèles de mélange

---

Chair/Président: Hugh Chipman (Acadia University)

Room/Salle: Thistle 259

### Abstracts/Résumés

---

**MARTIN BLOSTEIN** (McMaster University), **ANTONIO PUNZO** (University of Catania), **PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Robust High-Dimensional Modeling with the Contaminated Gaussian Distribution*  
*Modélisation robuste de grande dimension avec une distribution gaussienne contaminée*

The contaminated Gaussian is a robust elliptic distribution that allows for automatic detection of "bad points", i.e. outliers and noise. The contaminated Gaussian factor analysis model is proposed as an extension of the usual latent Gaussian factor analysis model. In turn, a mixture of these contaminated Gaussian factor analyzers is introduced, allowing robust data-reduction and detection of bad points even with high-dimensional data. The number of free parameters is controlled by specifying several parsimonious models with different constraints on covariance structure. For each model, a variant of the EM algorithm is implemented for parameter estimation.

La gaussienne contaminée est une distribution elliptique robuste qui permet la détection automatique de « mauvais points » à savoir les valeurs aberrantes et le bruit. Le modèle d'analyse factorielle de la gaussienne contaminée est proposé comme une extension du modèle habituel d'analyse factorielle gaussien latent. En retour, cela introduit un mélange de ces analyseurs de facteur gaussien, permettant une réduction robuste des données et une détection des mauvais points, même avec des données de grande dimension. Le nombre de paramètres libres est contrôlé en spécifiant plusieurs modèles parcimonieux avec différentes contraintes sur la structure de covariance. Pour chaque modèle, une variante de l'algorithme EM est mise en œuvre pour l'estimation des paramètres.

---

**MICHAEL PATRICK BRIAN GALLAUGHER & PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Extending Fractionally Supervised Classification to Non-Gaussian Mixture Models*  
*Étendre la classification fractionnaire supervisée aux modèles de mélanges non-gaussiens*

Fractionally-supervised classification has recently been considered in the context of the Gaussian mixture model. The approach allows for differing degrees of supervision by increasing or decreasing the respective influence of labelled and unlabelled observations when building a classifier. Using both real and simulated data, the performance of fractionally-supervised classification is considered in cases where the mixture component densities are not Gaussian.

La classification fractionnaire supervisée a récemment été prise en compte dans le cadre du modèle de mélanges gaussiens. L'approche permet divers degrés de supervision en augmentant ou en diminuant l'influence respective des observations étiquetées et non étiquetées lors de la construction d'un classificateur. À l'aide de données simulées et réelles, la performance de la classification fractionnaire supervisée est considérée dans les cas où les densités des composantes du mélange ne sont pas gaussiennes.

## 2B-C3: Mixture models Modèles de mélange

---

**YANG TANG** (McMaster University), **ANTONIO PUNZO** (University of Catania), **PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 11:05]

*Model-Based Clustering of Categorical Data with Extreme Patterns*

*Mise en grappes fondée sur un modèle de données catégoriques avec motifs extrêmes*

We propose a mixture of latent trait models with the contaminated Gaussian distribution for the clustering of binary data. A mixture of contaminated Gaussian distributions is implemented to capture the outliers in the latent space in order to enhance the clustering performance. A variational approximation to the likelihood is exploited to derive a fast algorithm for determining the model parameters. Real and simulated data are used to demonstrate this approach.

Nous proposons un mélange de modèles de traits latents avec une distribution normale contaminée pour la mise en grappes de données binaires. Un mélange de distributions normales contaminées sont mises à l'œuvre pour capturer les valeurs aberrantes dans l'espace latent de façon à relever la performance du groupement. Une approximation variationnelle de la vraisemblance est utilisée pour en tirer un algorithme rapide pour déterminer les paramètres du modèle. Des données réelles et simulées servent à décrire cette approche.

**UTKARSH J. DANG** (McMaster University), **RYAN BROWNE** (University of Waterloo), **PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 11:35]

*Power Exponential Mixtures and Skewed Extensions*

*Mélanges de puissance exponentielle et extensions biaisées*

A family of parsimonious mixtures of multivariate power exponential distributions is presented. The multivariate power exponential distribution is a flexible elliptical alternative to the Gaussian and Student t-distributions, allowing for dealing with both varying tail-weight (light or heavy) and peakedness of data. For particular values of the shape parameter, special and limiting cases of this distribution include the double-exponential, Gaussian, and the uniform distributions. Furthermore, an extension of these models is presented that can also model asymmetric data. Computational and inference challenges will be discussed. Lastly, the utility of the proposed models is illustrated using both toy and benchmark data.

Une famille de mélanges parcimonieux de distributions multivariées de puissance exponentielles est présentée. La distribution multivariée de puissance exponentielle est une variante elliptique flexible des distributions gaussiennes et t de Student, ce qui permet de traiter à la fois des poids de queues variables (légers ou lourds) et l'aplatissement des données. Pour les valeurs particulières du paramètre de forme, des cas particuliers et limitants de cette distribution comprennent les distributions double-exponentielle, gaussienne et uniforme. En outre, une extension de ces modèles qui peut également modéliser les données asymétriques est présentée. Les défis de calcul et d'inférence seront discutés. Enfin, l'utilité des modèles proposés est illustrée à l'aide à la fois de données accessoires et de référence.

**ANJALI SILVA, STEVEN ROTHSTEIN & SANJEENA DANG** (University of Guelph)

[Tuesday May 31 / mardi 31 mai, 11:50]

*Mixture Model Selection for Cluster Analysis of RNA Sequencing Data*

*Sélection de modèles de mélange pour l'analyse du regroupement des données de séquençage ARN*

Model-based clustering utilizes mixture models for clustering. It is a form of unsupervised learning where the group memberships of the observations

Les techniques de regroupement à l'aide de modèles utilisent des modèles de mélange. C'est une forme d'apprentissage non supervisé, où l'appartenance à un groupe d'ob-

## 2B-C3: Mixture models Modèles de mélange

---

are unknown. Hence, mixture models, commonly used for clustering, are fitted for a range of possible components and model selection is applied to determine the optimal number of components. Typically, each component corresponds to a cluster. Here we perform clustering of real and simulated RNA sequencing data, which is characterized as discrete, skewed, and high dimensional. Model selection through information criteria (BIC, ICL, AIC, AIC3) and slope heuristics (Djump and DDSE) are explored and compared.

servations est inconnue. Ainsi, les modèles de mélange, généralement utilisés pour le regroupement, sont ajustés pour une série de composantes du modèle possible, et une procédure de sélection de modèles est appliquée afin de déterminer le nombre optimal de composantes. En général, chaque composante correspond à un groupe. Dans cette étude, nous effectuons une analyse de regroupement pour des données de séquençage ARN réelles et simulées, caractérisées comme discrètes, asymétriques et de grandes dimensions. Nous explorons et comparons les différents modèles sélectionnés à l'aide des critères d'information (BIC, ICL, AIC, AIC3) et de l'heuristique de la pente (Djump et DDSE).

## 2B-C4: Multivariate Data Données multivariées

---

Chair/Président: Lisa Lix (University of Manitoba)

Room/Salle: Thistle 257

### Abstracts/Résumés

---

**BRIAN FRAN CZAK & PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 10:20]

*Discriminant Analysis via Parsimonious Multiple Scaled Mixtures*

*Analyse discriminante à l'aide de mélanges parcimonieux à multiples échelles*

Mixtures of multiple scaled distributions have garnered increased attention in the last few years. One issue with these mixtures is that the covariance structures become highly parameterized as the dimension of the data increases. Because these multiple scaled distributions are formulated using an eigen-decomposed scale matrix, we can introduce parsimony by constraining the constituent elements of this decomposition. We introduce a family of parsimonious multiple scaled mixtures where the component densities are a generalization of the multivariate-t density function. We utilize these models for discriminant analysis using real data sets, and compare their results to the state-of-the-art Gaussian alternative.

Les mélanges de distributions à multiples échelles ont suscité une attention accrue au cours des dernières années. Un problème avec ces mélanges est que les structures de covariance deviennent très paramétrées lorsque la dimension des données augmente. Du fait que ces distributions à multiples échelles sont formulées à partir d'une matrice d'échelle décomposée en vecteurs propres, nous pouvons introduire de la parcimonie en limitant les éléments constitutifs de cette décomposition. Nous introduisons une famille de mélanges parcimonieux à multiples échelles où les densités des composantes sont une généralisation de la fonction de densité t multivariée. Nous utilisons ces modèles pour l'analyse discriminante en utilisant de vrais jeux de données et en comparant leurs résultats à l'alternative gaussienne.

---

**HAOCHENG LI & YUKUN ZHANG** (University of Calgary), **SARAH KOZEY-KEADLE, JOSHUA SAMPSON & CHARLES MATTHEWS** (National Cancer Institute), **RAYMOND CARROLL** (Texas A&M University)

[Tuesday May 31 / mardi 31 mai, 10:35]

*Multivariate Longitudinal Data with Mixed Types of Measurements*

*Données longitudinales multivariées avec types de mesures mixtes*

We take a random effect modeling to longitudinal data with multiple measurements in different types. The outcomes can have continuous, binary, count and proportional variables measured at various time-points with random effects that have a hierarchical correlation structure. Via a quasi-likelihood type approximation for the non-Gaussian components, we transform all types of measurements into pseudo normal variables. The transformed responses are fitted by an ECME algorithm, which is efficient for high dimensional random effects. The method is applied to physical activity

Nous prenons une modélisation de l'effet aléatoire aux données longitudinales avec plusieurs mesures de différents types. Les résultats peuvent être des variables continues, binaires, de compte et proportionnelles mesurées à différents points dans le temps avec des effets aléatoires ayant une structure de corrélation hiérarchique. À l'aide d'une approximation de type quasi-vraisemblance pour les composantes non gaussiennes, nous transformons tous les types de mesures en variables pseudo-normales. Les réponses transformées sont ajustées par un algorithme de ECME qui est efficace pour des effets aléatoires de grande dimension. La méthode est appliquée aux données d'activité physique et

## 2B-C4: Multivariate Data Données multivariées

---

data, and is evaluated empirically by a simulation study.

est évaluée empiriquement par une étude de simulation.

---

**YUBIN (ÉRIC) LI** (University of Windsor), **SÉVÉRIEN NKURUNZIZA** (University of Windsor/Université de Sherbrooke)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Estimation in Multivariate Regression with Measurement Error and Change-Points*

*Estimation en régression multivariée avec erreur de mesure et points de rupture*

In this talk, we present an estimation problem in multivariate regression model with measurement error and unknown change-points. In particular, we consider the case where the target parameter is the matrix of the regression coefficients which is suspected to satisfy some restrictions. Under such an uncertainty, we propose shrinkage type estimators and study their asymptotic properties. The derived asymptotic properties generalize some recent findings in literature. We also prove that the proposed shrinkage estimators dominate the unrestricted estimator.

Dans cette présentation, il est question du problème d'estimation dans le modèle de régression multivarié avec erreur de mesure et points de rupture inconnus. En particulier, nous considérons le cas où le paramètre d'intérêt est la matrice des coefficients de régression qui est susceptible d'être sujette à certaines restrictions. Face à cette incertitude, nous proposons les estimateurs à rétrécissement et étudions leurs propriétés asymptotiques. Les propriétés asymptotiques établies généralisent certains résultats de la littérature récente. Nous prouvons également que les estimateurs à rétrécissement établis dominent l'estimateur sans restriction.

---

**MAKOSSO-KALLYTH SUN** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 11:05]

*Dimension Reduction of Histogram Data*

*Réduction de la dimension de variables de type histogramme*

We present the application of two extensions of principal component analysis to histogram data. Data where each entity and variable are histogram or empirical distribution are called histogram or distributional data. We present two approaches respectively based on the first order moments and the quantiles. We compare and we show the benefits of these methods using real data examples.

Nous présentons l'application de deux extensions de l'analyse en composantes principales de variables de type histogramme. Les tableaux de données distributionnelles correspondent aux tableaux de données pour lesquelles chaque individu et chaque variable sont décrits par une distribution empirique ou un histogramme. Nous présentons deux approches basées respectivement sur les moments d'ordre 1 et les quantiles. Nous comparons et montrons l'intérêt de ces méthodes à partir de données réelles.

---

**IRENE VRBIK** (McGill University)

[Tuesday May 31 / mardi 31 mai, 11:20]

*Cluster Analysis of Genetic Sequence Data via the Gap Procedure*

*Analyse de mise en grappes de données de séquence génétique en utilisant le procédé des écarts*

Phylogenetic clustering typically involves estimating a phylogenetic tree and identifying groups of sequences having small genetic distances and high clade support. We explore a simple distance-

La mise en grappes phylogénétique fait généralement appel à l'estimation d'un arbre phylogénétique et à l'identification de groupes de séquences de courte distance génétique et support de clade élevé. Nous explorons un algorithme de

## 2B-C4: Multivariate Data Données multivariées

---

based clustering algorithm, called the Gap Procedure, which uses gaps in sorted pairwise distances to suggest a natural divide between group members and non-members. We show that the clusters found using the Gap Procedure agree closely with computationally expensive gold standard techniques on well separated groups of HIV sequence data. Simulation studies are presented to illustrate scenarios in which this algorithm may be employed, and more importantly, when more sophisticated methods are required.

classification simple fondé sur la distance, appelé procédé des écarts, qui utilise des écarts en distances par paires ordonnées afin de suggérer une division naturelle entre les membres et non-membres d'un groupe. Nous montrons que les groupements obtenus par le procédé des écarts s'harmonisent étroitement avec ceux obtenus des techniques de référence computationnellement très coûteuses sur des groupes bien séparés de données relatives à la séquence du VIH. Des études de simulation sont présentées pour illustrer des scénarios dans lesquels cet algorithme peut être utilisé, et plus important encore, lorsque des méthodes plus complexes sont nécessaires.

---

**YANG ZHAO** (University of Regina)

[Tuesday May 31 / mardi 31 mai, 11:35]

*Classification and Factor Analysis for Big Data Inference: A Study of Gambling Behaviours and the Associate Factors*  
*Classification et analyse factorielle pour l'inférence de données massives : une étude des comportements de joueurs et des facteurs associés*

In this research we review statistical methods for big data inference. We focus on the k-means and the generalized linear models for big data classification and factor analysis. We are interested in studying of gambling behaviours and identifying problem gamblers. We consider Saskatchewan casinos player club card data base. We classify player into different risk groups based on multivariate observations and study factors that related to the gambling behaviours.

Cette recherche passe en revue les méthodes statistiques pour l'inférence de données massives. Nous nous attardons aux k-moyennes et aux modèles linéaires généralisés pour la classification et l'analyse factorielle des données massives. Nous nous intéressons à l'étude des comportements de joueurs et à l'identification des joueurs pour qui le jeu est problématique. Nous utilisons la base de données pour les jeux de cartes des casinos de la Saskatchewan. Nous classifions les joueurs dans divers groupes à risque en fonction d'observations multivariées et étudions les facteurs en rapport avec les comportements de joueurs.

## 2B-C5: Survey Methodology 2 Méthodologie d'enquête 2

---

Chair/Président: Mahmoud Torabi (University of Manitoba)

Room/Salle: Thistle 258

### Abstracts/Résumés

---

**HYUKJUN (JAY) GWEON & MATTHIAS SCHONLAU** (University of Waterloo), **LARS KACZMIREK & MICHAEL BLOHM** (GESIS)

[Tuesday May 31 / mardi 31 mai, 10:20]

*New Approaches to Automated Occupation Coding Using Statistical Learning*

*Nouvelles approches pour le codage automatisé de la profession à l'aide de l'apprentissage statistique*

Occupation coding refers to coding a respondent's text answer into one of hundreds occupation codes. Automated coding is a challenging problem because answers usually consist of only a few words while there are hundreds of possible categories. We develop several approaches that can be useful in the automatic coding context. The approaches include a hybrid method that combines a duplicate-based approach with a statistical learning algorithm, and a scoring method that is applied in a nearest neighbor approach. We illustrate the proposed approaches with occupational coding in the German ALLBUS panel survey based on the ISCO-88 standard.

Le codage de la profession se réfère au codage de la réponse d'un répondant à l'une des centaines de codes de profession. Le codage automatisé est un problème difficile car les réponses consistent habituellement en quelques mots seulement alors qu'il y a des centaines de catégories possibles. Nous développons plusieurs approches qui peuvent être utiles dans le contexte du codage automatisé. Les approches incluent une méthode hybride qui combine une approche basée sur les doublons et un algorithme d'apprentissage statistique et une méthode de cotation qui est appliquée à une approche du plus proche voisin. Nous illustrons les approches proposées au codage des professions de l'enquête par panel allemand ALLBUS basée sur la norme CITP-88.

---

**IBRAHIMA OUSMANE IDA & LOUIS-PAUL RIVEST** (Laval University)

[Tuesday May 31 / mardi 31 mai, 10:35]

*Balanced Sampling by Using the Cube Method and the Rejective Algorithm*

*L'échantillonnage équilibré par la méthode du cube et la méthode réjective*

In recent years, balanced sampling techniques have experienced a renewed interest. They allow to reproduce the structure of the population in samples in order to improve the efficiency of survey estimates. New procedures have been proposed. These include the cube method, an exact method presented by Deville and Tillé (2004), and an approximate method, the Fuller (2009) rejective algorithm. After a brief presentation of these methods as part of an angler survey, we compare using Monte Carlo simulations, the survey designs produced by these two sampling algorithms.

Au cours de ces dernières années, les techniques d'échantillonnage équilibré ont connu un regain d'intérêt. En effet, elles permettent de reproduire la structure de la population dans les échantillons afin d'améliorer l'efficacité des estimations. Récemment, des nouvelles procédures ont été proposées. Il s'agit notamment de la méthode du cube, une méthode exacte présentée par Deville et Tillé (2004), et une méthode approximative, l'algorithme réjectif de Fuller (2009). Après une brève présentation de ces deux méthodes dans le cadre d'un inventaire de pêcheurs, nous comparons à l'aide de simulations Monte-Carlo, les plans de sondage produits par ces deux méthodes d'échantillonnage.

## 2B-C5: Survey Methodology 2 Méthodologie d'enquête 2

---

**ISABELLE LEFEBVRE & DAVID HAZIZA** (Université de Montréal), **JEAN-FRANÇOIS BEAUMONT** (Statistique Canada)

[Tuesday May 31 / mardi 31 mai, 10:50]

*Simplified Variance Estimation for Complex Designs*

*Estimation simplifiée de la variance pour des plans complexes*

In a complex design framework, standard variance estimation methods entail substantial challenges. As we know, conventional variance estimators involve second order inclusion probabilities, which can be difficult to compute for some sampling designs. Based on Ohlsson's sequential Poisson sampling method (1998), we suggest a simplified estimator for which we only need first order inclusion probabilities. The idea is to approximate a survey strategy (which consists of a sampling design and an estimator) by an equivalent strategy for which Poisson sampling is used. We will discuss proportional to size sampling and two-stage sampling. Results of a simulation study will be presented.

En présence de plans de sondage complexes, les méthodes classiques d'estimation de la variance présentent certains défis. Celles-ci requièrent les probabilités d'inclusion d'ordre 2 qui peuvent être complexes à obtenir. En s'inspirant d'une approche développée par Ohlsson (1998) dans un contexte d'échantillonnage de Poisson séquentiel, nous proposons un estimateur ne requérant que les probabilités d'inclusion d'ordre 1. L'idée est d'approximer la stratégie de l'enquête (consistant d'un plan de sondage et d'un estimateur) par une stratégie équivalente qui considère le plan de Poisson. Nous discuterons d'échantillonnage proportionnel à la taille et à deux degrés. Une étude par simulation sera présentée.

**THUVA VANNIYASINGAM, CHUCK E CUNNINGHAM, GARY FOSTER & LEHANA THABANE** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 11:05]

*Determining the Impact of Different Design Features on Relative Design Efficiency in Discrete Choice Experiments*

*Déterminer l'effet de différents aspects d'un plan sur l'efficacité relative du plan dans les expériences avec choix discrets*

Discrete choice experiments (DCEs) are used to quantify preferences of patients and health care providers. Guidance is needed to avoid creating designs with low statistical efficiency, resulting in biased preference surveys. We simulated 3204 DCE designs to assess how varying DCE design characteristics including the number of attributes (2-20), attribute-levels (2-5), alternatives (2-5), and choice tasks (2-20) affect relative design efficiency. Across all designs, more optimal designs were achieved with fewer attributes, fewer attribute levels, and more alternatives per choice task. These results are widely applicable for creating designs to elicit individual preferences on health services, programs, and products.

Les expériences avec choix discrets (ECD) sont utilisées pour quantifier les préférences des patients et des fournisseurs de soins de santé. Des lignes directrices sont nécessaires pour éviter de créer des plans avec faible efficacité statistique, aboutissant à des sondages biaisés sur les préférences. Nous avons simulé 3204 plans d'ECD pour évaluer comment leurs diverses caractéristiques, y compris le nombre d'attributs (2-20), le niveau des attributs (2-5), les alternatives (2-5) et les tâches à choix multiples (2-20) influent sur l'efficacité relative d'un plan. Les plans les plus optimaux étaient ceux qui comportaient moins d'attributs et de niveaux d'attributs et plus d'alternatives par tâche à choix multiples. Ces résultats peuvent largement s'appliquer à la création de plans qui permettent d'obtenir les préférences individuelles en matière de services, programmes et produits de santé.



## 2B-C5: Survey Methodology 2 Méthodologie d'enquête 2

---

**WEI LIN & NANCY REID** (University of Toronto), **KARLA FOX** (Statistics Canada)

[Tuesday May 31 / mardi 31 mai, 11:20]

*Analysis of an Embedded Experiment in a Survey*

*Analyse d'une expérience intégrée dans le cadre d'une enquête*

We derive the Horvitz-Thompson estimator of the average treatment effect and its variance for a general design. In the presence of auxiliary information, a new model-assisted estimator for the average treatment effect is developed and the variance of the estimator is derived. We show that the new estimator is approximately design-unbiased when a general model is employed. Moreover, it doesn't require auxiliary variable information at the population level and is relatively easy to implement and compute. Simulations carried out indicate that the new estimator gains in efficiency and its relative bias is negligible.

Nous dérivons l'estimateur de Horvitz-Thompson de l'effet moyen du traitement et de sa variance pour un plan d'expérience général. Nous mettons au point un nouvel estimateur assisté par modèle de l'effet moyen du traitement en présence d'information auxiliaire, et nous dérivons la variance de l'estimateur. Nous démontrons que le nouvel estimateur est approximativement sans biais par rapport au plan d'expérience lorsqu'un modèle général est employé. De plus, il ne nécessite pas l'information de variables auxiliaires à l'échelle de la population et est relativement facile à mettre en œuvre et à calculer. Des simulations indiquent que le nouvel estimateur est plus efficace et que son biais relatif est négligeable.

## 2C-P: Poster Session Séance d'affichage

---

(Posters displayed 12:00-17:30. Presenters in attendance 13:30-15:30)

(Les affiches seront exposées de 12 h à 17 h 30. Les auteurs seront présents de 13 h 30 à 15 h 30.)

**Room/Salle: Thistle Corridor**

### Abstracts/Résumés

---

**HUDA AL-WAHSH, ABDULKADIR HUSSEIN & WEIWEI LIU** (University of Windsor)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Scan Statistics in Cox's Model with Frailty*

*Statistiques de balayage dans le modèle de Cox avec fragilité*

We propose a new method to detect clusters for survival data with frailty. It is based on Moran's I test for the Martingale residuals of Cox's model with frailty. Compared to the familiar scan statistic approach, the method maintains a reasonable type I error, has higher power and a reasonable sensitivity in detecting disease clusters under Cox's model with regional frailties.

In addition, the new approach is computationally easier to implement as it utilizes the asymptotic distribution Moran's I as opposed to the spatial scan statistics which resort to bootstrap sampling.

Nous proposons une nouvelle méthode pour détecter les grappes en matière de données de survie avec fragilité, méthode fondée sur le test I de Moran pour les résidus de martingale du modèle de Cox avec fragilité. Comparativement aux statistiques de balayage connues, cette méthode conserve une marge d'erreur de type I raisonnable, a une plus grande puissance et possède une sensibilité raisonnable pour la détection des grappes de maladies en vertu du modèle de Cox avec fragilités régionales.

De plus, cette nouvelle approche est plus facile à mettre en œuvre sur le plan informatique, étant donné qu'elle fait appel à la distribution asymptotique I de Moran contrairement aux statistiques de balayage spatiales qui ont recours à l'échantillonnage bootstrap.

---

**JUSTIN ANGEVAARE & ZENY FENG** (University of Guelph), **ROB DEARDON** (University of Calgary)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Phylogenetic Individual Level Models: Strategies for Simulation and Inference*

*Modèles phylodynamiques individuels : des stratégies de simulation et d'inférence*

Phylogenetics is an emergent field that explores the joint dynamics of disease spread and evolution. When epidemic and evolutionary processes occur on similar time scales, phylogenetic models can be used to improve our understanding of disease dynamics. A phylogenetic extension to the individual level models of infectious disease transmission of Deardon et al. (2010) is developed. Computational methods for stochastic simulation and for Bayesian inference are described, and simulation study results are presented. There are opportunities to better inform infectious disease control

La phylodynamique est un champ d'études émergent qui explore la dynamique conjointe de la dissémination et de l'évolution d'une maladie. Lorsque l'épidémie et le processus évolutif de la maladie se produisent sur des échelles de temps similaires, les modèles phylodynamiques peuvent servir pour mieux comprendre la dynamique de la maladie. Nous généralisons la phylodynamique aux modèles individuels de la transmission de maladies infectieuses de Deardon et coll. (2010). Des méthodes computationnelles pour la simulation stochastique et une inférence bayésienne sont décrites et les résultats d'études de simulation sont présentés. L'utilisation de ces modèles offre des possibilités d'éta-

## 2C-P: Poster Session Séance d'affichage

---

strategies through the use of these models, when, as is increasingly common, pathogen genetic sequence data are available.

blir de meilleures stratégies de contrôle des maladies infectieuses lorsque des données sur le séquençage génétique des agents pathogènes sont disponibles, ce qui est de plus en plus le cas.

---

**FAISAL ATAKORA & PO YANG** (University of Manitoba)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Construction of Optimal Foldover Designs with the General Minimum Lower-Order Confounding*

*Élaboration de plans de repliement optimal avec le critère de l'amalgame général minimal d'ordre inférieur*

Fractional factorial designs are widely used in industry and agriculture. Over the years much research work has been done to study these designs. Foldover fractional factorial designs can de-alias important factors (especially those of interest) so that their effects can be studied without ambiguities. We consider optimal foldover designs using general minimum lower-order confounding criterion. A catalogue of 16- and 32-run optimal foldover designs is constructed and tabulated for practical use. A comparison is made between the general minimum lower-order confounding optimal foldover designs and other optimal foldover designs.

Les plans factoriels partiels sont largement utilisés dans l'industrie et l'agriculture. Au fil des ans, bon nombre de travaux de recherche y ont été consacrés. Ce type de plans peut éliminer le repliement de facteurs importants (notamment d'un intérêt particulier) pour en étudier les effets sans aucune ambiguïté. Nous envisageons les plans de repliement optimal selon le critère de l'amalgame général minimal d'ordre inférieur. À des fins d'usage pratique, un catalogue de plans de repliement optimal de 16 et 32 suites est conçu et tabulé. Une comparaison est établie entre les plans de repliement optimal selon le critère de l'amalgame général minimal d'ordre inférieur à d'autres plans de repliement optimal.

---

**ANDREA BENEDETTI, MUHAMMAD MULLAH, RUSSELL STEELE & JAMES HANLEY** (McGill University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Evaluation of Estimation Techniques for Smoothing via Generalized Linear Mixed Models*

*Évaluation des techniques d'estimation pour le lissage à l'aide de modèles mixtes linéaires généralisés*

Semi-parametric (generalized linear) mixed models (SPMMs) provide a flexible approach to model non-linear associations for binary responses. To achieve a smooth curve, it uses regression splines, and shrinks the coefficients at knot points towards zero. Several approaches are used to estimate the SPMMs including penalized quasi-likelihood (PQL), Laplacian approximation (LA), Gauss-Hermite quadrature (GHQ), and Bayesian Markov Chain Monte Carlo (MCMC). Through simulations, we evaluate the performance of these techniques for estimating the SPMM where smoothing is the main purpose. We then apply these methods to obtain smooth curves in a real life dataset.

Les modèles mixtes semi-paramétriques (linéaires généralisés) (MMSP) offrent une approche souple à la modélisation d'associations non linéaires pour les réponses binaires. Afin d'obtenir une courbe lisse, ces modèles font appel à des splines de régression et rétrécissent les coefficients à des nœuds vers zéro. Plusieurs approches sont utilisées pour estimer les MMSP, y compris la quasi-vraisemblance pénalisée (QVP), l'approximation de Laplace (AL), la quadrature Gauss-Hermite (QGH) et la méthode bayésienne de Monte-Carlo par chaînes de Markov (MCMC). À l'aide de simulations, nous évaluons le rendement de ces techniques pour l'estimation des MMSP lorsque le but premier est le lissage. Nous appliquons ensuite ces méthodes pour obtenir des courbes lisses dans un ensemble de données réelles.

## 2C-P: Poster Session Séance d'affichage

---

**SHELLEY B. BULL** (University of Toronto), **JIHYUNG SHIN & RUIYANG YI** (Lunenfeld-Tanenbaum Research Institute of Mount Sinai Hospital)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Bootstrap Resampling in Genetic Association Analysis of Low-Frequency Variants*

*Rééchantillonnage bootstrap pour une analyse d'association génétique avec variantes peu fréquentes*

The winner's curse refers to upward bias in the magnitude of effect estimates obtained in genome-wide association studies. Performance of bootstrap resampling to reduce bias has been demonstrated for common variants (MAF>5%), but not for low-frequency variants. In standard logistic regression of a binary outcome, effect estimation can fail for low allele counts, particularly in bootstrap samples. The objective of our study is to assess the value of Firth's penalized logistic regression (PML) in bias reduction of effect estimates for variants with MAF=1-5%. In simulation studies we observed slightly larger PML shrinkage effects for lower compared to higher MAF variants.

La malédiction du vainqueur renvoie à un biais positif de la magnitude des estimations d'effet obtenues dans des études d'association pangénomique. La performance du rééchantillonnage bootstrap pour réduire le biais a été démontrée pour des variantes communes (fréquence des allèles mineurs > 5 %), mais non pour des variantes peu fréquentes. Dans la régression logistique standard d'une réponse binaire, l'estimation d'effet peut échouer lorsque le nombre d'allèles est peu élevé, en particulier avec des échantillons bootstrap. Le but de notre étude est d'évaluer la valeur de la régression logistique pénalisée (RLP) de Firth pour réduire le biais de l'estimation d'effet pour des variantes avec FAM =1-5 %. Des études de simulation ont permis d'observer des effets de rétrécissement RLP plus grands pour des variantes avec une FAM moins élevée plutôt que plus élevée.

**BO CHEN & ANDREA BENEDETTI** (McGill University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Quantifying Heterogeneity in Individual Participant Data Meta-Analysis with Binary Outcomes*

*Quantification de l'hétérogénéité dans la méta-analyse de données relatives au participant avec réponses binaires*

We consider how best to quantify heterogeneity in the context of individual participant data meta-analyses of binary data. Both two- and one-stage approaches are evaluated via simulation study. In the two-stage approach, we use the  $I^2$  and  $R^2$  statistics proposed by Higgins et al. In the one-stage approach, we adapt a simulation based intraclass correlation coefficient (ICC) proposed by Goldstein et al. to estimate the  $I^2$ . Preliminary results show that in the presence of effect modification, the estimated  $I^2$  from the one-stage model has better performance than that from the two-stage model.

Nous réfléchissons au meilleur moyen de quantifier l'hétérogénéité dans le contexte d'une méta-analyse de données binaires relatives au participant. Des études de simulation permettent d'évaluer des approches en une et deux étapes. Avec l'approche en deux étapes, nous utilisons les statistiques  $I^2$  et  $R^2$  proposées par Higgins et coll. Avec l'approche en une étape, nous adaptons une simulation fondée sur le coefficient de corrélation intra-classe (CCI) proposé par Goldstein et coll. pour estimer la statistique  $I^2$ . Des résultats préliminaires montrent qu'en présence d'une modification d'effet, la performance de la statistique  $I^2$  estimée dans le modèle en une étape est meilleure que dans le modèle en deux étapes.

**JONATHAN EARL & PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Ranking and Tanking: GLMs for Football*

*Classement et sous-performance délibérée : une approche MLG au football*

## 2C-P: Poster Session Séance d'affichage

---

The problem of ranking American college football teams using a statistical model is considered. A GLM approach is taken so that, for some team  $i$  with schedule  $S_i$ , the rating of team  $i$  is derived using  $S_i$  as predictor variables and a modified rule of succession as the response variable. The basic assumptions and benefits of the approach are discussed along with an explanation of how the ranking concepts can be abstracted.

Nous abordons le problème du classement des équipes collégiales américaines de football à l'aide d'un modèle statistique. Nous faisons appel à un modèle linéaire généralisé (MLG) de façon à classer une certaine équipe  $i$  avec un calendrier  $S_i$ , en utilisant  $S_i$  comme variable prédictive et une règle de succession modifiée comme variable réponse. Nous traitons des postulats de base et avantages de l'approche et expliquons les moyens d'en extraire des concepts de classement.

---

**YUANHAO LAI & MCLEOD A. IAN** (Western University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Computing Numerical Distribution Functions for Periodicity Tests in Microarray Time Series*

*Calcul numérique de fonctions de répartition pour des tests de périodicité dans des séries chronologiques en biopuce*

Identifying the periodicity in microarray time series data has become increasingly important in microarray technology. We propose an innovative likelihood ratio test based on fitting a four parameter harmonic regression and show that it is competitive compared with the other proposed tests of periodicity. A response surface regression approach is implemented to compute the P-values of the test statistic for any periodicity and sample size of the data. In addition, the computation is dramatically speeded up by utilizing the parallel computation through the Shared Hierarchical Academic Research Computer Network.

Il importe de plus en plus en technologie de biopuce d'identifier la périodicité dans les séries chronologiques en biopuce. Nous proposons un test novateur de rapport de vraisemblance basé sur la modélisation d'une régression harmonique à quatre paramètres et en montrons la compétitivité en le comparant à d'autres tests de périodicité. Une approche régressive de surface de réponse est mise en œuvre pour calculer les valeurs de P de la statistique de test pour toute périodicité et taille d'échantillonnage des données. De plus, le calcul est fortement accéléré en utilisant le calcul parallèle par l'entremise de SHARCNET (Shared Hierarchical Academic Research Computer Network).

---

**SEAN MCGRATH** (McGill University), **XIAOFEI ZHAO** (RI MUHC), **ANDREA BENEDETTI** (McGill University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Meta Analysis of Medians*

*Méta-analyse des valeurs médianes*

Our work focusses on meta-analyzing medians. Such data typically arises from meta-analyses that evaluate time-based outcomes (e.g. diagnostic delay, length of hospitalization), and is especially challenging given the variety of outcome data reported by the source studies (e.g. medians and percentiles, means and SDs, etc.). Via simulation, we compared various approaches to meta-analyzing medians including transforming medians and percentiles to means and SDs, and using linear quantile mixed models. We present results from a simulation study, demonstrate our approaches on several real life data sets and provide guidelines for

Nous nous intéressons à la méta-analyse des valeurs médianes. De telles données émanent généralement de méta-analyses qui évaluent des réponses d'ordre temporel (par ex. : délai de diagnostic, durée d'hospitalisation) et elles sont particulièrement intéressantes, compte tenu de la variété des réponses rapportées dans les études sources (par ex. : valeurs médianes et percentiles, moyennes et écarts-types, etc.). Nous avons comparé par simulation différentes approches de méta-analyse des valeurs médianes, y compris la transformation de médianes et percentiles en moyennes et écarts-types et l'utilisation de modèles linéaires mixtes de quantiles. En plus de présenter les résultats d'une étude de simulation, nous faisons état de nos approches à l'aide de

## 2C-P: Poster Session Séance d'affichage

---

data analysts.

plusieurs ensembles de données réelles et fournissons des lignes directrices aux analystes de données.

---

**PAUL NGUYEN & LENNON LI** (Public Health Ontario), **PATRICK BROWN** (Cancer Care Ontario), **STEVEN JOHNSON & JOHN MCLAUGHLIN** (Public Health Ontario)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Spatial Cluster Detection for Ontario Public Health Research*

*Détection de grappes spatiales aux fins d'une recherche de Santé publique Ontario*

Cluster detection is important for assessment and surveillance of disease patterns in public health; however, location data are reported with postal codes, which are dense in urban and sparse in rural areas. Misclassification errors are unavoidable when geocoding data to census regions. To overcome these difficulties, kernel smoothing local-EM algorithm is applied to data aggregated to forward sortation areas (geographical units based on the first 3 characters of postal codes). This method estimates the risk on a tessellation where cumulative intersections of multiple maps are distinct regions. Local-EM algorithm is demonstrated on simulated influenza and gonorrhoea cases in Southern Ontario.

La détection de grappes est importante pour l'évaluation et la surveillance du tableau réel de la morbidité en santé publique ; cependant, les données de localisation sont fournies par codes postaux, des valeurs denses en milieu urbain et éparses en zones rurales. Le géocodage des données par zones de recensement rend inévitables les erreurs de mauvaise classification. Afin de surmonter ces difficultés, un algorithme EM local de lissage par noyau est appliqué aux données agrégées pour arriver à des zones de triage (unités géographiques basées sur les trois premiers caractères du code postal). Cette méthode permet d'estimer le risque pour un pavage sur lequel les intersections cumulatives de cartes multiples sont des régions distinctes. L'algorithme EM local est mis en œuvre dans des cas simulés d'influenza et de gonorrhée dans le Sud ontarien.

---

**NIDHI PATEL, UTKARSH J. DANG & PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Longitudinal Clustering via Mixtures of Multivariate Power Exponential Distributions*

*Mise en grappes longitudinale par mélanges de lois de puissance exponentielles multivariées*

A mixture model approach for clustering longitudinal data is introduced. The approach, which is based on mixtures of multivariate power exponential distributions, allows for varying tail-weight and peakedness in data. In the longitudinal setting, this corresponds to more or less concentration around the most central time course in a component. The models utilize a modified Cholesky decomposition of the component scale matrices and the associated MLEs are derived via a generalized EM algorithm.

Un modèle de mélange pour une mise en grappes de données longitudinales est présenté. Cette approche fondée sur des mélanges de lois de puissance exponentielles multivariées permet de varier le relèvement des ailes et la pointi-cité des données. Dans un contexte longitudinal, cela correspond à une plus ou moins forte concentration autour de la chronologie la plus centrée dans une composante. Les modèles font appel à une décomposition de Cholesky des matrices d'échelle des composantes et les estimateurs du maximum de vraisemblance associés sont dérivés à l'aide d'un algorithme EM généralisé.

## 2C-P: Poster Session Séance d'affichage

---

**REUBEN PEREIRA** (University of Toronto), **PATRICK BROWN** (Cancer Care Ontario), **PAUL NGUYEN** (Public Health Ontario)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Comparing Spatial Models for Mapping Larynx Cancer: A Simulation Study*

*Comparaison de modèles spatiaux pour cartographier le cancer du larynx : une étude de simulation*

When dealing with aggregated spatial information, standard spatial methods can be inappropriate and result in misleading inference. The goal of this study was to assess the level of aggregation at which aggregated and location specific models yield comparable results, and the context under which this may be achieved. For this analysis the aggregated and location specific models were fitted to simulated larynx cancer events at the 2010 census block and county levels in Kentucky and California. The calculated Mean Integrated Squared Errors and ROC curves were used to investigate at which aggregation level the spatial models were comparable.

Lorsqu'il faut traiter des données spatiales agrégées, les méthodes spatiales standards peuvent être inappropriées et aboutir à une inférence trompeuse. Le but de cette étude était d'évaluer à la fois le niveau d'agrégation où les modèles agrégés et ceux de position précise donnent des résultats comparables et le contexte dans lequel cela se produit. Aux fins de cette analyse, les modèles agrégés et ceux de position précise ont été ajustés sur des cas de cancer du larynx simulés dans les îlots de recensement et comtés du Kentucky et de la Californie pour l'année 2010. Les erreurs quadratiques moyennes intégrées et courbes ROC calculées ont été utilisées pour savoir à quel niveau d'agrégation les modèles spatiaux sont comparables.

**GREGORY POND** (McMaster University), **CECILIA A. COTTON & RICHARD J. COOK** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 12:00]

*A Biostatistics Training Initiative for Cancer Research*

*Projet de formation en biostatistique pour la recherche sur le cancer*

The Biostatistics Training Initiative (BTI) is an important program sponsored by the Ontario Institute for Cancer Research. Its goals are to increase the capacity and capability for cancer research in the province through advanced, experiential training in biostatistics. To date 23 Master's students from the University of Waterloo have been placed in eight month internships with mentors at cancer centres in Ontario. In 2016, through additional support from CANSSI, a new and expanded BTI provides support for doctoral and postdoctoral trainees in the province, and features a monthly seminar series on biostatistical methods for, and applications to, cancer research.

Le projet de formation en biostatistique pour la recherche sur le cancer (acronyme anglais BTI) est un important programme parrainé par l'Institut ontarien de recherche sur le cancer. L'objectif est d'accroître la capacité de recherche sur le cancer dans cette province par une formation expérientielle avancée en biostatistique. À ce jour, 23 étudiants à la maîtrise de l'Université de Waterloo ont eu droit à un stage de huit mois avec un mentor dans des centres de cancérologie ontariens. En 2016, grâce à un soutien additionnel de l'Institut canadien des sciences statistiques, un nouveau projet BTI élargi est ouvert à des étudiants au doctorat et en formation postdoctorale dans la province. Il comporte une série de séminaires mensuels sur les méthodes biostatistiques et leurs applications en matière de recherche sur le cancer.

**ANNA THERESA SANTIAGO, MALCOLM BINNS & WENDY LOU** (University of Toronto)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Application of Dimension Reduction to Acquired Brain Injury Assessment*

*Application de la réduction dimensionnelle à l'évaluation de la lésion cérébrale acquise*

## 2C-P: Poster Session Séance d'affichage

---

The high-dimensional and inter-correlated nature of variables used in the evaluation of human brain function often calls for exploratory and confirmatory factor analyses that allow for the detailed exploration of structural relationships in neuroscience data. Motivated by a study involving patients with acquired brain injury (ABI), an overview of current statistical approaches, including principal component analysis, will be presented. The pros and cons of selected dimension reduction techniques will be discussed through real examples with joint consideration of neuropsychological, psychosocial, and functional data from these ABI patients.

La nature à dimensions élevées et à intercorrélations des variables utilisées dans l'évaluation de la fonction cérébrale humaine exige souvent des analyses factorielles exploratoires et confirmatoires qui permettent d'examiner en profondeur les relations structurelles des données en neurosciences. Motivée par une étude portant sur des patients atteints d'une lésion cérébrale acquise (LCA), on présentera un aperçu des approches statistiques courantes, y compris l'analyse en composantes principales. Les avantages et inconvénients de certaines techniques de réduction dimensionnelle seront discutés à l'aide d'exemples réels prenant en compte des données relatives aux patients avec une LCA à la fois d'ordre neuropsychologique, psychosocial et fonctionnel.

---

**SAMIRA SOLEYMANI & IAN MCLEOD** (Western University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Simulating with Box-Cox transformations*

*Étude de simulation avec transformations de Box-Cox*

A new more efficient algorithm is developed for simulating from truncated multivariate normal distributions. A new family of statistical distributions generated by Box-Cox transformations is introduced and the simulation method is data generation from these distributions. Our results have applications to space-time Kriging.

Un nouvel algorithme plus efficace est mis au point à des fins de simulation à partir de distributions normales multivariées. Une nouvelle famille de distributions statistiques générée par les transformations de Box-Cox est présentée et la méthode en simulation est une production de données à partir de ces distributions. Nos résultats comportent des applications en krigeage de l'espace-temps.

---

**MOHSEN SOLTANIFAR, ARMEND LOKKU, SHAHRIAR SHAMS & PAUL COREY** (University of Toronto)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Factors Associated to Brain Reaction Time: A Higher Ordered Complete Random Block Design Study*

*Facteurs associés au temps de réaction du cerveau : étude d'un plan en blocs aléatoires complet d'ordre supérieur*

Brain reactionary abilities in relation to individual and contextual factors have been of increasing interest among psychologists in past decades. The dataset for this study was collected from a completely randomized block design cross-over trial using American game developer Gary Darby's computer game on 48 university graduate students with oral and visual distractors to measure their association with reaction times. Results showed (i) a significant impairing association of both oral and visual distractor with reaction times, (ii) their significant effect modification on each other's association with outcome, and; (iii) significant association of arithmetic skills and driving experience with re-

L'intérêt des psychologues pour les capacités réactionnelles du cerveau liées aux facteurs individuels et contextuels s'est accru au cours des dernières décennies. L'ensemble de données utilisé pour cette étude a été collecté à partir de l'essai intercroisé en plan blocs aléatoires complet faisant appel au jeu informatique du développeur américain Gary Darby. L'essai a été mené auprès de 48 étudiants universitaires de cycle supérieur et comportait des distracteurs verbaux et visuels pour en mesurer l'association à leur temps de réaction. Les résultats indiquaient (i) une association significativement déficiente du distracteur à la fois verbal et visuel avec le temps de réaction ; (ii) une modification d'effet significative de l'association de l'un et l'autre à la réponse ; et (iii) une association significative des aptitudes arithmétiques et



## 2C-P: Poster Session Séance d'affichage

---

action times.

expérience de conduite avec le temps de réaction.

---

**AARON SPRINGFORD & DAVID THOMSON** (Queen's University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Predicting Long-Term Missing Ground-Level Ozone Using Precursor, Solar and Climate Data*

*Prévoir l'absence à long terme de l'ozone au sol à l'aide de données de précurseurs, solaires et climatiques*

Recent years have seen a general decline in ambient ground-level ozone concentrations in developed nations. However, Health Canada researchers are investigating what causes variation in ozone-related mortality risk in larger Canadian cities. As part of this investigation, we are examining the ground-level ozone records carefully, and generating predicted ozone concentrations when none are available. We present two case studies -- Halifax and Toronto -- in which local ozone concentrations appear to be primarily driven by different mechanisms, and develop methods of prediction for each. Predictions are based on data that are often measured at different locations and different sampling rates.

Depuis quelques années, on a observé un déclin général des concentrations ambiantes de l'ozone au sol. Cependant, les chercheurs de Santé Canada s'emploient à étudier les causes d'une variation du risque de mortalité lié à l'ozone dans de grandes villes canadiennes. Un volet de cette étude vise à examiner attentivement les rapports concernant l'ozone au sol, et à déterminer les concentrations d'ozone lorsqu'aucune prévision n'est disponible. Nous présentons deux études de cas – Halifax et Toronto – là où les concentrations locales d'ozone semblent être principalement causées par divers mécanismes et nous élaborons des méthodes de prévision dans chaque cas. Ces prévisions sont basées sur des données souvent mesurées en fonction de différents endroits et taux d'échantillonnage.

---

**IAN WAUDBY-SMITH** (Cancer Care Ontario; University of Waterloo), **ZHIHUI (AMY) LIU** (Cancer Care Ontario), **OLLI SAARELA** (University of Toronto)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Multi-state Models for Chronic Kidney Disease Prevalence Projections in Ontario*

*Modèles multi-états pour des prévisions sur la prévalence de la néphropathie chronique en Ontario*

We consider multi-state models aimed at projecting the prevalence of chronic kidney disease and the cost of its treatment in Ontario. Compared to using time-series methods to predict aggregate patient counts, both short and long term projections can benefit from an individual-level modeling approach by capturing the entire disease progression and treatment history and incorporating patient-level characteristics. We model the transitions from healthy state to pre-dialysis, through various dialysis modalities such as home and in-facility dialysis, to attrition. We then simulate each patient's future trajectories, which can be translated to measures such as patient years and treatment costs.

Nous faisons appel à des modèles multi-états visant à fournir des prévisions sur la prévalence de la néphropathie chronique et du coût de son traitement en Ontario. En les comparant aux méthodes de séries chronologiques pour prédire les nombres agrégés de patients, il peut être avantageux pour arriver à des prévisions à court et à long terme d'utiliser une modélisation individuelle en capturant la progression complète de la maladie et l'historique de traitement, tout en incorporant des caractéristiques propres aux patients. Nous modélisons le passage d'un bon état de santé au stade pré-dialytique à partir de diverses modalités de dialyse, à domicile ou en milieu médical, puis à l'attrition. Nous simulons ensuite les trajectoires ultérieures de chaque patient qui peuvent ensuite être transformées en mesures, comme le nombre de patients-années et les coûts de traitement.

## 2C-P: Poster Session Séance d'affichage

---

**DA ZHONG (DEXEN) XI** (Western University), **STEVE TAYLOR** (Pacific Forestry Centre)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Joint Modeling of Duration and Burn Area of Lightning Caused Forest Fires*

*Modélisation conjointe de la durée et de l'étendue des feux de forêt causés par les éclairs*

This talk discusses the development and the application of statistical methods to jointly model the duration and burn area of severe lightning-caused wildfires in BC. In the joint analysis of the survival processes corresponding to fire duration and the burn area, outcomes are linked through a shared latent random risk term. For such correlated outcomes, incorporating risks that reflect shared latent factors affecting outcomes may be useful for gaining precision for the estimation of parameters and, hence, the identification of risk factors. The use of copulas models to connect the two outcomes will also be considered.

Cet exposé fait état de la mise au point et de l'application de méthodes statistiques pour modéliser conjointement la durée et l'étendue des feux de forêt hors contrôle causés par des éclairs en Colombie-Britannique. Dans cette analyse conjointe des processus de survie correspondant à la durée et à l'étendue de l'incendie, les réponses sont liées par terme de risque aléatoire latents partagé. Pour de telles réponses corrélées, il peut être utile d'incorporer des risques reflétant des facteurs latents partagés influant sur les réponses, à des fins de précision accrue de l'estimation des paramètres et conséquemment, de l'identification des facteurs de risque. Le recours à des modèles de copules pour lier les deux réponses est aussi considéré.

**XU XUAN ZHANG & PAUL D. MCNICHOLAS** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 12:00]

*Non-Gaussian Mixture Model Averaging for Clustering*

*Approche par modèle moyen de mélange non gaussien pour la répartition en grappes*

Model averaging has recently been applied for Gaussian mixture model-based clustering, where either models or a posteriori probabilities are averaged. A similar approach is taken using non-Gaussian mixtures, where averaging models demands consideration of concentration and/or skewness parameters. In particular, merging will be applied to mixtures of skew t-distributions and the performance of the two averaging approaches – averaging a posteriori probabilities and model averaging --- will be compared using both real and simulated data.

L'approche par modèle moyen a récemment été appliquée à la répartition en grappes à l'aide de modèles de mélange gaussiens lorsque la moyenne soit des modèles, soit des probabilités a posteriori est calculée. Nous adoptons une approche similaire pour des modèles de mélange non gaussiens lorsque le calcul du modèle moyen exige de prendre en compte des paramètres de concentration ou d'asymétrie. En particulier, une fusion sera appliquée à des mélanges de lois t asymétriques (skew-t) et la performance des deux méthodes de calcul de moyenne – calcul de moyenne de probabilités a posteriori et aussi de modèles – sera comparée à l'aide de données réelles et simulées.

**2D-I1: Developing Curriculum Guidelines for Canadian Statistics Undergraduate Programs**  
**Comment mettre au point des directives pour les programmes de premier cycle en statistique au Canada**

---

Organizer and Chair / Responsable et présidente: Bethany White (University of Western Ontario)

**Room/Salle: Academic South 217**

**Description**

A working group formed at the 2015 "Advancing Innovation and Scholarship in Statistics Education in Canada" workshop is developing a set of recommendations along the lines of the ASA-endorsed US recommendations (<http://www.amstat.org/education/curriculumguidelines.cfm>). The goal is to create guidelines that Canadian departments can use to inform curriculum decisions. To initiate discussion, we will provide an overview of the development process and share the draft guidelines. Then, attendees and working group members will revise and refine them together. Copies of the draft guidelines will be circulated during the session. Advance copies are available from Bethany White ([bwhite@stats.uwo.ca](mailto:bwhite@stats.uwo.ca)).

Working Group: Karen Buro (MacEwan), Sotirios Damouras (Toronto-Scarborough), Alison Gibbs (Toronto), John Petkau (British Columbia), John Sheriff (Lethbridge), Jim Stallard (Calgary), and Bethany White (Western).

Un groupe de travail établi lors de l'atelier « Comment faire progresser l'innovation et l'érudition en éducation statistique au Canada » de 2015 est en train de mettre au point une série de recommandations inspirées des recommandations américaines approuvées par l'ASA (<http://www.amstat.org/education/curriculumguidelines.cfm>). L'objectif est de créer des directives que pourront utiliser les départements canadiens pour informer leurs décisions relatives aux programmes d'enseignement. Pour lancer la discussion, nous présenterons un survol du processus de développement et une première ébauche des directives. Ensuite, les participants et les membres du groupe de travail les réviseront et les affineront ensemble. Des copies des ébauches de directives seront distribuées lors de la session. Vous pouvez vous en procurer à l'avance auprès de Bethany White ([bwhite@stats.uwo.ca](mailto:bwhite@stats.uwo.ca)).

Groupe de travail : Karen Buro (MacEwan), Sotirios Damouras (Toronto-Scarborough), Alison Gibbs (Toronto), John Petkau (British Columbia), John Sheriff (Lethbridge), Jim Stallard (Calgary), et Bethany White (Western).

## 2D-I2: Genomic Data Analysis, Modelling and Testing Analyse, modélisation et tests pour des données de séquençage génomique

---

Organizer and Chair / Responsable et président: Brad McNeney (Simon Fraser University)

**Room/Salle: Welch - David Howes**

### Abstracts/Résumés

---

**ALI KARIMNEZHAD & DAVID R. BICKEL** (University of Ottawa)

[Tuesday May 31 / mardi 31 mai, 13:30]

*Robust Bayesian Multiple Hypothesis Testing with Application to Genetic Association Data*

*Test d'hypothèses multiples bayésien robuste avec applications aux données d'association génétiques*

An important objective in simultaneous hypothesis testing terminology is to control false discoveries. In most of studies in the literature, local false discovery rates (LFDRs) are estimated based on a combined analysis in which all features presented together are combined to be analyzed together. However, sometimes the nature of data suggests a separate analysis in which only features with the same property in common are analyzed together.

Un objectif important de la terminologie des tests d'hypothèses simultanés est le contrôle des fausses découvertes. Dans la plupart des études, le taux de fausses découvertes locales (LFDR) est estimé à partir d'une analyse combinée dans laquelle toutes les caractéristiques présentées sont combinées pour être analysées conjointement. Cependant, la nature des données suggère parfois une analyse séparée où seules les caractéristiques avec une même propriété en commun sont analysées conjointement.

Conducting both separate and combined analyses might lead to two different decisions which results in an uncertainty in making a unique decision. Overcoming with this kind of uncertainty, we introduce novel approaches including robust Bayes and information-theoretic ones. We analyze a genetic association data set to illustrate practical utility of the proposed methods.

Procéder aux analyses séparées et combinées peut conduire à deux décisions différentes ce qui amène une incertitude dans la prise d'une décision unique. Pour surmonter ce type d'incertitude, nous présentons des approches novatrices dont des approches robustes bayésiennes et théoriques de l'information. Nous analysons un ensemble de données d'association génétiques pour illustrer l'utilité pratique des méthodes proposées.

---

**DAVID A. STEPHENS, FARHAD SHOKOOHI & AURÉLIE LABBE** (McGill University)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Hidden Markov Models for Identifying Differentially Methylated Regions*

*Modèles de Markov cachés pour l'identification des régions différenciellement méthylées*

We develop specialized hidden Markov models (HMMs) to address the issue of detecting differentially methylated regions. These regions are potentially important in explaining phenotypic differences. HMM tools have only rarely been used in the analysis of methylation data, and our new model attempts to address issues of data corruption, and also the different sources of variability that methylation profiles typically exhibit. The computational challenge is, however, large and different

Nous développons des modèles de Markov cachés spécialisés (HMMs) pour traiter du problème de la détection des régions différenciellement méthylées. Ces régions sont potentiellement importantes pour expliquer les différences phénotypiques. Les outils HMM ont rarement été utilisés pour l'analyse de données de méthylation, et notre nouveau modèle tente d'aborder les questions de corruption de données et les différentes sources de variabilité que les profils de méthylation possèdent habituellement. Par contre, ce défi computationnel est vaste et différentes stratégies fondées sur

## 2D-I2: Genomic Data Analysis, Modelling and Testing

### Analyse, modélisation et tests pour des données de séquençage génomique

---

strategies based on EM, direct optimization and MCMC are examined. A test data set involving the BLK gene region, different cell types, is studied.

le EM, l'optimisation directe et MCMC sont étudiées. Un ensemble de données test impliquant la région génétique BLK, différents types de cellules, est étudié.

---

**ELENA SZEFER** (Simon Fraser University & The EMMES Corporation), **JINKO GRAHAM**, **FAISAL BEG & DONGHUAN LU** (Simon Fraser University), **FAROUK NATHOO** (University of Victoria)

[Tuesday May 31 / mardi 31 mai, 14:30]

*A Multivariate Imaging Genomics Application in Alzheimer's Disease*

*Une application d'imagerie génomique multivariée dans les cas de la maladie d'Alzheimer*

Both genetic variants and brain region abnormalities are recognized to play a role in cognitive decline. We explore the relationship between genome-wide variation and region-specific rates of decline in brain structure using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The high-dimensionality of the genomic data lends itself to regularized methods, and we use regularized multivariate methods to estimate sparse relationships between the tens of thousands of SNPs and dozens of brain imaging phenotypes. Because variable selection is challenging in this high dimensional setting, resampling techniques are incorporated to estimate the importance of each SNP in our sample. We also pursue validation of our results using data from the second phase of the ADNI study.

Les variantes génétiques et les anomalies de certaines régions du cerveau sont toutes deux reconnues comme jouant un rôle dans le déclin cognitif. Nous explorons la relation entre la variation pangénomique et les taux de décroissance dans la structure cérébrale en utilisant des données de l'initiative en neuroimagerie de la maladie d'Alzheimer (ADNI). La haute dimensionnalité des données génomiques convient aux méthodes régularisées et nous utilisons des méthodes régularisées multivariées pour estimer les rares relations entre les dizaines de milliers de SNPs et les douzaines de phénotypes d'imageries cérébrales. Parce que la sélection de variables est difficile dans ce cadre à grandes dimensions, des techniques de rééchantillonnage sont incorporées pour estimer l'importance de chaque SNP dans notre échantillon. Nous poursuivons aussi la validation de nos résultats en utilisant des données de la deuxième phase de l'étude ADNI.

## 2D-I3: Long Memory Time Series Séries chronologiques à mémoire longue

---

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

Room/Salle: Thistle 242

### Abstracts/Résumés

---

**CLIFFORD M. HURVICH, ROHIT DEO & JUN LIU** (New York University)

[Tuesday May 31 / mardi 31 mai, 13:30]

*The Slow Convergence of OLS Estimators of Alpha, Beta, and Portfolio Weights Under Long-Memory Stochastic Volatility*

*La lente convergence des estimateurs OLS d'alpha, de beta et des poids de portfolio selon la volatilité stochastique à mémoire longue*

We consider inference for the market model coefficients based on simple linear regression under a long memory stochastic volatility generating mechanism for the returns. We obtain limit theorems for the ordinary least squares (OLS) estimators of  $\alpha$  and  $\beta$  in this framework. These theorems imply that the convergence rate of the OLS estimators is typically slower than  $\sqrt{T}$  if both the regressor and the predictor have long memory in volatility, where  $T$  is the sample size. The traditional standard errors of the OLS-estimated intercept ( $\hat{\alpha}$ ) and slope ( $\hat{\beta}$ ), which disregard long memory in volatility, are typically too optimistic, and therefore the traditional  $t$ -statistic for testing, say,  $\alpha = 0$  or  $\beta = 1$ , will diverge under the null hypothesis.

Nous examinons l'inférence pour les coefficients des modèles de marché fondée sur la régression linéaire simple selon un mécanisme générateur de volatilité stochastique à mémoire longue pour les rendements. Nous obtenons des théorèmes limites pour les estimateurs moindres carrés réguliers (OLS) de  $\alpha$  et  $\beta$  dans ce cadre. Ces théorèmes impliquent que la vitesse de convergence des estimateurs OLS soit typiquement plus lente que  $\sqrt{T}$  si le régresseur et le prédicteur possèdent tous deux une volatilité à mémoire longue, où  $T$  est la taille de l'échantillon. L'erreur type traditionnelle de l'ordonnée ( $\hat{\alpha}$ ) et de la pente estimées des OLS, qui ne tiennent pas compte de la longue mémoire de la volatilité, sont typiquement trop optimistes et par conséquent, la statistique- $t$  traditionnelle pour tester, par exemple,  $\alpha = 0$  ou  $\beta = 1$ , va diverger sous l'hypothèse nulle.

---

**VALDERIO REISEN** (University of Espirito Santo), **CELINE LEVY-LEDUC** (AgroParisTech), **FABIO FAJARDO** (UFES), **MURAD TAQQU** (Boston University)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Long-memory Processes. Robust Spectral Estimators Based on Frequency Domain*

*Processus à mémoire longue. Estimateurs spectraux robustes fondés sur un domaine de fréquence*

This paper discusses the outlier effects on the estimation of a spectral estimator for long memory process under additive outliers and proposes robust spectral estimators. Some asymptotic properties of the proposed robust methods are derived and Monte Carlo simulations investigate their empirical properties. Pollution series, such as, PM (Particulate matter), SO<sub>2</sub> (Sulfur dioxide), are the applied examples investigated here to show the usefulness of the proposed robust methods in real applica-

Cet article discute des effets des valeurs aberrantes sur l'estimation d'un estimateur spectral pour un processus à mémoire longue selon des valeurs aberrantes additives et propose des estimateurs robustes spectraux. Quelques propriétés asymptotiques des méthodes robustes proposées sont dérivées et des simulations Monte-Carlo examinent leurs propriétés empiriques. Des séries de pollution, telles que PM (particules fines), SO<sub>2</sub> (dioxyde de soufre), sont les exemples appliqués examinés pour démontrer l'utilité des méthodes robustes proposées dans des applications réelles.

## 2D-I3: Long Memory Time Series Séries chronologiques à mémoire longue

---

tions. These pollutants present, in general, observations with high levels of pollutant concentrations which may produce sample densities with heavy tails and these high levels of concentrations can be identified as outliers which can destroy the statistical properties of sample functions such as the standard mean, covariance and the periodogram.

Ces polluants présentent, en général, des observations avec de hauts niveaux de concentration en polluants qui peuvent produire des densités d'échantillon avec des ailes lourdes et ces hauts niveaux de concentration peuvent être identifiés comme étant des valeurs aberrantes, ce qui peut détruire les propriétés statistiques des fonctions d'échantillon telles que la moyenne standard, la covariance et le périodogramme.

---

**AERA THAVANESWARAN** (University of Manitoba)

[Tuesday May 31 / mardi 31 mai, 14:30]

*Nonlinear and Long Memory Count Time Series Models and Inference Using Estimating Functions*

*Modèles de séries chronologiques de dénombrement non linéaires et à mémoire longue et inférence utilisant des fonctions d'estimation*

This paper introduces a class of generalized count models and shows that the Autoregressive Conditional Poisson (ACP) models and parameter driven count time series models discussed in the literature are special cases. It is also shown that count time series models such as quadratic stochastic count time series models and long memory count time series models are special cases. The moments of these count time series are derived. The combined martingale estimating functions approach, which provides a convenient framework for deriving optimal inference for nonlinear time series models (Thavaneswaran et al. (2015)), is described for the proposed generalized count time series models. Filtering and prediction for doubly stochastic long memory count time series models are also discussed in some detail.

Cet article présente une classe de modèles de dénombrement généralisés et démontre que les modèles autorégressifs conditionnels de Poisson (ACP) et les modèles de séries chronologiques de dénombrement axés sur les paramètres présentés dans la littérature sont des cas spéciaux. Il est aussi démontré que les modèles de séries chronologiques tels que les modèles chronologiques de dénombrement stochastiques quadratiques et les modèles chronologiques de dénombrement à mémoire longue sont des cas spéciaux. Les moments de ces séries chronologiques sont dérivés. L'approche combinée martingale des fonctions d'estimation, qui fournit un cadre pratique pour dériver l'inférence optimale pour les modèles chronologiques non linéaires (Thavaneswaran et al. (2015)), est décrite pour les séries chronologiques de dénombrement généralisées proposées. Le filtrage et la prédiction pour les modèles de séries chronologiques doublement stochastiques à mémoire longue sont aussi discutés en détail.

## 2D-I4: Methods in Survival Analysis Motivated by challenges in clinical data Méthodes d'analyse de survie motivées par les problèmes de données cliniques

---

Organizer and Chair / Responsable et présidente: Lisa Le (UHN Research)

**Room/Salle: Thistle 247**

### Abstracts/Résumés

---

**AUDREY BORUVKA** (University of Michigan), **YIDAN SHI**, **LEILEI ZENG** & **RICHARD J. COOK** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 13:30]

*Analysis of Clinical Trial Endpoints under Dual Censoring*

*Analyse de résultats d'essais cliniques sous double censure*

Progression-free survival (PFS), given by the earliest of disease progression and death, is a common endpoint in trials for accelerated approval. The adequacy of PFS as a surrogate for overall survival has long been under debate. Issues in the analysis PFS has, until recently, received relatively little attention. The literature offers a variety of approaches to account for the "dual" censoring schemes typical of progression-related endpoints. In this talk we show that the assumptions needed for unbiased estimation have substantive implications for trial design and use of nonparametric methods. To address this issue we devise some recommendations for sizing trials, the timing of assessments for progression, and endpoint evaluation.

La survie sans progression ou PFS (telle que définie par la première des éventualités suivantes : progression de la maladie ou décès) est une issue commune dans les essais pour autorisation accélérée. La pertinence de la PFS comme substitut de la survie globale fait depuis longtemps débat. Les problèmes concernant l'analyse de la PFS n'ont reçu que peu d'attention jusqu'à récemment. La littérature propose une variété d'approches pour tenir compte des mécanismes de censure « double » typiques des issues liées à la progression. Dans cette présentation, nous montrons que les hypothèses nécessaires à une estimation non biaisée ont des conséquences importantes pour les protocoles d'essais et l'utilisation de méthodes non paramétriques. Pour répondre à ce problème, nous formulons des recommandations quant à la taille des essais, au moment de l'évaluation de la progression et à l'évaluation des issues.

---

**MELANIA PINTILIE** (Princess Margaret Cancer Centre), **HARMKE GROOT** & **MICHAEL SCHAAPVELD** (The Netherlands Cancer Institute)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Power for a Case-cohort Study in the Presence of Competing Risks*

*Puissance d'une étude de cas-cohorte en la présence de risques concurrents*

When the number of events is small relative to the size of a cohort, an efficient design is the case-cohort. In this way, all the events and a fraction of the records without events are included in the study, minimising the resources necessary to obtain the complete data for each record. In the presence of competing risks the power of the case-cohort design decreases. The calculation of power for a case-cohort design in the presence of competing risks will be presented. The results of simu-

Lorsque le nombre d'événements est petit par rapport à la taille de la cohorte, un plan d'étude efficace est l'étude cas-cohorte. Celle-ci permet d'inclure dans l'étude tous les événements ainsi qu'une fraction des enregistrements sans événement, minimisant ainsi les ressources nécessaires à l'obtention des données complètes pour chaque enregistrement. Cependant, en la présence de risques concurrents, la puissance d'un tel plan d'essai diminue. Nous calculons la puissance d'un plan cas-cohorte en la présence de risques concurrents et présentons les résultats de simulations. Ces



## 2D-I4: Methods in Survival Analysis Motivated by challenges in clinical data Méthodes d'analyse de survie motivées par les problèmes de données cliniques

---

lations will also be shown. This work was inspired by a case-cohort study seeking to investigate the effect of testicular cancer treatment on urogenital cancer risk in the presence of mortality as a competing risk.

travaux ont été inspirés par une étude cas-cohorte visant à étudier l'effet du traitement du cancer testiculaire sur le risque de cancer urogénital, avec la mortalité comme risque concurrent.

---

**ELEANOR PULLENAYEGUM** (The Hospital for Sick Children)

[Tuesday May 31 / mardi 31 mai, 14:30]

*Using Recurrent Event Models to Handle Longitudinal Data Subject to Irregular Follow-Up*

*Analyse de données longitudinales à suivi irrégulier par modèles d'événements récurrents*

Recurrent event models are helpful in analysing longitudinal data subject to irregular follow-up. Irregular follow up can happen when working with clinic-based cohorts where all visits happen as part of usual care. If sicker patients are seen more often, this risks overestimating the burden of disease. One solution is to weight observations by the inverse of the visit intensity, which requires a recurrent event model for the visit process. Traditionally, this has been done using a Cox proportional hazards model. This talk will highlight examples where frailty models and accelerated failure time models are of interest.

Les modèles d'événements récurrents peuvent être utiles pour l'analyse de données longitudinales à suivi irrégulier. Le suivi peut en effet être irrégulier pour les cohortes en clinique où les visites se font dans le cadre des soins habituels. Puisque les patients plus malades sont examinés plus souvent, la charge de morbidité peut en être surestimée. Une solution est de pondérer les observations par l'inverse de l'intensité des visites, ce qui exige la création d'un modèle d'événements récurrent pour le processus de visite. Traditionnellement, on utilise pour cela un modèle des risques proportionnels de Cox. Nous présentons ici des exemples où les modèles de fragilité et du temps de défaillance accéléré peuvent être plus intéressants.

## 2D-I5: On Some Applications of James-Stein Shrinkage Estimation À propos de quelques applications des estimateurs de rétrécissement à la James Stein

---

Organizer and Chair / Responsable et présidente: Melody Ghahramani (University of Winnipeg)

Room/Salle: Thistle 246

### Abstracts/Résumés

---

**SYED EJAZ AHMED** (Brock University)

[Tuesday May 31 / mardi 31 mai, 13:30]

*Shrinkage Strategy: From the Margin to the Centre of Universe*

*Stratégie de rétrécissement : de la bordure au centre de l'univers*

Stein revealed the most stunning result that the classical maximum likelihood estimator is inadmissible to the suggested shrinkage estimator. Since its inception, the Stein-type shrinkage estimation strategies have received considerable attention from researchers. Modern regularization estimation strategies extend Stein's procedures powerfully. A lot of research is going on in high dimensional data, where the number of variables is greater than the observations. Interestingly, in 1995, in the newsletter of the Royal Statistical Society, Efron predicted that shrinkage and empirical Bayes methodology would be a major area of statistical research for the early 21st century. Shrinkage strategy continues to be useful tools for efficient estimation. I will give some perspectives and historical developments of shrinkage strategy and its applications in big data analytics.

Stein a révélé le résultat le plus étonnant : l'estimateur du maximum de vraisemblance classique est inapplicable à l'estimateur de rétrécissement suggéré. Depuis toujours, les stratégies d'estimation de rétrécissement de type Stein ont bénéficié d'une grande attention de la part des chercheurs. La plupart des stratégies d'estimation par régularisation étendent puissamment les procédures de Stein. De nombreux travaux de recherche sont en cours dans le domaine des données de grande dimension, où le nombre de variables est supérieur à celui des observations. Fait intéressant, en 1995, dans le bulletin de nouvelles de la Royal Statistical Society, Efron a prédit que le rétrécissement et les méthodes bayésiennes empiriques constitueraient un domaine de recherche important au début du 21e siècle. La stratégie de rétrécissement continue d'offrir des outils utiles pour une estimation efficace. Je donnerai un survol et une mise à jour sur la stratégie de rétrécissement et ses applications en analyse des données volumineuses.

---

**SHAKHAWAT HOSSAIN & TREVOR THOMSON** (University of Winnipeg)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Shrinkage Estimation for Generalized Linear Mixed Models*

*Estimation de rétrécissement pour les modèles linéaires mixtes généralisés*

We proposed the pretest and shrinkage estimation methods in the generalized linear mixed models when some of the regression parameters are restricted to a linear subspace. We develop the properties of the pretest and shrinkage estimators including asymptotic distributional biases and risks. We show that these estimators have a significantly higher relative efficiency than the classical estimator. We also consider the LASSO, and numerically compare its relative performance with the proposed

Nous proposons des méthodes d'estimation de prétest et de rétrécissement pour les modèles mixtes linéaires généralisés où certains des paramètres de régression sont limités à un sous-espace linéaire. Nous développons les propriétés des estimateurs de prétest et de rétrécissement, y compris les biais de distribution et les risques asymptotiques. Nous montrons que l'efficacité relative de ces estimateurs est beaucoup plus élevée que celle de l'estimateur classique. Nous examinons aussi le LASSO et comparons numériquement sa performance relative à celle des estima-

## 2D-I5: On Some Applications of James-Stein Shrinkage Estimation

### À propos de quelques applications des estimateurs de rétrécissement à la James Stein

---

estimators. A Monte Carlo simulation study is conducted to evaluate the performance of these estimators with respect to the classical estimators. The study shows that the proposed estimation methods are comparable to the LASSO. A real data example is applied to illustrate the practical usefulness of the proposed estimation methods.

teurs proposés. Nous effectuons une étude par simulation de Monte-Carlo pour évaluer la performance de ces estimateurs par rapport aux estimateurs classiques. L'étude montre que les méthodes d'estimation proposées sont comparables au LASSO. Nous proposons un exemple de données réelles pour illustrer l'utilité pratique des méthodes d'estimation proposées.

---

**BAHADIR YÜZBAŞI** (Inonu University), **SYED EJAZ AHMED** (Brock University)

[Tuesday May 31 / mardi 31 mai, 14:30]

*Pretest, Shrinkage, Positive Shrinkage Ridge Estimators in Linear Regression Models*

*Estimateurs de prétest, de rétrécissement et de rétrécissement positif pseudo-orthogonal dans les modèles de régression linéaire*

We suggest pretest, shrinkage and positive part shrinkage estimators based on ridge estimation in the context of a linear regression model. We compare their performance with some penalty estimators namely lasso, adaptive lasso and SCAD. Monte Carlo studies are conducted to compare the relative performance of the estimators and a real data example is given to illustrate the usefulness of the suggested methods. Further, we investigate the asymptotic properties of suggested estimators.

Nous suggérons des estimateurs de prétest, de rétrécissement et de rétrécissement de partie positive basés sur l'estimation pseudo-orthogonale (ridge) dans le contexte d'un modèle de régression linéaire. Nous en comparons la performance avec d'autres estimateurs par pénalité comme le lasso, le lasso adaptif et le SCAD. Nous effectuons des études de Monte-Carlo pour comparer la performance relative des estimateurs et donnons un exemple de données réelles pour illustrer l'utilité des méthodes suggérées. Enfin, nous étudions les propriétés asymptotiques des estimateurs suggérés.

## 2D-I6: Risk Measures in Actuarial Science Mesures du risque en actuariat

---

Organizer and Chair / Responsable et président: Jun Cai (University of Waterloo)

Room/Salle: Thistle 256

### Abstracts/Résumés

---

**EDWARD FURMAN** (York University), **RICARDAS ZITIKIS** (Western University)

[Tuesday May 31 / mardi 31 mai, 13:30]

*A New Risk Measure for Heavy Tailed Risks*

*Nouvelle mesure de risque pour les risques à queues lourdes*

I will introduce a new tail-based risk measure, the Tail Gini (TG) risk measure, and discuss its properties and links to Solvency II. The TG risk measure aims to catch the variability along the (right) tail of the risk's distribution, but unlike the Tail Standard Deviation risk measure ([Furman, E. and Landsman, Z. (2006). Tail variance premium with applications for Elliptical portfolio of risks. ASTIN Bulletin, 36(2), 433 - 462]), the TG risk measure only requires the finiteness of the first moment. I will suggest an economic capital allocation rule induced by the TG and show explicit expressions in the context of risk portfolios with jointly elliptical risk components. This is a joint work with Ricardas Zitikis of Western University.

J'introduis une nouvelle mesure du risque à queue, dite Tail Gini (TG), et discute de ses propriétés et de ses liens avec le régime Solvabilité II. La mesure du risque TG vise à capter la variabilité sur la queue (droite) de la distribution du risque, mais différemment de la mesure du risque d'écart-standard à queue (Tail Standard Deviation) ([Furman, E. et Landsman, Z. (2006). Tail variance premium with applications for Elliptical portfolio of risks. ASTIN Bulletin, 36(2), 433 - 462]), la mesure de risque TG n'exige qu'une finitude du premier moment. Je suggère une règle de répartition du capital économique induite par la mesure TG et montre des expressions explicites dans le contexte des portefeuilles de risques avec des éléments de risque conjointement elliptiques. Il s'agit d'un projet conjoint avec Ricardas Zitikis de la Western University.

---

**MÉLINA MAILHOT** (Concordia University), **MHAMED MESFIOUI** (Université du Québec à Trois-Rivières), **MANUEL MORALES** (Université de Montréal), **NICHOLAS BECK** (Concordia University)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Multivariate TVaR Risk Decomposition Techniques*  
*Allocation de capital basée sur la TVaR multivariée*

In this talk, different methods of calculating the contribution of each risk within a portfolio with dependent business lines that are not aggregated will be presented. From an enterprise wide risk management point of view, it has become important to calculate the contribution of each risk and provide conservative provisions. The multivariate Value-at-Risk and Tail-Value-at-Risk will be presented, and we focus on three different methods to calculate optimal finite sets for the contribution of each risk within the sums of random vectors to the overall portfolio, which could particularly apply in actuar-

Différentes méthodes afin de calculer la contribution de chaque risque d'un portefeuille ayant des lignes d'affaire dépendantes seront présentées. Il est maintenant extrêmement important pour une compagnie d'assurance ou institution financière d'allouer du capital d'une manière conservatrice, afin de protéger adéquatement les assurés ou investisseurs. Les VaR et TVaR multivariées seront présentées et trois différentes techniques afin d'obtenir un vecteur de même dimension que le nombre de lignes d'affaire du portefeuille en question. Des méthodes d'approximation et estimateurs empiriques seront présentés.

## 2D-I6: Risk Measures in Actuarial Science Mesures du risque en actuariat

---

ial science. Approximations methods and empirical estimators will also be presented.

---

**RUODU WANG** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 14:30]

*Risk Aversion in Monetary Risk Measures*

*Aversion au risque dans les mesures du risque monétaire*

We incorporate the notion of risk aversion favoring prudent decisions from financial institutions into the concept of monetary risk measures. The class of monetary risk measures representing this risk aversion is referred to as consistent risk measures. We characterize the class of consistent risk measures by establishing an Expected Shortfall-based representation. The results obtained suggest that for the determination of regulatory capital, every regulator in favor of risk-averse financial decisions is essentially using a combination of Expected Shortfalls up to some adjustments. This reveals important advantages of the Expected Shortfall for prudent regulation as compared to the Value-at-Risk, a topic very much under discussion in Basel III.

Nous intégrons la notion d'aversion au risque qui encourage les institutions financières à prendre des décisions prudentes dans le concept des mesures du risque monétaire. Nous appelons cette classe de mesures du risque monétaire qui représente cette aversion au risque, mesures de risque cohérentes. Nous caractérisons cette classe de mesures de risque cohérentes en établissant une représentation basée sur le déficit prévu. Les résultats obtenus suggèrent que pour la détermination du capital réglementaire, tous les régulateurs en faveur de décisions financières minimisant les risques utilisent essentiellement une combinaison de déficits prévus avec certains ajustements. Cela révèle les importants avantages que présente le déficit prévu pour une réglementation prudente par rapport à la valeur exposée au risque, sujet très discuté dans le cadre de Bâle III.

## 2D-C1: Actuarial Science and Finance 2 Actuariat et finance 2

---

Chair/Président: Cary Tsai (Simon Fraser University)

Room/Salle: Thistle 255

### Abstracts/Résumés

---

**PETER ADAMIC** (Laurentian University), **JENNA GUSE** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 13:30]

*LOESS Density Estimation for Multivariate Survival Data Subject to Censoring and Masking*

*Estimation de densité LOESS pour données de survie multivariées soumises à de la censure et à du masquage*

Actuaries often encounter censored and masked survival data when constructing multiple decrement tables. In this presentation, we propose estimators for the cause-specific failure time density using LOESS smoothing techniques that are employed in the presence of left-censored data, while still allowing for right-censored and exact observations in addition to masked causes of failure. The smoothing mechanism is incorporated as part of an Expectation-Maximization (EM) algorithm. The proposed models are applied to a bivariate African trypanosomiasis data set.

Les actuaires rencontrent souvent des données de survie censurées et masquées lors de la construction de tables de décrets multiples. Dans cette présentation, nous proposons des estimateurs pour la densité des temps de défaillance par cause en utilisant les techniques de lissage LOESS qui sont utilisées en présence de données censurées à gauche tout en autorisant les observations exactes et censurées à droite en plus des causes masquées de défaillance. Le mécanisme de lissage est incorporé dans le cadre d'un algorithme de maximisation de l'espérance (EM). Les modèles proposés sont appliqués à un ensemble de données à deux variables sur la trypanosomiase africaine.

---

**TAEHAN BAE** (University of Regina), **ALEX KREININ** (IBM)

[Tuesday May 31 / mardi 31 mai, 13:45]

*A Backward Construction and Simulation of Correlated Poisson Processes*

*Une construction en arrière et simulation de processus Poisson corrélés*

In this talk, I will discuss a generalisation of the backward simulation method by Duch et al. (2014) to build and simulate bivariate Poisson processes with specific time correlation structures. The proposed backward construction uses the Marshall-Olkin bivariate binomial distribution for the conditional law and some well-known families of bivariate copulas for joint success probability in lieu of the typical conditional independence assumption. With an appropriate choice of copula function and parameters, the resulting bivariate Poisson processes can incorporate various time correlation structures which are commonly observed in real data.

Dans cet exposé, je discuterai d'une généralisation de la méthode de simulation en arrière par Duch et al. (2014) pour construire et simuler des processus bidimensionnels Poisson avec des structures de corrélation chronologique spécifiques. La construction en arrière proposée utilise la loi binomiale bidimensionnelle Marshall-Olkin pour la loi conditionnelle et quelques familles connues de copules bidimensionnelles pour la probabilité à succès conjointe à la place de l'hypothèse d'indépendance conditionnelle typique. Avec le choix d'une fonction de copule et de paramètres appropriés, les processus bidimensionnels Poisson résultants peuvent incorporer différentes structures de corrélation chronologique qui sont communément observées dans des données réelles.

## 2D-C1: Actuarial Science and Finance 2 Actuariat et finance 2

---

**YIZHOU FANG & MARTIN LYSY** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Stochastic Volatility Modeling With Non-linear Filtering*

*Modélisation de la volatilité stochastique avec filtrage non-linéaire*

Since Heston model, various stochastic volatility modeling methods become very popular in the financial field and multiple assets joint modeling also attracts some attentions. However, popular models incur a heavy computational burden even in the single asset case. We try to incorporate filtering methods to models with latent variables, in order to achieve fast estimation of the joint likelihood. This approach will potentially make multiple assets stochastic volatility joint modeling possible. Primary inference results will be compared with Heston model as a benchmark.

Depuis le modèle Heston, diverses méthodes de modélisation de la volatilité stochastique deviennent très populaires dans le domaine financier et la modélisation conjointe de multiples actifs attire également l'attention. Cependant, les modèles populaires subissent un lourd fardeau informatique même dans le cas d'actifs uniques. Nous essayons d'intégrer des méthodes de filtrage aux modèles à variables latentes afin d'obtenir une estimation rapide de la vraisemblance conjointe. Cette approche pourrait potentiellement rendre possible la modélisation conjointe de la volatilité stochastique de plusieurs actifs. À titre de référence, les résultats d'inférence primaire seront comparés au modèle Heston.

**RAHIM M KHORASANY** (Western University), **R. MARK REESOR** (SHARCNET Research Chair in Financial Mathematics Department of Statistical and Actuarial Sciences Richard Ivey School of Busin)

[Tuesday May 31 / mardi 31 mai, 14:15]

*Monte Carlo Valuation of Multiple Exercise Options: A Structured Regression Approach*

*Évaluation de Monte-Carlo de multiples levées d'options : une approche de régression structurée*

Recent work by Letourneau and Stentoft (2014) shows American option price estimator bias is reduced by imposing additional structure on the regressions used in Monte Carlo pricing algorithms. We extend their methodology to the Monte Carlo valuation of multiple exercise options by requiring additional structure on the regressions used to estimate continuation values. The resulting price estimators have reduced bias, particularly for small sample sizes, and results hold across a variety of option types, maturities, moneyness.

Les travaux récents de Letourneau et Stentoft (2014) montrent que le biais de l'estimateur américain du prix de l'option est réduit en imposant une structure supplémentaire sur les régressions utilisées dans les algorithmes de prix de Monte-Carlo. Nous étendons leur méthodologie à l'évaluation de Monte Carlo de multiples levées d'options en exigeant plus de structure sur les régressions utilisées pour estimer les valeurs de continuation. Les estimateurs de prix résultant ont un biais réduit, en particulier pour les petites tailles d'échantillon, et les résultats tiennent à travers une variété de types d'options, d'échéances et de liquidités.

**ETIENNE MARCEAU & HÉLÈNE COSSETTE** (Université Laval)

[Tuesday May 31 / mardi 31 mai, 14:30]

*A Note on Extreme Negative Dependence*

*Une note sur la dépendance négative extrême*

In this talk, we examine the aggregation of pair of countermonotonic random variables. Notably, we derive closed-form expressions for the cumulative distribution function or other risk measures for sum

Dans cet exposé, on examinera l'agrégation de deux variables aléatoires antimonotones. Notamment, on développera les expressions fermées de la fonction de répartition et de mesures de risque pour la somme de variables aléatoires

## 2D-C1: Actuarial Science and Finance 2 Actuariat et finance 2

---

of two continuous positive countermonotonic random variables in specific cases. A general approach is also proposed to derive the VaR of the sum of two continuous positive countermonotonic random variables. Results related to extreme negative dependence for vector of more than two random variables are also presented.

continues positives antimonotones. Une approche générale sera aussi abordée. Des résultats reliés à la dépendance négative extrême pour des vecteurs de plus de 2 variables aléatoires seront aussi présentés.

---

**NATALIA NOLDE** (University of British Columbia), **JOHANNA ZIEGEL** (University of Bern)

[Tuesday May 31 / mardi 31 mai, 14:45]

*Elicitability and Backtesting*

*Élicitabilité et contrôle ex-post*

Conditional forecasts of risk measures are of importance in internal risk management and regulatory capital calculations. Backtesting is a common way to assess performance of a risk measurement procedure. While traditional backtests are concerned with assessing some optimality property of a set of risk measure estimates, they are not appropriate to compare different procedures. We investigate the proposal of comparative backtests, which are better suited for method comparisons, but necessitate an elicitable risk measure. The discussion focuses on three risk measures, value-at-risk, expected shortfall and expectiles, and is supported by simulation studies and a data analysis of financial time series.

Les prévisions conditionnelles des mesures de risque sont importantes dans la gestion du risque interne et dans les calculs de capital réglementaire. Le contrôle ex-post est une méthode commune pour évaluer l'efficacité d'une procédure de mesure du risque. Bien que les contrôles ex-post classiques servent à évaluer certaines propriétés d'optimalité d'un ensemble d'estimateurs de mesure de risque, ils ne sont pas appropriés pour comparer les différentes procédures. Nous étudions une proposition de contrôles ex-post comparatifs, mieux adaptés aux comparaisons de méthodes, mais qui nécessitent une mesure de risque. L'exposé met l'accent sur trois mesures de risques, à savoir la valeur à risque, la perte de valeur espérée et les expectiles. Il est appuyé par des études de simulation et une analyse de données de séries chronologiques financières.



## 2D-C2: Biostatistics: Treatment Comparisons Biostatistique : comparaisons de traitements

---

Chair/Président: Gyanendra Pokharel (University of Calgary)

Room/Salle: Thistle 254

### Abstracts/Résumés

---

**SIU HUNG CHEUNG** (The Chinese University of Hong Kong), **WENFU XU** (Renmin University), **FEIFANG HU** (George Washington University)

[Tuesday May 31 / mardi 31 mai, 13:30]

*Adaptive Treatment Allocation for Non-inferiority Trials with Multiple Experimental Treatments*

*Allocation de traitement adaptatif pour les essais de non-infériorité avec traitements expérimentaux multiples*

Statistical procedures are available for treatment comparisons in non-inferiority clinical trials that have multiple experimental treatments. An ethical concern for non-inferiority trials is that some patients would be treated by the less effective treatments, and this problem is more serious when multiple experimental treatments are included in a balanced design, where sample sizes are the same for all experimental treatments. With the aim to send fewer patients to inferior treatments, we propose an adaptive treatment allocation scheme which is also superior to a balanced design in terms of test power.

Les procédures statistiques sont disponibles pour les comparaisons de traitement dans les essais cliniques de non-infériorité qui ont des traitements expérimentaux multiples. Un souci éthique dans les essais de non-infériorité fait en sorte que certains patients seraient traités avec des traitements moins efficaces et ce problème est plus sérieux lorsque plusieurs traitements expérimentaux sont inclus dans le plan équilibré où les tailles d'échantillon sont les mêmes pour tous les traitements expérimentaux. Dans le but d'envoyer moins de patients à des traitements inférieurs, nous proposons un schéma d'allocation de traitement adaptatif qui est également supérieur à un modèle équilibré en termes de puissance du test.

---

**JAMES HANLEY** (McGill University), **ZHIHUI (AMY) LIU** (Cancer Care Ontario), **OLLI SAARELA** (University of Toronto)

[Tuesday May 31 / mardi 31 mai, 13:45]

*Fitting a Model of the Mortality Reductions Produced by One/Several Rounds of Cancer Screening: Time and Sample Size Considerations*

*Ajuster un modèle de réduction de la mortalité produit pour un ou plusieurs cycles de dépistage du cancer : considérations de temps et de taille d'échantillon*

Previously (J Med Scr 2013, Int Stat Inst Review 2014) we described: how the earlier treatments prompted by repeated screenings halt/alter the time courses of otherwise-fatal cancers; why screening-induced reductions in time-specific mortality rates exhibit a bathtub-shaped rate ratio curve; and how a 3-parameter model characterizing the impact of a single round of screening can be fitted to the observed year-specific mortality data. We now address the time and sample size requirements (numbers and time locations of deaths) to

Auparavant (J Med Scr 2013, Revue Int Stat Inst 2014), nous avons décrit : comment les traitements antérieurs provoqués par des dépistages répétés stoppent / modifient les cours du temps de cancers autrement fatals ; pourquoi les réductions induites par dépistage des taux de mortalité à des temps spécifiques présentent une courbe du ratio des taux en forme de baignoire ; et comment un modèle à 3 paramètres caractérisant l'impact d'un seul cycle de dépistage peut être ajusté aux données de mortalité observées spécifiques à chaque année. Nous abordons maintenant les exigences de temps et de taille d'échantillon (nombres et em-

## 2D-C2: Biostatistics: Treatment Comparisons Biostatistique : comparaisons de traitements

---

fit this rate ratio curve with a given degree of precision. This approach also applies to other non-proportional-hazards settings.

placements des temps des décès) pour adapter cette courbe du rapport des taux avec un degré de précision donné. Cette approche est également valable pour d'autres contextes de risques non-proportionnels.

---

**TADDELE CHERINET KIBRET, TIAN FENG & REGINA S. KAMPO** (McMaster University), **JEMILA S. HAMID** (St. Michael's Hospital, Toronto), **JOSEPH BEYENE** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Bayesian Network Meta-analysis (NMA) using MCMC and Integrated Nested Laplace Approximation (INLA): A Comparative Study*

*Méta-analyse bayésienne par réseau à l'aide de chaînes de Markov et approximation de Laplace emboîtée intégrée (ALEI) : une étude comparative*

Network meta-analysis (NMA) is an emerging methodology and has become very popular because it allows synthesis of evidence across multiple treatments. The method combines direct and indirect evidence. A MCMC-based Bayesian approach is widely used to fit NMA models, but MCMC is generally computationally intensive. A computationally efficient alternative approach to MCMC is the so-called Integrated Nested Laplace Approximation (INLA). Using comprehensive simulations, we compared the performance of INLA and MCMC with respect to key NMA results including relative treatment effect estimates, rank probabilities for the treatments, and goodness-of-fit measures. The two approaches are also illustrated using real data sets.

La méta-analyse par réseau (MAR) est une méthode émergente qui est devenue très populaire car elle permet la synthèse de preuves à travers plusieurs traitements. La méthode combine une preuve directe et indirecte. Une approche bayésienne basée sur les chaînes de Markov est largement utilisée pour ajuster les modèles MAR, mais les chaînes de Markov sont généralement computationnelles intensives. Une autre approche computationnelle intensive aux chaînes de Markov est la soi-disant approximation de Laplace emboîtée intégrée (ALEI). À l'aide de simulations complètes, nous avons comparé la performance de l'ALEI et des chaînes de Markov par rapport aux résultats clés de la MAR, y compris des estimations de l'effet relatif du traitement, les probabilités de rang pour les traitements et les mesures de validité de l'ajustement. Les deux approches sont également illustrées à l'aide de jeux de données réels.

---

**SHUN FU LEE** (Population Health Research Institute), **ALLAN DONNER & NEIL KLAR** (University of Western Ontario)

[Tuesday May 31 / mardi 31 mai, 14:15]

*Adjusted  $I^2$  Statistic and its Confidence Intervals*

*Statistique  $I^2$  corrigée et ses intervalles de confiance*

The  $I^2$  statistic has been adopted by many researchers for quantifying heterogeneity across trials for its intuitive interpretations allowing comparisons across meta-analyses. In case of the meta-analyses of cluster randomized trials, it is important to adjust the  $I^2$  statistic for clustering and modify existing approaches for constructing confidence intervals including MOVER, Q distribution, test-based and nonparametric bootstrap. A simulation study is used to evaluate the performance of the confidence intervals in terms of coverage,

De nombreux chercheurs ont adopté la statistique  $I^2$  pour quantifier l'hétérogénéité entre les essais, en raison de ses interprétations intuitives qui permettent de comparer les méta-analyses. Dans le cas de méta-analyses d'essais randomisés par grappes, il est important de corriger la statistique  $I^2$  pour la structure de grappes et de modifier les approches existantes pour créer des intervalles de confiance, y compris le modèle MOBILE, la distribution q, et le test de type bootstrap non paramétrique. Nous utilisons une étude de simulation pour évaluer l'efficacité des intervalles de confiance en termes de taux de couverture, d'erreur de

## 2D-C2: Biostatistics: Treatment Comparisons Biostatistique : comparaisons de traitements

---

tail errors and interval width. Data from to a meta-analysis of four cluster randomization trials is used for illustration.

queues et de taille de l'intervalle. Nous illustrons nos travaux à l'aide de données d'une méta-analyse de quatre essais de randomisation par grappes.

---

**TONG-YU LU** (China Jiliang University), **WAI-YIN POON**, **SIU HUNG CHEUNG** & **PING YANG** (The Chinese University of Hong Kong)

[Tuesday May 31 / mardi 31 mai, 14:30]

*Covariates-Adjusted Comparison of Treatments with Ordinal Responses*

*Comparaison de traitements par covariables ajustées, avec des réponses ordinales*

Treatment comparison is frequently performed in clinical studies. When the responses are measured with an ordinal scale, the traditional Wilcoxon-Mann-Whitney test is a popular tool. However, it is less effective than the latent variable methods which have recently been proposed. For treatment comparisons, the inclusion of covariates can increase estimation precision and test power. However, there is a lack of effective methods for the covariates-adjusted comparison of treatments with ordinal responses. Our research gives a new identification procedure for ordinal regression models. These latent ordinal regression models can be conveniently used to conduct covariates-adjusted comparison of treatments with ordinal responses.

Les études cliniques visent souvent à comparer des traitements. Lorsque les réponses sont mesurées au moyen d'une échelle ordinale, on utilise souvent le test traditionnel de Wilcoxon-Mann-Whitney. Cependant, ce test est moins efficace que les méthodes de variables latentes qui ont été récemment proposées. Pour les comparaisons de traitements, l'inclusion de covariables peut augmenter la précision de l'estimation et la puissance du test. Toutefois, il y a un manque de méthodes efficaces pour comparer les traitements par covariables ajustées dans le cadre de réponses ordinales. Notre recherche implémente une nouvelle procédure d'identification pour les modèles de régression ordinale. Ces modèles latents de régression ordinale peuvent facilement être utilisés pour comparer les traitements par covariables ajustées.

---

**BINOD NEUPANE** & **JOSEPH BEYENE** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 14:45]

*Performance of Bayesian Network Meta-analysis for Sparse Data*

*Efficacité de la méta-analyse en réseau bayésienne pour des données rares*

Bayesian network meta-analysis (NMA) is increasingly used to combine evidence across multiple treatments. The method has been used in a range of scenarios including sparse data settings (e.g., rare adverse outcomes, several small trials, etc.). However, comparative performance of fixed-effects (FE) and random-effects (RE) models, and impact of discarding or including trials with zero-event in all arms are not well understood. Using extensive simulations, we showed that overall RE performs better unless event rates are very rare, average trial sizes are small or heterogeneity is low. Discarding zero total event trials did not have marked impact on the results.

La méta-analyse en réseau bayésienne est de plus en plus utilisée pour combiner les données scientifiques de nombreux traitements. Cette méthode a été utilisée dans divers scénarios, y compris dans des ensembles de données rares (ex., dans l'analyse de résultats indésirables rares, dans plusieurs essais cliniques de taille modeste, etc.). Cependant, l'efficacité comparative des modèles à effets fixes et à effets aléatoires, ainsi que les répercussions de l'exclusion de ou de l'inclusion d'essais sans événement dans toutes les branches ne sont pas bien compris. Grâce à plusieurs simulations, nous avons démontré, qu'en général, les modèles à effets aléatoires fonctionnent mieux, à moins que les taux d'événement soient très rares, que la taille moyenne des essais soit modeste ou que l'hétérogénéité soit faible. L'exclusion des essais ayant un nombre d'événements total de zéro n'a pas de répercussions importantes sur les résultats.

## 2D-C3: Statistical Computing and Visualization Calcul statistique et visualisation

---

Chair/Président: Martin Lysy (University of Waterloo)

Room/Salle: Thistle 257

### Abstracts/Résumés

---

**SIMON BONNER** (University of Western Ontario), **MATTHEW SCHOFIELD** (University of Otago), **STEVEN PRICE** (University of Kentucky)

[Tuesday May 31 / mardi 31 mai, 13:30]

*Mark-recapture, Misidentification, and Markov Bases*  
*Capture-recapture, identification erronée et bases de Markov*

Link et al. (2010) introduced the latent multinomial as a general tool for modelling mark-recapture data in which individuals are partially, or incorrectly, identified. Key to this approach is an MCMC algorithm that samples from the set of possible true data sets that could have generated the observed data. We link this model and the field of algebraic statistics, showing that the original algorithm can produce reducible Markov chains that fail to cover the entire sample space and providing a solution using (dynamic) Markov bases. As an example, we present data from a mark-recapture study of queen snakes in Kentucky.

Link et al. (2010) ont présenté le multinomial latent comme un outil général pour la modélisation des données de capture-recapture dans lequel les individus sont partiellement ou mal identifiés. La clé de cette approche est un algorithme MCMC qui échantillonne à partir de l'ensemble des vrais jeux de données possibles qui auraient pu générer les données observées. Nous relient ce modèle au domaine des statistiques algébriques, montrant que l'algorithme original peut produire des chaînes de Markov réductibles ne parvenant pas à couvrir tout l'espace de l'échantillon et fournissant une solution à l'aide de bases de Markov (dynamiques). À titre d'exemple, nous présentons des données d'une étude de capture-recapture de couleuvres royales au Kentucky.

---

**PHILIPPE GAGNON & MYLÈNE BÉDARD** (Université de Montréal), **ALAIN DESGAGNÉ** (Université du Québec à Montréal)

[Tuesday May 31 / mardi 31 mai, 13:45]

*Weak Convergence and Automation of the Reversible Jump Algorithm*  
*Convergence faible et automatisation de l'algorithme à sauts réversibles*

The reversible jump algorithm is a useful MCMC method that allows switches between subspaces of differing dimensionality and therefore model determination. This method is now increasingly used in key areas of human activity (e.g., biology). Its main drawback is the difficulty to implement it practically and efficiently. In this paper, we focus on a simple sampling context to obtain theoretical results that lead to optimisation of the algorithm. The main contribution of this paper is the weak convergence of the stochastic process engendered by the algorithm, which is used to design optimally the functions needed to do the implementation.

L'algorithme à sauts réversibles est une méthode MCMC qui permet de passer d'un sous-espace ayant une certaine dimension à un autre n'ayant pas nécessairement la même dimension, et alors, de sélectionner un modèle. Cette méthode est utilisée dans plusieurs sphères de l'activité humaine (par exemple, la biologie). Il est cependant difficile d'implémenter cette méthode de manière à obtenir un algorithme efficace. Dans cet article, on étudie un contexte d'échantillonnage relativement simple afin d'obtenir des résultats théoriques qui mènent à l'optimisation de l'algorithme. La contribution principale de cet article est la convergence faible du processus stochastique engendré par l'algorithme.

## 2D-C3: Statistical Computing and Visualization Calcul statistique et visualisation

---

**YUN LING & MARTIN LYSY** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Superfast Gaussian Likelihood Solver and Application*

*Solveur très rapide du maximum de vraisemblance gaussien et application*

Gaussian distributions with Toeplitz variance matrices commonly occur. Toeplitz variance matrices admit 'fast' likelihood evaluations by the famous Durbin-Levinson algorithm, scaling as  $\mathcal{O}(N^2)$  instead of the usual  $\mathcal{O}(N^3)$ . We adapt a 'superfast' algorithm of Ammar and Gragg (1988) to calculate the likelihood in  $\mathcal{O}(N \log^2 N)$ . We also use it to derive formulas for the score function and Hessian matrix, thereby obtaining superfast versions of common statistical inference algorithms. Our current implementation in the *R* package *SuperGaussian* beats Durbin-Levinson for  $N > 512$ . We present an application to drift estimation of sub-diffusive pathogen trajectories in pulmonary mucus.

Les lois gaussiennes ayant comme matrice de variance une matrice de Toeplitz sont communes. Les matrices de variance de Toeplitz permettent des évaluations « rapides » du maximum de vraisemblance au moyen de l'algorithme bien connu de Durbin-Levinson d'ordre  $\mathcal{O}(N^2)$  au lieu de  $\mathcal{O}(N^3)$ . Nous adaptons l'algorithme « très rapide » de Ammar et Gragg (1988) afin de calculer le maximum de vraisemblance d'ordre  $\mathcal{O}(N \log^2 N)$ . Nous l'utilisons également afin d'obtenir des formules pour la fonction de caractérisation et la matrice de Hessian, ce qui nous permet d'obtenir des versions très rapides d'algorithmes d'inférence statistiques communs. Notre mise en œuvre actuelle dans le paquet *R SuperGaussian* surpasse l'algorithme de Durbin-Levinson lorsque  $N > 512$ . Nous présentons une application de l'estimation du changement des trajectoires d'un pathogène à sous-diffusion dans la muqueuse pulmonaire.

**WAYNE OLDFORD** (University of Waterloo), **ADRIAN WADDELL** (Roche, Basel Switzerland)

[Tuesday May 31 / mardi 31 mai, 14:15]

*Loon – the Interactive R Visualization Toolkit for Exploration*

*Loon : une boîte à outils interactive pour la visualisation avec R à des fins exploratoires*

The power and extendibility of loon will be demonstrated with new tools for exploring data interactively. These include graph-based exploration of low dimensional spaces, interactive parallel and radial axes displays and glyphs, and interactive maps for geographic spatial data. Loon complements and supports statistical exploration: of data, of models, even of methods. The principal example will be visual clustering and exploration of high dimensional data. An exceedingly brief demonstration of the ease with which loon may be adapted to serve almost any visual exploration will also be given.

Nous démontrerons la puissance et l'extensibilité de Loon avec de nouveaux outils d'exploration interactive de données. Ces outils permettent l'exploration graphique d'espaces à basse dimension, l'affichage d'axes parallèles et radiaux et de glyphes interactifs, ainsi que de cartes interactives pour les données géographiques spatiales. Loon complète et permet l'exploration statistique de données, de modèles et même de méthodes. Notre principal exemple sera le regroupement et l'exploration visuelle de données de grande dimension. Nous démontrerons très brièvement la facilité avec laquelle Loon peut être adapté pour servir à presque toutes les explorations visuelles.

**YUKUN ZHANG** (Yukun Zhang), **HAOCHENG LI** (University of Calgary), **SARAH KOZEY-KEADLE & CHARLES MATTHEWS** (National Institutes of Health/National Cancer Institute), **RAYMOND CARROLL** (Texas A&M University)

[Tuesday May 31 / mardi 31 mai, 14:30]

*PAactivPAL: An R Package for Statistical Analysis on Physical Activity Data*

*« PAactivPAL » : un paquet R pour l'analyse statistique de données sur l'activité physique*

## 2D-C3: Statistical Computing and Visualization Calcul statistique et visualisation

---

Researches for the links between physical activity and cancer risk have been revolutionized by the availability of accelerometers. ActivPAL<sup>TM</sup> is a popular wearable device in application which classifies the person's activity and estimates daily energy expenditure. The device could produce over 10,000 observations per person per day on a second-by-second basis which makes the raw activPAL<sup>TM</sup> data be complicated to analyze. We develop an R package PAactivPAL to solve this problem. The functions embedded summarize variables recorded in physical activity. The package also includes functions for plotting and group comparison, which facilitates users to analyze the data for different perspectives.

La disponibilité des accéléromètres a révolutionné les recherches sur les liens entre l'activité physique et le risque de cancer. ActivPAL<sup>TM</sup> est un appareil portable populaire en application qui catégorise l'activité d'une personne et calcule sa dépense d'énergie quotidienne. L'appareil peut fournir de seconde en seconde plus de 10 000 observations par personne et par jour, ce qui complique l'analyse des données brutes activPAL<sup>TM</sup>. Nous avons élaboré le paquet R PAactivPAL pour résoudre ce problème. Les fonctions intégrées résumant les variables enregistrées de l'activité physique. Le paquet comporte aussi des fonctions pour le traçage et la comparaison de groupes, ce qui permet aux utilisateurs d'analyser facilement les données selon différentes perspectives.

---

**ANDREJ BLEJEC** (National Institute of Biology, Ljubljana, Slovenia)

[Tuesday May 31 / mardi 31 mai, 14:45]

*Presentation of Statistical Concepts with Animated Graphics and Simulations in R*

*Présentation des concepts statistiques à l'aide de graphiques animés et de simulations en R*

Understanding of statistical concepts is important for proper use of statistics. The idea of using simulations and dynamic graphics to foster understanding of statistical concepts is not new. In recent years, R became the *lingua franca* for statistical data analysis. While R graphical devices are not meant for display of animated graphics, my aim is to use base R graphics for display of animated sequences. This is enabled with a R package *animator*, which supports smooth transitions of graphical elements and simplifies preparation of animated displays. I will present package features and animations that might be useful for statistics teaching.

La compréhension des concepts statistiques est importante pour une utilisation adéquate des statistiques. L'idée d'utiliser des simulations et des graphiques dynamiques pour favoriser la compréhension des concepts statistiques n'est pas nouvelle. Au cours des dernières années, R est devenu le *lingua franca* de l'analyse de données statistiques. Alors que les dispositifs graphiques de R ne sont pas destinés à l'affichage de graphiques animés, mon but est d'utiliser des graphiques de base de R pour l'affichage de séquences animées. Cette option est activée avec un paquet de R *animator*, qui prend en charge des transitions en douceur d'éléments graphiques et simplifie la préparation des affichages animés. Je présenterai les caractéristiques du paquet et des animations qui pourraient être utiles pour l'enseignement statistique.

## 2D-C4: Statistical Genetics Génétique statistique

---

Chair/Président: David Campbell (Simon Fraser University)

Room/Salle: Thistle 259

### Abstracts/Résumés

---

**BO CHEN, RADU V. CRAIU & LEI SUN** (University of Toronto)

[Tuesday May 31 / mardi 31 mai, 13:30]

*Bayesian Model Averaging Method for X Chromosome Inactivation Problem*

*Méthode de la moyenne bayésienne des modèles pour l'inactivation du chromosome X*

In genetic association studies it is not known whether the X-chromosome is inactivated or not. In the absence of evidence in favour of one specific model, we consider a Bayesian model averaging framework that provides a principled way to account for model uncertainty. Given an established association, we then use data-based evidence for model selection via Bayes factors. We examine the inferential properties of the proposed methods via simulations and applications and compare with frequentist solutions.

Dans les études d'association génétique, on ne sait pas si le chromosome X est inactif ou pas. En absence d'évidence en faveur d'un modèle ou l'autre, nous considérons la méthode de la moyenne bayésienne des modèles qui nous offre une alternative de principe pour intégrer, dans l'estimation, l'incertitude concernant le modèle correct. Après qu'une association génétique ait été établie, nous considérons l'évidence fournie par le facteur Bayes pour choisir un modèle. Nous examinons les propriétés de l'analyse statistique que nous proposons en utilisant des simulations et des données réelles et nous les comparons avec des solutions fréquentistes.

---

**LU CHENG & MU ZHU** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 13:45]

*Epistasis Detection by Reducing Disease Model Space using Clustering*

*Détection de l'épistasie en réduisant l'espace du modèle de maladie à l'aide de la classification*

Failure of replication for single locus effects in GWAS motivates interaction effect (epistasis) detection. Determining the interaction form, or "disease model" is important for the essentially variable selection task. There are theoretically  $2^9$  interaction forms for each pair of SNPs and assessing all is not statistically or computationally efficient. We thus define a novel 'distance' metric to measure how different they are, do a clustering of them, and then select representatives to guide the epistasis detection. Our method provides competitive power when compared to two most relevant existing method, i.e., MDR and Wan et al. (2013)'s complete compositional epistasis detection method.

L'échec de réplication des effets du locus simple dans les GWAS motive la détection de l'effet d'interaction (épistasie). Il est important de déterminer la forme de l'interaction ou « modèle de maladie » pour la tâche de sélection des variables. Il existe des formes d'interaction théoriques  $2^9$  pour chaque paire de SNP et il n'est pas statistiquement ou informatiquement efficace de toutes les évaluer. Nous définissons alors une nouvelle métrique de « distance » pour mesurer dans quelle mesure elles sont différentes, en faire une classification et ensuite sélectionner des représentants pour guider la détection de l'épistasie. Notre méthode fournit une puissance compétitive en comparaison avec les deux méthodes existantes les plus pertinentes, soit la MDR et la méthode de détection de l'épistasie par composition complète de Wan et al. (2013).

## 2D-C4: Statistical Genetics Génétique statistique

---

**GUN HO JANG, LINCOLN STEIN & JOHN BARTLETT** (Ontario Institute for Cancer Research)

[Tuesday May 31 / mardi 31 mai, 14:00]

*Locally Constant Signal Segmentation using a Relative Belief Inference*

*Segmentation de signal localement constant utilisant une inférence de croyance relative*

Considerable amount of copy number alterations are observed in many tumor samples. As tumor develops chromosomal regions are involved in deletions and amplifications, which are often assumed to follow a locally constant signal model in many developed methods. However weak signals affected by low abnormal tumor fraction and/or subclonal heterogeneous tumor are still hard to be identified.

We propose a segmentation method using a relative belief inference on a locally constant model. The performance of the proposed method is presented and compared with several segmentation algorithms including circular binary segmentation, allele-specific piecewise constant fitting and SCAN algorithms.

Des quantités considérables de modifications du nombre de copies sont observées dans de nombreux échantillons de tumeurs. Lorsque la tumeur se développe, les régions chromosomiques sont impliquées dans des suppressions et des amplifications. Cela est souvent supposé suivre un modèle de signal localement constant dans de nombreux modèles développés. Cependant, les signaux faibles affectés par une fraction de la tumeur anormalement faible et/ou une tumeur hétérogène subclonale sont encore difficiles à identifier.

Nous proposons une méthode de segmentation qui utilise une inférence de croyance relative sur un modèle localement constant. La performance de la méthode proposée est présentée et comparée à plusieurs algorithmes de segmentation, y compris la segmentation binaire circulaire, l'ajustement par morceaux constants d'allèles spécifiques et les algorithmes SCAN.

**MATEEN SHAIKH, JEN STEARNS, MICHAEL SURETTE, SONIA ANAND, RUSSELL DE SOUZA & JOSEPH BEYENE** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 14:15]

*The Beta Negative Binomial Distribution for the Analysis of Microbiome Data*

*La distribution bêta-binomiale négative pour l'analyse des données sur le microbiome*

Microbiome data is a high-dimensional count data that has gained considerable attention in recent years, both biologically and statistically. The data is typically noisy and exhibits complex structure. Appropriately modelling these data is important for a variety of applications as human microbiomes in particular, influences and is influenced by the health of the host. In this study, we compare the performance of the beta negative binomial distribution to model microbiome data in comparison with other discrete distributions. We assessed performance of alternative distributions using both simulated data as well as real data generated by studies investigating developmental origins of diseases.

Les données sur le microbiome sont des données de dénombrement de grandes dimensions qui ont suscité beaucoup d'intérêt biologique et statistique au cours des dernières années. Les données sont généralement bruyantes et présentent une structure complexe. Modéliser ces données de façon appropriée est important pour de nombreuses applications puisque le microbiome humain est influencé par – et influence – la santé de l'hôte. Dans cette étude, nous comparons l'efficacité de la distribution bêta-binomiale négative pour modéliser les données sur le microbiome, par rapport à d'autres distributions discrètes. Nous évaluons l'efficacité des distributions alternatives au moyen de données simulées ainsi que de données réelles provenant des études sur les origines développementales de maladies.



## 2D-C4: Statistical Genetics Génétique statistique

---

**JIHYUNG SHIN** (Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto), **SHELLEY B. BULL** (Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Dalla Lana School of Public Health, University of Toronto)

[Tuesday May 31 / mardi 31 mai, 14:30]

*On the Analysis of Low-Frequency Genetic Variants in the Presence of Quantitative Covariates Using Prospective and Retrospective Penalized Logistic Regression Models*

*À propos de l'analyse de variants génétiques rares en présence de covariables quantitatives utilisant des modèles prospectifs et rétrospectifs par régression logistique pénalisée*

In genetic association studies of binary outcomes, standard logistic regression analysis of low-frequency variants can violate large sample assumptions, yielding invalid conclusions about genetic effects. Firth's penalized logistic regression (PLR) is an alternative. We evaluate empirical estimation bias, LR test size and power of PLR under two analytic approaches: (i) prospective model with disease outcome as response variable and genotype as predictor; (ii) retrospective model with an additively constrained genotype as response variable and disease outcome as predictor. We recommend prospective PLR for analysis in the presence of a quantitative covariate and suggest that the retrospective model be applied cautiously.

Dans les études d'association génétique de phénotypes binaires, les modèles de régression logistique standards qui incluent des variants rares peuvent ne pas être conformes aux hypothèses de grand échantillon, ce qui rend les conclusions sur les effets génétiques invalides. La régression logistique pénalisée de Firth constitue une solution de rechange. Nous évaluons le biais d'estimation empirique, la taille du test du ratio de vraisemblance et la puissance de la régression logistique pénalisée dans le cadre de deux approches analytiques : i) un modèle prospectif présentant une maladie comme variable-réponse et un génotype comme prédicteur ; ii) un modèle rétrospectif présentant un génotype additivement contraint comme variable-réponse et une maladie comme prédicteur. Nous recommandons l'analyse prospective lorsque la covariable est quantitative, et nous suggérons que le modèle rétrospectif soit appliqué avec prudence.

## 2D-C5: Statistical Inference and Applications 1 Inférence statistique et applications 1

---

Chair/Président: Jeffrey Rosenthal (University of Toronto)

Room/Salle: Thistle 258

### Abstracts/Résumés

---

**MIKELIS BICKIS** (University of Saskatchewan)

[Tuesday May 31 / mardi 31 mai, 13:30]

*Statistical Relativity*

*La relativité en statistique*

Einstein's theory of relativity is based on the understanding that certain intrinsic properties should remain invariant under changes to the frame of reference. A statistical model is also described relative to a frame of reference. Meaningful inferences ought not to depend on this often arbitrary frame of reference. Nonetheless, commonly used statistical techniques such as principal components, model selection, and design optimality are affected by reparametrization. While this failing is implicitly acknowledged by practitioners it raises the question about whether such methods can be described in a way that gives meaning to their application

La théorie de la relativité d'Einstein est basée sur l'idée que certaines caractéristiques intrinsèques doivent rester invariantes lorsque le système de référence est modifié. Un modèle statistique est également décrit par rapport à un système de référence. Les inférences raisonnables ne doivent pas dépendre d'un système de référence souvent arbitraire. Néanmoins, des techniques statistiques d'usage courant, comme les composantes principales, la sélection des modèles, et l'optimalité des plans expérimentaux sont affectés par la paramétrisation. Quoique les praticiens reconnaissent ce défaut, il pose la question à savoir si on peut décrire de telles méthodes d'une façon qui donne un sens à leur application.

---

**EDIT GOMBAY** (University of Alberta)

[Tuesday May 31 / mardi 31 mai, 13:45]

*Change Detection Algorithms*

*Algorithmes de détection du changement*

``A contributed session in Honor of Professor Ian MacNeill."`

In this talk we give a brief review of the different type of change detection algorithms. Various models are considered and the common features and the common underlying theory are explained. Change detection is a field of statistics that is fast growing in importance. Its applications include Economics and Finance, climate change studies, clinical studies, just to name a few major areas.

« Une séance libre en l'honneur du professeur Ian MacNeill. »

Dans cet exposé, nous donnons un bref aperçu des différents types d'algorithmes de détection du changement. Différents modèles sont considérés et les caractéristiques et la théorie sous-jacente communes sont expliquées. La détection du changement est un champ de la statistique dont l'importance s'accroît rapidement. Ses applications comprennent l'économie et la finance, les études sur les changements climatiques et les études cliniques pour ne citer que plusieurs grands domaines.

## 2D-C5: Statistical Inference and Applications 1

### Inférence statistique et applications 1

---

---

**KEVIN MONGEON** (Brock University)

[Tuesday May 31 / mardi 31 mai, 14:00]

*The Process of Winning and Losing Sporting Contests*

*Le processus de gagner ou perdre des évènements sportifs*

Bayes Rule is used to derive an in-game probabilistic model of winning and losing sporting contests. We use Bayesian hierarchical regression models to estimate the unknown model parameters and iteratively sample the conditional posterior distributions to obtain the posterior expected win probabilities. We decompose the expected in-game win probabilities into their prior and conditional likelihood components to examine team behaviours under various game-state conditions.

Le théorème de Bayes est utilisé pour dériver un modèle probabiliste de gagner ou perdre des évènements sportifs lors du déroulement du match. Les paramètres inconnus du modèle sont estimés à partir de modèles de régression hiérarchique de Bayes et les probabilités de gagner escomptées sont obtenues par le biais d'un échantillonnage aléatoire des distributions conditionnelles a posteriori. Les probabilités escomptées de gagner sont décomposées en composantes a posteriori et conditionnelle pour examiner les comportements d'équipe sous diverses conditions lors du déroulement du match.

---

**SHAUN ZHENG SUN** (University of the Fraser Valley)

[Tuesday May 31 / mardi 31 mai, 14:15]

*Improved Tests for Independence in Ordinal Contingency Tables*

*Tests améliorés pour l'indépendance des tableaux de contingence ordinaux*

In a contingency table, the standard test for independence among variables is the Pearson's Chi-squared. When the variable(s) are ordinal, such as those in Likert scale survey data, the Chi-squared statistic does not take into account the natural orderings.

Dans un tableau de contingence, le test standard d'indépendance entre les variables est le khi-deux de Pearson. Lorsque la(les) variable(s) sont ordinales, telles que celles des données d'enquête selon une échelle de Likert, la statistique du khi-deux ne tient pas compte des ordres naturels.

In this talk, an ordinal sensitive test is proposed and large sample theory is given. Monte Carlo studies show that the new test is more powerful than some classical tests, including the Chi-squared and the likelihood ratio tests. An example from an education survey will be given. The test can be extended to higher dimensional tables.

Dans cet exposé, un test ordinal sensible est proposé et la théorie des grands échantillons est donnée. Des études de Monte-Carlo montrent que le nouveau test est plus puissant que certains tests classiques, y compris les tests du khi-deux et des ratios de vraisemblance. Un exemple à partir d'une enquête sur l'éducation sera présenté. Le test peut être étendu à des tableaux de dimensions supérieures.

---

**GARY J. UMPHREY** (University of Guelph)

[Tuesday May 31 / mardi 31 mai, 14:30]

*Aesthetics and Measurement: The Measurement of Aesthetics and the Aesthetics of Measurement*

*Esthétique et mesure : la mesure de l'esthétique et l'esthétique de la mesure*

An object (such as an artwork, a piece of music, a mineral specimen, or an ant) that is perceived as having aesthetic qualities is referred to here as an "aesthetic object". Investigating what

Un objet (œuvre d'art, pièce musicale, spécimen minéralogique ou fourmi) perçu comme ayant des qualités esthétiques sera appelé dans le présent contexte « objet esthétique ». L'analyse et l'explication des propriétés d'un objet

## 2D-C5: Statistical Inference and Applications 1

### Inférence statistique et applications 1

---

properties contribute to the perception of an object as aesthetic and why is often facilitated by taking measurements. But are measurements themselves aesthetic objects? This talk will explore the idea that measurements – particularly collections of measurements we call data sets – have aesthetic qualities. A question is posed: Does the act of analyzing data cast the statistician in dual roles of data scientist and data artist?

qui le font percevoir comme esthétique sont souvent simplifiées en faisant appel à des mesures. Ces mesures sont-elles aussi des objets esthétiques? Cet exposé porte sur l'idée que les mesures – en particulier les séries de mesures que nous appelons ensembles de données – ont elles-mêmes des qualités esthétiques. Une question se pose : Par la simple analyse de données, le statisticien remplit-il à la fois les rôles de scientifique et d'artiste des données?

---

**XIKUI WANG, YOU LIANG & LYSA PORTH** (University of Manitoba)

[Tuesday May 31 / mardi 31 mai, 14:45]

*Dynamic Risk Measures and Agricultural Insurance*

*Mesures de risque dynamiques et assurance agricole*

We discuss important ideas of dynamic risk measures and the use of Markov decision processes to characterize and derive such measures. We also discuss how such measures may be used in the study of agricultural insurance.

Nous discutons des idées importantes de mesures de risque dynamiques et de l'utilisation des processus de décision de Markov pour caractériser et dériver de telles mesures. Nous discutons aussi comment ces mesures peuvent être utilisées dans l'étude de l'assurance agricole.

## 2E-I1: Adaption and Approximation in Markov Chain Monte Carlo

### Adaptation et approximation pour la méthode de Monte-Carlo par chaînes de Markov

---

Organizer and Chair / Responsable et président: Aaron Smith (University of Ottawa)

Room/Salle: Thistle 242

#### Abstracts/Résumés

---

**MYLÈNE BÉDARD** (Université de Montréal)

[Tuesday May 31 / mardi 31 mai, 15:30]

*Hierarchical Models: Local Proposal Variances for the RWM-within-Gibbs*

*Modèles hiérarchiques: variances instrumentales locales pour le RWM-dans-Gibbs*

We study the performance of RWM-within-Gibbs algorithms for sampling from hierarchical models. Using existing scaling analyses, we develop asymptotically optimal tunings for that sampler. This leads to locally optimal proposal variances that depend on the mixing components of the hierarchical model and that correspond to the classical asymptotically optimal acceptance rate of 0.234. Ignoring the local character of the optimal scaling leads to an optimal proposal variance that remains fixed for the duration of the algorithm; the corresponding asymptotically optimal acceptance rate is then lower than 0.234. We provide results for location and scale hierarchies, and illustrate the findings through numerical studies. We compare these local and constant approaches to RWM with diagonal covariance matrix and Adaptive Metropolis samplers.

Nous étudions la performance d'un algorithme RWM-dans-Gibbs pour échantillonner de modèles hiérarchiques. En nous basant sur des analyses existantes, nous obtenons des variances instrumentales locales et optimales pour cet algorithme. Celles-ci dépendent du paramètre de mélange dans le modèle hiérarchique, et mènent au taux d'acceptation optimal classique de 0.234. Il est également possible d'obtenir une variance instrumentale optimale et constante pour la durée de l'algorithme; le taux d'acceptation asymptotiquement optimal est alors inférieur à 0.234. Nous fournissons des résultats dans le cadre de modèles hiérarchiques avec paramètres de position et/ou d'échelle, et illustrons nos résultats à l'aide d'études de simulation. Nous comparons les approches locale et constante aux algorithmes RWM (avec matrice de covariance instrumentale diagonale) et Metropolis adaptatif.

---

**DANIEL JERISON** (Cornell University)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Honest MCMC Convergence Guarantees*

*Garanties honnêtes de convergence MCMC*

Is MCMC estimation as trustworthy as sampling directly from the target probability distribution (if that were feasible)? Usually not: Monte Carlo standard errors, which purport to measure the uncertainty introduced by the Markov chain, are asymptotically valid but provide no finite-time guarantees. The few nonasymptotic results are difficult to apply in practice. I will discuss new MCMC estimation theorems for Markov chains with a regenerative structure. These theorems give accuracy guarantees of

L'estimation MCMC est-elle aussi fiable que le serait un échantillonnage direct de la distribution de probabilité cible (si cela était possible)? Généralement pas: les écarts-types Monte-Carlo, qui prétendent mesurer l'incertitude introduite par la chaîne de Markov, sont asymptotiquement valides mais n'offrent aucune garantie en temps fini. Les quelques résultats non asymptotiques sont difficiles à appliquer en pratique. Je discuterai de nouveaux théorèmes d'estimation MCMC pour les chaînes de Markov avec une structure régénératrice. Ces théorèmes offrent des garanties d'exactitude

## 2E-I1: Adaption and Approximation in Markov Chain Monte Carlo

### Adaptation et approximation pour la méthode de Monte-Carlo par chaînes de Markov

---

the same type that sampling directly from the target distribution would provide. I will illustrate the results using a Gibbs sampler for a Bayesian hierarchical model.

du même type qu'offrirait un échantillonnage direct de la distribution cible. J'illustrerai les résultats à l'aide d'un échantillonneur de Gibbs pour un modèle hiérarchique bayésien.

---

**MATTI VIHOLA** (University of Jyväskylä)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Unbiased Estimators and Multilevel Monte Carlo*

*Estimateurs non biaisés et Monte-Carlo multiniveaux*

Multilevel Monte Carlo (MLMC) and recently proposed debiasing schemes are closely related methods which can be applied in scenarios where exact simulation methods are difficult to implement, but biased estimators are easily available. An important example of such a scenario is the inference with continuous-time diffusion processes, where the process is difficult to simulate exactly but time-discretized approximations are available. I will present a new general class of unbiased estimators which admits earlier debiasing schemes as special cases, and new lower variance estimators which behave asymptotically like MLMC, both in terms of variance and cost, under general conditions. This suggests that bias can often be eliminated entirely with arbitrarily small extra cost. (arXiv:1512.01022)

Les méthodes Monte-Carlo multiniveaux (MLMC) et les récentes techniques de réduction du biais sont des méthodes étroitement liées applicables à des scénarios dans lesquels il est difficile de mettre en œuvre une méthode de simulation exacte, mais des estimateurs biaisés sont disponibles. Un bon exemple : une inférence avec des processus de diffusion en temps continu, scénario dans lequel il est difficile de simuler le processus de manière exacte, mais il existe des approximations en temps discrétisé. Je présenterai une nouvelle classe générale d'estimateurs non biaisés qui permet certaines techniques de réduction du biais à titre de cas spéciaux, ainsi que de nouveaux estimateurs à variance réduite qui présentent un comportement asymptotique similaire au MLMC (en ce qui concerne la variance et le coût) dans des conditions générales. Cela suggère que le biais peut souvent être entièrement éliminé pour un très léger surcoût. (arXiv :1512.01022)

## 2E-I2: Building a Research Career: From New Investigator to Present Comment bâtir une carrière en recherche : de nouveau chercheur à aujourd'hui

---

Organizer and Chair / Responsable et président: Martin Lysy (University of Waterloo)

Room/Salle: Welch - David Howes

### Abstracts/Résumés

---

**JAMES RAMSAY** (McGill University)

[Tuesday May 31 / mardi 31 mai, 15:30]

*The Art of the Awkward Question (and Some Psychology, Too)*

*De l'art de la question délicate (et de la psychologie)*

Sciences advances by questions, not answers, and a recent biography of James Clerk Maxwell makes this case superbly. This talk will focus on the awkward questions that shaped my career, including an early one in the Princeton bookstore that still haunts me.

But accolades in science are given for answers, and those who ask awkward questions can expect consequences for their relationships with their colleagues. And perhaps even for their careers. One way out is to avoid asking questions, but there are better ways, most of which I have yet to discover.

Les sciences avancent grâce aux questions plutôt qu'aux réponses, ainsi que le montre brillamment une récente biographie de James Clerk Maxwell. Cette présentation mettra l'accent sur les questions délicates qui ont influencé ma carrière, dont une – il y a bien longtemps – dans une librairie de Princeton – qui me hante encore.

Mais, en sciences, ce sont les réponses qui sont félicitées – et ceux qui posent des questions délicates peuvent souvent s'attendre à des conséquences en ce qui concerne les relations avec leurs collègues, voire leur carrière. Une solution est d'éviter de poser des questions, mais il en existe assurément de meilleures, dont certaines que je n'ai pas encore découvertes.

---

**JEFFREY ROSENTHAL** (University of Toronto)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Lessons From a Twisted Career Path*

*Leçons d'un chemin de carrière tordu*

In this talk, I will reflect upon the unique twists and turns of my own academic career path, which took me from enthusiastic pure-math undergrad, to worried Harvard PhD student, to struggling junior probability professor, to ultimately receiving the COPSS Presidents' Award for outstanding contributions to statistics. I will try to use my story to provide lessons and insights for budding statistical scholars and researchers.

Dans cette présentation, je méditerai sur les tours et détours de mon propre chemin de carrière universitaire, de l'étudiant enthousiaste en mathématiques pures, au doctorant inquiet de Harvard, au professeur adjoint en probabilité en difficulté, au fier récipiendaire du prix COPSS pour mes contributions à la statistique. Je tâcherai de tirer de mon histoire des leçons et des conclusions pour les jeunes scientifiques et chercheurs en statistique d'aujourd'hui.

---

**MARY THOMPSON** (University of Waterloo)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Early Career Research Back Then and Now*

*Recherches de début de carrière : hier et aujourd'hui*

## **2E-I2: Building a Research Career: From New Investigator to Present**

### **Comment bâtir une carrière en recherche : de nouveau chercheur à aujourd'hui**

---

In many ways the research culture has changed radically from the early 1970s, when reading was done in the library, papers were handwritten for a typist, and manuscripts were submitted through the post. In fact in some ways it is quite a challenge to find the commonalities between then and now. I will discuss how the environment at the University of Waterloo influenced the course of my research career, following graduate studies at a US university. For a beginning researcher today it is certainly necessary to employ different strategies, but the fundamentals would be the same: follow what is fascinating, and find one's collaborative strengths.

À de nombreux égards, la culture de recherche a radicalement changé depuis le début des années 1970, où on lisait en bibliothèque, on rédigeait ses articles à la main avant de les faire dactylographier et on soumettait ses manuscrits par la poste. En fait, il est assez difficile de trouver des points communs entre cette époque et la nôtre. Je discuterai de la façon dont l'environnement de Waterloo a influencé la trajectoire de ma carrière en recherche, après des études de cycle supérieur aux États-Unis. Un jeune chercheur d'aujourd'hui emploiera certainement d'autres stratégies, mais les principes de base n'ont pas changé : suivez ce qui vous fascine et découvrez vos talents de collaborateur.



## 2E-I3: Design and Methodological Issues in Health Studies of Vulnerable Populations Problèmes de conception et de méthodologie des études sur la santé des populations vulnérables

---

Chair/Président: John Petkau (University of British Columbia)  
Organizer/Responsable: Lehana Thabane (McMaster University)

**Room/Salle: Thistle 247**

### Abstracts/Résumés

---

**LISA LIX** (Ottawa University of Manitoba)

[Tuesday May 31 / mardi 31 mai, 15:30]

*Methodological Challenges in Aboriginal Health Research*

*Problèmes méthodologiques en recherche en santé autochtone*

Aboriginal populations in Canada, which include First Nations, Inuit, and Metis people, often have poorer health outcomes and a greater frequency of health risk behaviors than non-Aboriginal populations. Comparative information about the health and healthcare use of Aboriginal and non-Aboriginal populations is therefore important in the design and delivery of healthcare services. Misclassification bias in ascertaining ethnicity, small sample sizes, incomplete data about on-reserve populations, and differential item functioning of self-reported health status measures are challenges that can affect the validity of comparative studies. This talk will describe the opportunities for biostatisticians to contribute to high-quality Aboriginal health research in Canada, by addressing these methodological challenges while remaining sensitive to data ownership and control principles and information privacy legislation.

Les populations autochtones du Canada, qui incluent les Premières Nations, les Inuit et les Métis, présentent souvent des résultats de santé moins bons et une plus grande fréquence de comportements à risque que les populations non autochtones. Il est donc important de disposer d'informations comparatives sur la santé et l'utilisation des soins de santé par ces deux catégories de populations pour concevoir et fournir des services de santé adaptés. Le biais d'erreur de classification dans la détermination de l'ethnicité, la petite taille des échantillons, les données incomplètes sur les populations des réserves et le fonctionnement différentiel des items sur l'état de santé auto-déclaré sont autant de défis susceptibles d'affecter la validité des études comparatives. Je décrirai les possibilités pour les biostatisticiens de contribuer à une recherche en santé autochtone de qualité au Canada en relevant ces défis méthodologiques tout en demeurant sensibles aux principes de propriété et de contrôle des données et à la législation sur la protection des renseignements personnels.

---

**ROBERT W. PLATT** (McGill University)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Randomized Trials in Obstetric Populations*

*Essais randomisés en milieu obstétrique*

Pregnancy is a vulnerable period, in which both mother and fetus are at risk for benefit and harm. I discuss three problems arising in studies during pregnancy: 1) Treatments can have opposite effects for the fetus and the mother, and some important outcomes are very rare. Composite outcomes are a solution, but raise other problems. 2) Timing during pregnancy of an intervention (e.g., induction

La grossesse est une période vulnérable pendant laquelle mère et enfant à naître courent certains risques. Je discuterai de trois problèmes qui se posent dans les études menées pendant la grossesse : 1) Les traitements peuvent avoir des effets opposés pour le fœtus et la mère et certains événements importants sont très rares. Les résultats composites constituent une solution, mais ils soulèvent d'autres problèmes. 2) Le moment d'une intervention pendant la gros-

## 2E-I3: Design and Methodological Issues in Health Studies of Vulnerable Populations Problèmes de conception et de méthodologie des études sur la santé des populations vulnérables

---

of pregnancy) is a common research interest. This raises statistical and logistical challenges; the optimal statistical design can be at odds with feasible care plans. 3) Pregnancy is a highly dynamic process, in which the condition of the mother and fetus can change rapidly. Consideration must be given to timing of recruitment and randomization.

sesse (p. ex., déclenchement) est un domaine d'intérêt commun en recherche. Ceci soulève des problèmes statistiques et logistiques ; le plan statistique optimal ne cadre pas toujours avec un plan de soins réaliste. 3) La grossesse est un processus très dynamique, dans lequel l'état de la mère et du fœtus peut évoluer rapidement. Il faut prendre en considération le moment de recrutement et de randomisation.

---

**LEHANA THABANE** (McMaster University)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Planning for the Twilight Years: Why We Need More Statisticians Involved in Trials in the Elderly Living in Long-Term Care Settings*

*Planification du crépuscule de la vie: pourquoi il nous faut plus de statisticiens dans les essais menés auprès des personnes âgées vivant en centres de soins de longue durée*

Providing optimal healthcare of the elderly living in long-term care (LTC) settings requires deliberate planning and good evidence to support decision-making. Conducting trials in LTC to generate evidence or apply evidence in practice can present many design, implementation, methodological, ethical and analytical challenges. First, the population is quite frail which can present serious challenges; second, the environment is hard to work in because of the overcommitted staff; third, outcome assessment can be challenging; fourth, determining the appropriate unit of randomization, analysis or inference can be complicated. In this presentation, I will use experiences from some of our studies to illustrate the issues and highlight the role of biostatisticians in addressing them.

Pour fournir des soins de santé optimaux aux personnes âgées vivant en centres de soins de longue durée (SLD), il faut une planification délibérée et de bonnes données à l'appui de la prise de décisions. Or l'organisation d'essais sur les SLD pour générer des données et la mise en pratique de ces données peut présenter de nombreux défis de conception, de mise en application, d'éthique et d'analyse. Tout d'abord, il s'agit d'une population fragile, ce qui peut poser de graves défis ; deuxièmement, l'environnement est difficile à aborder en raison de la surcharge en travail du personnel ; troisièmement, l'évaluation des issues peut être difficile ; quatrièmement, il peut être compliqué de déterminer l'unité de randomisation, d'analyse ou d'inférence appropriée. Dans cette présentation, j'utiliserai des exemples de nos études pour illustrer les problèmes et souligner le rôle que peuvent jouer les biostatisticiens pour y répondre.

## 2E-I4: Inference and Prediction from Distributed Data Inférence et prévisions à partir de données distribuées

---

Organizer and Chair / Responsable et président: Jean-Francois Plante (HEC Montreal)

**Room/Salle: Thistle 258**

### Abstracts/Résumés

---

**DOINA CARAGEA, LI HONGMIN, SOPOVA OLEKSANDRA & HERNDON NIC** (Kansas State University)

[Tuesday May 31 / mardi 31 mai, 15:30]

*Learning Domain Adaptation Classifiers from Multiple Distributed Sources*

*Classificateurs d'adaptation de domaines d'apprentissage à partir de sources multiples distribuées*

For many inference and classification problems, the amount of labeled data available is limited, making it impossible to learn accurate supervised classifiers. To address this challenge, we propose a domain adaptation approach that leverages labeled data from multiple distributed source domains to learn classifiers for a target domain. Our proposed approach, which is based on expectation maximization and the naïve Bayes classifier, identifies information that needs to be extracted from source domains and transferred to the target domain. This information includes relevant features and instances in each source, sufficient statistics to be transferred to the target, as well as importance weights for each source. We evaluate our approach in the context of disaster management and response based on microblogging data.

Pour plusieurs problèmes d'inférence et de classification, la quantité de données étiquetées disponible est limitée, ce qui rend impossible d'apprendre des classificateurs supervisés précis. Pour remédier à ce problème, nous proposons une approche d'adaptation de domaine qui tire profit de données étiquetées provenant de multiples domaines sources distribués pour apprendre les classificateurs pour un domaine cible. Notre approche, qui se base sur la maximisation de l'espérance et sur le classificateur naïf bayésien, identifie l'information qui doit être extraite des domaines sources et transférée au domaine cible. Cette information inclut les caractéristiques et les cas pertinents de chaque source, les statistiques exhaustives à être transférées à la cible, ainsi que des pondérations importantes de chaque source. Nous évaluons notre approche dans le contexte de la gestion et de la réponse en cas de catastrophe sur la base de données de microblogging.

---

**GEORGE OSTROUCHOV** (Oak Ridge National Laboratory & University of Tennessee)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Statistical Computing with R on Distributed and Multicore Platforms*

*Calcul statistique avec R sur des plateformes distribuées et multicœur*

The Programming with Big Data in R (pbdR) project (see pbdR.org) aims to teach and simplify parallel programming in R and to harness parallel numerical libraries relevant to statistical computing. In this talk, I will give a very brief overview of parallel computing platforms and introduce the pbdR project. I will give two examples. One, where parallelism is introduced directly in R and another, where parallel numerical linear algebra libraries are engaged.

Le projet de programmation avec mégadonnées dans le logiciel R (voir pbdR.org) vise à enseigner et à simplifier la programmation parallèle dans R ainsi qu'à exploiter les bibliothèques numériques pertinentes pour les calculs statistiques. Dans cet exposé, je donnerai un aperçu des plateformes de parallélisme et je présenterai le projet de programmation avec des mégadonnées dans R. Je donnerai deux exemples : l'un où le parallélisme est effectué directement dans R et l'autre où les bibliothèques numériques d'algèbre linéaire sont utilisées.

## 2E-I4: Inference and Prediction from Distributed Data Inférence et prévisions à partir de données distribuées

---

**RUIBIN XI** (Peking University), **NAN LIN** (Washington University in St. Louis)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Statistical Aggregation for Big Data Analysis*

*Agrégation statistique pour l'analyse de données volumineuses*

Big data provides great new opportunities for many fields. Deep analysis of big data may reveal important unknown phenomena. At the same time, big data also presents new statistical and computational challenges. In this talk, I will discuss the statistical aggregation technique that we developed in the last few years to perform statistical analysis in big data setting. This technique is based on the divide-and-conquer strategy. We first divide the big data to smaller subsets, perform statistical analysis on each subsets and aggregate the analysis results of each subsets to get a solution for the entire data set. Distributed computation of the resulted estimators is straightforward. We also show that the new estimators are statistically equivalent to the original estimator.

Les données volumineuses offrent de nouvelles possibilités pour de nombreux domaines. Une analyse en profondeur de ces données pourrait révéler un phénomène important inconnu. De plus, les données volumineuses présentent également des défis statistiques et informatiques. Dans cet exposé, je parlerai de la technique d'agrégation que nous avons mise au point au cours des dernières années pour effectuer une analyse statistique dans un contexte de données volumineuses. Cette technique est basée sur la stratégie de diviser pour régner. Nous divisons tout d'abord les données volumineuses en sous-ensembles plus petits. Ensuite, nous effectuons une analyse statistique de chaque sous-ensemble et nous agrégeons les résultats de l'analyse pour chaque sous-ensemble afin d'obtenir une solution du jeu entier des données. Le calcul distribué des estimateurs qui en résultent est direct. Nous démontrons également que les nouveaux estimateurs sont statistiquement équivalents à l'estimateur original.

## 2E-I5: Statistical Challenges in Large-Scale Inference Défis statistiques en inférence à grande échelle

---

Organizer and Chair / Responsable et président: Kun Liang (University of Waterloo)

Room/Salle: Thistle 246

### Abstracts/Résumés

---

**CHUNMING ZHANG** (University of Wisconsin-Madison)

[Tuesday May 31 / mardi 31 mai, 15:30]

*Single-Index Modulated Multiple Testing*  
*Tests multiples modulés à indice unique*

We present a SIM multiple testing procedure, by assuming the availability of a bivariate p-value,  $(p_1, p_2)$ , for each hypothesis. To find the optimal rejection region for the bivariate p-value, we propose a criteria based on the ratio of probability density functions of  $(p_1, p_2)$  under the true null and non-null. This criteria in the bivariate normal setting motivates us to project the bivariate p-value to a single-index,  $p(\theta)$ , for a wide range of directions  $\theta$ . The true null distribution of  $p(\theta)$  is estimated via parametric and nonparametric approaches. To derive the optimal projection direction  $\theta$ , we propose a new approach, which is shown to be consistent under some mild conditions. Simulation evaluations and analysis of a real dataset will be illustrated.

Nous présentons une procédure de tests multiples SIM en supposant que la valeur p bidimensionnelle,  $(p_1, p_2)$ , est disponible pour chaque hypothèse. Pour trouver la région de rejet optimale pour la valeur p bidimensionnelle, nous proposons un critère basé sur le rapport des fonctions de fréquences de  $(p_1, p_2)$  sous la vraie nulle et non nulle. Ce critère dans le cadre normal bidimensionnel nous motive à projeter la valeur p bidimensionnelle sur un indice unique,  $p(\theta)$ , pour un large éventail de directions  $\theta$ . La vraie loi nulle de  $p(\theta)$  est estimée à l'aide d'approches paramétriques et non paramétriques. Pour dériver la direction de projection optimale  $\theta$ , nous proposons une nouvelle approche, qui est convergente sous de légères conditions. Des études de simulation et l'analyse d'un ensemble de données réelles seront illustrées.

---

**WENGUANG SUN** (University of Southern California), **TONY CAI** (University of Pennsylvania), **WEINAN WANG** (University of Southern California)

[Tuesday May 31 / mardi 31 mai, 16:00]

*CARS: Covariate Assisted Ranking and Screening for Large-Scale Two-Sample Inference*

*CARS: Classement et triage assistés de covariables pour inférence à deux échantillons à grande échelle*

The conventional two-sample inference framework often leads to suboptimal multiple testing procedures due to the loss of information in the data reduction step. This article studies how to construct an auxiliary variable from the original data matrix and incorporate the variable in the inference procedure to improve the power in multiple testing. We study the problem in a decision-theoretic framework and develop oracle and data-driven procedures for false discovery rate (FDR) control. The proposed oracle procedure employs a covariate-assisted ranking and screening (CARS) scheme,

Le cadre conventionnel de l'inférence à deux échantillons mène souvent à des procédures de tests multiples sous-optimales en raison de la perte d'information à l'étape de la réduction de données. Cet article étudie la façon de construire une variable auxiliaire à partir de la matrice de données originales et incorpore la variable dans la procédure d'inférence pour améliorer la puissance dans les tests multiples. Nous étudions le problème dans un cadre de décision théorique et nous développons une procédure oracle ainsi qu'une procédure fondée sur les données pour le contrôle du taux de fausses découvertes (FDR). La méthode oracle proposée utilise un schéma de classement et

## 2E-I5: Statistical Challenges in Large-Scale Inference Défis statistiques en inférence à grande échelle

---

which is shown to be optimal in the sense that it has the smallest missed discovery rate among all valid FDR procedures. We also develop a data-driven procedure and establish its asymptotic properties.

de triage assistés de covariables (CARS) qui s'avère être optimal en ce sens qu'il a le plus petit taux de fausses découvertes parmi toutes les procédures valides FDR. Nous développons aussi une procédure fondée sur les données et établissons ses propriétés asymptotiques.

---

**RUTH HELLER** (Tel-Aviv University), **MARINA BOGOMOLOV** (Technion), **DANIEL YEKUTIELI** (Tel-Aviv University)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Assessing Replicability Across High-Dimensional Studies*

*Évaluer la reproductibilité entre les études à grande dimension*

There is need in modern applications for procedures to identify features with two (or more) findings. For example, in GWAS, there is interest in generalization, where the aim is to identify the SNPs that are associated with a phenotype in more than one population, as well as in identifying the loci with pleiotropic effects. We propose two formal statistical approaches for this purpose. In the empirical Bayes approach, we estimate the optimal rejection rule for discovering replicated findings. In the frequentist approach, we first select the promising features for each type of finding (e.g., study) separately, and then we test for replicability only the features that were selected for both types of findings (e.g., studies).

Il existe un besoin dans les applications modernes pour des procédures qui identifient les caractéristiques dans deux résultats (ou plus). Par exemple, dans GWAS, il y a un intérêt dans la généralisation lorsque le but est d'identifier les SNP associés à un phénotype dans plus d'une population, ainsi que dans l'identification des locus avec des effets pléiotropes.

À cet effet, nous proposons deux approches statistiques formelles. Dans l'approche empirique de Bayes, nous estimons la règle de rejet optimal pour découvrir les résultats répliqués. Dans l'approche fréquentiste, nous choisissons d'abord les caractéristiques prometteuses pour chaque type de résultat (par exemple, étude) séparément, puis nous testons pour la reproductibilité uniquement les caractéristiques qui ont été sélectionnées pour les deux types de résultats (par exemple, études).

## 2E-I6: Survey Methodology in a Big Data World Méthodes d'enquête dans un monde de mégadonnées

---

Organizer and Chair / Responsable et présidente: Nancy Reid (University of Toronto)

Room/Salle: Academic South 217

### Abstracts/Résumés

---

**ANNE DE MOLINER** (Université de Montréal), **HERVÉ CARDOT & CAMELIA GOGA** (Université de Bourgogne)  
[Tuesday May 31 / mardi 31 mai, 15:30]

*Estimating a Mean Electricity Consumption Curve in the Presence of Missing Data Using Survey Sampling Techniques*

*Estimation de courbe moyenne de consommation électrique par sondage en présence de valeurs manquantes*

In the near future, millions of electricity load curves of French households measured at a fine scale will be available. All these collected load curves represent a huge amount of information difficult to store due to technical and budgetary constraints. In these situations, survey sampling techniques are attractive alternatives to signal compression to estimate mean consumption curves for different groups of clients. Unfortunately, data collection may undergo technical problems resulting in missing values. This problem reduces the accuracy of the estimators and may generate bias. Different approaches can be adapted to deal with missing data in this functional framework: nearest neighbor imputation, kernel smoothing or linear interpolation to the differences around the mean. We compare these methods on real data.

Dans un futur proche, des millions de courbes de consommation électrique de ménages français, mesurées à un pas de temps fin seront disponibles. Ces données constitueront une masse d'information considérable qui pourrait être exploitée grâce à des techniques d'échantillonnage pour estimer des courbes de consommation moyennes à différents périmètres. Malheureusement, des aléas techniques pourraient survenir, générant des valeurs manquantes qui risqueraient de détériorer la précision des estimateurs, voire de créer des biais. Différentes méthodes d'estimation en présence de valeurs manquantes pourraient être adaptées à notre contexte de données fonctionnelles : l'imputation par le plus proche voisin, le lissage par noyau ou encore l'interpolation linéaire des différences à la moyenne. Nous comparons ces approches sur des données réelles.

---

**JOHN ELTINGE** (U.S. Bureau of Labor Statistics)

[Tuesday May 31 / mardi 31 mai, 16:15]

*Characterization, Modeling and Management of Data Quality, Risk and Cost in the Integration of Surveys with Alternative Data Sources*

*Caractérisation, modélisation et gestion de la qualité, du risque et du coût de données dans l'intégration des enquêtes avec sources de données alternatives*

Government statistical agencies have missions centered on providing the public with high-quality information on a sustainable and cost-effective basis. This requires the agencies to use practical design tools that allow them to characterize, model and manage multiple dimensions of data quality, risk and cost that are inherent in statistical production processes. Historically, those tools have cen-

Les organismes de statistique gouvernementaux ont pour mission de fournir au public de l'information de haute qualité sur une base durable et rentable. Ceci exige que les organismes utilisent des outils pratiques qui leur permettent de caractériser, de modéliser et de gérer plusieurs dimensions de qualité, de risque et de coût de données qui sont inhérentes aux processus de production de statistiques. Historiquement, ces outils étaient centrés sur les enquêtes par son-

## 2E-I6: Survey Methodology in a Big Data World Méthodes d'enquête dans un monde de mégadonnées

---

tered on sample surveys and some administrative record systems. However, agencies now need to extend these tools to cover a broad suite of alternative data sources (sometimes called "big data," "non-designed data" or "organic data"). This paper develops a framework for such tools, with special emphasis on (1) the level of precision required for a given design decision; and (2) use of available empirical information.

dage et sur quelques systèmes de dossiers administratifs. Par contre, ces organismes ont maintenant besoin d'élargir ces outils pour couvrir un large éventail de sources de données alternatives (quelques fois appelées «données volumineuses», «données non-planifiées» ou «données organiques»). Cet article développe un cadre pour de tels outils, avec un accent particulier sur (1) le niveau de précision requis pour une décision de conception donnée ; et (2) l'utilisation de l'information empirique disponible.



## 2E-C1: Biostatistics: New Inferential Strategies Biostatistique : nouvelles stratégies d'inférence

---

Chair/Président: Karen Kopciuk (University of Calgary)

Room/Salle: Thistle 255

### Abstracts/Résumés

---

**CHRISTOPHER GRAVEL & ROBERT W. PLATT** (McGill University)

[Tuesday May 31 / mardi 31 mai, 15:30]

*Inverse Probability Weighted Estimation in Confounded Binary Data with Outcome Misclassification*

*Estimation pondérée selon la probabilité inverse pour les données binaires confondues avec classification erronée du résultat*

We consider a study with a binary exposure and outcome, subject to confounding and misclassification of the outcome variable. We use inverse probability weighting and internal validation sampling to rebalance covariates across treatment groups while mitigating misclassification bias. We discuss several validation sampling schemes and a Monte Carlo approach to approximate optimal sample size determination. A parametric bootstrap is used for variance estimation. We explore finite sample properties of the weighted estimators via simulation, with particular attention to relative efficiency of different sampling schemes for validation. We demonstrate the use of the methods through an example using administrative data.

Nous considérons une étude avec une exposition et un résultat binaires, sous réserve de confusion et de classification erronée de la variable de résultat. Nous utilisons la pondération selon la probabilité inverse et un échantillonnage interne de validation pour rééquilibrer les covariables entre les groupes de traitement tout en atténuant le biais attribuable à la classification erronée. Nous discutons de plusieurs plans d'échantillonnage de validation et d'une approche de Monte-Carlo pour approximer la détermination de la taille d'échantillon optimale. Un bootstrap paramétrique est utilisé pour l'estimation de la variance. Nous explorons les propriétés des échantillons finis des estimateurs pondérés à l'aide de simulations, avec une attention particulière pour l'efficacité relative des différents plans d'échantillonnage pour la validation. Nous démontrons l'utilisation des méthodes à l'aide d'un exemple qui utilise des données administratives.

---

**QIUGUANG SANG, GUOHUA YAN & RENJUN MA** (University of New Brunswick)

[Tuesday May 31 / mardi 31 mai, 15:45]

*Bayesian Compound Poisson Mixed Models for Longitudinal Semi-Continuous Data with Non-Ignorable Missingness*

*Modèles composés mixtes bayésiens de Poisson pour des données longitudinales semi-continues manquantes de façon non-ignorable*

Longitudinal semi-continuous data, which usually contain a fair amount of zeros, is common in biomedical, econometric and health studies. The presence of missing values creates additional difficulty in data analysis. In this talk, we propose a Tweedie compound Poisson mixed model to incorporate positive observations and zeros in an integral way. We use a selection model for missing data. A simulation study is carried out to evaluate the performance of the proposed method. The

Les données longitudinales semi-continues, qui contiennent généralement un bon nombre de zéros, sont courantes dans les études biomédicales, économétriques et sur la santé. La présence de valeurs manquantes pose des difficultés supplémentaires dans l'analyse de données. Dans cet exposé, nous proposons un modèle composé mixte de Poisson de Tweedie afin d'incorporer de façon intégrative des observations positives et des zéros. Nous utilisons un modèle de sélection pour données manquantes. Nous effectuons une étude de simulation pour évaluer l'efficacité de la méthode

## 2E-C1: Biostatistics: New Inferential Strategies Biostatistique : nouvelles stratégies d'inférence

---

method is also illustrated through the reanalysis of a fluoride intake dataset.

proposée et l'illustrons par une nouvelle analyse d'un jeu de données sur l'apport en fluorure.

---

**FARHAD SHOKOOHI, ABBAS KHALILI & MASOUD ASGHARIAN** (McGill University), **SHILI LIN** (The Ohio State University)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Feature Selection in High-Dimensional Heterogeneous Time-to-Event Data; A Study on Ovarian Cancer*

*Sélection de traits pour des données de temps d'événement hétérogènes et de grande dimension; étude sur le cancer de l'ovaire*

Ovarian cancer is among the leading causes of cancer deaths. DNA methylation may be part of a genetic signature for survival time. It is of interest to study how recurrence time is related to methylations of gene promoters. A study on relationship between genes and recurrence time of ovarian cancer is carried out. The observations are subject to right censoring and indicate signs of heterogeneity. We consider a finite mixture of accelerated failure time model. We apply screening and penalization methods. Properties of the proposed method are studied theoretically and via simulation. The results of data analysis are discussed.

Le cancer de l'ovaire est parmi les causes principales de décès par cancer. La méthylation de l'ADN peut faire partie d'une signature génétique pour le temps de survie. Il est pertinent d'étudier comment la récurrence de la maladie est reliée à la méthylation des promoteurs de gènes. Nous avons mené une étude sur la relation entre les gènes et le temps de récurrence du cancer de l'ovaire. Les observations sont censurées à droite et présentent des signes d'hétérogénéité. Nous considérons un mélange fini du modèle de temps de défaillance accéléré et nous appliquons des méthodes de dépistage et de pénalisation. Les propriétés de la méthode proposée sont étudiées de façon théorique ainsi que par des simulations. Nous présentons aussi les résultats de l'analyse de données.

---

**MENG ZHAO & RUSSELL STEELE** (McGill University), **IAN SHRIER** (Lady Davis Institute, Jewish General Hospital and McGill University)

[Tuesday May 31 / mardi 31 mai, 16:15]

*Application of the Trend-Renewal Process to Recurrent Sports Injury*

*Application du processus de tendance-renouvellement (PTR) aux blessures sportives récurrentes*

The two classical frameworks in intensity-based modelling of recurrent events are the renewal process and the Poisson process, respectively describing "perfect repair" and "minimal repair" after each failure. When modelling recurrent sports injury, these two extremes are often inadequate. The trend-renewal process, developed by B.H. Lindqvist, is a mathematically elegant compromise between the two and includes them as special cases. Until recently, it has only been applied in reliability. We discuss a new practical interpretation of the parameters in a parametric TRP and evaluate its finite sample inference, focusing on its relevance in modelling recurrent sports injury.

Les deux cadres classiques de la modélisation basée sur l'intensité des événements récurrents sont le processus de renouvellement et le processus de Poisson, chacun décrivant respectivement la « réparation intégrale » et la « réparation minimale » après chaque défaillance. Ces deux extrêmes sont souvent inadéquats pour la modélisation des blessures sportives récurrentes. Mis au point par B.H. Lindqvist, le processus de tendance-renouvellement constitue un élégant compromis mathématique entre les deux et les inclut comme cas particuliers. Jusqu'à tout récemment, il n'a été appliqué qu'en matière de fiabilité. Nous traitons d'une nouvelle interprétation pratique des paramètres dans un PTR paramétrique et en évaluons l'inférence en échantillon fini, mettant l'accent sur sa pertinence pour la modélisation des blessures sportives récurrentes.

## 2E-C2: Probability and Estimation Probabilité et estimation

---

Chair/Président: Fernando Camacho (Damos Inc.)

Room/Salle: Thistle 257

### Abstracts/Résumés

---

**SHANOJA NAIK & TAEHAN BAE** (University of Regina)

[Tuesday May 31 / mardi 31 mai, 15:30]

*On Convolution of Inverse Gaussian and Gamma Distribution and Auto Regressive Model*

*A propos de la convolution des lois inverse gaussienne et Gamma pour le modèle autorégressif*

This paper investigate a new two parameter inverse Gaussian Gamma distribution obtain by convoluting an Inverse Gaussian and Gamma random variable. The goal of the article is to provide characteristics of the new model, to obtain certain statistical function related to the model, to estimate the parameters and obtain the auto regressive models of order one (AR1). The new distribution is promising fit for data set with positive support, compared to some well known distributions like Inverse Gaussian, Gamma ect. The applicability of the model has established for real-world data on US/Canada disasters.

Nous étudions une nouvelle loi inverse gaussienne Gamma à deux paramètres, obtenue par la convolution d'une variable aléatoire inverse gaussienne et d'une variable aléatoire Gamma. L'objectif de cet exposé est d'établir les caractéristiques du nouveau modèle, et d'obtenir une certaine fonction statistique liée au modèle, afin d'estimer les paramètres et obtenir des modèles autorégressifs d'ordre 1 (AR1). La nouvelle loi est prometteuse pour modéliser des jeux de données avec un support positif, par rapport à certaines lois bien connues, comme la loi inverse gaussienne, la loi Gamma, etc. L'applicabilité du modèle a été établie pour des données réelles sur les catastrophes aux États-Unis et au Canada.

---

**STEVE PUGLIA & XU XIAOJIAN** (Brock University)

[Tuesday May 31 / mardi 31 mai, 15:45]

*D-optimal Design of Linear Measurement Error Models*

*Système optimal D de modèles linéaires de mesure d'erreur*

The design of any experiment impacts the precision of model parameter estimates, and future predictions. Utilizing a D-optimal design ensures that the joint confidence regions for true model parameters will be as small as possible for a fixed sample size. Moreover, measurement error is present in the majority of models, and should be considered in the design stage of an experiment. Considering a simple linear model with measurement error in the response, we have investigated the properties of exact and approximate D-optimal designs, which appear to be two-point balanced designs. When measurement error is symmetric, so is the D-optimal design.

La planification de toute expérience a des répercussions sur la précision de l'estimation des paramètres du modèle et sur les prédictions futures. L'utilisation d'un système optimal D permet aux régions de confiance conjointes des paramètres du modèle d'être aussi petites que possible, pour une taille d'échantillon fixe. De plus, l'erreur de mesure est présente dans la plupart des modèles et elle devrait être considérée dans l'étape de planification. Dans le cadre d'un modèle linéaire simple avec une erreur de mesure dans la réponse, nous avons étudié les propriétés des systèmes D optimaux exacts et approximatifs, qui se présentent comme étant des systèmes à deux points équilibrés. Lorsque l'erreur de mesure est asymétrique, le système optimal D l'est également.

## 2E-C2: Probability and Estimation Probabilité et estimation

---

**JEAN VAILLANCOURT** (École Nationale d'Administration Publique), **BRUNO RÉMILLARD** (HEC Montréal)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Loser Takes All!*

*Qui perd gagne!*

In this talk, we show how one can combine losing games of chance into a winning one. It illustrates the famous phenomenon known as Parrondo's Paradox. This phenomenon is extended to random combinations (or regime switchings) of simple random walks in random environments. The paradoxical behaviour of the resulting stochastic process can be explained through considerations regarding the random environment. This is joint work with Bruno Rémillard.

Dans cet exposé, on montre que l'on peut combiner des jeux de hasard perdants en un jeu gagnant. C'est une illustration du célèbre phénomène connu sous le vocable de paradoxe de Parrondo. Ce phénomène est étendu à des combinaisons aléatoires (ou changements de régime) de marches aléatoires simples dans des environnements aléatoires. Le comportement paradoxal du processus stochastique résultant est expliqué par l'effet de l'environnement aléatoire. Ce travail est conjoint avec Bruno Rémillard.

**MYRON HLYNKA & ZISHAN HUANG** (University of Windsor)

[Tuesday May 31 / mardi 31 mai, 16:15]

*Probabilistic Laplace Transform Inversion*

*Transformée inverse probabiliste de Laplace*

We use the probabilistic interpretation of Laplace transforms to create a technique to invert Laplace transforms. The method requires the solution of a quadratic programming problem.

Nous utilisons l'interprétation probabiliste des transformées de Laplace pour créer une technique afin d'inverser la transformation de Laplace. La méthode nécessite la résolution d'un problème de programmation quadratique.

**CHEL HEE LEE & MIKELIS BICKIS** (University of Saskatchewan)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Geometrical View of Imprecise Prior for Imprecise Inference*

*Vision géométrique des lois a priori imprécises pour l'inférence imprécise*

We are concerned with the problem of constructing an imprecise inferential framework. The concept of imprecise probabilities introduced by Walley (1991) is adopted. A conjugate family of priors is formulated in the form of three-parameter exponential family. Two examples with Binomial and Poisson samplings models are provided to show that the advantage of standard conjugate family (i.e. Bayesian updating) still holds. Geometrically, it can also be viewed as translation on the hyperparameter space. We also discuss general strategies of characterizing imprecise prior under the state of near or complete prior ignorance and show stochastic independence and focusing behaviors.

Nous sommes préoccupés par le problème de construction d'un cadre d'inférence imprécise. Nous adoptons le concept des probabilités imprécises présenté par Walley (1991). Nous formons une famille de lois a priori conjuguées sous la forme d'une famille exponentielle à trois paramètres. Nous présentons deux exemples de modèles d'échantillonnage binomial et de Poisson afin de démontrer que la famille conjuguée standard (à savoir, l'actualisation bayésienne) continue de procurer un avantage. Sur le plan géométrique, cette famille peut aussi être considérée comme une translation sur l'espace des hyperparamètres. Nous discutons également des stratégies générales de caractérisation des lois a priori imprécises lorsqu'on ne connaît pas ou presque pas l'information a priori, et nous montrons l'indépendance stochastique ainsi que certains comportements spécifiques.

## 2E-C2: Probability and Estimation Probabilité et estimation

---

**DELI LI** (Lakehead University), **HAN-YING LIANG** (Tongji University, China), **ANDREW ROSALSKY** (University of Florida, USA)

[Tuesday May 31 / mardi 31 mai, 16:45]

*A Note on the Symmetrization Procedures for the Laws of Large Numbers*

*Remarque sur les procédures de symétrisation des lois des grands nombres*

In this talk, general symmetrization procedures for both the weak and strong laws of large numbers concerning the row sums from arrays of rowwise independent Banach space valued random variables are established. Necessary and sufficient conditions are provided for the weak and strong laws of large numbers to hold in terms of corresponding laws of large numbers for a symmetrized version of the array. Several corollaries of the main result are provided.

Dans cet exposé, nous établissons des procédures de symétrisation des lois faibles et fortes des grands nombres concernant la somme des lignes de tableaux de variables aléatoires indépendantes d'espace de Banach. Nous présentons les conditions nécessaires et suffisantes des lois faibles et fortes des grands nombres selon les lois correspondantes des grands nombres pour une version symétrisée du tableau. Nous présentons plusieurs corollaires du résultat principal.

## 2E-C3: Regression Models Modèles de régression

---

Chair/Président: Juli Atherton (UQAM)

Room/Salle: Thistle 254

### Abstracts/Résumés

---

**MOHAMED AMEZZIANE** (Central Michigan University), **SYED EJAZ AHMED** (Brock University)

[Tuesday May 31 / mardi 31 mai, 15:30]

*Variable Selection via Individual Shrinkage of Regression Coefficients*

*Sélection de variable par le rétrécissement individuel des coefficients de régression*

The proposed estimator is obtained by independently shrinking the parameters of some regression estimator towards zero. The mean square error of the estimator is minimized with respect to the shrinkage coefficients to simultaneously select relevant variables and estimate their associated regression coefficients. This results in a soft threshold estimator similar to the non-negative garrote, but does not employ a regularization parameter. The asymptotic properties of the estimator are derived and its performance is compared to that of penalized regression estimators through Monte Carlo simulation experiments.

L'estimateur proposé est obtenu en rétrécissant indépendamment les paramètres d'un estimateur de régression vers zéro. L'erreur carrée moyenne de l'estimateur est minimisée par rapport aux coefficients de rétrécissement pour sélectionner simultanément des variables pertinentes et pour estimer leurs coefficients de régression correspondants. Le résultat est un estimateur à seuil faible similaire au garrot non négatif mais qui n'utilise pas un paramètre de régularisation. Les propriétés asymptotiques de l'estimateur sont dérivées et sa performance est comparée à celle des estimateurs de régression pénalisés à l'aide d'expériences de simulation Monte Carlo.

---

**MAI GHANNAM** (University of Windsor), **SÉVÉRIEN NKURUNZIZA** (University of Windsor/Université de Sherbrooke)

[Tuesday May 31 / mardi 31 mai, 15:45]

*Some Stein Rules Methods in Tensor Regression Models with High and Ultrahigh Dimensional Data*

*Certaines règles de Stein dans les modèles de régression tensorielle avec des données de dimensions supérieures*

In this talk, we consider an estimation problem in tensor regression models with high and ultrahigh dimensional data, when the target parameters may or may not satisfy some restrictions. In particular, we construct James Stein-type estimators for the target parameter. Further, we derive the asymptotic distributional risk of the proposed estimators and prove that shrinkage estimators dominate the unrestricted estimator. The proposed method has some interesting applications as for example in neuroimaging data analysis.

Dans cette présentation, on s'intéresse au problème d'estimation dans les modèles de régression tensorielle avec des données de dimensions supérieures et ultra-supérieures, lorsque les paramètres d'intérêt sont susceptibles de satisfaire certaines restrictions. En particulier, nous construisons les estimateurs de type James-Stein pour le paramètre en question. Par ailleurs, nous élaborons le risque distributionnel asymptotique des estimateurs proposés et montrons que les estimateurs à rétrécissement dominent l'estimateur sans restriction. La méthode proposée a des applications intéressantes comme par exemple dans l'analyse des données de neuroimagerie.

## 2E-C3: Regression Models Modèles de régression

---

**BASHIR KHAN** (Saint Mary's University), **A.K.MD. EHSANES SALEH** (Carleton University)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Estimation in a Regression Model with Autocorrelated Errors*

*Estimation dans un modèle de régression avec erreurs autocorrélées*

We consider the model  $\mathbf{y} = \mathbf{XB} + \mathbf{e}$  where,  $\mathbf{X} = (X_1, X_2)$ ,  $\mathbf{B} = (\beta'_1, \beta'_2)$ , and  $\mathbf{e}$  satisfy the first-order stationary autocorrelated model. We assume that  $\mathbf{e}$  is normal with mean  $\mathbf{0}$  and covariance  $\sigma_z^2 \Sigma_\rho$ , where  $\Sigma_\rho$  is the correlation matrix. We propose four estimators for  $\beta_1$  when it is suspected that  $\beta_2 = \mathbf{0}$ . The performances of the estimators are compared through quadratic bias and risk criteria.

Nous considérons le modèle  $\mathbf{y} = \mathbf{XB} + \mathbf{e}$  où  $\mathbf{X} = (X_1, X_2)$ ,  $\mathbf{B} = (\beta'_1, \beta'_2)$ , et  $\mathbf{e}$  satisfont le modèle autocorrélé stationnaire de premier ordre. Nous supposons que  $\mathbf{e}$  est normal avec moyenne  $\mathbf{0}$  et covariance  $\sigma_z^2 \Sigma_\rho$ , où  $\Sigma_\rho$  est la matrice de corrélation. Nous proposons quatre estimateurs pour  $\beta_1$  lorsque l'on soupçonne que  $\beta_2 = \mathbf{0}$ . La performance des estimateurs est comparée à l'aide des critères du biais quadratique et du risque.

**SHAHEDUL KHAN** (University of Saskatchewan)

[Tuesday May 31 / mardi 31 mai, 16:15]

*Exponentiated Weibull Regression Model for Time-to-Event Data*

*Modèle de régression de Weibull exponentié pour des données de temps d'événement*

The exponentiated Weibull distribution has demonstrated considerable potential to model time-to-event data. It is a generalization of the Weibull distribution, and can be used to model data exhibiting increasing, decreasing, unimodal or bathtub-shaped hazard function. However, the main focus of many studies is to understand the relationship between lifetime and covariates. This leads to a consideration of regression models. In this study, we develop regression methodology based on the exponentiated Weibull distribution. Analyses of four data sets and a simulation study reveal that the exponentiated Weibull regression can be valuable in adequately describing different types of time-to-event data.

La distribution de Weibull exponentiée a démontré un potentiel considérable pour modéliser des données de temps d'événement. Il s'agit d'une généralisation de la distribution de Weibull et elle peut être utilisée pour modéliser des données présentant une fonction de défaillance croissante, décroissante, unimodale ou en forme de baignoire. Cependant, l'objectif principal de plusieurs études est de comprendre la relation entre la durée de vie et les covariables. Cela conduit à un examen des modèles de régression. Dans cette étude, nous développons une méthode de régression qui s'appuie sur la distribution de Weibull exponentiée. Les analyses des quatre jeux de données et une étude de simulation montrent que la régression de Weibull exponentiée peut être utile pour décrire adéquatement différents types de données de temps d'événement.

**KA HO WU** (The Chinese University of Hong Kong)

[Tuesday May 31 / mardi 31 mai, 16:45]

*Error Model Fitting in Benchmarking*

*Ajustement du modèle d'erreur à l'étalonnage*

Two sources of data with different precisions and collecting frequencies may be available for a target socio-economic variable. Benchmarking is a process which uses less frequent and more reliable data to adjust more frequent and less reliable data. The regression method of benchmarking may lead

La disponibilité de deux sources de données comportant différentes précisions et fréquences de collecte est possible pour une variable socioéconomique cible. L'étalonnage est un processus qui fait appel à des données de moindre fréquence et de plus grande fiabilité pour ajuster des données plus fréquentes mais moins sûres. La méthode de régres-

## 2E-C3: Regression Models Modèles de régression

---

to better results than widely used numerical methods. By properly choosing an AR(1) model as the "working model" for the error of monthly data, the regression method may work well in most real situations. In this paper, some new AR(1)-modelling procedures via inside-data-period benchmark forecasts are proposed. The performance of various modelling procedures is compared.

sion de l'étalonnage peut produire de meilleurs résultats que les méthodes numériques largement utilisées. En choisissant judicieusement un processus AR(1) comme « modèle fonctionnel » pour l'erreur de données mensuelles, la méthode de régression peut fonctionner adéquatement dans la plupart des situations réelles. Cet article propose quelques nouveaux processus de modélisation AR(1) à l'aide de prévisions pour une période de référence de collecte interne de données. Les performances de diverses procédures de modélisation sont aussi comparées.



## 2E-C4: Statistical Applications and Methods Méthodes et applications statistiques

---

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

Room/Salle: Thistle 256

### Abstracts/Résumés

---

**BAKTIAR HASAN** (European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium), **CONG CHEN** (Merck & Co., USA), **LAURENCE COLLETTE** (European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium), **XANTHI PEDELI** (Frontier Science Foundation - Hellas, Athens, Greece), **YUE SHENTU** (Merck & Co., USA), **JESSICA MENIS** (European Organisation for Research and Treatment of Cancer (EORTC), Brussels, Belgium), **URANIA DAFNI** (Frontier Science Foundation - Hellas, Athens, Greece.)

[Tuesday May 31 / mardi 31 mai, 15:30]

*Adaptive Study Design: Dealing with Differential Treatment Effect across Subgroups, a Case Study from an EORTC-ETOP Randomized, Phase III Trial in Early Stage Non-Small Cell Lung Cancer (NSCLC)*

*Plan d'étude adapté à un effet différentiel de traitement selon une classification en sous-groupes: une phase III de EORTC-ETOP en exemple*

Previous studies suggested that Pembrolizumab, a checkpoint inhibitor agent, might work better in patients expressing higher PD-L1. Nevertheless, patients with low/no PD-L1 expression may still benefit from treatment. The challenge is to incorporate the level of PD-L1 expression [negative(-): 0 staining, positive(+): 1 – 49% staining, strong positive(++):  $\geq 50\%$  staining] into a phase III study design in adjuvant NSCLC. Taking into account potential heterogeneity of treatment effect across subgroups, the co-primary endpoints of disease-free survival (DFS) in the whole population and DFS in the ++ subgroup are used. Study population enrichment is allowed by a futility interim analysis for the – subgroup.

Des essais récents suggèrent que le Pembrolizumab, un inhibiteur de point de contrôle, pourrait être plus efficace contre les cancers du poumon à petites cellules qui expriment fortement le PD-L1 (PDL1++ :  $\geq 50\%$  de cellules expriment le PDL1) que contre ceux qui ne l'expriment pas (PDL1-) ou peu (PDL1+ : 1 – 49% de cellules positives). Un design de phase III avec deux critères principaux conjoints est utilisé pour intégrer ces hypothèses : la survie sans maladie pour la population complète et pour la sous-population PD-L1++. L'enrichissement de la population en cours d'étude est opéré via une analyse intermédiaire de futilité dans le sous-groupe PD-L1-.

---

**MICHAEL MCISAAC** (Queen's University)

[Tuesday May 31 / mardi 31 mai, 15:45]

*Overcoming Missing Data in Physical Activity Studies Involving Accelerometers*

*Surmonter le problème des données manquantes dans les études d'activité physique impliquant des accéléromètres*

Accelerometers are an increasingly important measure of physical activity in studies of healthy lifestyles, such as our Active Play Study. However, accelerometers sometimes go unworn. Typical approaches involving deleting or averaging over non-wear periods can undermine study findings, particularly since nonwear can coincide with participa-

Les accéléromètres sont de plus en plus utilisés pour mesurer l'activité physique dans les études sur les modes de vie sains, tels que notre étude sur les jeux actifs. Cependant, les accéléromètres ne sont pas toujours portés par les sujets. Des approches classiques qui impliquent la suppression ou le calcul de moyennes sur des périodes où les accéléromètres ne sont pas portés, peuvent nuire aux résultats

## 2E-C4: Statistical Applications and Methods Méthodes et applications statistiques

---

tion in organized sport. This talk will discuss employing modern missing data methodology and using self-report journals to make missing-at-random assumptions more plausible in this important setting. Focus will be on multiple imputation and the unique challenges arising in these incomplete high-dimensional count data that are to be analyzed at the aggregate level.

des études, notamment lorsque le fait de ne pas porter les accéléromètres coïncide avec la participation à un sport organisé. Dans cet exposé, nous discuterons de l'emploi de méthodes modernes permettant de traiter les données manquantes, ainsi que de l'utilisation de registres auto-déclarés qui permettent de rendre les hypothèses de « manquant entièrement dû au hasard » plus plausibles dans ce contexte important. Nous mettrons l'accent sur l'imputation multiple et sur les défis uniques liés à ces données incomplètes de dénombrement à grande dimension qui doivent être analysées au niveau agrégé.

---

**KHOKAN C. SIKDAR** (University of Calgary), **GEORGIY BOBASHEV** (RTI International), **HUDE QUAN, DANIELLE SOUTHERN & MATTHEW THOMAS JAMES** (University of Calgary)

[Tuesday May 31 / mardi 31 mai, 16:00]

*Shifting Epidemiologic Paradigm from Population-level to Individual-level: Cardiovascular Mortality and Morbidity in Hypertensive Patients*

*Déplacer le paradigme épidémiologique du niveau de la population au niveau individuel : mortalité et morbidité cardiovasculaires chez des patients hypertendus*

Population-based epidemiological studies provide evidence of risk factors for disease at a population level. Although focusing on a specific individual is not of interest, inferences from population-based research findings are often used as reference to provide patient-specific care. This is an issue in predicting individual-level outcomes. We propose to use various predictive models, including logistic and survival, to outline the advances as well as newer fundamental findings as an improvement for within-individual epidemiology. We will model cardiovascular outcomes using data from a hypertension cohort. This work may provide a direction for future improvement in big data analytics for precision medicine.

Les études épidémiologiques basées sur la population montrent l'existence de facteurs de risques de maladie à l'échelle de la population. Bien qu'il soit futile de se concentrer sur un individu spécifique, les inférences issues des résultats de recherche basés sur la population servent souvent de références pour fournir des soins spécifiques à un patient. Ceci pose un problème pour la prédiction de résultats à l'échelle individuelle. Nous proposons l'utilisation de divers modèles de prédiction, y compris des modèles logistiques et de survie, afin de mettre en évidence les avancées ainsi que les résultats fondamentaux les plus récents, soulignant l'amélioration épidémiologique intra-sujet. Nous modéliserons des données de maladies cardiovasculaires provenant d'une étude de cohorte sur l'hypertension. Cette étude pourrait fournir une orientation pour des améliorations futures dans les analyses de mégadonnées pour une médecine de précision.

---

**JENNA SYKES** (St. Michael's Hospital), **SANJA STANOJEVIC** (The Hospital for Sick Children), **CHRISTOPHER H. GOSS** (University of Washington), **BRADLEY S. QUON** (University of British Columbia), **ANNE L. STEPHENSON** (St. Michael's Hospital)

[Tuesday May 31 / mardi 31 mai, 16:15]

*A Standardized Approach to Estimating Survival Statistics for Population-Based Registry Cohorts*

*Approche standardisée pour estimer les statistiques de survie de cohortes de registres basées sur des populations*

## 2E-C4: Statistical Applications and Methods Méthodes et applications statistiques

---

It is often of interest to compare disease epidemiology among countries; however, direct comparison of independently-produced registry reports can be challenging due to differences in data processing and statistical methodology. Simulating data from the Canadian Cystic Fibrosis registry, the impact of various registry pitfalls (loss to follow-up, missing death dates) on the estimation of the median age of survival was quantified. The median age of survival was calculated and compared using the common life-table and Cox proportional-hazards methods. We illustrate these effects in comparing Canada to the US and present guidelines to help standardize reporting of disease survival outcomes.

Il est souvent pertinent de comparer l'épidémiologie des maladies entre les pays. Cependant, la comparaison directe de rapports de registres produits de façon indépendante peut présenter des défis en raison des différences dans le traitement des données et de la méthodologie statistique. En simulant des données à partir du registre canadien de la fibrose kystique, nous avons pu quantifier les répercussions de différents problèmes liés aux registres (perte de suivis et dates de décès manquantes) sur l'estimation de l'âge médian de survie. L'âge médian de survie a été calculé et comparé au moyen de la table de survie classique et des modèles de risques proportionnels de Cox. Nous illustrons ces effets en comparant le Canada et les États-Unis et présentons des lignes directrices pour aider à standardiser la communication des résultats de survie à la maladie.

---

**ERNEST DANKWA, TARANEH ABARIN & HONG WANG** (Memorial University)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Bayesian Approach to a Linear Regression model with Measurement Error in Covariate*

*Approche bayésienne pour un modèle de régression linéaire avec erreur de mesure dans la covariable*

This research presents a Bayesian perspective to measurement error problems. We consider a linear regression model with one continuous predictor that is subject to measurement error and additional precisely measured covariates. Common designs used in the presence of measurement errors such as validation and replication designs are discussed in this work. We conduct simulation studies to demonstrate how the Bayesian approach accounts for measurement error in explanatory variables using Markov Chain Monte Carlo methods.

Cette recherche présente une perspective bayésienne face aux problèmes des erreurs de mesure. Nous considérons un modèle de régression linéaire avec une variable indépendante continue qui est sujette à des erreurs de mesure et avec des covariables supplémentaires mesurées précisément. Les modèles courants utilisés en présence d'erreurs de mesure tels que les modèles de validation et de réplication sont abordés dans cette étude. Nous réalisons des études de simulation pour démontrer la façon dont l'approche bayésienne tient compte des erreurs de mesure dans les variables explicatives en utilisant des méthodes Monte-Carlo par chaînes de Markov.

Keywords: Bayesian approach, Linear regression, Monte Carlo, Measurement error.

Mots-clés : Approche bayésienne, régression linéaire, Monte-Carlo, erreur de mesure.

---

**YISUB KYE & EDWARD FURMAN** (York University)

[Tuesday May 31 / mardi 31 mai, 16:45]

*Economic Capital Analysis within a Portfolio of Dependent and Heavy-tailed Risks*

*Analyse du capital économique d'un portefeuille de risques dépendants et à ailes lourdes*

The problem of evaluating the economic capital (EC) for the aggregate risk of a portfolio as well as of consequently allocating the aforementioned EC to risk sources is fundamental in many areas of actuarial science, be it risk management or pricing.

Le problème de l'évaluation du capital économique (CE) pour le risque agrégé d'un portefeuille ainsi que, par conséquent, l'allocation du CE mentionné précédemment aux sources de risque, est fondamental dans de nombreux domaines de la science actuarielle, que ce soit en gestion ou en

## 2E-C4: Statistical Applications and Methods Méthodes et applications statistiques

---

ing. In this talk, I will discuss a complete solution to this problem within a portfolio of Paretian risks with exchangeable and non-exchangeable underlying copulas. As the EC allocation rule, I will use some popular members of the class of weighted EC allocation functionals.

This is a joint work with Ed Furman of York University, Toronto, Canada.

tarification des risques. Dans cet exposé, je discuterai d'une solution complète à ce problème dans un portefeuille de risques parétiens avec copules sous-jacentes échangeable et non-échangeables. Comme règle de répartition du CE, j'utiliserai certains membres populaires de la classe des allocations fonctionnelles pondérées du CE.

Ce travail est réalisé conjointement avec Ed Furman de l'Université York, Toronto, Canada.

## 2E-C5: Statistical Inference and Applications 2 Inférence statistique et applications 2

---

Chair/Président: Ali Karimnezhad (University of Ottawa)

Room/Salle: Thistle 259

### Abstracts/Résumés

---

**JULES DE TIBEIRO** (Université de Moncton), **LUIGI D'AMBRA** (Dipartimento di Economia, Management Istituzioni, University of Naples, Italy.), **ANTONELLO D'AMBRA** (Dipartimento di Economia, Second University of Naples, Italy)  
[Tuesday May 31 / mardi 31 mai, 15:30]

*Different Constrained Ordination Methods for Contingency Table with External Information*

*Différentes méthodes d'ordination sous contraintes pour des tableaux de contingence à variables instrumentales*

We study the information contained in dataset  $K = (X, Z)$  using the statistical method named Correspondence Analysis. A descriptive approach is applied on the whole data set before we partition the dataset into two parts: the first part is data subset  $X$  for which we focus on the explanation and prediction; the second part is the data subset  $Z$  which consists of additional qualitative variables. We also perform Non-Symmetrical Correspondence Analysis, Canonical Correspondence Analysis, and Canonical Non-Symmetrical Correspondence Analysis on  $K$ . We predict expenses of goods and services knowing the global amount and the profile of a new subject.

L'analyse des correspondances trouve toute sa place lors de la phase exploratoire du tableau  $K = (X, Z)$ . Une approche descriptive est appliquée sur l'ensemble des données avant de partitionner ce tableau en deux parties : la première partie  $X$  est un sous-tableau analysé pour expliquer et prédire ; la deuxième partie  $Z$  est un sous-tableau à variables qualitatives additionnelles. Nous effectuons sur  $K$ , une analyse non symétrique des correspondances, une analyse canonique des correspondances et une analyse canonique des correspondances non symétriques. Nous voulons enfin prévoir les dépenses des biens et des services, le montant global et le profil du nouveau sujet.

---

**OTHMANE KORTBI** (UAE University), **AZIZ L'MOUDDEN & ÉRIC MARCHAND** (Université de Sherbrooke), **WILLIAM E. STRAWDERMAN** (Rutgers University)

[Tuesday May 31 / mardi 31 mai, 15:45]

*On Predictive Density Estimation with Constraints under Kullback-Leibler Loss*

*Estimation de densité prédictive avec des contraintes sous la perte Kullback-Leibler*

We consider the estimation of a predictive density for  $Y \sim \text{Gamma}(\alpha_2, \beta)$ , based on  $X \sim \text{Gamma}(\alpha_1, \beta)$ , with  $\beta \in C = (a, b)$ . We obtain representations for Bayes and the MRE predictive densities in the unconstrained problem. We show that the generalized Bayes against the truncation of the non-informative prior onto  $C$  dominates the MRE predictive density and is minimax whenever  $a = 0$  or  $b = \infty$ . Comparisons of plug-in  $\text{Gamma}(\alpha_2, \hat{\beta})$ , which include the predictive mle density, are obtained with results applying as well for point estimation under dual entropy loss.

Nous considérons l'estimation d'une densité prédictive pour  $Y \sim \text{Gamma}(\alpha_2, \beta)$  basée sur  $X \sim \text{Gamma}(\alpha_1, \beta)$ , avec  $\beta \in C = (a, b)$ . Nous obtenons des représentations des densités prédictives de Bayes et MRE pour un problème sans contraintes. Nous montrons que le Bayes généralisé contre la troncation de la non-information a priori sur  $C$  domine la densité prédictive MRE et est un minimax lorsque  $a = 0$  ou  $b = \infty$ . Nous comparons les plugs-in  $\text{Gamma}(\alpha_2, \hat{\beta})$ , qui incluent la densité prédictive de l'EMV, obtenus avec des résultats appliqués également à l'estimation ponctuelle sous la double perte d'entropie.

## 2E-C5: Statistical Inference and Applications 2 Inférence statistique et applications 2

---

**CHANTAL LAROSE** (State University of New York at New Paltz), **OFER HAREL** (University of Connecticut), **KATARZYNA KORDAS** (University of Bristol), **DIPAK DEY** (University of Connecticut)

[Tuesday May 31 / mardi 31 mai, 16:00]

*An Entropy-Based Method for Choosing a Number of Classes for Incomplete Data LCA*

*Méthode fondée sur l'entropie pour choisir le nombre de structures dans l'analyse de structures latentes*

The current method for performing latent class analysis on incomplete data requires the number of classes to be specified before imputation of missing values. We present an entropy-based model selection criterion that extends incomplete data LCA to consider more than one number of classes. The performance of our new criterion is demonstrated in a simulation study, and its effect on the analysis of a family studies data set is presented.

La méthode actuelle pour réaliser une analyse de structures latentes sur des données incomplètes nécessite la spécification du nombre de structures avant l'imputation des valeurs manquantes. Nous présentons un critère de sélection de modèle fondé sur l'entropie qui généralise l'analyse de structures latentes pour données incomplètes et qui considère plus qu'un nombre de classes. Nous démontrons l'efficacité de notre nouveau critère dans une étude de simulation et présentons son effet sur l'analyse d'un jeu de données sur une étude de familles.

**TIAGO MAGALHÃES, DENISE BOTTER & MONICA SANDOVAL** (University of São Paulo)

[Tuesday May 31 / mardi 31 mai, 16:15]

*A Wald-type Test in Proper Dispersion Models*

*Test de type Wald dans les modèles de dispersion propre*

The proper dispersion models (PDM) contain several important non-exponential models, for instance the von Mises regression model for data distributed along the unit circle and the simplex model for data distributed in the standard unit interval  $(0, 1)$ . In this work, we find the second-order covariance matrix of the bias-corrected maximum likelihood estimator of the regression parameter  $\beta$  in PDM. Based on the obtained second-order covariance matrix, we modify the Wald test, improving its performance. We evaluate the result by using a Monte Carlo simulation and apply the expression of the modified Wald test to a real data set.

Les modèles de dispersion propre contiennent plusieurs modèles importants non exponentiels, comme le modèle de régression de von Mises pour des données distribuées le long du cercle d'unité, et le modèle simplexe pour les données distribuées dans l'intervalle unitaire standard  $(0, 1)$ . Dans ces travaux, nous trouvons la matrice de covariance de second ordre de l'estimateur du maximum de vraisemblance corrigé pour le biais du paramètre de régression  $\beta$  dans les modèles de régression propre. En se basant sur cette matrice estimée, nous modifions le test de Wald et améliorons son efficacité. Nous évaluons le résultat au moyen d'une étude de simulation de Monte-Carlo et appliquons l'expression du test de Wald modifié à un jeu de données réelles.

**FAHIMEH MORADI & ALI KARIMNEZHAD** (University of Ottawa)

[Tuesday May 31 / mardi 31 mai, 16:30]

*Bayesian Parameter Learning in Diagnostic Networks*

*Apprentissage de paramètres bayésiens dans les réseaux de diagnostiques*

In most studies in the literature, parameter learning in diagnostic networks is conducted through maximizing the posterior probability function, but the resulting maximum a posteriori estimator, as a Bayes estimator under the zero-one loss function, gives

Dans la plupart des études dans la littérature, l'apprentissage de paramètres dans des réseaux de diagnostiques est réalisé en maximisant la fonction de probabilité a posteriori. Cependant, l'estimateur du maximum a posteriori qui en résulte, tel que l'estimateur bayésien basé sur la fonc-

## 2E-C5: Statistical Inference and Applications 2

### Inférence statistique et applications 2

---

no credit to over/under-learning. To take the effect of over/under-learning into account, we suggest using the Bayes estimator under the entropy loss function. Addressing a real world problem, we conduct a simulation study to measure performance of the proposed estimator. Further, we carry out a sensitivity analysis w.r.t. hyperparameters of a chosen prior to examine effects of changing the hyperparameters on the learning procedure.

tion de perte de zéro à un, ne prend pas en considération le sur-apprentissage ni le sous-apprentissage. Afin de prendre en compte ces effets de sur-apprentissage et de sous-apprentissage, nous suggérons d'utiliser l'estimateur bayésien basé sur la fonction de perte entropique. Dans le cadre d'un problème concret, nous menons une étude de simulation afin de mesurer l'efficacité de l'estimateur proposé. De plus, nous effectuons une analyse de sensibilité des hyperparamètres d'une loi a priori choisie afin d'examiner les effets de changement de ces hyperparamètres sur la procédure d'apprentissage.

---

**CHEN XU** (University of Ottawa), **SHAOBO LIN & JIAN FANG** (Xi'an Jiaotong University), **RUNZE LI** (The Pennsylvania State University)

[Tuesday May 31 / mardi 31 mai, 16:45]

*Prediction-Based Termination Rule for Greedy Learning with Massive Data*

*Règle de terminaison fondée sur la prévision pour l'apprentissage glouton avec des données volumineuses*

In massive-data analysis, the orthogonal greedy algorithm (OGA) has been revitalized as an efficient method for accurate prediction. To implement OGA, an appropriate termination is crucial. In this talk, I introduce a new termination rule for OGA via investigating its predictive performance. The proposed rule is conceptually simple and convenient for implementation, which suggests an  $O(\sqrt{n/\log n})$  number of essential updates in an OGA process. It therefore provides an appealing route to conduct efficient learning for massive-data. Under mild conditions, the proposed method is strongly consistent with an  $O(\sqrt{\log n/n})$  convergence rate to the oracle prediction.

Dans l'analyse de données volumineuses, l'algorithme glouton orthogonal (OGA) a été réactualisé comme méthode efficace de prévision exacte. Pour mettre à l'œuvre un OGA, une terminaison appropriée est essentielle. Dans cet exposé, je présente une nouvelle règle de terminaison pour OGA en examinant sa performance prévisionnelle. La règle proposée est d'une conceptualisation simple et pratique pour sa mise en œuvre, ce qui suggère un nombre  $O(\sqrt{n/\log n})$  de mises à jour essentielles dans un processus OGA. Par conséquent, elle est d'un grand intérêt pour offrir un apprentissage efficace en matière de données volumineuses. Sous des hypothèses faibles, la méthode proposée est fortement convergente avec un taux de convergence  $O(\sqrt{\log n/n})$  à la prévision oracle.

**3A-A1: CRM-SSC Prize in Statistics Address**  
**Allocution du récipiendaire du Prix CRM-SSC en statistique**

---

Chair/Président: Jose Garrido (Concordia University)

**Room/Salle: Thistle 325**

**Abstract/Résumé**

---

**RADU V. CRAIU** (University of Toronto)

[Wednesday June 1 / mercredi 1er juin, 8:45]

*Design Strategies for Adaptive MCMC*

*Stratégies de conception pour le MCMC adaptatif*

Markov chain Monte Carlo (MCMC) algorithms have become essential for modern statistical computation. However, the proper tuning of MCMC samplers can be a complicated affair. An Adaptive MCMC (AMCMC) algorithm is designed to gain information about the target from the very samples it produces and to use it in order to optimize the simulation parameters on the fly. This approach, while effective, can be challenging to validate theoretically. We will discuss and illustrate with examples recent theoretical results that simplify the validation task and make the design of adaptive strategies considerably easier.

Les algorithmes de Monte-Carlo par chaînes de Markov (MCMC) sont devenus des outils essentiels pour le calcul statistique moderne. Cependant, l'ajustement des échantillonneurs MCMC peut s'avérer compliqué. Un algorithme de MCMC adaptatif permet d'obtenir des informations sur la cible à partir des échantillons mêmes qu'il produit et d'utiliser ces informations afin d'optimiser les paramètres de simulation à la volée. Cette approche, bien qu'efficace, peut être difficile à valider en théorie. Nous discutons et illustrons par des exemples de récents résultats théoriques qui simplifient la tâche de validation et facilitent nettement la conception de stratégies adaptatives.



**3A-A2: SSC Impact Award Session**  
**Allocution du récipiendaire du Prix pour impact de la SSC**

---

Chair/Président: Rick Routledge (Simon Fraser University)

**Room/Salle: Welch - David Howes**

**Abstract/Résumé**

---

**SHELLEY B. BULL** (University of Toronto)

[Wednesday June 1 / mercredi 1er juin, 8:45]

*Statistical Issues and Opportunities in Molecular and Genetic Epidemiology*

*Problèmes et possibilités statistiques en épidémiologie moléculaire et génétique*

Molecular and genetic association studies conducted in well-characterized longitudinal cohorts offer a powerful approach to investigate factors influencing disease course or complex trait expression. As measurement technologies continue to develop and evolve, studies based on existing cohorts raise methodological questions that I will illustrate in two long-term inter-disciplinary collaborations. In one, we are investigating molecular genetic prognostic factors in the natural history of node-negative breast cancer using a combination of hypothesis-testing and hypothesis-generating molecular approaches. In the other, we aim to identify genes for multiple traits in participants of a therapeutic RCT in type 1 diabetes using genome-wide genetic analysis of extended follow-up data. I will close with brief comments on emerging challenges for insight into biological mechanisms.

Les études d'associations moléculaires et génétiques menées sur des cohortes longitudinales bien caractérisées permettent une étude approfondie des facteurs qui influencent l'évolution de la maladie ou l'expression de caractères complexes. Alors que les technologies de mesure continuent de progresser et d'évoluer, les études fondées sur des cohortes existantes soulèvent des questions méthodologiques que j'illustrerai dans le cadre de deux collaborations interdisciplinaires à long terme. Dans l'une, nous étudions les facteurs pronostiques génétiques moléculaires dans l'histoire naturelle du cancer du sein sans atteinte des ganglions à l'aide d'une combinaison d'approches moléculaires de génération et de vérification d'hypothèses. Dans l'autre, nous tentons d'identifier les gènes qui contrôlent plusieurs caractères chez les participants à un ERC thérapeutique sur le diabète de type 1 via une analyse génétique sur l'ensemble du génome de données de suivi prolongé. Je conclurai par quelques mots sur les nouveaux défis que présentent l'étude des mécanismes biologiques.

## 3B-I1: Analysis of Complex Survey Data Analyse de données d'enquête complexes

---

Chair/Président: Karla Fox (Statistics Canada)  
Organizer/Responsable: Lenka Mach (Statistics Canada)

Room/Salle: Thistle 258

### Abstracts/Résumés

---

**YVES BERGER** (University of Southampton), **MELIKE OGUZ-ALPER** (Statistics Norway)

[Wednesday June 1 / mercredi 1er juin, 10:20]

*An Empirical Likelihood Approach for Complex Sampling*

*Une approche par vraisemblance empirique pour l'échantillonnage complexe*

Survey data are often collected with unequal probabilities from a stratified population. We propose a new empirical likelihood approach for sample data selected with unequal probabilities. We show that the empirical likelihood ratio statistic follows a chi-squared distribution asymptotically. The approach proposed does not rely on variance estimates, re-sampling or joint-inclusion probabilities, even when the parameter of interest is not linear. Standard confidence intervals based on variance estimates may give poor coverages, when normality does not hold. This can be the case with skewed data and outlying values. The empirical likelihood confidence interval proposed has good coverages, even when the sampling distribution of the point estimator is not normal.

Les données d'enquêtes sont souvent collectées à probabilités inégales à partir de populations stratifiées. Nous proposons une nouvelle approche basée sur la vraisemblance empirique pour des données sélectionnées avec des probabilités inégales. Nous montrons que le ratio de vraisemblance empirique suit une distribution chi-carré asymptotiquement. L'approche proposée ne repose pas sur des estimations de la variance, le re-échantillonnage ou les probabilités d'inclusion jointes, même lorsque le paramètre d'intérêt n'est pas linéaire. Les intervalles de confiance standards basés sur l'estimation de variance peuvent donner des couvertures pauvres en l'absence de normalité. Cela peut être le cas avec des données asymétriques et des valeurs aberrantes. L'intervalle de confiance proposé a des bonnes couvertures, même si l'estimateur ponctuel n'est pas normal.

---

**ELISABETH NEUSY & HAROLD MANTEL** (Statistics Canada)

[Wednesday June 1 / mercredi 1er juin, 10:50]

*Confidence Intervals for Proportions Estimated from Complex Survey Data*

*Intervalles de confiance pour des proportions estimées à partir de données d'enquêtes complexes*

A large number of methods of constructing confidence intervals for proportions have been proposed and studied for non-survey data. A few of these methods have been adapted for complex survey data. A simulation study was undertaken to evaluate the performance of the logit transformation interval, the modified Clopper-Pearson interval, the modified Wilson interval and the bootstrap percentile interval in the context of stratified sampling and two-stage sampling. The paper will briefly review these methods, and present the results and conclusions of the study.

Un grand nombre de méthodes ont été proposées et étudiées pour la construction d'intervalles de confiance pour les données non issues d'enquêtes. Certaines d'entre elles ont été adaptées aux données d'enquêtes complexes. Une étude par simulation a été menée pour évaluer la performance de l'intervalle par transformation logit, l'intervalle modifié de Clopper-Pearson, l'intervalle modifié de Wilson et l'intervalle bootstrap par percentiles dans le contexte de l'échantillonnage stratifié et à deux degrés. J'examinerai brièvement ces méthodes et présenterai les résultats et les conclusions de l'étude.

### 3B-I1: Analysis of Complex Survey Data Analyse de données d'enquête complexes

---

**CHANGBAO WU** (University of Waterloo), **J.N.K. RAO** (Carleton University)

[Wednesday June 1 / mercredi 1er juin, 11:20]

*Empirical Likelihood for Public-Use Survey Data*

*Vraisemblance empirique pour données de sondage d'utilité publique*

In this paper we develop empirical likelihood methods for analyzing public-use survey data that contain only the variables of interest and the final adjusted and calibrated survey weights along with final replication weights. Asymptotic distributions of the empirical likelihood ratio statistics are derived for parameters defined through estimating equations. Finite sample performances of the empirical likelihood ratio confidence intervals, with comparisons to methods based on the estimating equation theory, are investigated through simulation studies. The proposed approaches make empirical likelihood a practically useful tool for users of complex survey data.

Dans ce travail, nous mettons au point des méthodes de vraisemblance empirique pour l'analyse de données de sondage d'utilité publique qui ne contiennent que les variables d'intérêt et les pondérations d'enquête ajustées et calibrées finales, ainsi que les pondérations de réplification finales. Nous dérivons les distributions asymptotiques du rapport des vraisemblances empiriques pour des paramètres définis par des équations d'estimation. Nous étudions les propriétés des intervalles de confiance du rapport des vraisemblances empiriques dans le cas d'échantillons finis et les comparons aux méthodes basées sur la théorie des équations d'estimation par des études en simulation. Les approches proposées font de la vraisemblance empirique un outil pratique pour les utilisateurs de données de sondage complexes.

## 3B-I2: Analysis of Genomic Sequencing Data in Families Analyse des données de séquençage génomique dans les familles

---

Organizer and Chair / Responsable et présidente: Jinko Graham (Simon Fraser University)

Room/Salle: Welch - David Howes

### Abstracts/Résumés

---

**M'HAMED LAJMI LAKHAL CHAIEB, MARTIN LECLERC & JACQUES SIMARD** (Université Laval)

[Wednesday June 1 / mercredi 1er juin, 10:20]

*SNP-Set Association Testing for Survival Outcomes in the Presence of Intrafamilial Correlation*

*Tests d'association d'ensembles de SNPs pour des réponses de survie en présence de corrélation intra-familiale*

In this talk, we propose a SNP-set association test for censored phenotypes in the presence of a family-based design. The proposed test is valid for both common and rare variants. A proportional hazards Cox model is specified for the marginal distribution of the trait and the familial dependence is modeled via a Gaussian copula. Censored values are treated as partially missing data and a multiple imputation procedure is proposed in order to compute the test statistics. The P-value is deduced analytically. The empirical properties of the proposed method are evaluated and compared to existing competitors by simulations and its use is illustrated using a breast cancer data set from the Consortium of Investigators of Modifiers of BRCA1 and BRCA2.

Dans cet exposé, nous proposons un test d'association d'ensembles de SNP pour des phénotypes censurés en présence de données familiales. Le test proposé est valide pour les variants courants et rares. Un modèle des risques proportionnels de Cox est spécifié pour la distribution marginale du trait, et la dépendance familiale est modélisée par une copule gaussienne. Les valeurs censurées sont traitées comme des données partiellement manquantes, et une procédure d'imputation multiple est proposée afin de calculer des statistiques de tests. La valeur P est déduite de façon analytique. Les propriétés empiriques de la méthode proposée sont évaluées et comparées par des simulations avec des méthodes concurrentes existantes. Nous illustrons l'utilisation de cette méthode par un jeu de données sur le cancer du sein provenant du Consortium of Investigators of Modifiers of BRCA1 and BRCA2.

---

**MARIE-HÉLÈNE ROY-GAGNON, KELLY BURKETT, JEAN-FRANÇOIS LEFEBVRE, CHENG WANG, BÉNÉDICTE FONTAINE-BISSON & LISE DUBOIS** (University of Ottawa)

[Wednesday June 1 / mercredi 1er juin, 10:50]

*A Comparison of Statistical Methods for the Discovery of Genetic Risk Factors Using Longitudinal Family Study Designs*

*Comparaison de méthodes statistiques pour l'identification de facteurs de risque génétiques dans des devis d'études familiales longitudinales*

Phenotypes measured at one time point fail to capture time-varying genetic disease mechanisms. Longitudinal studies can shed light on complex disease mechanisms but require specialized statistical methods, especially in family studies where multiple sources of correlation must be modeled. Using simulations we compared analytical methods for investigating genetic associations in longitudinal twin data modeled on the Québec Newborn

Les phénotypes mesurés une seule fois ne capturent pas les mécanismes de maladies génétiques variant dans le temps. Les études longitudinales peuvent aider à comprendre ces maladies complexes, mais requièrent des méthodes d'analyses spécialisées, surtout les études familiales incluant plusieurs sources de corrélation. À l'aide de simulations, nous avons comparé des méthodes d'analyse d'association génétique pour des données longitudinales de jumeaux basées sur l'étude des jumeaux nouveau-nés du Québec, incluant

## 3B-I2: Analysis of Genomic Sequencing Data in Families Analyse des données de séquençage génomique dans les familles

---

Twin Study, including family-based methods applied to summaries of the observations over time, longitudinal-based methods simplifying the familial correlation and Bayesian family-based methods simplifying the temporal correlation. For estimation of genetic effects, all methods gave estimates close to simulated values with similar power. Thus, simpler approaches may likely be adequate to detect genetic effects but their interpretation is more challenging.

des méthodes d'analyse familiale de mesures phénotypes résumées à travers le temps, longitudinale simplifiant la corrélation familiale, et familiale bayésienne simplifiant la corrélation temporelle. Toutes les méthodes donnent des estimés d'effets génétiques proches des valeurs simulées avec puissance comparable. Ainsi, les méthodes simples pourraient être adéquates, mais interprétables plus difficilement.

---

**JIANPING SUN** (McGill), **KARIM OUALKACHA** (Université du Québec à Montréal), **CELIA GREENWOOD** (Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal)

[Wednesday June 1 / mercredi 1er juin, 11:20]

*Multivariate Association Test for Rare Variants Controlling for Cryptic and Family Relatedness*

*Test d'association à plusieurs variables pour variantes rares contrôlées pour parenté familiale et cryptique*

Since rare causal genetic variants are often enriched in families containing multiple affected individuals, family-based study designs are back in vogue for studying rare genetic variants. Recently, we have developed an approach for multivariate-phenotype rare-variant analysis in families by combining univariate test statistics for each trait. However, it is usually infeasible to derive the distribution of the optimal combination due to the complex unknown covariance structure among these univariate test statistics. Hence, we propose two different approaches by using either a copula model or a perturbation method to approximate the distribution of the combined test statistic. In this talk, we will present this work together with some simulations describing the performance of these two options.

Puisque les variantes génétiques rares de causalité sont souvent enrichies dans les familles ayant plusieurs individus affectés, les plans d'étude basés sur la famille reviennent en vogue pour étudier les variantes génétiques rares. Récemment, nous avons développé une approche pour l'analyse des phénotypes multivariés et des variantes rares dans les familles en combinant les statistiques de test univarié pour chaque trait. Cependant, il est habituellement impossible de dériver la loi de la combinaison optimale en raison de la structure complexe de la covariance inconnue parmi ces statistiques de test univarié. Par conséquent, nous proposons deux approches différentes par l'utilisation soit d'un modèle de copules ou d'une méthode de perturbation pour approximer la loi des statistiques de test combiné. Dans cet exposé, nous présenterons ce travail ainsi que des simulations pour décrire la performance de ces deux options.

## 3B-I3: High-dimensional Statistics: Challenges and Recent Developments Statistiques en haute dimension : défis et progrès récents

---

Organizer and Chair / Responsable et président: Yingli Qin (University of Waterloo)

Room/Salle: Thistle 242

### Abstracts/Résumés

---

**SHOJA'EDDIN CHENOURI, JIAXI LIANG & CHRISTOPHER G. SMALL** (University of Waterloo)

[Wednesday June 1 / mercredi 1er juin, 10:20]

*Nonlinear Dimension Reduction: Performance Analysis and Robustness*

*Réduction de la dimension non linéaire : analyse de l'efficacité et robustesse*

Nonlinear dimension reduction has been of much interest in the recent years and many methods have been introduced. Each method works only under certain underlying assumptions and their robustness is generally unknown. In the example of PCA, a few robust alternatives have been introduced. In the nonlinear framework, the robustness of dimension reduction methods has not been explored thoroughly. In this talk we attempt to tackle this problem. We introduce a goodness measure called local rank correlation for assessing the performance of dimension reduction methods. We discuss how the local rank correlation can be used to select tuning parameters of dimension reduction algorithms. We also introduce influence functions to measure local robustness and discuss outlier detection in dimension reduction framework.

La réduction de la dimension non linéaire a suscité beaucoup d'intérêt au cours des dernières années et de nombreuses méthodes ont été présentées. Chaque méthode fonctionne seulement sous certaines hypothèses sous-jacentes et la robustesse est généralement inconnue. Dans l'exemple de l'analyse en composantes principales, quelques méthodes robustes alternatives ont été proposées. Dans le cadre non linéaire, la robustesse des méthodes de réduction de la dimension n'a pas encore été examinée en profondeur. Dans cet exposé, nous tentons de résoudre ce problème. Nous présentons une mesure de l'adéquation appelée corrélation locale de rang, permettant d'évaluer l'efficacité des méthodes de réduction de la dimension. Nous discutons de la façon dont la corrélation locale de rang peut être utilisée pour sélectionner les paramètres de réglage des algorithmes de réduction de la dimension. Nous présentons également les fonctions d'influence pour mesurer la robustesse locale et discutons de la détection des valeurs aberrantes dans le cadre de la réduction de la dimension.

---

**JEFF YAO** (University of Hong Kong), **ZENG LI** (The University of Hong Kong), **WANG QINWEN** (University of Pennsylvania)

[Wednesday June 1 / mercredi 1er juin, 10:50]

*Identifying the Number of Factors from Singular Values of a Large Sample Auto-Covariance Matrix*

*Identification du nombre de facteurs à l'aide des valeurs singulières d'une grande matrice d'autocovariance empirique*

This paper first proposes a complete theory for singular values of lagged sample autocovariance matrices from a high-dimensional factor model covering both the factor part and the noise part assuming that the dimension and the sample size proportionally grow to infinity. In particular, we provide an exact description of the phase transition phenomenon

Cet article propose d'abord une théorie complète des valeurs singulières d'une matrice d'autocovariance empirique dans un modèle à facteurs de grande dimension, couvrant à la fois la partie facteurs et la partie bruit et lorsque la dimension et la taille d'échantillon croissent proportionnellement. En particulier, nous déterminons la frontière exacte de transition pour qu'un facteur soit détectable asymptotiquement.

### 3B-I3: High-dimensional Statistics: Challenges and Recent Developments Statistiques en haute dimension : défis et progrès récents

---

that determines whether a factor is strong enough to be detected asymptotically. Next, we propose a new and strongly consistent estimator for the number of significant factors including both weak and pervasive factors. In all tested cases, the new estimator largely outperforms existing estimators using the same ratios of singular values.

This is a joint work with Zeng Li and Qinwen Wang.

Ensuite, nous proposons un nouvel estimateur du nombre de facteurs incluant à la fois les facteurs faibles et forts qui est fortement convergent. Dans tous les cas testés, ce nouvel estimateur est nettement meilleur que les estimateurs existants qui utilisent ces mêmes valeurs singulières.

C'est un travail joint avec Zeng Li et Qinwen Wang.

---

**PING-SHOU ZHONG** (Michigan State University), **RUNZE LI** (Pennsylvania State University)

[Wednesday June 1 / mercredi 1er juin, 11:20]

*Homogeneity Test of Covariance Matrices and Change-Points Identification with High-Dimensional Longitudinal Data*  
*Test d'homogénéité de matrices de covariance et l'identification de points de rupture avec des données longitudinales de grande dimension*

High-dimensional longitudinal data are widely available. One important feature of such data is that, for each individual, high-dimensional measurements are repeatedly collected over time and they are spatially and temporally dependent. This paper focuses on testing the homogeneity of covariance matrices of high-dimensional measurements over time against the change-point type alternatives. We allow the dimension of measurements to be much larger than the number of individuals. An estimator of the location of the change point is also given whose rate of convergence is established. The proposed method is then extended to locate multiple change points by applying a binary segmentation approach, which is shown to be consistent. Simulation studies and an application are presented to demonstrate the proposed methods.

Des données longitudinales de grandes dimensions sont largement disponibles. Une caractéristique importante de ces données est que, pour chaque individu, des mesures de grande dimension sont recueillies à maintes reprises au fil du temps et celles-ci sont dépendantes du temps et de l'espace. Cet article se concentre sur les tests d'homogénéité des matrices de covariance des mesures de grande dimension au fil du temps, par rapport aux alternatives de type points de rupture. Nous permettons à la dimension des mesures d'être beaucoup plus grande que le nombre d'individus. Un estimateur de l'emplacement du point de rupture est aussi donné dont la vitesse de convergence est établie. La méthode proposée est ensuite élargie pour inclure la localisation de multiples points de rupture en appliquant une approche de segmentation binaire, qui est démontrée comme étant cohérente. Des études de simulation et une application sont présentées pour démontrer les méthodes proposées.

## 3B-I4: Statistical Theory and Modeling for Functional Data Analysis Théorie et modélisation statistiques pour l'analyse de données fonctionnelles

---

Organizer and Chair / Responsable et président: Haocheng Li (University of Calgary)

**Room/Salle: Thistle 246**

### Abstracts/Résumés

---

**KEXIN JI & JOEL A. DUBIN** (University of Waterloo)

[Wednesday June 1 / mercredi 1er juin, 10:20]

*Semiparametric Stochastic Mixed Models for Bivariate Periodic Longitudinal Data*

*Modèles stochastiques mixtes semi-paramétriques pour des données longitudinales périodiques bidimensionnelles*

We propose and consider inference for a semiparametric stochastic mixed model for bivariate periodic repeated measures data. The bivariate model uses parametric fixed effects for modelling covariate effects and periodic smooth nonparametric functions for each of the two underlying time effects. In addition, the between-subject and within-subject correlations are modelled using separate but correlated random effects and a bivariate Gaussian random field, respectively. We derive maximum penalized likelihood estimators for both the fixed effects regression coefficients and the nonparametric functions. The smoothing parameters and all variance components are estimated simultaneously using restricted maximum likelihood. We investigate the proposed methodology through simulation. We also illustrate the model by analyzing bivariate longitudinal female hormone data collected daily over multiple consecutive menstrual cycles.

Nous proposons et étudions un modèle stochastique mixte semi-paramétrique pour des données de mesures répétées périodiques bidimensionnelles. Le modèle bidimensionnel utilise des effets paramétriques fixes pour modéliser les effets de covariables et les fonctions périodiques lissées non paramétriques pour chacun des deux effets chronologiques sous-jacents. De plus, les corrélations entre sujets et intra-sujets sont modélisées en utilisant, respectivement, des effets aléatoires séparés mais corrélés et un champ aléatoire gaussien bidimensionnel. Nous dérivons des estimateurs de maximum de vraisemblance pénalisée pour les coefficients de régression à effets fixes et pour les fonctions non paramétriques. Les paramètres lissants et toutes les composantes de la variance sont estimés simultanément en utilisant le maximum de vraisemblance restreint. Nous examinons la méthodologie proposée par la simulation. Nous illustrons aussi le modèle en analysant des données longitudinales bidimensionnelles d'hormones féminines recueillies quotidiennement durant de multiples cycles menstruels.

---

**BEI JIANG** (University of Alberta), **EVA PETKOVA** (New York University), **THADDEUS TARPEY** (Wright State University), **R. TODD OGDEN** (Columbia University)

[Wednesday June 1 / mercredi 1er juin, 10:50]

*Latent Class Modeling Using Matrix-valued Covariates with Application to Identifying Early Placebo Responders Based on EEG Signals*

*Modélisation de classes latentes utilisant des covariables à valeurs matricielles pour identifier les répondeurs hâtifs au placebo basée sur les signaux EEG*

In this talk, we extend existing latent class models to incorporate matrix covariates. The proposed method is built upon a low rank Candecomp/Parafac (CP) decomposition to express the target coefficient matrix through low-dimensional latent variables, which effectively reduces the

Dans cet exposé, nous élargissons les modèles de classes latentes existantes pour incorporer les covariables matricielles. La méthode proposée est fondée sur une décomposition Candecomp/Parafac (CP) à rang inférieur pour exprimer la matrice de coefficients par des variables latentes à basses dimensions, ce qui peut efficacement réduire la



### 3B-I4: Statistical Theory and Modeling for Functional Data Analysis

#### Théorie et modélisation statistiques pour l'analyse de données fonctionnelles

---

model dimensionality, and utilizes a Bayesian hierarchical modeling approach to estimating these latent variables, which provides a way to incorporate prior knowledge on the patterns of covariate effect heterogeneity and provides a data-driven method of regularization. Our simulation studies suggest that our proposed method is robust against potentially misspecified rank in the CP decomposition. We show that the proposed method allows us to extract useful information from baseline EEG measurements that explains the likelihood of belonging to the early responder subgroup.

dimension du modèle, et utilise une approche de modélisation bayésienne hiérarchique pour estimer ces variables latentes, ce qui donne une façon d'incorporer des connaissances préalables sur les modèles de l'hétérogénéité des effets de covariables et fournit une méthode de régularisation basée sur les données. Nos études de simulation suggèrent que notre méthode est robuste face aux rangs potentiellement mal spécifiés dans la décomposition CP. Nous démontrons que la méthode proposée nous permet d'extraire de l'information importante des mesures de références EEG qui explique la vraisemblance d'appartenir à un sous-groupe de réponders hâtifs.

---

**ZHENHUA LIN & FANG YAO** (University of Toronto), **HANS G. MÜLLER** (University of California at Davis)  
[Wednesday June 1 / mercredi 1er juin, 11:20]

*Mixture Models and Densities for Functional Data*  
*Modèles de mélange et de densités pour données fonctionnelles*

We propose a novel perspective to represent infinite-dimensional functional data as mixtures, where the number of basis functions that constitutes the mixture may be arbitrarily large, but the number of included mixture components is finite and specifically adapted for each random trajectory. Within this framework, we show that a probability density can be well defined without the need for finite truncation or approximation. Dimensions of individual trajectories are treated as latent random variables and a modified expectation-maximization algorithm is adopted for model fitting. Simulations confirm that comparing to traditional FPCA the proposed method achieves similar or better data recovery while using fewer components on average. The practical merits of functional mixture modeling are demonstrated by analyzing egg-laying trajectories for medflies.

Nous proposons un point de vue novateur pour la représentation des données fonctionnelles à dimension infinie par des mélanges, où le nombre de fonctions de base qui constituent le mélange peut être arbitrairement élevé, mais le nombre de composantes incluses dans le mélange est fini et spécifiquement adapté à chaque trajectoire aléatoire. Dans ce cadre, nous démontrons qu'une densité de probabilité peut être bien définie sans avoir besoin d'une approximation ou d'une troncature finie. Les dimensions des trajectoires individuelles sont traitées comme des variables aléatoires latentes et un algorithme modifié de maximisation d'espérance est adopté pour l'ajustement du modèle. Des simulations confirment qu'en la comparant au FPCA traditionnel, la méthode proposée réalise une récupération de données similaire ou meilleure tout en utilisant en moyenne moins de composantes. L'intérêt pratique de la modélisation de mélange fonctionnel est démontré par l'analyse des trajectoires de pontes d'oeufs de la mouche méditerranéenne.

### **3B-I5: Staying Up-To-Date As a Practicing Biostatistician Comment rester à la page en tant que biostatisticien praticien**

---

Organizer and Chair / Responsable et présidente: Eleanor Pullenayegum (Hospital for Sick Children)

**Room/Salle: Thistle 325**

#### **Description**

A medical colleague recently pointed out that when he cares for his patients, he is expected to be up to date with all the relevant medical research. So, when he goes to see a biostatistician, that person should be up-to-date with all recent developments in biostatistics. Perhaps you're thinking, "That's impossible"? This roundtable will discuss: a) As a practicing biostatistician, what do you need to stay up to date with? b) How do you stay up to date? c) Are there ways we could organize ourselves as a discipline to make it easier to stay up to date?

Un collègue du milieu médical faisait remarquer dernièrement qu'on s'attendait à ce qu'il soit à jour sur les études de recherche pertinentes lorsqu'il prodigue des soins aux patients. Par conséquent, s'il s'adresse à un biostatisticien, celui-ci doit être au fait des plus récents développements dans son domaine. Impossible, croyez-vous ? Dans le cadre de cette table ronde, nous nous demanderons a) De quel contenu informatif un biostatisticien praticien doit-il être au fait ? b) Comment mettre à jour nos connaissances ? c) De quelle façon les praticiens de notre discipline peuvent-ils s'organiser pour faciliter l'actualisation de leur savoir ?

## 3B-16: The Theory of Dispersion Models with Applications La théorie de la dispersion avec des applications

---

Organizer and Chair / Responsable et président: Renjun Ma (University of New Brunswick)

Room/Salle: Academic South 217

### Abstracts/Résumés

---

**RODRIGO LABOURIAU** (Aarhus University)

[Wednesday June 1 / mercredi 1er juin, 10:40]

*Multivariate Generalized Linear Mixed Models with High Complexity*

*Modèles linéaires généralisés mixtes multivariés avec complexité élevée*

The theory of exponential dispersion models (EDM), for which Bent Jørgensen made substantial contributions, provides a flexible framework of models alternative to the classic Gaussian linear models (e.g. generalized linear models and additive models). I will review some multivariate extensions of those models that allow the distribution of the different dimensions to belong to different EDMs. As an illustration, I will present some applications in quantitative genetics with high complexity (several hundreds of thousand observations and deep pedigrees). In all the presented applications, it is crucial to understand the underlying stochastic process related to the EDMs used to represent well and interpret the biological questions of interest. Bent Jørgensen advocated similar ideas in his work since the 1980s.

La théorie des modèles de dispersion exponentielle (EDM), à laquelle Bent Jørgensen a apporté d'importantes contributions, fournit un cadre flexible de modèles alternatifs aux modèles linéaires gaussiens classiques (par exemple, les modèles linéaires et les modèles additifs). J'examinerai quelques extensions multivariées de ces modèles qui permettent à la loi des différentes dimensions d'appartenir à différents EDM. À titre d'exemple, je présenterai quelques applications à la génétique quantitative avec complexité élevée (plusieurs centaines de milliers d'observations et des pedigrees profonds). Dans toutes les applications présentées, il est essentiel de comprendre le processus stochastique sous-jacent associé aux EDMs pour bien représenter et interpréter les questions biologiques d'intérêt. Bent Jørgensen prônait des idées similaires dans son travail depuis les années 1980.

---

**PETER XK SONG** (University of Michigan)

[Wednesday June 1 / mercredi 1er juin, 11:20]

*Dispersion Models and Beyond*

*Les modèles de dispersion et au-delà*

Seminal work by Dr. Bent Jørgensen for the theory of dispersion models has made profound impact on the statistical distribution theory, in particular providing the novel extension of the classical theory of generalized linear models. In this talk, I will present a review on several important developments of multivariate dispersion models useful to analyze multi-dimensional data, longitudinal data and spatially correlated data. My review will focus on multivariate dispersion models generated by copulas and the method of generalized estimating

Le travail précurseur du Dr Bent Jørgensen sur la théorie des modèles de dispersion a eu un profond impact sur la théorie des lois statistiques, particulièrement avec l'extension novatrice de la théorie classique des modèles linéaires généralisés. Dans cet exposé, je présenterai une revue de plusieurs développements importants dans les modèles de dispersion multivariés qui sont utiles pour l'analyse de données multidimensionnelles, de données longitudinales et de données corrélées spatialement. Ma revue se concentrera sur les modèles de dispersion multivariés générés par des copules et sur la méthode des fonctions d'estimation généra-

### **3B-16: The Theory of Dispersion Models with Applications**

#### **La théorie de la dispersion avec des applications**

---

functions, as well as their applications in regression analysis. A few R packages will be introduced to demonstrate the usefulness of dispersion models in practical studies. In addition, some data applications will be discussed.

lisées, ainsi que leurs applications dans l'analyse de régression. Quelques paquets R seront présentés pour démontrer l'utilité des modèles de dispersion dans des études pratiques. De plus, je discuterai de quelques applications des données.

### 3B-C1: Biostatistics: Survival Analysis Biostatistique : analyse de survie

---

Chair/Président: John Koval (University of Western Ontario)

Room/Salle: Thistle 256

#### Abstracts/Résumés

---

**SAMANTHA-JO CAETANO, GREGORY POND, PETER ELLIS & DAVID DAWE** (McMaster University)

[Wednesday June 1 / mercredi 1er juin, 10:20]

*Proposing and Analysing Methods which Improve the Kaplan-Meier Estimates of Altered Time-to-Event Data*

*Proposer et analyser des méthodes qui améliorent les estimations Kaplan-Meier de données altérées de temps d'événement*

Patient registry databases must be de-identified prior to analysis to ensure confidentiality. This de-identification alters the data, which can impact the results, and potentially lead to incorrect conclusions. A recent analysis of lung cancer patients with data obtained from registry databases was observed to contain a missing time bias due to de-identification. Five statistical methods were proposed and investigated to examine how to reduce this bias when calculating time-to-event estimates. Through statistical simulation, the bias, mean-squared error and coverage probability of each method was assessed. The optimal method was then applied to the original dataset and estimates re-evaluated.

Les bases de données de registre des patients doivent être dépersonnalisées avant l'analyse pour assurer la confidentialité. Cette dépersonnalisation modifie les données, ce qui peut avoir une incidence sur les résultats, et potentiellement conduire à des conclusions erronées. Une récente analyse des patients atteints d'un cancer du poumon avec des données obtenues à partir des bases de données de registre contenait un biais de temps manquant en raison de la dépersonnalisation. Cinq méthodes statistiques ont été proposées et étudiées pour examiner comment réduire ce biais lors du calcul des estimations de temps à l'événement. Grâce à la simulation statistique, le biais, l'erreur quadratique moyenne et la probabilité de couverture de chaque méthode ont été évalués. La méthode optimale a ensuite été appliquée au jeu de données initial et les estimations ont été réévaluées.

---

**NOORI AKHTAR-DANESH** (McMaster University), **LINA JANSEN & HERMANN BRENNER** (German Cancer Research Centre)

[Wednesday June 1 / mercredi 1er juin, 10:35]

*Estimating Complete Life Tables from Abridged Life Tables in Relative Survival Analysis*

*Estimer des tables de mortalité complètes dans une analyse de survie relative à partir de tables de mortalité abrégées*

Complete life tables (CLT) are required for the computation of relative survival in cancer patients. Different methods are used to estimate CLT from abridged life tables (ALT). However, the impact of these methods on relative survival estimates is unknown. We used Elandt-Johnson method, improved Akima method, Brass logit model, Kostaki method, and a simple arithmetic progression method (APM) to estimate the CLT from an ALT. The maximum difference in relative survival estimates within 10 years of follow-up was less than 2 percent in the

Des tables complètes de mortalité (CLT) sont nécessaires pour calculer la survie relative des patients atteints d'un cancer. Différentes méthodes sont utilisées pour l'estimation de CLT à partir de tables de mortalité abrégées (ALT). Par contre, l'impact de ces méthodes sur les estimations de survie relative est inconnu. Nous avons utilisé la méthode Elandt-Johnson, amélioré la méthode Akima, le modèle Brass logit, la méthode Kostaki, et une méthode simple de progression arithmétique (APM) pour estimer une CLT à partir d'une ALT. La différence maximale dans les estimations de survie relative lors des 10 années de suivi fut de

### 3B-C1: Biostatistics: Survival Analysis Biostatistique : analyse de survie

---

oldest age group. Most importantly, the APM was as good as the other complex methods.

moins de 2 pour cent dans la tranche d'âge la plus âgée et qui plus est, la APM était aussi bonne que les autres méthodes complexes.

---

**CORALINE DANIELI & MICHAL ABRAHAMOWICZ** (McGill University)

[Wednesday June 1 / mercredi 1er juin, 10:50]

*Competing Risk Modelling of Time-Varying Exposures*

*Modélisation du risque concurrent d'expositions variant dans le temps*

We propose a new method to address two important challenges, often encountered when modeling time-varying exposures in time-to-event analyses: the need to account for cumulative effects of past exposures and competing risks of alternative endpoints (e.g. different sources of mortality). Our approach combines flexible regression spline-based weighted cumulative exposure methodology, with competing risks extension of time-dependent Cox model, that adapts the Lunn-McNeil data augmentation approach. We also propose a systematic approach for testing significance of the difference between the effects of the same exposure on alternative outcomes. Simulations demonstrate the accuracy of the proposed estimates and tests.

Nous proposons une nouvelle méthode pour faire face à deux défis importants, souvent rencontrés lors de la modélisation des expositions variant dans le temps dans les analyses de temps d'événement : la nécessité de tenir compte des effets cumulatifs des expositions passées et des risques concurrents des terminaux alternatifs (par exemple, différentes sources de mortalité). Notre approche combine une méthodologie flexible d'exposition pondérée cumulative de régression spline, avec une extension des risques concurrents du modèle de Cox dépendant du temps qui adapte l'approche d'augmentation de données de Lunn-McNeil. Nous proposons également une approche systématique pour tester la signification de la différence entre les effets d'une même exposition sur les résultats alternatifs. Les simulations démontrent l'exactitude des estimations et des tests proposés.

---

**HUA SHEN** (University of Calgary)

[Wednesday June 1 / mercredi 1er juin, 11:05]

*Methods for Incomplete Covariates in Clustered Survival Data with Left Truncation*

*Méthodes pour les covariables incomplètes dans des données de survie regroupées avec troncation à gauche*

In studies of chronic diseases individuals are often routinely sampled subject to certain conditions on an event time of interest. These conditions yield response-biased samples featuring left-truncated event times. The failure times of the individuals within some groups can be correlated due to natural common features or shared environmental factors. Incomplete covariate data are widely encountered in these settings. The fact that the covariate distribution is affected by the left truncation selection criterion and the clustering is often ignored in standard methods leading to biased estimates. An algorithm is developed to deal with these challenges and applied to case studies.

Dans les études de maladies chroniques, les individus sont souvent échantillonnés régulièrement en fonction de certaines conditions reliées à un temps d'événement d'intérêt. Ces conditions engendrent des échantillons dont les réponses sont biaisées et avec des temps d'événement tronqués à gauche. Les temps de défaillance des individus dans certains groupes peuvent être corrélés en raison de caractéristiques communes naturelles ou de facteurs environnementaux partagés. On retrouve régulièrement des covariables incomplètes dans ces contextes. La plupart du temps, les méthodes standards ne tiennent pas compte du fait que la distribution des covariables est affectée par le critère de sélection à troncation à gauche ainsi que le regroupement, ce qui biaise les estimations. Un algorithme est conçu pour relever ces défis et est appliqué dans des études de cas.

### 3B-C1: Biostatistics: Survival Analysis Biostatistique : analyse de survie

---

**YINGYING XU, JOON LEE & JOEL A. DUBIN** (University of Waterloo)

[Wednesday June 1 / mercredi 1er juin, 11:20]

*Personalized Time-to-Event Prediction in Intensive Care*

*Prévision personnalisée des occurrences dans le temps pour les soins intensifs*

Predicting the time to clinical outcomes for patients in intensive care units (ICU's) helps to support doctor's treatment decisions. The time could be, for example, survival time or time to recovery from a disease within the ICU. We developed methodology that extends current survival analysis models and improves the prediction accuracy by systematically identifying similar patients and making personalized predictions. Both time-fixed and time-dependent covariates can be incorporated in our proposed model. Our method is validated in the Multi-Parameter Intelligent Monitoring for Intensive Care (MIMIC-III) database, and we will also present properties of our methodology through a comprehensive simulation study.

La prévision du temps aux réponses cliniques chez les patients admis dans une unité de soins intensifs (USI) aide à appuyer les traitements décidés par le médecin. Le facteur temps pourrait être, par exemple, la durée de survie ou de rétablissement à la suite d'une maladie pendant la période passée aux soins intensifs. Nous avons mis au point une méthodologie qui étend la portée des modèles courants d'analyse de survie et améliore l'exactitude prévisionnelle en identifiant systématiquement les cas similaires et en avançant des prévisions personnalisées. Le modèle que nous proposons peut comporter à la fois des covariables à temps fixe et dépendantes du temps. Notre méthode est validée par la base de données de veille sanitaire intelligente multiparamétrique en soins intensifs (MIMIC-III). Nous présenterons également les propriétés de notre méthodologie à l'aide d'une étude de simulation approfondie.

**YAYUAN ZHU, JERALD F. LAWLESS & CECILIA A. COTTON** (University of Waterloo)

[Wednesday June 1 / mercredi 1er juin, 11:35]

*Non- and Semiparametric Analysis of Dependently Interval-Censored Failure Time Data Due to Intermittent Visits at Disease Clinics*

*Analyse non paramétrique et semi-paramétrique de données de temps de défaillance avec censure par intervalle dépendante en raison de visites intermittentes à une clinique de traitement de maladies*

Event history studies based on disease clinic data often have intermittent clinic visits. Irregular inspection times may depend on the history of disease-related variables; this can cause event times or failure times to be dependently interval-censored. A so-called inverse-intensity-of-visit (IIV) weighting method was developed for irregularly observed longitudinal or repeated measures data. We extend this method to failure time data analysis with dependent interval censoring. This presentation focuses on non- and semiparametric estimation of failure time distributions with monotone smoothing techniques applied. As an illustration, the proposed methodologies are applied to the Toronto Psoriatic Arthritis (PsA) Cohort Study.

Les études événementielles fondées sur les données de cliniques de traitement de maladies tiennent souvent compte des visites intermittentes à la clinique. Les périodes irrégulières d'inspection peuvent dépendre de l'historique des variables liées à la maladie, ce qui peut faire en sorte que les temps d'événements ou de défaillance soient censurés par intervalle de manière dépendante. Une soi-disant méthode de pondération en fonction de l'intensité inverse des visites (IIV) a été conçue pour les mesures longitudinales ou répétées observées de façon irrégulière. Nous élargissons cette méthode à l'analyse de données sur le temps des défaillances avec censure par intervalle dépendante. Cette présentation est axée sur une estimation non paramétrique et semi-paramétrique des distributions des temps de défaillance appliquant des techniques de lissage monotone. Pour illustrer les méthodologies proposées, nous les appliquons à l'étude de cohorte de la clinique de polyarthrite psoriasique de Toronto.

## 3B-C2: Time Dependent Data Données dépendantes du temps

---

Chair/Président: Edward Chen (Statistics Canada)

Room/Salle: Thistle 255

### Abstracts/Résumés

---

**FUQI CHEN & ROGEMAR MAMON** (Western University), **SÉVÉRIEN NKURUNZIZA** (University of Windsor)

[Wednesday June 1 / mercredi 1er juin, 10:20]

*Estimation and Inference for a Change-point Problem under a Generalised Ornstein-Uhlenbeck Setting*

*Estimation et inférence pour un problème de point de changement sous une situation généralisée Ornstein-Uhlenbeck*

We develop two methods in locating the change points of a generalised Ornstein-Uhlenbeck process. The proposed methods were motivated by the least sum of squared error (LSSE) and the maximum log-likelihood (MLL) approaches. We show further that our methods provide consistent estimators, as well as the asymptotic normality of the MLE for the drift parameters. The case where both the number and the location of the change points are unknown is investigated and an informational methodology is employed to address these issues. We also provide some numerical simulations to illustrate the performance of the proposed methods.

Nous développons deux méthodes pour localiser les points de changement d'un processus généralisé d'Ornstein-Uhlenbeck. Les méthodes proposées ont été motivées par les approches de la somme des erreurs des moindres carrés (SEMCC) et du maximum de la log-vraisemblance (MLV). Nous démontrons en outre que les méthodes fournissent des estimateurs convergents de même que la normalité asymptotique du EMV pour les paramètres de dérive. Le cas où le nombre et l'emplacement des points de changement sont inconnus est étudié et une méthodologie d'information est utilisée pour répondre à ces questions. Nous fournissons également des simulations numériques afin d'illustrer la performance des méthodes proposées.

---

**FRANÇOIS MARSHALL** (Queen's University)

[Wednesday June 1 / mercredi 1er juin, 10:35]

*Robust Spectrum Estimation: Application to Relative Ionospheric Opacity Meters*

*Estimation spectrale robuste : application à des instruments de mesure de l'opacité ionosphérique relative*

Relative ionospheric opacity meters (riometers) are antennae which measure the effects of the solar wind on the upper atmosphere. Their response is a voltage series, and because they are outdoors measuring natural phenomena, riometers are prone to the full range of noise sources and contaminating agents. In this presentation, a review will be given of robust techniques for computing an estimate of the spectral density. In particular, techniques are employed which make use of a base spectrum and a detail spectrum, in an effort to account for the non-centrality of the chi-squared distribution commonly associated with the noise.

Les instruments de mesure de l'opacité ionosphérique relative (riomètres) sont des antennes qui mesurent les effets du vent solaire sur la haute atmosphère. Ils réagissent par des séries de voltage, et parce qu'ils se trouvent à l'extérieur pour mesurer les phénomènes naturels, les riomètres sont soumis à toute une série de sources de bruit et d'agents de contamination. Dans cet exposé, nous examinerons les techniques robustes de calcul d'une estimation de la densité spectrale. Les techniques employées font notamment appel à un spectre de base et à un spectre détaillé afin de prendre en compte la non-centralité de la loi du  $\chi^2$  souvent associée au bruit.



## 3B-C2: Time Dependent Data Données dépendantes du temps

---

**FRÉDÉRIC PICARD & STEVE MATTHEWS** (Statistics Canada)

[Wednesday June 1 / mercredi 1er juin, 10:50]

*The Addition of Trend-Cycle Estimates to Selected Publications at Statistics Canada*

*L'ajout de la tendance-cycle à quelques publications de Statistique Canada*

This talk will cover a number of aspects related to the trend-cycle estimates that Statistics Canada has recently included in selected publications. Statistics Canada uses a trend estimation method proposed by Dagum and Luati (2009). We will briefly describe the method and explain why it was chosen. In particular, an evaluation study to compare a number of options for estimating the trend-cycle will be presented and the findings will be discussed. As well, several issues related to the use of the trend-cycle estimates will be explored, including how they are presented and recommendations for data users in interpreting the results.

Dans cette présentation, plusieurs aspects reliés à l'estimation de la tendance-cycle, que Statistique Canada a récemment incluse dans certaines de ses publications, seront traités. Statistique Canada utilise une méthode d'estimation de tendance proposée par Dagum et Luati (2009). La méthode sera brièvement décrite et les raisons pour lesquelles elle a été choisie seront expliquées. En particulier, une étude évaluative pour comparer quelques méthodes d'estimation de tendance sera présentée et suivie d'une discussion des résultats. Aussi, plusieurs enjeux reliés à l'utilisation de la tendance-cycle seront explorés. Ceci inclut la façon de la présenter et les recommandations aux utilisateurs.

**REZA RAMEZAN** (California State University, Fullerton), **PAUL MARRIOTT & SHOJA'EDDIN CHENOURI** (University of Waterloo)

[Wednesday June 1 / mercredi 1er juin, 11:05]

*Time Scale in Neuronal Data: A Multiscale Model*

*Échelle de temps dans les données neuronales : un modèle à multi-échelles*

A large portion of our knowledge about the brain relies on the statistical models developed for the analysis of neuronal data. In this talk, we show that our conclusions depend on the time scales in which spike trains are analyzed. Two flexible models within the inhomogeneous Poisson process framework are discussed. Both of these formulations account for simultaneous biological phenomena, from multiple time scales, such as neural refractory period, burst spiking activity, brain rhythms, and learning. Simulations and real data analyses highlight the promising results of these computationally efficient models.

Une grande partie de nos connaissances sur le cerveau se fonde sur les modèles statistiques conçus pour l'analyse de données neuronales. Dans cet exposé, nous démontrons que nos conclusions dépendent des échelles de temps utilisées dans l'analyse des trains d'impulsions. Nous discutons de deux modèles flexibles dans le cadre du processus non homogène de Poisson. Chacune de ces formulations prend en compte le phénomène biologique simultané à partir de plusieurs échelles de temps, comme par exemple la période réfractaire neuronale, les pics d'activité, les rythmes cérébraux et l'apprentissage. Des simulations et des analyses de données réelles mettent en évidence les résultats prometteurs de ces modèles informatiques efficaces.

## 3C-I1: Isobel Loutit Lecture Allocution invitée Isobel Loutit

---

Organizer and Chair / Responsable et président: Hugh Chipman (Acadia)

**Room/Salle: Thistle 242**

### **Abstract/Résumé**

---

**CHRISTINE M. ANDERSON-COOK** (Los Alamos National Laboratory)

[Wednesday June 1 / mercredi 1er juin, 13:30]

*Realistic Optimal Designs in a Resource-Constrained World*

*Plans optimaux réalistes dans un monde à ressources limitées*

Historically designed experiments were chosen from a collection of designs itemized in the statistics literature, with limited size and design region shape options. These designs had good general properties with some robustness across applications. Next came computer-generated designs optimal for a single criterion. These designs allowed greater size and region flexibility, but lacked overall robustness. An emerging area seeks to combine the best of both worlds: well-rounded performance with increased flexibility. Priorities of experimenters differ and economic pressures on experiments are increasing. Hence, identifying leading candidate designs based on user-specified design criteria allows for better designs. A flexible two-stage Pareto front approach is described, which objectively finds promising candidates using experimenter priorities. Diverse examples illustrate the approach.

Historiquement, les expériences étaient conçues d'après une collection de plans énumérés dans la littérature statistique, avec des choix limités quant à la taille et à la forme de la région du plan. Ces plans possédaient de bonnes propriétés générales et une certaine robustesse d'une application à l'autre. Par la suite, la conception assistée par ordinateur a permis d'optimiser les expériences pour un critère donné. Ces plans offraient plus de souplesse en ce qui concerne la taille et la région, mais manquaient de robustesse globale. Un nouveau domaine de recherche cherche à combiner le meilleur des deux mondes : performance homogène et souplesse accrue. Les expérimentateurs n'ont pas tous les mêmes priorités et les pressions économiques sur les expériences ne font qu'augmenter. L'identification des meilleures options de plans en fonction des critères de conception de l'utilisateur est donc essentielle. Je décris une approche de front de Pareto souple à deux étapes qui permet objectivement de trouver des candidats prometteurs en fonction des priorités de l'expérimentateur. J'illustrerai cette approche avec divers exemples.

**3C-12: Survey Methods Section Presidential Invited Address**  
**Allocution de l'invité du président du Groupe des méthodes d'enquête**

---

Organizer and Chair / Responsable et présidente: Karla Fox (Statistics Canada)

**Room/Salle: Academic South 204**

**Abstract/Résumé**

---

**MICHAEL HIDIROGLOU** (Statistics Canada)

[Wednesday June 1 / mercredi 1er juin, 13:30]

*Data Collection*

*Collecte de données*

A large-scale survey can be divided into several stages. Given that a survey is selected from a good quality frame, a survey methodologist should be in good control of all stages. One of the most critical steps is data collection, and it is increasingly more difficult to control. There are several reasons for this. Data collection costs are increasing and response is decreasing. In addition, the mode of collection has changed in recent years.

This talk will provide an overview of some the difficulties encountered in data collection. It will also outline a number of solutions that include methodological innovations, as well as the use of alternative data sources to model (substitute) the data of interest.

Une enquête à grande échelle peut être divisée en plusieurs étapes. Un méthodologiste d'enquête devrait être en mesure d'avoir une bonne maîtrise de toutes les étapes. Une des étapes les plus cruciales est la collecte de données, et elle devient de plus en plus difficile à contrôler. Il y a plusieurs raisons pour ceci. Les coûts de collecte de données sont en augmentation et la réponse est en baisse. En plus, le mode de collecte a beaucoup changé pendant ces dernières années.

Cette présentation donnera un aperçu des difficultés rencontrées lors de la collecte des données. Il décrira également un certain nombre de solutions qui incluent des innovations méthodologiques, ainsi que l'utilisation d'autres sources de données.

**3C-A1: CJS Award Address**  
**Allocution du récipiendaire du Prix de la RCS**

---

Chair/Président: Richard Lockhart (Simon Fraser University)

**Room/Salle: Welch - David Howes**

**Abstract/Résumé**

---

**GRACE Y. YI** (University of Waterloo)

[Wednesday June 1 / mercredi 1er juin, 13:30]

*Variable Selection and Inference Procedures for Marginal Analysis of Longitudinal Data with Missing Observations and Covariate Measurement Error*

*Sélection de variables et procédures d'inférence pour l'analyse marginale de données longitudinales avec observations manquantes et erreur de mesure des covariables*

Model selection for univariate data has received extensive attention whereas research on model selection for longitudinal data remains largely unexplored. This is particularly true for data subject to missingness and measurement error. To address this important problem, we propose marginal methods that simultaneously carry out model selection and estimation for longitudinal data with missing responses and error-prone covariates. Our methods have several appealing features: broad applicability because development is for a unified framework with marginal generalized linear models; minimal model assumptions in that no full distribution is required for the response process and the distribution of the mismeasured covariates is left unspecified; straightforward implementation. To justify the proposed methods, we provide both theoretical properties and numerical assessments.

La sélection de modèles pour les données univariées a fait l'objet de nombreuses recherches, contrairement à la sélection de modèles pour les données longitudinales. Ceci est particulièrement vrai pour les données avec observations manquantes et erreur de mesure. Pour remédier à ce problème important, nous proposons des méthodes marginales qui permettent à la fois la sélection de modèles et l'estimation pour les données longitudinales avec réponses manquantes et covariables sujettes à erreur. Nos méthodes présentent plusieurs caractéristiques intéressantes : elles sont largement applicables car développées pour un cadre unifié avec des modèles linéaires généralisés marginaux ; les modèles ne nécessitent que des hypothèses minimales dans la mesure où le processus de réponse n'exige pas une distribution complète et la distribution des covariables mal mesurées n'est pas spécifiée ; ces méthodes sont faciles à mettre en oeuvre. À titre de justification, nous présentons les propriétés théoriques et des évaluations numériques des méthodes proposées.

**3C-A2: Pierre Robillard Award Address**  
**Allocution du récipiendaire du Prix Pierre-Robillard**

---

Chair/Président: Juli Atherton (UQAM)  
Organizer/Responsable: Gordon (Fick)

**Room/Salle: Thistle 325**

**Abstract/Résumé**

---

**YUJIE ZHONG** (Cambridge Institute of Public Health)

[Wednesday June 1 / mercredi 1er juin, 13:30]

*Inference About Within-Family Associations in Disease Onset Times Under Biased Sampling*

*Inférence sur les associations familiales dans les moments d'apparition de la maladie en cas d'échantillonnage biaisé*

The heritability of disease is often studied by examining within-family associations in onset times. Families for such studies typically employ biased sampling schemes in which affected individuals are sampled, along with relatives providing right-censored or current status information on disease onset times. We develop likelihood and composite likelihood methods for modeling within-family associations using copula functions and second-order regression models in which dependencies are characterized by Kendall's tau. Auxiliary data from independent individuals are exploited by augmenting composite likelihoods to increase precision of marginal parameter estimates and consequently increase efficiency in dependence parameter estimation. An application to a motivating family study in psoriatic arthritis illustrates the method and provides evidence of excessive paternal transmission of risk.

L'héritabilité des maladies s'étudie souvent via un examen des associations familiales en ce qui concerne le moment d'apparition de la maladie. Dans de telles études, les schémas d'échantillonnage sont souvent biaisés dans la mesure où les individus sont échantillonnés avec leurs proches qui fournissent des informations de statut censurées à droite ou actuelles concernant le moment d'apparition de la maladie. Nous mettons au point des méthodes de vraisemblance et de vraisemblance composée pour modéliser les associations familiales à l'aide de fonctions de copule et de modèles de régression de second ordre où les dépendances sont caractérisées par le tau de Kendall. Nous exploitons des données auxiliaires d'individus indépendants en augmentant les vraisemblances composées pour améliorer la précision des estimations des paramètres marginaux et l'efficacité de l'estimation des paramètres de dépendance. Nous illustrons la méthode par une application à une étude motivante familiale en arthrite psoriasique et prouvons la transmission paternelle excessive.

## 3D-I1: Financial and Actuarial Mathematics Mathématique actuarielle et financière

---

Organizer and Chair / Responsable et président: Cody Hyndman (Concordia University)

**Room/Salle: Thistle 242**

### Abstracts/Résumés

---

**ALEX BADESCU** (University of Calgary), **JUAN-PABLO ORTEGA** (CNRS, Université de Franche-Comté), **ZHENYU CUI** (Stevens Institute of Technology)

[Wednesday June 1 / mercredi 1er juin, 15:00]

*Non-Affine GARCH Option Pricing Models, Variance Dependent Kernels, and Diffusion Limits*

*Modèles d'évaluation d'options GARCH non affine, noyaux dépendant de la variance et limites de diffusion*

This paper investigates the pricing and weak convergence of an asymmetric non-affine, non-Gaussian GARCH model when the risk-neutralization is based on a variance dependent exponential linear pricing kernel with stochastic risk aversion parameters. The risk-neutral dynamics are obtained for a general setting and its weak limit is derived. We show how several GARCH diffusions, martingalized via well-known pricing kernels, are obtained as special cases and we derive necessary and sufficient conditions for the presence of financial bubbles. An extensive empirical analysis using both historical returns and options data illustrates the advantage of coupling this pricing kernel with non-Gaussian innovations.

J'étudie l'établissement du prix et la faible convergence d'un modèle GARCH non affine non gaussien asymétrique dans lequel la neutralisation du risque est basée sur un noyau d'établissement du prix linéaire exponentiel dépendant de la variance avec des paramètres d'aversion au risque stochastiques. La dynamique risque neutre est obtenue pour des conditions générales et sa limite faible est dérivée. Je montre comment plusieurs diffusions GARCH, martingalisées via des noyaux d'établissement du prix bien connus, sont obtenues à titre de cas spéciaux ; j'en dérive les conditions nécessaires et suffisantes pour la présence d'une bulle financière. Une analyse empirique approfondie de rendements historiques et de données relatives aux options illustre l'avantage de combiner ce noyau d'établissement du prix avec des innovations non gaussiennes.

---

**CODY HYNDMAN & SAMIRA SHIRGIR** (Concordia University)

[Wednesday June 1 / mercredi 1er juin, 15:30]

*Issuing a Convertible Bond with Call-Spread Overlay: Incorporating the Effects of Convertible Arbitrage*

*Émission d'une obligation convertible avec superposition d'opérations mixtes sur option d'achat : comment inclure les effets de l'arbitrage convertible*

Firms may attempt to mitigate some of the negative impacts of issuing convertible bonds, such as the dilution of existing shares, by concurrently entering into "call-spread overlays" or other transactions. Previous empirical studies show the stock prices of convertible bond issuers drops on the issue announcement date due to the activities of convertible bond arbitrageurs. We explore the motivation for using these combined transactions and price the convertible bonds with call-spread subject to default risk. We propose a model to estimate the

Les entreprises tentent parfois de mitiger certains effets négatifs de l'émission d'obligations convertibles, comme la dilution d'actions existantes, en effectuant simultanément des transactions de « call-spread overlays » (superposition d'opérations mixtes sur option d'achat) ou autres transactions. Des études empiriques ont montré que le cours des actions des émetteurs d'obligations convertibles chute à la date de l'annonce de l'émission en raison des activités des arbitragistes en obligations convertibles. Nous explorons la motivation de l'utilisation de ces transactions combinées et établissons le prix des obligations convertibles avec « call-

### 3D-I1: Financial and Actuarial Mathematics Mathématique actuarielle et financière

---

drop in the stock price due to convertible bond arbitrage activities at the time of planning and designing the security to be offered but before announcement. We examine the features of the model with simulated and real-world data.

spread » sous réserve du risque de défaut. Nous proposons un modèle permettant d'estimer la réduction du cours de l'action résultant des activités d'arbitrage en obligations convertibles au moment de la planification et de la conception de la valeur mobilière, avant son annonce. Nous examinons les caractéristiques du modèle avec des données simulées et réelles.

---

**ADAM METZLER** (Wilfred Laurier University), **ALEXANDRE SCOTT** (BMO)

[Wednesday June 1 / mercredi 1er juin, 16:00]

*Importance Sampling for Portfolio Loss Probabilities under Conditional Independence*

*Échantillonnage préférentiel pour probabilités de perte sur portefeuille avec indépendance conditionnelle*

We develop a novel importance sampling algorithm for estimating the probability of large portfolio losses in the conditional independence framework. We apply exponential tilts to (i) the distribution of the natural sufficient statistic of the systematic risk factors and (ii) the conditional distribution of portfolio loss, given the simulated values of the systematic risk factors, and select parameter values by minimizing the Kullback-Leibler divergence of the resulting parametric family from the ideal (zero-variance) importance density. Optimal parameter values are shown to satisfy intuitive moment-matching conditions, and the asymptotic behaviour of large portfolios is used to approximate the requisite moments. In a sense we extend Glasserman and Li (2005) to allow for heavy-tailed risk factors and/or PD-LGD correlation.

Nous développons un nouvel algorithme d'échantillonnage préférentiel permettant d'estimer la probabilité de lourdes pertes sur portefeuille dans un cadre d'indépendance conditionnelle. Nous appliquons des inclinaisons exponentielles à (i) la distribution de la statistique exhaustive naturelle des facteurs de risque systémiques et (ii) la distribution conditionnelle de la perte sur portefeuille, compte tenu des valeurs simulées des facteurs de risque systémiques, puis nous sélectionnons des valeurs de paramètres en minimisant la divergence de Kullback-Leibler de la famille paramétrique résultante de la densité d'importance idéale (variance zéro). Nous montrons que les valeurs de paramètres optimales respectent des conditions de mise en correspondance des moments intuitives et utilisons le comportement asymptotique de gros portefeuilles pour évaluer les moments nécessaires. Dans un certain sens, nous avons généralisé Glasserman et Li (2005) pour permettre des facteurs de risque « à ailes lourdes » et/ou une corrélation PD-LGD.

## 3D-I2: High Frequency Data Données de haute fréquence

---

Organizer and Chair / Responsable et président: Rafal Kulik (University of Ottawa)

Room/Salle: Thistle 325

### Abstracts/Résumés

---

**SEBASTIAN JAIMUNGAL** (University of Toronto), **ALVARO CARTEA** (Oxford University), **LUHUI GAN** (University of Toronto)

[Wednesday June 1 / mercredi 1er juin, 15:00]

*Option Pricing and Hedging with Limit and Market Orders*

*Fixation des prix en fonction de la valeur d'option et couverture avec ordres à cours limité et ordres au cours du marché*

Traditionally, option valuation is carried out using continuous time models for the underlier which are based on diffusive models, sometimes augmented with jump processes. However, such models are far from the reality, e.g., stock prices are inherently discrete and there is a bid-ask spread. Spreads act as costs for market orders, but income for limit orders. In this paper, we introduce a class of pure jump models, in continuous time, develop a calibration methodology, and demonstrate how to value and hedge options using an optimal mix of limit and market orders. Finally, we investigate the qualitative features of the strategies, and demonstrate how they differ from the naïve ones generated from continuous time models.

En général, l'évaluation des options est effectuée au moyen de modèles en temps continu pour l'assureur de base. Ces modèles sont basés sur des modèles de diffusion, parfois amplifiés par des processus de saut. Cependant, de tels modèles sont loin d'être réalistes. Par exemple, les cours des actions sont en soi discrets et il y a un écart acheteur-vendeur. Ces écarts agissent comme des coûts pour les ordres au cours du marché, mais comme des revenus pour les ordres à cours limité. Dans cet exposé, nous présentons une classe de modèles purs de saut en temps continu. Nous élaborons une méthode de calage, et nous démontrons comment évaluer et couvrir des options au moyen d'un mélange optimal d'ordres à cours limité et d'ordres au cours du marché. Enfin, nous étudions les caractéristiques qualitatives des stratégies et nous démontrons comment elles diffèrent des caractéristiques naïves produites à partir des modèles de temps continu.

---

**FREDERI VIENS & BARON LAW** (Purdue University)

[Wednesday June 1 / mercredi 1er juin, 15:30]

*A Pure-Jump Market-Making Model for High-Frequency Trading using Constrained Forward-Backward Stochastic Differential Equations*

*Un modèle de sauts pur pour teneur de marché faisant usage des équations différentielles stochastiques avant-arrière avec contrainte*

We propose a new market-making model which incorporates a number of realistic features relevant for high-frequency trading. We model the dependency structure of prices and order arrivals with cross-exciting point processes. We formulate the market maker's decisions as an optimal problem for discrete switching, allowing for over-trading

Nous proposons un nouveau modèle de tenue de marché qui intègre des caractéristiques réalistes relatives aux transactions financières à haute fréquence, avec des processus ponctuels à excitation croisée. Les décisions du teneur de marché forment un problème de commutation optimale, avec risque lié à un excès de transactions, et usage des ordres de marché pour se défaire de l'excès



## 3D-I2: High Frequency Data Données de haute fréquence

---

risk, with market orders modeled as impulse control, avoiding excessive inventory. Because of the stochastic intensities, the optimality condition falls outside of the scope of classical Hamilton-Jacobi-Bellman quasi-variational inequalities, so we leverage a newly developed constrained forward backward stochastic differential equation (FBSDE) to solve the optimal control problem. The method's implementability, which includes a Monte-Carlo requirement, is illustrated thanks to full-scale simulations. This is joint work in progress with Mr. Baron Law.

d'inventaire. La condition d'optimalité est hors du champ d'application des inégalités quasi-variationnelles classiques de Hamilton-Jacobi-Bellman. Nous nous appuyons sur un concept d'équation différentielle stochastique avant-arrière (FBSDE) pour résoudre le problème de la commande optimale, dont l'applicabilité comprend une composante de Monte-Carlo. Ceci est un travail commun en cours avec M. Baron Law.

---

**IBRAHIM ABDELRAZEQ** (Rhodes College)

[Wednesday June 1 / mercredi 1er juin, 16:00]

*Model Checking for the Stochastic Volatility Model: Lévy Driven CARMA(2,1)*

*Vérification de modèle pour modèle de volatilité stochastique : processus CARMA(2,1) régi par un processus de Lévy*

The Lévy-driven CARMA(2,1) process is a popular one with which to model stochastic volatility. However, there has been little development of statistical tools to verify this model assumption and assess the goodness-of-fit of real world data (Realized Volatility). When this process is observed at high frequencies, the unobserved driving process can be approximated from the observed process. Since, under general conditions, the Lévy-driven CARMA(2,1) process can be written as a sum of two dependent Lévy-driven Ornstein-Uhlenbeck processes, the methods developed for Ornstein-Uhlenbeck can be employed in order to use the approximated increments of the driving process to test the assumption that the process is Lévy-driven. Performance of the test is illustrated through simulation assuming that the model parameters are known.

Le processus CARMA(2,1) régi par un processus de Lévy est un processus populaire pour la modélisation de la volatilité stochastique. Cependant, peu d'outils statistiques ont été développés pour vérifier cette hypothèse de modèle et évaluer l'adéquation à des données réelles (volatilité réalisée). Lorsque ce processus est observé à haute fréquence, le processus gouvernant non observé peut être estimé à partir du processus observé. Puisque, dans des conditions générales, le processus CARMA(2,1) régi par un processus de Lévy peut s'écrire comme la somme de deux processus d'Ornstein-Uhlenbeck dépendants régis par un processus de Lévy, les méthodes développées pour Ornstein-Uhlenbeck peuvent être utilisées pour utiliser les incréments approximatifs du processus gouvernant pour tester l'hypothèse selon laquelle le processus est régi par un processus de Lévy. La performance du test est illustrée par une simulation sous l'hypothèse que les paramètres du modèle sont connus.

### 3D-I3: Network Meta-Analysis Méta-analyse de réseaux

---

Organizers/Responsables: Jemila Hamid (St. Michael's Hospital) and/et Lehana Thabane (McMaster University)

**Room/Salle: Thistle 246**

#### Abstracts/Résumés

---

**CAITLIN DALY** (McMaster University) (dalych@mcmaster.ca), **SHARON STRAUS** (Li Ka Shing Knowledge Institute, St. Michael's Hospital), **LEHANA THABANE** (McMaster University), **JEMILA S. HAMID** (Li Ka Shing Knowledge Institute, St. Michael's Hospital)

[Wednesday June 1 / mercredi 1er juin, 15:00]

*Ranking Probabilities and Surface Under Cumulative Ranking Probabilities (SUCRA) in Network Meta-Analysis (NMA)*

*Probabilités de classement et surface selon le classement cumulatif de probabilités (SUCRA) dans la méta-analyse en réseau (NMA)*

Bayesian NMA has gained more popularity over the frequentist approach since, among other mathematical and computational advantages, it leads to a framework that supports decision making through probabilities, rankings and predictions. The Bayesian method provides probabilities, indicating which of the treatments is the most effective or safest for each outcome. In particular, it provides the probability that a treatment is the first, second, third,  $n^{th}$  best and allows a ranking of treatments with respect to their comparative effectiveness. In this talk we will consider ranking probabilities in NMA, cumulative ranking probabilities, and SUCRA. NMA of chronic obstructive pulmonary disease interventions will be used to illustrate each method and their sensitivity to the type and amount of evidence available.

La NMA bayésienne est devenue plus populaire que l'approche fréquentiste puisque, entre autres avantages mathématiques et informatiques, elle conduit à un cadre qui supporte la prise de décision fondée sur les probabilités, les classements et les prédictions. La méthode bayésienne donne des probabilités, indiquant ainsi quel traitement est le plus efficace ou le plus sûr pour chaque résultat. Notamment, elle donne la probabilité qu'un traitement soit le premier, deuxième, troisième ou  $n^{ième}$  meilleur et permet un classement des traitements par rapport à leur efficacité relative. Dans cet exposé, nous examinerons les probabilités de classement dans la NMA, les probabilités cumulatives de classement et SUCRA. Les NMA d'interventions pour les maladies pulmonaires obstructives chroniques seront utilisées pour illustrer chaque méthode ainsi que leur sensibilité au type et à la quantité de preuves disponibles.

---

**JEMILA S. HAMID** (St. Michael's Hospital) (jhamid@mcmaster.ca), **CAITLIN DALY** (McMaster University), **SHARON STRAUS** (St. Michael's Hospital)

[Wednesday June 1 / mercredi 1er juin, 15:30]

*Integrating Ranking Probabilities in Network Meta-Analysis*

*Intégrer les probabilités de classement dans la méta-analyse en réseau*

Network meta-analysis (NMA) is a powerful statistical framework for synthesizing evidence from multiple treatments. NMA is often performed at an outcome level and results are provided for each outcome separately. Stakeholders and knowledge

La méta-analyse en réseau (NMA) est un cadre statistique puissant pour synthétiser des preuves provenant de traitements multiples. La NMA est souvent accomplie à un niveau d'objectifs et les résultats sont donnés pour chaque objectif séparément. De classements de sécurité et d'efficacité

### 3D-I3: Network Meta-Analysis Méta-analyse de réseaux

---

users are provided with multiple and at times conflicting comparative effectiveness and safety rankings of the treatments, leading to difficulties in decision making. Researchers are, therefore, faced with the challenge of how to integrate the information from multiple NMAs to aid decision making. The overall objective of this talk is to present conceptual, statistical and methodological frameworks for effectively integrating results from multiple NMAs. We will use data from Randomized Controlled Trials comparing treatments for Alzheimer's Dementia to illustrate the methods and show application of the methods.

multiples et parfois contradictoires sont présentés aux parties concernées et aux utilisateurs des connaissances, ce qui rend difficile la prise de décision. Les chercheurs font donc face au défi d'intégrer l'information de multiples NMAs pour aider la prise de décision. L'objectif global de cet exposé est de présenter des cadres conceptuels, statistiques et méthodologiques pour une intégration efficace des résultats de multiples NMAs. Nous utiliserons des données d'essais contrôlés aléatoires comparant des traitements de la démence d'Alzheimer pour illustrer les méthodes et pour démontrer les applications des méthodes.

---

**ARGIE VERONIKI** (St. Michael's Hospital) (veronikiA@smh.ca)

[Wednesday June 1 / mercredi 1er juin, 16:00]

*Addressing Consistency in Networks of Randomized Trials*

*Traiter de la cohérence dans les réseaux des essais aléatoires*

Network Meta-Analysis (NMA) is most justifiable if some underlying assumptions are satisfied including consistency between different sources of evidence (eg. direct and indirect treatment comparisons) in the network. In this presentation we will examine the various prerequisite assumptions in network meta-analysis with particular focus on the evaluation of consistency necessary to obtain valid results. Methodologies to evaluate consistency in networks of interventions with complex evidence structure, along with their properties will be presented. Empirical and simulation findings on inconsistency will be also discussed, including the identification of potential factors that might control its statistical detection.

La méta-analyse en réseau (NMA) est justifiée si quelques hypothèses sous-jacentes sont satisfaites, incluant la cohérence entre différentes sources de preuves (par exemple, des comparaisons de traitements directes et indirectes) dans le réseau. Dans cette présentation, nous examinerons les différentes hypothèses préalables dans la méta-analyse en réseau en se concentrant sur l'évaluation de la cohérence nécessaire pour l'obtention de résultats valides. Des méthodologies pour évaluer la cohérence dans des réseaux d'interventions avec une structure de la preuve complexe ainsi que leurs propriétés seront présentées. Des conclusions empiriques et de simulation sur l'incohérence seront aussi discutées, incluant l'identification d'autres facteurs qui pourraient contrôler sa détection statistique.

## 3D-I4: Statistical Methods for Analysing Cancer Data Méthodes statistiques pour analyse des données sur le cancer

---

Chair/Président: Gregory Pond (McMaster University)  
Organizer/Responsable: Amy Liu (Cancer Care Ontario)

Room/Salle: Welch - David Howes

### Abstracts/Résumés

---

**PATRICK BROWN & JAMIE STAFFORD** (University of Toronto), **HEDY JIANG** (Cancer Care Ontario), **PAUL NGUYEN & LENNON LI** (Public Health Ontario), **REUBEN PEREIRA & LAURA FELDMAN** (University of Toronto)  
[Wednesday June 1 / mercredi 1er juin, 15:00]

*Spatial Statistics with Cancer Registry Data*

*Les statistiques spatiales avec les données des registres du cancer*

Spatial data in cancer registries is messy; data quality and confidentiality mean only area-level data are available and low counts are suppressed. Area-level spatial models can give results which change with the spatial resolution, a phenomenon known as the 'modifiable area unit problem'. Treating aggregated spatial data as censored point data eliminates this problem, with aggregation reducing power without incurring bias. This approach will be shown with: a spatial point process model for cancer data with exact locations; a kernel-smoothing local-EM algorithm for aggregated case counts; and a spatially discrete model for case counts with censoring.

Les données spatiales dans les registres du cancer sont malcommodes; la qualité des données et la confidentialité rend nécessaire que les données soient disponibles au niveau régional seulement et les données avec faibles fréquences sont supprimées. Les modèles spatiaux au niveau de la région peuvent donner des résultats changeant avec la résolution spatiale, un phénomène nommé « problème de l'unité de la zone modifiable ». Considérer les données spatiales agrégées comme points censurés élimine ce problème, l'agrégation réduisant la puissance sans induire de biais. Cette approche sera illustrée avec : un modèle de processus ponctuel pour les données sur le cancer avec emplacements exacts ; un lissage « local-EM » pour les dénombrements de cas agrégés ; et un modèle spatialement discret pour dénombrements censurés.

---

**KAREN KOPCIUK** (University of Calgary), **LAURENT BRIOLLAIS** (Lunenfeld-Tanenbaum Research Institute), **YUN-HEE CHOI** (The University of Western Ontario), **PATRICK PARFREY & JANE GREEN** (Memorial University)  
[Wednesday June 1 / mercredi 1er juin, 15:30]

*Risk Estimation in Family Data*

*Estimation du risque dans les données familiales*

Estimating risk in families who harbour a genetic mutation predisposing them to several types of cancer presents many statistical challenges: (a) families are identified and selected directly from population disease registries or high risk cancer clinics or from two-stage sampling designs, (b) missing genetic information is common or may be completely unknown for putative genes, (c) residual familial correlation exists when additional risk genes or environmental factors are shared, and (d) sequential

L'estimation du risque chez les familles porteuses d'une mutation génétique qui les prédispose à certains types de cancer présente divers défis statistiques : (a) les familles sont identifiées et sélectionnées directement dans les registres des maladies, les centres anticancéreux à haut risque ou à l'aide de plans d'échantillonnage à deux degrés, (b) des informations génétiques manquent ou sont inconnues pour certains gènes putatifs, (c) il existe une corrélation familiale résiduelle si d'autres gènes à risque ou facteurs environnementaux sont partagés, et (d) les cancers séquentiels sont

### 3D-I4: Statistical Methods for Analysing Cancer Data

#### Méthodes statistiques pour analyse des données sur le cancer

---

cancers are frequent. Models need to take into account these features for age-at-onset outcomes, as well as direct interventions to the disease process. Genetic and disease risk models for family data we have developed will be described as well as our new R package, FamEvent.

fréquents. Les modèles doivent tenir compte de ces éléments pour l'étude d'événements de type « âge au diagnostique » et tenir compte des interventions directes dans le processus de la maladie. Nous décrivons des modèles de risque génétique et de maladie que nous avons développé pour les données familiales et notre nouveau progiciel R, FamEvent.

---

**OLLI SAARELA** (University of Toronto), **ZHIHUI (AMY) LIU** (Cancer Care Ontario)

[Wednesday June 1 / mercredi 1er juin, 16:00]

*A New Weighted Partial Likelihood Method for Estimating Marginal Structural Hazard Models*

*Nouvelle méthode de vraisemblance partielle pondérée pour l'estimation des modèles de risques structurels marginaux*

Parameters in marginal structural Cox models can be estimated through an inverse probability of treatment and censoring (IPTC) weighted Cox partial likelihood. We propose an alternative weighted partial likelihood for estimating flexible parametric marginal structural hazard models, based on case-base sampling of person-moments, resulting in a weighted logistic regression form estimating function. This enables estimation of absolute hazards, and can accommodate continuous-time IPTC weights. In terms of computational convenience, the proposed method resembles the conventional discrete time pooled logistic regression method, but works in continuous time. We study the properties of the resulting estimator theoretically and through simulations, and illustrate the method in modeling the effects of repeated treatment procedures in cancer patients.

Dans les modèles structurels marginaux de Cox, les paramètres peuvent être estimés par une vraisemblance partielle de Cox pondérée selon la probabilité de traitement et de censure (IPTC). Nous proposons une autre vraisemblance partielle pondérée pour l'estimation flexible des modèles de risques structurels marginaux paramétriques, fondée sur un échantillonnage des personnes-moments selon une approche basée sur les cas, résultant en une fonction d'estimation sous la forme de régression logistique. Cela permet une estimation des risques absolus et l'utilisation de poids IPTC en temps continu. En ce qui concerne la commodité computationnelle, la méthode proposée ressemble à la méthode de régression logistique groupée en temps discret conventionnelle, mais elle fonctionne en temps continu. Nous étudions les propriétés de l'estimateur résultant à la fois en théorie et via des simulations et illustrons la méthode en modélisant les effets de procédures de traitement répétées chez des patients atteints de cancer.

## Author List • Liste des auteurs

- Abarin, Taraneh, 57, 234 (2E-C4)  
 Abdelrazeq, Ibrahim, 65, 264 (3D-I2)  
 Abdollel, Mohamed, 61 (3B-I5)  
 Abrahamowicz, Michal, 61, 253 (3B-C1)  
 Acar, Elif, 26, 87 (1B-C3)  
 Adamic, Peter, 50, 197 (2D-C1)  
 Aeberhard, William, 41, 152 (2B-I2)  
 Ahmed, Syed Ejaz, 27 (1B-C4), 49 (2D-I5), 56 (2E-C3),  
 91 (1B-C4), 193, 194 (2D-I5), 229 (2E-C3)  
 A. Ian, McLeod, 47, 180 (2C-P)  
 Ainsworth, Jill, 25, 83 (1B-C2)  
 Akhtar-Danesh, Noori, 61, 252 (3B-C1)  
 Alam, Shomoita, 29, 96 (1C-D1)  
 Albert-Green, Alisha, 22, 69 (1B-I1)  
 Allen, Brian, 22 (1A-A1)  
 Almutiry, Waleed, 23, 73 (1B-I3)  
 Alvo, Mayer, 32, 112 (1D-C2)  
 Al-Wahsh, Huda, 46, 177 (2C-P)  
 Al-Yassin, Julan, 43, 165 (2B-C2)  
 Amanda, Cross, 43, 162 (2B-C1)  
 Amezziane, Mohamed, 56, 229 (2E-C3)  
 Anand, Sonia, 52, 207 (2D-C4)  
 Anderson-Cook, Christine M., 33 (1D-C3), 63 (3C-I1),  
 116 (1D-C3), 257 (3C-I1)  
 Andrews, Jessica, 27, 93 (1B-C5)  
 Angevaare, Justin, 46, 177 (2C-P)  
 Archibald, Samuel, 39, 146 (1E-C5)  
 Asgharian, Masoud, 38 (1E-C3), 55 (2E-C1), 141 (1E-  
 C3), 225 (2E-C1)  
 Atakora, Faisal, 29 (1C-D1), 46 (2C-P), 96 (1C-D1), 178  
 (2C-P)  
 Atherton, Juli, 37, 136 (1E-C1)  
 Auger, Nathalie, 39, 148 (1E-C5)  
 Ayilara, Olawale Fatai, 29, 96 (1C-D1)  
 Bacon, Simon L., 37, 136 (1E-C1)  
 Badescu, Alex, 64, 261 (3D-I1)  
 Bae, Taehan, 50 (2D-C1), 55 (2E-C2), 197 (2D-C1), 226  
 (2E-C2)  
 Bahamyrou, Asma, 36, 134 (1E-C1)  
 Bahraoui, Tarik, 26, 87 (1B-C3)  
 Bahuguna, Manoj, 26, 87 (1B-C3)  
 Balamchi, Shabnam, 29, 96 (1C-D1)  
 Balka, Jeremy, 34, 121 (1D-C5)  
 Banijamali, Ershad, 29, 98 (1D-I1)  
 Bartlett, John, 52, 206 (2D-C4)  
 Bartolucci, Francesco, 30, 102 (1D-I3)  
 Beaumont, Jean-François, 45, 174 (2B-C5)  
 Beck, Nicholas, 31 (1D-C1), 49 (2D-I6), 109 (1D-C1),  
 195 (2D-I6)  
 Becker, Devan, 39, 146 (1E-C5)  
 Bédard, Mylène, 51 (2D-C3), 53 (2E-I1), 203 (2D-C3),  
 212 (2E-I1)  
 Bee, Marco, 39, 146 (1E-C5)  
 Beg, Faisal, 48, 188 (2D-I2)  
 Bégin, Jean-François, 31, 109 (1D-C1)  
 Benedetti, Andrea, 46, 47, 178--180 (2C-P)  
 Bentoumi, Rachid, 32, 112 (1D-C2)  
 Berger, Yves, 59, 241 (3B-I1)  
 Bergeron, Pierre-Jérôme, 39, 146 (1E-C5)  
 Bernard, Isabelle, 29, 97 (1C-D2)  
 Bertarelli, Gaia, 30, 102 (1D-I3)  
 Beyene, Joseph, 25 (1B-C1), 32 (1D-C2), 37 (1E-C2),  
 51 (2D-C2), 52 (2D-C4), 81 (1B-C1), 113 (1D-  
 C2), 137, 139 (1E-C2), 201, 202 (2D-C2), 207  
 (2D-C4)  
 Bian, He, 29, 97 (1C-D2)  
 Bickel, David R., 48, 187 (2D-I2)  
 Bickis, Mikelis, 52 (2D-C5), 55 (2E-C2), 209 (2D-C5),  
 227 (2E-C2)  
 Bilder, Christopher R., 20  
 Bilodeau, Martin, 26, 90 (1B-C4)  
 Bingham, Derek, 22, 69 (1B-I1)  
 Binns, Malcolm, 47, 182 (2C-P)

- Blais, Lucie, 24 (1B-C1), 36 (1E-C1), 81 (1B-C1), 134 (1E-C1)
- Blejec, Andrej, 51, 205 (2D-C3)
- Blohm, Michael, 45, 174 (2B-C5)
- Blostein, Martin, 44, 168 (2B-C3)
- Bobashev, Georgiy, 57, 233 (2E-C4)
- Bogomolov, Marina, 54, 221 (2E-I5)
- Bonner, Ashley, 32 (1D-C2), 37 (1E-C2), 113 (1D-C2), 137 (1E-C2)
- Bonner, Simon, 51, 203 (2D-C3)
- Borg-Debono, Victoria, 25, 85 (1B-C2)
- Bornn, Luke, 29, 98 (1D-I1)
- Boruvka, Audrey, 49, 191 (2D-I4)
- Botter, Denise, 57, 237 (2E-C5)
- Boudreault, Mathieu, 31, 109 (1D-C1)
- Bouezmarni, Taoufik, 26 (1B-C3), 33 (1D-C4), 38 (1E-C4), 89 (1B-C3), 118 (1D-C4), 144 (1E-C4)
- Braun, John, 20, 22, 69 (1B-I1)
- Braun, W. John, 39, 146 (1E-C5)
- Bremner, Karen, 39, 147 (1E-C5)
- Brenner, Hermann, 61, 252 (3B-C1)
- Bretz, Frank, 42, 159 (2B-I5)
- Briollais, Laurent, 37 (1E-C2), 65 (3D-I4), 138 (1E-C2), 267 (3D-I4)
- Brown, Patrick, 47 (2C-P), 65 (3D-I4), 181 (2C-P), 267 (3D-I4)
- Browne, Ryan, 44, 169 (2B-C3)
- Bruce, Dunham, 34, 121 (1D-C5)
- Buckley, Matthew, 27, 93 (1B-C5)
- Bull, Shelley B., 37 (1E-C2), 46 (2C-P), 52 (2D-C4), 59 (3A-A2), 137 (1E-C2), 178 (2C-P), 207 (2D-C4), 240 (3A-A2)
- Burkett, Kelly, 60, 243 (3B-I2)
- Caetano, Samantha-Jo, 61, 252 (3B-C1)
- Cai, Jun, 32 (1D-C1), 42 (2B-I4), 111 (1D-C1), 156 (2B-I4)
- Cai, Song, 35, 129 (1E-I4)
- Cai, Tony, 54, 220 (2E-I5)
- Campbell, David, 41, 152 (2B-I2)
- Campbell, Harlan, 43, 165 (2B-C2)
- Cantoni, Eva, 41, 152 (2B-I2)
- Cao, Jingjing, 29, 97 (1C-D2)
- Caragea, Doina, 54, 218 (2E-I4)
- Cardot, Hervé, 54, 222 (2E-I6)
- Carroll, Rachel, 23, 73 (1B-I3)
- Carroll, Raymond, 44 (2B-C4), 51 (2D-C3), 171 (2B-C4), 204 (2D-C3)
- Cartea, Alvaro, 65, 263 (3D-I2)
- Chaieb, M'Hamed Lajmi Lakhal, 60, 243 (3B-I2)
- Charest, Anne-Sophie, 24 (1B-I6)
- Charpentier, Arthur, 23, 75 (1B-I4)
- Chen, Bo, 29 (1C-D2), 47 (2C-P), 52 (2D-C4), 97 (1C-D2), 179 (2C-P), 206 (2D-C4)
- Chen, Cong, 56, 232 (2E-C4)
- Chen, Fuqi, 62, 255 (3B-C2)
- Chen, Li, 32, 110 (1D-C1)
- Chen, Lichen, 32, 109 (1D-C1)
- Chen, Tinglei, 43, 165 (2B-C2)
- Chen, Xinfu, 42, 156 (2B-I4)
- Chen, Zhengfei, 25, 84 (1B-C2)
- Cheng, Lu, 52, 206 (2D-C4)
- Chenouri, Shoja'eddin, 60 (3B-I3), 62 (3B-C2), 245 (3B-I3), 256 (3B-C2)
- Cheung, Siu Hung, 50, 51, 200, 202 (2D-C2)
- Chipman, Hugh, 35, 126 (1E-I2)
- Choi, Yun-Hee, 65, 267 (3D-I4)
- Chowdhury, Monsur, 38, 144 (1E-C4)
- Chu, Jingjia, 26, 90 (1B-C4)
- Ciampi, Antonio, 25, 81 (1B-C1)
- Clarke, Stephen, 36, 131 (1E-I5)
- Coia, Vincenzo, 23, 71 (1B-I2)
- Collette, Laurence, 56, 232 (2E-C4)
- Cook, Richard J., 24 (1B-C1), 31 (1D-I6), 47 (2C-P), 49 (2D-I4), 80 (1B-C1), 182 (2C-P), 191 (2D-I4)
- Cooper-Parks, Andee, 27, 94 (1B-C5)
- Corey, Paul, 47, 183 (2C-P)
- Cossette, H el ene, 50, 198 (2D-C1)
- Cotton, Cecilia A., 32 (1D-C2), 47 (2C-P), 62 (3B-C1), 114 (1D-C2), 182 (2C-P), 254 (3B-C1)
- Craiu, Radu V., 22 (1B-I1), 26 (1B-C3), 37 (1E-C2), 52 (2D-C4), 59 (3A-A1), 70 (1B-I1), 88 (1B-C3), 137 (1E-C2), 206 (2D-C4), 239 (3A-A1)
- Cui, Zhenyu, 64, 261 (3D-I1)
- Cunningham, Chuck E, 45, 175 (2B-C5)
- Czarna, Irmina, 42, 158 (2B-I4)
- Dafni, Urania, 56, 232 (2E-C4)
- D'Al , Michele, 30, 102 (1D-I3)
- Daly, Caitlin, 65, 265 (3D-I3)
- D'Ambra, Antonello, 57, 236 (2E-C5)
- D'Ambra, Luigi, 57, 236 (2E-C5)
- Dang, Sanjeena, 44, 169 (2B-C3)
- Dang, Utkarsh J., 44 (2B-C3), 47 (2C-P), 169 (2B-C3), 181 (2C-P)
- Danieli, Coraline, 61, 253 (3B-C1)

- Dankwa, Ernest, 57, 234 (2E-C4)  
 Daoust, Pierre, 27, 93 (1B-C5)  
 David, Lieberman, 43, 162 (2B-C1)  
 Davis, Karelyn, 34, 124 (1E-I1)  
 Davison, Matt, 24, 77 (1B-I5)  
 Dawe, David, 61, 252 (3B-C1)  
 Dean, Charmaine, 22 (1B-I1), 32 (1D-C2), 69 (1B-I1), 112 (1D-C2)  
 Deardon, Rob, 23 (1B-I3), 31 (1D-I4), 43 (2B-C1), 46 (2C-P), 73 (1B-I3), 105 (1D-I4), 163 (2B-C1), 177 (2C-P)  
 Dehghan, Golsa, 25, 81 (1B-C1)  
 de Leon, Alex, 38, 143 (1E-C4)  
 Demedash, Susan, 27, 93 (1B-C5)  
 de Moliner, Anne, 54, 222 (2E-I6)  
 Deo, Rohit, 48, 189 (2D-I3)  
 Desgagné, Alain, 51, 203 (2D-C3)  
 de Souza, Russell, 52, 207 (2D-C4)  
 de Tibeiro, Jules, 57, 236 (2E-C5)  
 Dette, Holger, 42, 159 (2B-I5)  
 Devereaux, Philip J., 25, 85 (1B-C2)  
 Dey, Dipak, 57, 236 (2E-C5)  
 Doljanu, Delia Alexandra, 31, 109 (1D-C1)  
 Dong, Yuan, 38, 141 (1E-C3)  
 Donner, Allan, 51, 201 (2D-C2)  
 Duanmu, Haosui, 29 (1C-D2), 33 (1D-C4), 97 (1C-D2), 118 (1D-C4)  
 Dubin, Joel A., 36 (1E-C1), 60 (3B-I4), 62 (3B-C1), 135 (1E-C1), 247 (3B-I4), 253 (3B-C1)  
 Dubois, Lise, 60, 243 (3B-I2)  
 Duchesne, Thierry, 26, 91 (1B-C4)  
 Dupuis, Debbie, 39, 146 (1E-C5)
- Earl, Jonathan, 47, 179 (2C-P)  
 Elizabeth, Jacobs, 43, 162 (2B-C1)  
 Ellis, Peter, 61, 252 (3B-C1)  
 El-Shaarawi, Abdel, 41, 154 (2B-I3)  
 Eltinge, John, 55, 222 (2E-I6)  
 Espin-Garcia, Osvaldo, 37, 137 (1E-C2)  
 Essis-Breton, Nicolas, 41, 156 (2B-I4)  
 Esterby, Sylvia R., 41, 154 (2B-I3)  
 Estrella, Bertha, 25, 84 (1B-C2)  
 Ewusie, Joycelyne Efua, 25, 83 (1B-C2)
- Fajardo, Fabio, 48, 189 (2D-I3)  
 Falconi, Tania Alarcon, 25, 84 (1B-C2)  
 Fang, Jian, 57, 238 (2E-C5)  
 Fang, Yizhou, 50, 197 (2D-C1)
- Farooqi, Bisma, 29, 96 (1C-D1)  
 Feldman, Laura, 65, 267 (3D-I4)  
 Feng, Cindy, 31, 104 (1D-I4)  
 Feng, Tian, 37 (1E-C2), 51 (2D-C2), 139 (1E-C2), 201 (2D-C2)  
 Feng, Zeny, 46, 177 (2C-P)  
 Field, Chris, 41, 152 (2B-I2)  
 Fithian, Will, 40, 150 (2B-I1)  
 Flemming, Joanna Mills, 41, 153 (2B-I2)  
 Fontaine-Bisson, Bénédicte, 60, 243 (3B-I2)  
 Fortin, Daniel, 26, 91 (1B-C4)  
 Foster, Gary, 45, 175 (2B-C5)  
 Fox, Karla, 45, 175 (2B-C5)  
 Franczak, Brian, 44, 171 (2B-C4)  
 Freiman, Michael, 27, 93 (1B-C5)  
 Furman, Edward, 49 (2D-I6), 57 (2E-C4), 195 (2D-I6), 234 (2E-C4)
- Gagnon, Philippe, 51, 203 (2D-C3)  
 Gaillardetz, Patrice, 41, 156 (2B-I4)  
 Gallagher, Michael Patrick Brian, 44, 168 (2B-C3)  
 Gan, Luhui, 65, 263 (3D-I2)  
 Gauthier, Geneviève, 31, 109 (1D-C1)  
 Geagea, Maikel, 29, 97 (1C-D2)  
 Geerdens, Candida, 26, 87 (1B-C3)  
 George, Edward, 35, 126 (1E-I2)  
 Ghadimi, Elnaz, 32, 113 (1D-C2)  
 Ghannam, Mai, 56, 229 (2E-C3)  
 Ghodsi, Ali, 29, 98 (1D-I1)  
 Ghosh, Debashis, 30 (1D-I2), 37 (1E-C1), 100 (1D-I2), 136 (1E-C1)  
 Gibbs, Alison, 20, 34, 122 (1D-C5)  
 Girard, Claude, 20  
 Goga, Camelia, 54, 222 (2E-I6)  
 Golparvar, Leila, 36, 134 (1E-C1)  
 Gombay, Edit, 52, 209 (2D-C5)  
 Goss, Christopher H., 57, 233 (2E-C4)  
 Graham, Jinko, 31 (1D-I5), 48 (2D-I2), 106 (1D-I5), 188 (2D-I2)  
 Gravel, Christopher, 55, 224 (2E-C1)  
 Green, Jane, 65, 267 (3D-I4)  
 Greenlaw, Keelin, 31, 106 (1D-I5)  
 Greenwood, Celia, 25 (1B-C1), 60 (3B-I2), 81 (1B-C1), 244 (3B-I2)  
 Greenwood, Priscilla (Cindy), 36, 131 (1E-I5)  
 Groot, Harmke, 49, 191 (2D-I4)  
 Gruber, Susan, 30, 101 (1D-I2)  
 Guerra, Steve Ferreira, 24, 81 (1B-C1)



- Guindani, Michele, 31, 106 (1D-I5)  
 Guo, Danqiao, 29, 96 (1C-D1)  
 Guo, Yang Guan Jian, 29, 97 (1C-D2)  
 Guo, Yishan, 29, 96 (1C-D1)  
 Gupta, Samir, 43, 162 (2B-C1)  
 Guse, Jenna, 50, 197 (2D-C1)  
 Gwen, Murphy, 43, 162 (2B-C1)  
 Gweon, Hyukjun (Jay), 45, 174 (2B-C5)
- Haddadian, Rojjar, 38, 143 (1E-C4)  
 Hamid, Jemila S., 25 (1B-C2), 37 (1E-C2), 51 (2D-C2),  
 65 (3D-I3), 83 (1B-C2), 139 (1E-C2), 201 (2D-  
 C2), 265 (3D-I3)  
 Han, Peisong, 30, 100 (1D-I2)  
 Hanley, James, 46 (2C-P), 50 (2D-C2), 178 (2C-P), 200  
 (2D-C2)  
 Haocheng, Li, 32, 112 (1D-C2)  
 Harel, Daphna, 43, 161 (2B-C1)  
 Harel, Ofer, 57, 236 (2E-C5)  
 Hasan, Baktiar, 56, 232 (2E-C4)  
 Hasler, Caren, 26, 88 (1B-C3)  
 Haziza, David, 35 (1E-I4), 45 (2B-C5), 129 (1E-I4), 174  
 (2B-C5)  
 He, Fang, 43, 165 (2B-C2)  
 He, Peiyang, 29, 97 (1C-D2)  
 He, Wenqing, 27 (1B-C4), 33 (1D-C3), 92 (1B-C4), 116  
 (1D-C3)  
 He, Zhoushanyue, 29, 96 (1C-D1)  
 Heller, Ruth, 54, 221 (2E-I5)  
 Henrey, Andrew, 35, 126 (1E-I2)  
 Hidiroglou, Michael, 63, 258 (3C-I2)  
 Hlynka, Myron, 55, 227 (2E-C2)  
 HongMin, Li, 54, 218 (2E-I4)  
 Hoque, Md. Erfanul, 29, 96 (1C-D1)  
 Hossain, Shakhawat, 49, 193 (2D-I5)  
 Hosseini, Reyhaneh, 26, 90 (1B-C4)  
 Hu, Feifang, 50, 200 (2D-C2)  
 Hu, Hongyang, 29 (1C-D1), 29 (1C-D2), 96 (1C-D1), 97  
 (1C-D2)  
 Hu, Joan, 38, 142 (1E-C3)  
 Huang, Huang, 29, 97 (1C-D2)  
 Huang, Longlong, 37, 140 (1E-C3)  
 Huang, Mei Ling, 23, 71 (1B-I2)  
 Huang, Zishan, 55, 227 (2E-C2)  
 Hurvich, Clifford M., 48, 189 (2D-I3)  
 Hussein, Abdulkadir, 43 (2B-C2), 46 (2C-P), 165 (2B-  
 C2), 177 (2C-P)  
 Hyndman, Cody, 64, 261 (3D-I1)
- Ida, Ibrahima Ousmane, 45, 174 (2B-C5)
- Jaimungal, Sebastian, 65, 263 (3D-I2)  
 James, Matthew Thomas, 57, 233 (2E-C4)  
 Jandhyala, Venkata Krishna, 41, 155 (2B-I3)  
 Jang, Gun Ho, 52, 206 (2D-C4)  
 Jansen, Lina, 61, 252 (3B-C1)  
 Janssen, Paul, 26, 87 (1B-C3)  
 Jazi, Omidali Aghababaei, 38, 141 (1E-C3)  
 Jerison, Daniel, 53, 212 (2E-I1)  
 Ji, Kexin, 60, 247 (3B-I4)  
 Jian, Yang, 32, 112 (1D-C2)  
 Jiang, Bei, 61, 247 (3B-I4)  
 Jiang, Hedy, 65, 267 (3D-I4)  
 Jiang, Tianpei, 29, 96 (1C-D1)  
 Jiang, Wenjun, 32, 110 (1D-C1)  
 Jingjing, Wu, 32, 112 (1D-C2)  
 Joe, Harry, 23 (1B-I2), 26 (1B-C3), 71 (1B-I2), 88 (1B-C3)  
 John, Baron, 43, 162 (2B-C1)  
 Johnson, Steven, 47, 181 (2C-P)  
 Judd, Patricia, 44, 167 (2B-C2)
- Kaczmirek, Lars, 45, 174 (2B-C5)  
 Kalbfleisch, John D., 22, 68 (1A-A2)  
 Kampo, Regina S., 32 (1D-C2), 51 (2D-C2), 113 (1D-C2),  
 201 (2D-C2)  
 Kang, Sohee, 34, 121 (1D-C5)  
 Karim, Mohammad Ehsanul, 42, 161 (2B-C1)  
 Karimnezhad, Ali, 48 (2D-I2), 57 (2E-C5), 187 (2D-I2),  
 237 (2E-C5)  
 Kastner, Monika, 25, 83 (1B-C2)  
 Kaviul, Khan Mohammad, 25, 84 (1B-C2)  
 Kenmogne, Romain Kadje, 33, 118 (1D-C4)  
 Khalili, Abbas, 25 (1B-C2), 38 (1E-C3), 55 (2E-C1), 83  
 (1B-C2), 141 (1E-C3), 225 (2E-C1)  
 Khan, Bashir, 56, 229 (2E-C3)  
 Khan, Kaviul Mohammad, 29, 96 (1C-D1)  
 Khan, Shahedul, 38 (1E-C3), 56 (2E-C3), 140 (1E-C3),  
 230 (2E-C3)  
 Khan, Steven, 24 (1B-I6)  
 Khattree, Ravindra, 26, 87 (1B-C3)  
 Khorasany, Rahim M, 50, 198 (2D-C1)  
 Khosa, Saima, 38, 140 (1E-C3)  
 Kibret, Taddele Cherinet, 51, 201 (2D-C2)  
 Kim, Sangook, 29, 96 (1C-D1)  
 Klar, Neil, 51, 201 (2D-C2)  
 Knight, Keith, 23, 71 (1B-I2)  
 Koch, Magaly, 25, 84 (1B-C2)

- Kolkiewicz, Adam, 32, 109 (1D-C1)  
 Kong, Linglong, 42, 159 (2B-I5)  
 Kopciuk, Karen, 37 (1E-C3), 65 (3D-I4), 140 (1E-C3), 267 (3D-I4)  
 Kordas, Katarzyna, 57, 236 (2E-C5)  
 Kortbi, Othmane, 57, 236 (2E-C5)  
 Kozey-Keadle, Sarah, 44 (2B-C4), 51 (2D-C3), 171 (2B-C4), 204 (2D-C3)  
 Krahn, Murray, 39, 147 (1E-C5)  
 Kreinin, Alex, 50, 197 (2D-C1)  
 Kulperger, Reg, 26, 90 (1B-C4)  
 Kustra, Rafal, 25, 84 (1B-C2)  
 Kwakkenbos, Linda, 43, 161 (2B-C1)  
 Kye, Yisub, 57, 234 (2E-C4)
- Labbe, Aurélie, 25 (1B-C1), 48 (2D-I2), 81 (1B-C1), 187 (2D-I2)  
 Labelle, Samuel, 29, 97 (1C-D2)  
 Labouriau, Rodrigo, 61, 250 (3B-I6)  
 Lai, Hang, 29, 96 (1C-D1)  
 Lai, Yuanhao, 29 (1C-D1), 47 (2C-P), 96 (1C-D1), 180 (2C-P)  
 Lam, Gemini, 29, 96 (1C-D1)  
 Landriault, David, 42, 156 (2B-I4)  
 Larose, Chantal, 57, 236 (2E-C5)  
 Lauger, Amy, 27, 93 (1B-C5)  
 Law, Baron, 65, 263 (3D-I2)  
 Lawless, Jerald F., 62, 254 (3B-C1)  
 Lawson, Andrew, 23, 73 (1B-I3)  
 Leclerc, Martin, 60, 243 (3B-I2)  
 Lee, Chel Hee, 55, 227 (2E-C2)  
 Lee, David, 26, 88 (1B-C3)  
 Lee, Joon, 62, 253 (3B-C1)  
 Lee, Melissa, 34, 121 (1D-C5)  
 Lee, Shun Fu, 51, 201 (2D-C2)  
 Lefebvre, Geneviève, 37, 136 (1E-C1)  
 Lefebvre, Isabelle, 45, 174 (2B-C5)  
 Lefebvre, Jean-François, 60, 243 (3B-I2)  
 Lei, Jing, 40, 150 (2B-I1)  
 Lemieux, Christiane, 43, 166 (2B-C2)  
 Lemyre, Félix Camirand, 33, 118 (1D-C4)  
 Lesage, Éric, 35, 129 (1E-I4)  
 Lesperance, Mary, 31, 106 (1D-I5)  
 Levis, Alexander, 29 (1C-D2), 43 (2B-C1), 97 (1C-D2), 161 (2B-C1)  
 Levy-Leduc, Celine, 48, 189 (2D-I3)  
 Li, Bin, 42, 156 (2B-I4)  
 Li, Deli, 56, 227 (2E-C2)  
 Li, Dongchen, 42, 156 (2B-I4)  
 Li, Dongdong, 38, 142 (1E-C3)  
 Li, Haocheng, 44 (2B-C4), 51 (2D-C3), 171 (2B-C4), 204 (2D-C3)  
 Li, Lennon, 47 (2C-P), 65 (3D-I4), 181 (2C-P), 267 (3D-I4)  
 Li, Mufan, 29, 97 (1C-D2)  
 Li, Runze, 57 (2E-C5), 60 (3B-I3), 238 (2E-C5), 246 (3B-I3)  
 Li, Yubin (Éric), 44, 172 (2B-C4)  
 Li, Zeng, 60, 245 (3B-I3)  
 Liang, Han-Ying, 56, 227 (2E-C2)  
 Liang, Jiaxi, 60, 245 (3B-I3)  
 Liang, You, 52, 211 (2D-C5)  
 Lillo, Angelo, 24 (1B-I6)  
 Lin, Luyao, 32, 110 (1D-C1)  
 Lin, Nan, 54, 218 (2E-I4)  
 Lin, Shaobo, 57, 238 (2E-C5)  
 Lin, Shili, 55, 225 (2E-C1)  
 Lin, Tzuling, 24, 76 (1B-I4)  
 Lin, Wei, 29 (1C-D1), 29 (1C-D2), 45 (2B-C5), 96 (1C-D1), 97 (1C-D2), 175 (2B-C5)  
 Lin, Yisen, 29, 97 (1C-D2)  
 Lin, Zhenhua, 61, 248 (3B-I4)  
 Ling, Yun, 51, 203 (2D-C3)  
 Liu, Haiyan, 42, 156 (2B-I4)  
 Liu, Jun, 48, 189 (2D-I3)  
 Liu, Kuan, 29, 96 (1C-D1)  
 Liu, Lin, 43, 162 (2B-C1)  
 Liu, Weiwei, 46, 177 (2C-P)  
 Liu, Xin, 33, 116 (1D-C3)  
 Liu, Zhihui (Amy), 47 (2C-P), 50 (2D-C2), 66 (3D-I4), 184 (2C-P), 200 (2D-C2), 268 (3D-I4)  
 Lix, Lisa, 31 (1D-I6), 53, 216 (2E-I3)  
 Lkabous, Mohamed Amine, 42, 158 (2B-I4)  
 L'Moudden, Aziz, 57, 236 (2E-C5)  
 Lockhart, Richard, 40 (2A-A), 40 (2B-I1), 149 (2A-A), 150 (2B-I1)  
 Loepky, Jason, 20  
 Lok, Judith, 30, 101 (1D-I2)  
 Lokku, Armend, 47, 183 (2C-P)  
 Longtin, Andre, 36, 131 (1E-I5)  
 Lou, Wendy, 47, 182 (2C-P)  
 Loughin, Thomas M., 20  
 Loughin, Tom, 35, 126 (1E-I2)  
 Lu, Donghuan, 48, 188 (2D-I2)  
 Lu, Lin, 29, 96 (1C-D1)  
 Lu, Shanshan, 38, 141 (1E-C3)

- Lu, Tong-Yu, 51, 202 (2D-C2)  
 Lu, Xuewen, 24 (1B-C1), 37, 38 (1E-C3), 80 (1B-C1), 140, 141 (1E-C3)  
 Lu, Yi, 32, 110 (1D-C1)  
 Lundy, Erin, 32, 112 (1D-C2)  
 Luo, Kexin, 29, 97 (1C-D2)  
 Luo, Wei, 30, 100 (1D-I2)  
 Lysy, Martin, 36 (1E-C1), 50 (2D-C1), 51 (2D-C3), 135 (1E-C1), 197 (2D-C1), 203 (2D-C3)
- Ma, Jin, 29, 97 (1C-D2)  
 Ma, Renjun, 31 (1D-I4), 37 (1E-C2), 55 (2E-C1), 61 (3B-I6), 104 (1D-I4), 138 (1E-C2), 224 (2E-C1)  
 Madhanagopal, Banoo, 29, 97 (1C-D2)  
 Magalhães, Tiago, 57, 237 (2E-C5)  
 Mailhot, Mélina, 31 (1D-C1), 49 (2D-I6), 109 (1D-C1), 195 (2D-I6)  
 Maler, Len, 36, 131 (1E-I5)  
 Mamon, Rogemar, 62, 255 (3B-C2)  
 Mandal, Saumen, 38, 144 (1E-C4)  
 Mantel, Harold, 59, 241 (3B-I1)  
 Marceau, Etienne, 50, 198 (2D-C1)  
 Marchand, Éric, 33 (1D-C3), 57 (2E-C5), 115 (1D-C3), 236 (2E-C5)  
 Marriott, Paul, 62, 256 (3B-C2)  
 Marshall, François, 62, 255 (3B-C2)  
 Martsynyuk, Yuliya, 38, 143 (1E-C4)  
 Matthews, Charles, 44 (2B-C4), 51 (2D-C3), 171 (2B-C4), 204 (2D-C3)  
 Matthews, Steve, 62, 255 (3B-C2)  
 Mbuagbaw, Lawrence, 25, 85 (1B-C2)  
 McCulloch, Robert, 35, 126 (1E-I2)  
 McGrath, Sean, 47, 180 (2C-P)  
 McIsaac, Michael, 56, 232 (2E-C4)  
 McLaughlin, John, 47, 181 (2C-P)  
 McLeod, Ian, 47, 183 (2C-P)  
 McNicholas, Paul D., 29 (1D-I1), 43 (2B-C2), 44 (2B-C3), 44 (2B-C4), 47 (2C-P), 98 (1D-I1), 166 (2B-C2), 168, 169 (2B-C3), 171 (2B-C4), 179, 181, 185 (2C-P)  
 Menis, Jessica, 56, 232 (2E-C4)  
 Merener, Nicolas, 24, 77 (1B-I5)  
 Mesfioui, Mhamed, 32 (1D-C2), 49 (2D-I6), 112 (1D-C2), 195 (2D-I6)  
 Metzler, Adam, 64, 262 (3D-I1)  
 Millar, Anne Michele, 34, 122 (1D-C5)  
 Mills Flemming, Joanna, 41, 152 (2B-I2)  
 Mingchen, Ren, 38, 143 (1E-C4)
- Mireuta, Matei, 27, 93 (1B-C5)  
 Mitsakakis, Nicholas, 39, 147 (1E-C5)  
 Möllenhoff, Kathrin, 42, 159 (2B-I5)  
 Mongeon, Kevin, 52, 209 (2D-C5)  
 Moodie, Erica, 25, 83 (1B-C2)  
 Moon, Michael, 29, 96 (1C-D1)  
 Moradi, Fahimeh, 57, 237 (2E-C5)  
 Morales, Manuel, 49, 195 (2D-I6)  
 Moshin, Faizan, 29 (1C-D1), 29 (1C-D2), 96 (1C-D1), 97 (1C-D2)  
 Mullah, Muhammad, 46, 178 (2C-P)  
 Müller, Hans G., 61, 248 (3B-I4)  
 Murray, Lori, 34, 122 (1D-C5)
- Naik, Shanoja, 55, 226 (2E-C2)  
 Naimi, Ashley, 39, 148 (1E-C5)  
 Nangue, Aurélien Guetsop, 26, 90 (1B-C4)  
 Nasri, Bouchra, 38, 144 (1E-C4)  
 Nathoo, Farouk, 31 (1D-I5), 48 (2D-I2), 106 (1D-I5), 188 (2D-I2)  
 Naumova, Elena, 25 (1B-C2), 41 (2B-I3), 84 (1B-C2), 154 (2B-I3)  
 Negeri, Zelalem, 25, 81 (1B-C1)  
 Neupane, Binod, 51, 202 (2D-C2)  
 Neusy, Elisabeth, 59, 241 (3B-I1)  
 Nguyen, Christine, 23, 71 (1B-I2)  
 Nguyen, Paul, 47 (2C-P), 65 (3D-I4), 181 (2C-P), 267 (3D-I4)  
 Nic, Herndon, 54, 218 (2E-I4)  
 Nicolieris, Theodoros, 33, 115 (1D-C3)  
 Nicosia, Aurélien, 26, 91 (1B-C4)  
 Nie, Zixin, 27, 94 (1B-C5)  
 Nkurunziza, Sévérien, 44 (2B-C4), 56 (2E-C3), 62 (3B-C2), 172 (2B-C4), 229 (2E-C3), 255 (3B-C2)  
 Nolde, Natalia, 23 (1B-I2), 50 (2D-C1), 71 (1B-I2), 199 (2D-C1)
- Ogden, R. Todd, 61, 247 (3B-I4)  
 Oguz-Alper, Melike, 59, 241 (3B-I1)  
 Oldford, Wayne, 33 (1D-C3), 51 (2D-C3), 115 (1D-C3), 204 (2D-C3)  
 Oleksandra, Sopova, 54, 218 (2E-I4)  
 Omar, Zayd, 29, 96 (1C-D1)  
 Ortega, Juan-Pablo, 64, 261 (3D-I1)  
 Ostrouchov, George, 54, 218 (2E-I4)  
 Oualkacha, Karim, 25 (1B-C1), 60 (3B-I2), 81 (1B-C1), 244 (3B-I2)  
 Oyarzun, Javier, 27, 94 (1B-C5)

- Pang, Mengian, 29, 97 (1C-D2)  
Panju, Maysum, 38, 144 (1E-C4)  
Parfrey, Patrick, 65, 267 (3D-I4)  
Parker, Gary, 32, 110 (1D-C1)  
Patel, Nidhi, 47, 181 (2C-P)  
Pedeli, Xanthi, 56, 232 (2E-C4)  
Pei, Jingchun, 29, 97 (1C-D2)  
Pereira, Reuben, 47 (2C-P), 65 (3D-I4), 181 (2C-P), 267 (3D-I4)  
Perron, François, 33, 118 (1D-C4)  
Petkova, Eva, 61, 247 (3B-I4)  
Picard, Frédéric, 62, 255 (3B-C2)  
Piché, Alexandre, 43, 162 (2B-C1)  
Pigeon, Mathieu, 23, 75 (1B-I4)  
Pintilie, Melania, 49, 191 (2D-I4)  
Platt, Robert W., 31 (1D-I6), 36 (1E-C1), 42 (2B-C1), 53 (2E-I3), 55 (2E-C1), 134 (1E-C1), 161 (2B-C1), 216 (2E-I3), 224 (2E-C1)  
Pokharel, Gyanendra, 31, 105 (1D-I4)  
Pond, Gregory, 47 (2C-P), 61 (3B-C1), 182 (2C-P), 252 (3B-C1)  
Poon, Wai-Yin, 51, 202 (2D-C2)  
Porth, Lysa, 52, 211 (2D-C5)  
Pratola, Matthew, 35, 126 (1E-I2)  
Price, Steven, 51, 203 (2D-C3)  
Provost, Serge, 27 (1B-C4), 41 (2B-I3), 91 (1B-C4), 154 (2B-I3)  
Puglia, Steve, 55, 226 (2E-C2)  
Pullenayegum, Eleanor, 25 (1B-C2), 49 (2D-I4), 85 (1B-C2), 192 (2D-I4)  
Punzo, Antonio, 44, 168 (2B-C3)  
  
Qian, Wei, 40, 151 (2B-I1)  
Qin, Yingli, 33, 115 (1D-C3)  
Qinwen, Wang, 60, 245 (3B-I3)  
Quach, Harris, 29, 97 (1C-D2)  
Quan, Hude, 57, 233 (2E-C4)  
Quessy, Jean-François, 25 (1B-C1), 33 (1D-C4), 82 (1B-C1), 118 (1D-C4)  
Quon, Bradley S., 57, 233 (2E-C4)  
  
Rabhi, Yassir, 26, 89 (1B-C3)  
Rahman, Adam, 33, 115 (1D-C3)  
Rai, Kamal, 36, 135 (1E-C1)  
Rak, Leonard, 24 (1B-I6)  
Ramezan, Reza, 62, 256 (3B-C2)  
Ramsay, James, 53, 214 (2E-I2)  
Ranalli, M. Giovanna, 30, 102 (1D-I3)  
  
Rao, J.N.K., 27 (1B-C5), 30 (1D-I3), 35 (1E-I4), 59 (3B-I1), 95 (1B-C5), 102 (1D-I3), 129 (1E-I4), 241 (3B-I1)  
Reesor, R. Mark, 50, 198 (2D-C1)  
Reid, Nancy, 45, 175 (2B-C5)  
Reisen, Valderio, 48, 189 (2D-I3)  
Reiter, Jerome, 27, 93 (1B-C5)  
Rémillard, Bruno, 55, 226 (2E-C2)  
Ren, Jiandong, 32, 110 (1D-C1)  
Renaud, Jean-François, 42, 158 (2B-I4)  
Riegert, David, 39, 147 (1E-C5)  
Rivest, Louis-Paul, 25 (1B-C1), 26 (1B-C3), 26 (1B-C4), 38 (1E-C4), 45 (2B-C5), 82 (1B-C1), 88 (1B-C3), 91 (1B-C4), 145 (1E-C4), 174 (2B-C5)  
Romanescu, Razvan, 43, 163 (2B-C1)  
Rosalsky, Andrew, 56, 227 (2E-C2)  
Rosenthal, Jeffrey, 53, 214 (2E-I2)  
Rossi, Amanda M., 37, 136 (1E-C1)  
Rostamiforooshani, Mehdi, 31, 104 (1D-I4)  
Rosychuk, Rhonda, 23 (1B-I3), 31 (1D-I6), 61 (3B-I5), 74 (1B-I3)  
Rothstein, Steven, 44, 169 (2B-C3)  
Rowat, Peter, 36, 131 (1E-I5)  
Roy, Daniel, 33 (1D-C4), 35 (1E-I2), 118, 119 (1D-C4), 127 (1E-I2)  
Roy-Gagnon, Marie-Hélène, 60, 243 (3B-I2)  
Ruiz-Trujillo, Jorge-Enrique, 29, 97 (1C-D2)  
  
Saarela, Olli, 47 (2C-P), 50 (2D-C2), 66 (3D-I4), 184 (2C-P), 200 (2D-C2), 268 (3D-I4)  
Saha, Krishna, 43, 163 (2B-C1)  
Saleh, A.K.Md. Ehsanes, 56, 229 (2E-C3)  
Sampson, Joshua, 44, 171 (2B-C4)  
Sandoval, Monica, 57, 237 (2E-C5)  
Sang, Qiuguang, 55, 224 (2E-C1)  
Santiago, Anna Theresa, 47, 182 (2C-P)  
Sattar, Abdus, 34, 124 (1E-I1)  
Savage, Jennifer, 37, 136 (1E-C1)  
Schaapveld, Michael, 49, 191 (2D-I4)  
Schnitzer, Mireille, 24 (1B-C1), 30 (1D-I2), 36 (1E-C1), 81 (1B-C1), 101 (1D-I2), 134 (1E-C1)  
Schofield, Matthew, 51, 203 (2D-C3)  
Schonlau, Matthias, 45, 174 (2B-C5)  
Schorning, Kirsten, 42, 159 (2B-I5)  
Scott, Alexandre, 64, 262 (3D-I1)  
Sempértegui, Fernando, 25, 84 (1B-C2)  
Sen, Arusharka, 32, 113 (1D-C2)  
Sendova, Kristina, 42, 157 (2B-I4)

- Shaikh, Mateen, 25 (1B-C1), 52 (2D-C4), 81 (1B-C1), 207 (2D-C4)
- Shams, Shahriar, 25 (1B-C2), 47 (2C-P), 85 (1B-C2), 183 (2C-P)
- She, Xichen, 35, 129 (1E-I4)
- Shen, Hua, 62, 253 (3B-C1)
- Shentu, Yue, 56, 232 (2E-C4)
- Shi, Peng, 23, 75 (1B-I4)
- Shi, Yidan, 49, 191 (2D-I4)
- Shin, Jihyung, 46 (2C-P), 52 (2D-C4), 178 (2C-P), 207 (2D-C4)
- Shirgir, Samira, 64, 261 (3D-I1)
- Shokoohi, Farhad, 48 (2D-I2), 55 (2E-C1), 187 (2D-I2), 225 (2E-C1)
- Shrier, Ian, 39 (1E-C5), 43 (2B-C1), 55 (2E-C1), 148 (1E-C5), 162 (2B-C1), 225 (2E-C1)
- Shu, Di, 36, 135 (1E-C1)
- Sikdar, Khokan C., 57, 233 (2E-C4)
- Silva, Anjali, 44, 169 (2B-C3)
- Simard, Jacques, 60, 243 (3B-I2)
- Singh, Radhey, 38, 141 (1E-C3)
- Sinha, Sanjoy, 34, 124, 125 (1E-I1)
- Slavchev, Pavel, 39, 148 (1E-C5)
- Slimene, Latifa Ben Hadj, 33, 115 (1D-C3)
- Small, Christopher G., 60, 245 (3B-I3)
- Solari, Fabrizio, 30, 102 (1D-I3)
- Soleymani, Samira, 29 (1C-D1), 47 (2C-P), 96 (1C-D1), 183 (2C-P)
- Soltanifar, Mohsen, 47, 183 (2C-P)
- Song, Peter XK, 61, 250 (3B-I6)
- Southern, Danielle, 57, 233 (2E-C4)
- Spiegelhalter, David, 36, 133 (1E-I6)
- Springford, Aaron, 39 (1E-C5), 47 (2C-P), 147 (1E-C5), 184 (2C-P)
- Sriskandarajah, Suj, 29, 96 (1C-D1)
- Stafford, Jamie, 65, 267 (3D-I4)
- Stallard, Jim, 48 (2D-I1)
- Stanojevic, Sanja, 57, 233 (2E-C4)
- Stearns, Jen, 52, 207 (2D-C4)
- Steele, Russell, 39 (1E-C5), 43 (2B-C1), 46 (2C-P), 55 (2E-C1), 148 (1E-C5), 162 (2B-C1), 178 (2C-P), 225 (2E-C1)
- Stein, Lincoln, 52, 206 (2D-C4)
- Stephens, David A., 36 (1E-C1), 48 (2D-I2), 134 (1E-C1), 187 (2D-I2)
- Stephenson, Anne L., 57, 233 (2E-C4)
- Stevens, Nathaniel, 33, 116 (1D-C3)
- St-Hilaire, André, 38, 144 (1E-C4)
- Straus, Sharon, 25 (1B-C2), 65 (3D-I3), 83 (1B-C2), 265 (3D-I3)
- Strawderman, William E., 57, 236 (2E-C5)
- Study group, BeAMS, 42, 161 (2B-C1)
- Sun, Jianping, 60, 244 (3B-I2)
- Sun, Lei, 52, 206 (2D-C4)
- Sun, Makosso-Kallyth, 44, 172 (2B-C4)
- Sun, Shaun Zheng, 52, 210 (2D-C5)
- Sun, Wenguang, 54, 220 (2E-I5)
- Surette, Michael, 52, 207 (2D-C4)
- Sutradhar, Rinku, 61 (3B-I5)
- Swishchuk, Anatoliy, 24, 77 (1B-I5)
- Sykes, Jenna, 57, 233 (2E-C4)
- Szefer, Elena, 31 (1D-I5), 48 (2D-I2), 106 (1D-I5), 188 (2D-I2)
- Taback, Nathan, 34, 122 (1D-C5)
- Talbot, Denis, 37, 136 (1E-C1)
- Tan, Ken Seng, 42, 157 (2B-I4)
- Tan, Tingbin, 29, 97 (1C-D2)
- Tang, Yanbo, 29, 97 (1C-D2)
- Tang, Yang, 44, 168 (2B-C3)
- Taniguchi, Yoshihiro, 43, 166 (2B-C2)
- Taqqu, Murad, 48, 189 (2D-I3)
- Tarpey, Thaddeus, 61, 247 (3B-I4)
- Taylor, Jonathan, 40, 150 (2B-I1)
- Taylor, Steve, 47, 184 (2C-P)
- Thabane, Lehana, 25 (1B-C2), 45 (2B-C5), 54 (2E-I3), 65 (3D-I3), 85 (1B-C2), 175 (2B-C5), 217 (2E-I3), 265 (3D-I3)
- Thavaneswaren, Aera, 48, 190 (2D-I3)
- Thombs, Brett, 43, 161 (2B-C1)
- Thomplinson, George, 25, 83 (1B-C2)
- Thompson, Mary, 53, 214 (2E-I2)
- Thomson, David, 39 (1E-C5), 47 (2C-P), 147 (1E-C5), 184 (2C-P)
- Thomson, Trevor, 49, 193 (2D-I5)
- Tibshirani, Robert, 40, 150 (2B-I1)
- Tibshirani, Ryan, 40, 150 (2B-I1)
- Torabi, Mahmoud, 30, 103 (1D-I3)
- Toupin, Marie-Hélène, 25, 82 (1B-C1)
- Trapin, Luca, 39, 146 (1E-C5)
- Tremlett, Helen, 42, 161 (2B-C1)
- Tsai, Cary Chi-Liang, 24, 76 (1B-I4)
- Tshitumbu, Danny, 29, 96 (1C-D1)
- Tuenter, Hans, 24, 78 (1B-I5)
- Turgeon, Maxime, 25, 81 (1B-C1)
- Umphrey, Gary J., 52, 210 (2D-C5)

- Vaillancourt, Jean, 55, 226 (2E-C2)  
van der Laan, Mark, 30, 101 (1D-I2)  
Vanniyasingam, Thuva, 45, 175 (2B-C5)  
Veitch, Victor, 20, 33, 119 (1D-C4)  
Veroniki, Argie, 65, 266 (3D-I3)  
Vidyshankar, Anand N., 40, 150 (2B-I1)  
Viens, Frederi, 65, 263 (3D-I2)  
Vihola, Matti, 53, 213 (2E-I1)  
Vincent, Kyle, 27, 94 (1B-C5)  
Volgushev, Stanislav, 42, 159 (2B-I5)  
Vrbik, Irene, 45, 172 (2B-C4)
- Waddell, Adrian, 51, 204 (2D-C3)  
Wang, Cheng, 60, 243 (3B-I2)  
Wang, Guanbo, 29, 97 (1C-D2)  
Wang, Hong, 57, 234 (2E-C4)  
Wang, Liqun, 38, 143 (1E-C4)  
Wang, Lu, 30, 100 (1D-I2)  
Wang, Ruodu, 42 (2B-I4), 50 (2D-I6), 156 (2B-I4), 196 (2D-I6)  
Wang, Shouao, 29, 96 (1C-D1)  
Wang, Weinan, 54, 220 (2E-I5)  
Wang, Xiaotian, 29, 97 (1C-D2)  
Wang, Xiaoxiao, 29, 97 (1C-D2)  
Wang, Xikui, 52, 211 (2D-C5)  
Wang, Ying, 32, 111 (1D-C1)  
Ward, Lawrence, 36, 131 (1E-I5)  
Wasserman, Larry, 40, 150 (2B-I1)  
Waudby-Smith, Ian, 47, 184 (2C-P)  
Wedderkopp, Niels, 43, 162 (2B-C1)  
Wei, Qian, 27, 95 (1B-C5)  
Wei, Yuhong, 43, 166 (2B-C2)  
Weng, Chengguo, 42, 157 (2B-I4)  
White, Bethany, 48 (2D-I1)  
Wiens, Doug, 42, 159 (2B-I5)  
Wirjanto, Tony, 32, 109 (1D-C1)  
Wu, Changbao, 35 (1E-I4), 59 (3B-I1), 129 (1E-I4), 241 (3B-I1)  
Wu, Jingjing, 24 (1B-C1), 38 (1E-C3), 80 (1B-C1), 141 (1E-C3)  
Wu, Ka Ho, 56, 230 (2E-C3)  
Wu, Pan, 32, 114 (1D-C2)  
Wu, Yilei, 33, 115 (1D-C3)  
Wu, Ying, 24, 80 (1B-C1)
- Xi, Da Zhong (Dexen), 47, 184 (2C-P)  
Xi, Ruibin, 54, 218 (2E-I4)  
Xi, Yao, 29, 97 (1C-D2)
- Xiaojian, Xu, 55, 226 (2E-C2)  
Xie, Yuying, 32, 114 (1D-C2)  
Xu, Chen, 57, 238 (2E-C5)  
Xu, Jingxiong, 37, 138 (1E-C2)  
Xu, Wei, 37, 138 (1E-C2)  
Xu, Wenfu, 50, 200 (2D-C2)  
Xu, Ximing, 41, 152 (2B-I2)  
Xu, Yingying, 62, 253 (3B-C1)
- Yan, Guohua, 37 (1E-C2), 55 (2E-C1), 138 (1E-C2), 224 (2E-C1)  
Yan, Jiajun, 29, 97 (1C-D2)  
Yang, Jun, 33, 119 (1D-C4)  
Yang, Lu, 23, 75 (1B-I4)  
Yang, Ping, 51, 202 (2D-C2)  
Yang, Po, 46, 178 (2C-P)  
Yang, Yi, 40, 151 (2B-I1)  
Yao, Fang, 61, 248 (3B-I4)  
Yao, Jeff, 60, 245 (3B-I3)  
Yapa, Gaitri, 34, 121 (1D-C5)  
Yauck, Mamadou, 38, 145 (1E-C4)  
Yeasmin, Fahmida, 29, 96 (1C-D1)  
Yekutieli, Daniel, 54, 221 (2E-I5)  
Yi, Grace Y., 35 (1E-I3), 36 (1E-C1), 64 (3C-A1), 128 (1E-I3), 135 (1E-C1), 259 (3C-A1)  
Yi, Ruiyang, 46, 178 (2C-P)  
Yilmaz, Yildiz, 44, 167 (2B-C2)  
Ying, Yan, 38, 143 (1E-C4)  
Yu, Hao, 20, 26, 90 (1B-C4)  
Yu, Qing, 37, 138 (1E-C2)  
Yu, Ruoqi, 29 (1C-D1), 29 (1C-D2), 96 (1C-D1), 97 (1C-D2)  
Yuen, Wai Kong, 22 (1A-A1), 23, 71 (1B-I2)  
Yüzbaşı, Bahadır, 49, 194 (2D-I5)
- Zacks, Shelemyahu, 41, 155 (2B-I3)  
Zareamoghaddam, Hossein, 27, 91 (1B-C4)  
Zarepour, Mahmoud, 26, 90 (1B-C4)  
Zeng, Leilei, 49, 191 (2D-I4)  
Zhang, Chunming, 54, 220 (2E-I5)  
Zhang, Gong, 29, 97 (1C-D2)  
Zhang, Jingong, 42, 157 (2B-I4)  
Zhang, Ruixi, 42, 157 (2B-I4)  
Zhang, Runchu, 33, 120 (1D-C4)  
Zhang, Xu Xuan, 47, 185 (2C-P)  
Zhang, Yukun, 44 (2B-C4), 51 (2D-C3), 171 (2B-C4), 204 (2D-C3)  
Zhao, Bangxin, 27, 92 (1B-C4)

Zhao, Meng, 55, 225 (2E-C1)  
Zhao, Xiaofei, 47, 180 (2C-P)  
Zhao, Xu, 29, 96 (1C-D1)  
Zhao, Yang, 45, 173 (2B-C4)  
Zhong, Ping-Shou, 60, 246 (3B-I3)  
Zhong, Wenyan, 24 (1B-C1), 29 (1C-D1), 80 (1B-C1), 96  
(1C-D1)  
Zhong, Yujie, 64, 260 (3C-A2)  
Zhou, Julie, 42, 159 (2B-I5)  
Zhou, Zhiyang, 33, 120 (1D-C4)  
Zhu, Mu, 33 (1D-C3), 52 (2D-C4), 115 (1D-C3), 206 (2D-  
C4)  
Zhu, Yayuan, 62, 254 (3B-C1)  
Zhu, Yeying, 30 (1D-I2), 32 (1D-C2), 37 (1E-C1), 100  
(1D-I2), 114 (1D-C2), 136 (1E-C1)  
Ziegel, Johanna, 50, 199 (2D-C1)  
Zitikis, Ricardas, 49, 195 (2D-I6)  
Zou, Hui, 40, 151 (2B-I1)  
Zou, Yuxin, 29 (1C-D1), 29 (1C-D2), 96 (1C-D1), 97 (1C-  
D2)  
Zuo, Fei, 29, 97 (1C-D2)

# Welcome to Brock

Conference Services  
905-688-5550 x 3369  
confdesk@brocku.ca



ACCOMMODATION  
PARKING  
(Must have parking pass)

SANKEY CHAMBERS

ACCOMMODATION  
CHECK-IN

POND INLET

DECEW DINING HALL

THISTLE COMPLEX

REGISTRATION

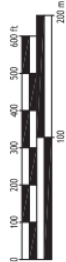
DAILY PARKING  
(\$7 per entry)

MARKET HALL

ACADEMIC SOUTH

WELCH HALL

D HOWE'S THEATRE



BROCK UNIVERSITY

