

Table of Contents • Table des matières

Bienvenue • Welcome	2
Sponsors • Commanditaires	4
Organizers • Organisateurs	5
Exhibitors • Exposants	5
Language of Presentations • Langue des présentations	6
On-site Registration • Inscription sur place	6
General Information • Information générale	6
Committees and Meetings • Comités et réunions	13
List of Sessions • Liste des sessions	17
Program • Programme	23
Abstracts • Résumés	70
Author Index (per session) • Index des auteurs (par session)	221

Bienvenue • Welcome

Bienvenue au campus Point Grey de l'Université de Colombie-Britannique, situé sur le territoire traditionnel de la nation Musqueam. Le campus Point Grey occupe 402 hectares et est entouré d'un parc forestier de 763 hectares. La vue à partir du côté nord du campus, donnant sur la baie Anglaise et les montagnes de la rive nord, est spectaculaire. En 2008, plus de 36 000 étudiants de premier cycle et 8 000 étudiants de cycles supérieurs y étudiaient.

Vancouver est une ville cosmopolite avec un mélange de plusieurs cultures. Suivant l'anglais et le chinois, les langues maternelles y étant les plus courantes sont le pendjabi, l'allemand, l'italien, le français, le tagalog et l'espagnol. Plus de la moitié des enfants d'âge scolaire parlent une autre langue que l'anglais à la maison.

Vancouver est située dans le territoire traditionnel des nations salish côtières, qui avaient déjà peuplé la région en 500 av J.-C. La ville fut baptisée Vancouver en 1886, en l'honneur du capitaine George Vancouver, un capitaine anglais qui explora la région en 1792. Vancouver est la troisième plus grande ville du Canada, et 2,1 millions de personnes résident dans la région métropolitaine de Vancouver. La région est de climat doux tout au long de l'année, et les températures maximales quotidiennes moyennes pour mai et juin sont de 18 et 21 degrés Celsius (65 et 69 degrés Fahrenheit) respectivement. En 2010, Vancouver accueillera les Jeux olympiques d'hiver et le Congrès de statistique conjoint de l'American Statistical Association.

Une courte liste des attractions offertes par Vancouver comprend : du magasinage le long de l'élégante rue Robson, une variété époustouflante de restaurants à travers la ville, l'Aquarium dans l'impressionnant parc Stanley -tout près du centre-ville-, une galerie d'art moderne et une variété de quartiers et de commerce éclectiques et variés à visiter, tels Kitsilano, ou l'île Granville, avec son marché public et ses commerces artisanaux. Il y a aussi plusieurs kilomètres de pistes cyclables et de sentiers de randonnées urbains et forestiers. Par exemple, il est possible de partir du campus UBC et de longer la rive à pied ou en vélo jusqu'au parc Stanley et de contourner celui-ci pour atteindre le côté nord de la péninsule du centre-ville. Nous vous encourageons à saisir cette opportunité d'explorer notre ville et notre province, et souhaitons que le temps que vous passerez ici soit à la fois productif et agréable.

Welcome to the Point Grey campus of The University of British Columbia, located in traditional territory of the Musqueam Nation. The Point Grey campus is 402 hectares large and is surrounded by 763 hectares of forested park. The north end of the campus provides spectacular views of English Bay and the North Shore Mountains. In 2008 there were over 36,000 undergraduate students, and over 8,000 graduate students enrolled at this campus.

Vancouver is a vibrant cosmopolitan city with a mix of many cultures. After English and Chinese the most common mother tongue languages are Punjabi, German, Italian, French, Tagalog and Spanish. More than half of the city's school-age children have been raised speaking a language other than English.

Vancouver is located in traditional territory of the Coast Salish Nations, who had settled the area by 500 BC. The city was incorporated in 1886 and named Vancouver after Captain George Vancouver, a British naval captain who explored the area in 1792. Vancouver is the third largest city in Canada, with a population of approximately 2.1 million in the Greater Vancouver metropolitan region and is surrounded by stunning natural beauty. The region enjoys a mild climate year-round with average daily maximum temperatures for May and June of 18 and 21 degrees Celcius (65 and 69 degrees Fahrenheit) respectively. In 2010 Vancouver will host the Winter Olympic Games and the Joint Statistical Meetings of the American Statistical Association.

A short list of the amenities Vancouver has to offer includes: elegant street shopping along Robson street, an amazing variety of restaurant options across the city, an exciting Aquarium inside the impressive

Stanley Park –right next to the downtown core– a modern art gallery and a number of different and eclectic neighbourhoods and shopping areas to visit like Kitsilano, or Granville Island, with its Public Market and artisan shops. There are also many kilometres of urban and forest hiking and bicycling trails. For example, it is possible to walk or cycle along the sea shore starting from the UBC campus, all the way to Stanley Park, then around it to reach the north side of the downtown peninsula.

We encourage you to take this opportunity to explore our city and province, and sincerely hope you have a productive and at the same time relaxing time here.

Sponsors • Commanditaires

The 2009 Annual meeting of the SSC thanks the following sponsors for their generous contributions towards the success of this meeting: the Department of Statistics at The University of British Columbia, Pacific Institute for the Mathematical Sciences (PIMS), Centre de Recherches Mathématiques (CRM), Fields Institute for Research in Mathematical Sciences, Mathematics of Information Technology and Complex Systems (MITACS), Michael Smith Laboratories, the Department of Mathematics and Statistics at the University of Victoria, the Faculty of Science at The University of British Columbia, the School of Population and Public Health at The University of British Columbia, the Department of Statistics and Actuarial Science at Simon Fraser University, Capital One, FPIInnovations, STATA and Syreon Corporation.

Le congrès annuel 2009 de la SSC remercie les commanditaires suivants pour leurs généreuses contributions au succès de ce congrès : le département de statistique de l'Université de Colombie-Britannique, le Pacific Institute for the Mathematical Sciences (PIMS), le Centre de Recherches Mathématiques (CRM), le Fields Institute for Research in Mathematical Sciences, Mathematics of Information Technology and Complex Systems (MITACS), Michael Smith Laboratories, le département de mathématiques et de statistique de l'Université Victoria, la faculté des sciences de l'Université de Colombie-Britannique, l'école de santé publique de l'Université de Colombie-Britannique, le département de statistique et d'actuariat de l'Université Simon Fraser, Capital One, FPIInnovations, STATA et Syreon Corporation.

Organizers • Organisateurs

Local Arrangements Committee • Comité d'organisation

The Local Arrangements Committee would like to acknowledge the help and support received from a number of people and organizations. Many people contributed to what you will see, hear, touch and taste during these short three days.

Individuals that were of particular help with the preparations for this meeting include the host Department's hard-working, resourceful and fantastic administrative staff, faculty and graduate students; Adam Wojtowicz at the Pacific Institute for the Mathematical Sciences (PIMS); Mike Danilov (our webmaster); The Ha (for IT support); and Aline Tabet (our Graduate Students Representative). We would also like to thank PIMS for their generous in-kind contributions (including very valuable time of their helpful staff), and Simon Peacock the Dean of Science. Finally, we have enjoyed the extremely professional and kind services of the UBC Conference Services, particularly Heidi Dale-Johnson.

Le comité d'organisation souhaite souligner l'aide et le soutien apporté par plusieurs individus et organismes. Plusieurs personnes ont contribué à ce que vous allez voir, entendre, toucher et goûter durant ces trois courtes journées.

Les individus ayant apporté une aide particulière aux préparations de ce congrès comprennent le personnel administratif du département de statistique d'UBC, ainsi que son personnel enseignant et ses étudiants diplômés ; Adam Wojtowicz du Pacific Institute for the Mathematical Sciences (PIMS) ; Mike Danilov (notre webmestre) ; The Ha (pour le soutien informatique) ; et Aline Tabet (notre représentante des étudiants de cycles supérieurs). Nous souhaitons également remercier PIMS pour le temps très précieux de leur personnel, et Simon Peacock le doyen des sciences. Finalement, nous avons bénéficié des services très professionnels des services de conférence d'UBC, particulièrement de Heidi Dale-Johnson.

Exhibitors • Exposants

Displays for examination and purchase will be available in the lobby of the IRC building. Confirmed exhibitors at the time this program went to press are: Nelson Education, Springer, STATA and Wiley-Blackwell.

Des kiosques de vente seront situés dans le hall du pavillon IRC. Les exposants confirmés au moment de mettre ce programme sous presse sont : Nelson Education, Springer, STATA and Wiley-Blackwell.

Language of Presentations • Langue des présentations

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

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 (F), (E), alors que (F), (E) indiquent le support visuel. La lettre à l'intérieur identifie la langue : F pour français et E pour anglais (English).

On-site Registration • Inscription sur place

Registration will be in the lobby of the Leonard S. Klink building (LSK) on Sunday May 31st from 8 am to 12 noon, and at the Walter Gage Residence, Fort Camp Lounge from 12 to 6 pm. The registration desk will also be available at the Woodward IRC Upper Lobby on Monday (June 1st), from 7 am to 5 pm; and on Tuesday and Wednesday (June 2nd and 3rd) from 8 am to 5 pm.

L'inscription sur place se fera dans le hall du pavillon Leonard S. Klink (LSK) de 8h00 à 12h00 et au hall Fort Camp de la Walter Gage Residence de 12h00 à 18h00, le dimanche 31 mai. Le bureau d'inscription sera aussi ouvert au hall supérieur du pavillon Woodward IRC le lundi (1er juin) de 7h00 à 17h00, ainsi que mardi et mercredi (2 et 3 juin) de 8h00 à 17h00.

General Information • Information générale

Directions • Emplacement

Campus and city maps are included in your registration package. You will also find lists of food outlets on campus or within walking distance, along with their opening hours. Directions from off-campus hotels are available at the conference website. Please visit http://www.ssc.ca/2009/hotels_e.html

Des cartes du campus et de la ville sont incluses dans les sacs du congrès. Vous y trouverez aussi une liste de restaurants sur le campus ou à proximité, ainsi que leurs heures d'ouverture. Des directives pour se

rendre au campus à partir des hôtels hors campus sont disponibles sur le site web du congrès. Visitez la page http://www.ssc.ca/2009/hotels_f.html

Campus Security • Sécurité sur le campus

Please note that certain areas of the UBC campus may be sparsely populated, particularly after sunset, and so, for your personal safety, care should be taken. In particular, we urge you to not travel alone in deserted areas. In case of an emergency, dial 911. Campus security can be reached at 604 822-2222. There are also blue light phones located on campus, which provide a direct connection to security. These can be used for emergencies and for general assistance or information, and are identified on the customized campus map in your registration package.

There are 3 community shuttle routes at UBC that travel between Alma and Broadway along Spanish Banks to UBC, and connect Totem Park, the Botanical Gardens, Nitobe Gardens, the Museum of Anthropology and the Chan Centre, among others. All routes begin and terminate at the north side of the Student Recreation Centre, adjacent to the diesel bus loop, and provide connections with existing services at UBC Loop and service to UBC Hospital. For more information, please visit http://www.trek.ubc.ca/programs/community_shuttle/index.html

Veuillez noter que certains endroits sur le campus peuvent être presque déserts, surtout après le coucher du soleil. Il est donc nécessaire, pour votre sûreté, d'être vigilant. En particulier, il n'est pas recommandé de se déplacer seul dans les endroits déserts. En cas d'urgence, composez le 911. La sécurité sur le campus peut être rejointe au 604 822-2222. Plusieurs téléphones bleus sur le campus sont en connexion directe avec la sécurité. Ils peuvent être utilisés en cas d'urgence, ou pour obtenir de l'information ou assistance générale, et sont identifiés sur la carte du campus du sac du congrès. Il y a trois navettes sur le campus circulant entre Alma et Broadway le long de Spanish Banks jusqu'à l'université et reliant (entre autres) Totem Park, les jardins botaniques et Nitobe, le Musée d'anthropologie et le Chan Centre. Les circuits des navettes débutent et se terminent au côté nord du Student Recreation Centre, près du circuit d'autobus diesel. Les circuits font liaison avec des services existants du UBC Loop et desservent l'Hôpital UBC. Pour plus d'information, veuillez visiter http://www.trek.ubc.ca/programs/community_shuttle/index.html

Rooms • Salles

Most sessions will be held in the Woodward IRC building (IRC), while others are located in the following buildings: Food Nutrition and Health (FNH), Wesbrook (WES), Chemistry (CHE), Michael Smith Labs (MSL) and Scarfe (SCR). Please refer to the customized campus map in your registration package.

La plupart des sessions auront lieu dans le pavillon Woodward IRC (IRC), mais certaines seront situées dans les édifices suivants : Food Nutrition and Health (FNH), Wesbroom (WES), Chemistry (CHE), Michael Smith Labs (MSL) et Scarfe (SCR). Veuillez vous référer à la carte du campus dans votre sac du congrès.

Workshops • Ateliers

The workshops organized by the Biostatistics, Survey Methods and Industrial Statistics sections of the SSC will take place on Sunday May 31st. The Biostatistics Workshop will take place in MSL 101, whereas the workshops of the Survey Methods and Industrial Statistics sections will take place in LSK 201 and 202.

Les ateliers organisés par les groupes de biostatistique, de méthodes d'enquête, et de statistique industrielle et de gestion auront lieu le dimanche 31 mai. L'atelier de biostatistique aura lieu au MSL 101, alors que les ateliers de méthodes d'enquête et de statistique industrielle et de gestion seront au LSK201 et 202, respectivement.

Poster Session • Séance d'affichage

Regular poster sessions will be held in the Woodward IRC Lobby. Special poster sessions will be held in Woodward IRC G65/66. Please refer to the program and the customized campus map.

Les séances d'affichage régulières auront lieu dans le hall du pavillon Woodward IRC. Des séances d'affichage particulières auront lieu au Woodward IRC G65/66. Veuillez vous référer au programme et à la carte du campus. Elles seront installées le matin pour n'être enlevées qu'à la fin de la journée.

NSERC News Session and Workshop • Séance de nouvelles et atelier offerts par le CRSNG

Monday, June 1st • lundi 1er juin, 12:00–13:30, IKB 185

Representatives from NSERC and members of the Statistical Sciences Grant Selection Committees (GSC-14) will make a presentation to familiarize researchers with the peer review process and how to prepare an application. While the workshop will be most helpful to new faculty members and those preparing applications this fall, all researchers are welcome to attend.

Des représentants du CRSNG et des membres des comités de sélection des subventions (CSS) en statistique feront une présentation afin de familiariser les chercheurs avec le processus d'évaluation par les pairs et donnera des conseils sur la manière de préparer une demande. Bien que l'atelier s'adresse plus particulièrement aux nouveaux chercheurs et à ceux qui prévoient présenter une demande cet automne, tous les chercheurs sont invités.

Monday, June 1st • lundi 1er juin, 18:15–19:15, SCR 100

Open meeting on the 2008/9 NSERC competition and the new evaluation model.

Réunion publique sur la compétition et le nouveau modèle d'évaluation du CRSNG 2008/9.

Exclusive Museum Night • Soirée au musée

Conference participants are cordially invited to an exclusive evening at the Museum of Anthropology (<http://www.moa.ubc.ca>) on Monday (June 1st) from 6 to 10pm. The Museum of Anthropology is world-famous for its collections, its spectacular architecture designed by Arthur Erickson, and its unique setting on the cliffs overlooking English Bay. The Museum's collection includes objects from around the world, but it is most famous for its collections from the Northwest Coast of British Columbia. Highlights of the Museum are its Great Hall with its 15 metre high glass walls and massive totem poles, and "The Raven

and the First Men” sculpture by Bill Reid. This event is included in your registration fee. Guided tours of the museum will be available. Please refer to the customized campus map to locate the museum.

Les participants à la conférence sont cordialement invités à une soirée exclusive au Musée d’anthropologie (<http://www.moa.ubc.ca>) lundi (le 1er juin) de 18h00 à 20h00. Le Musée d’anthropologie est reconnu internationalement pour ses collections, son architecture spectaculaire planifiée par Arthur Erickson et son emplacement unique sur les falaises surplombant la baie Anglaise. La collection du musée comprend des objets de diverses provenances autour du monde, mais il est spécifiquement reconnu pour ses collections de la côte nord-ouest de Colombie-Britannique.

Le musée est aussi reconnu pour son grand hall avec ses murs de verre de 15 mètres de haut et ses totems massifs, et pour la sculpture "The Raven and the First Men" de Bill Reid.

Le coût de cet événement est compris dans les frais d’Inscription au congrès. Des visites guidées du musée seront disponibles. Veuillez vous référer à la carte du campus pour vous rendre au musée.

Welcome Reception • Réception de bienvenue

We cordially invite all conference participants to a welcome reception at the Sage Restaurant on Sunday May 31st from 5 pm to 7 pm. Please refer to the customized campus map to locate the restaurant.

Nous invitons cordialement tous les participants à la conférence à une réception de bienvenue au Sage Restaurant le dimanche 31 mai de 17h00 à 19h00. Veuillez vous référer à la carte du campus pour localiser le restaurant.

Graduate Students Mixer • Coquetel des étudiants diplômés

If you are a student, your registration fee includes a ticket for the Graduate Students Mixer, to be held on Monday June 1st from 6 to 11 pm, at the Ballroom of the Graduate Student Society (across the road from the Museum of Anthropology). There will be light food, a cash bar, music and a chance to mingle with the young crowd. Please refer to the customized campus map to locate the Graduate Student Society building.

Si vous êtes étudiant, vos frais d’inscription comprennent billet pour le Coquetel des étudiants diplômés, le lundi 1er juin de 18h00 à 23h00, à la salle de bal de la Graduate Student Society (en face du Musée d’anthropologie). Il y aura des hors-d’oeuvre, un bar à la carte, de la musique et la chance de tisser des liens. Veuillez vous référer à la carte du campus pour localiser le pavillon de la Graduate Student Society.

Banquet

All participants are cordially invited to attend the Conference Banquet on Tuesday June 2nd at the Life Sciences Centre West Atrium. A cash bar will be available starting at 6:45 pm, and food will be served at 7:30 pm. You will need to present your banquet ticket to attend this event. Please see the staff at the Registration Desk for any questions, and refer to the customized campus map to locate the Life Sciences Centre.

Tous les participants sont cordialement invités au banquet du congrès, mardi de 2 juin à l’atrium ouest du Life Sciences Centre. Un bar à la carte sera ouvert à partir de 18h45, et le repas sera servi à 19h30. Vous devrez présenter votre billet du banquet pour assister à cet événement. Veuillez vous référer au personnel du bureau d’inscription pour toute question, et à la carte du campus pour localiser le Life Sciences Centre.

Greater Vancouver Science Fair Teams • Équipes de l'expo-sciences du grand Vancouver

The Statistical Society of Canada is proud to invite several Science Fair teams to the 2009 Annual Meeting. These teams will show their projects during the afternoon poster sessions, in the lobby of the Instructional Resources Centre on Monday June 1 from 3-4:30pm. Two of the teams won UBC Science's Statistical Methods Awards in their divisions. These awards are for the best use of statistical methods in data collection and analysis. Two teams have earned the honour to progress to the Canada Wide Science Fair.

La Société Statistique du Canada est fière d'inviter plusieurs équipes de l'expo-sciences régionale de Vancouver au congrès annuel de 2009. Ces équipes montreront leurs projets au cours des séances d'affichage dans le hall du centre IRC le lundi 1er juin de 15h00 à 16h30. Deux des équipes ont remporté le prix des méthodes statistiques d'UBC Science pour leur division. Ces prix sont accordés aux équipes faisant le meilleur usage de méthodes statistiques en collecte et analyse de données. Deux équipes vont continuer à l'Expo-science pancanadienne.

Session in honour of Jack Kalbfleisch's 65th birthday • Séance en l'honneur du 65e anniversaire de Jack Kalbfleisch

The session Analysis of Life History Data Arising from Non-standard Selection and Observation Schemes, in Honour of Jack Kalbfleisch's 65th Birthday will be held on Wednesday 10:30 to 12:00 noon in room IRC 6 (session 12E). There will also be an informal celebration following the session at a nearby restaurant (to be announced). For further information please contact: Richard Cook, University of Waterloo.

La séance Analyse de données de cycles de vie provenant de schémas de sélection et d'observation non standards, en l'honneur du 65e anniversaire de Jack Kalbfleisch aura lieu mercredi de 10h30 à 12h00 dans la salle IRC 6 (séance 12E). Il y aura également une célébration informelle après la séance dans un restaurant à proximité (à annoncer). Pour plus de détails, veuillez contacter Richard Cook à l'Université de Waterloo.

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

Please join the Caucus for Women in Statistics and the Committee on Women in Statistics of the SSC for an informal get-together over dinner on Monday June 1st, at 7:30 pm at the Candia Taverna Restaurant (4510 West 10th Ave., Phone: 604-228-9512, Web: <http://www.foodvancouver.com/restaurant.phprestaurant1182>). The average dinner entrée costs between CAD \$10.00 and \$18.00, and Candia's accepts both cash and credit cards. How to get there: take 99-B line bus from UBC, drop off at Sasamat St. (the 2nd stop), then walk back toward UBC about five stores after re-crossing Sasamat Street.

All people (both men and women) are invited and students are especially welcome.

Veuillez vous joindre au caucus des femmes en statistique et au comité de la promotion de la femme en statistique de la SSC pour une rencontre informelle le lundi 1er juin à 19h30 au Candia Taverna Restaurant (4510 West 10th Ave., téléphone : 604-228-9512, web : <http://www.foodvancouver.com/restaurant.phprestaurant1182>). Le prix moyen d'un plat est entre 10,00\$ et 18,00\$ CAD, et le restaurant accepte l'argent comptant et les

cartes de crédit. Pour s'y rendre : prenez la ligne d'autobus 99-B sur le campus jusqu'à la rue Sasamat (le 2e arrêt), et retraversez la rue Sasamat.

Tous (hommes et femmes) sont invités, et les étudiants sont particulièrement bienvenus.

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

The Second Annual SSC New Investigators Dinner is intended for new investigators including upper level PhD students, recent graduates working in government or industry positions, post-doctoral fellows, and newly appointed faculty members. It provides a chance to meet and network with other junior researchers and to learn more about the transition from student life to your independent career. Brief talks will be given by senior researchers including Dr. Paul Gustafson, the 2008 winner of the CRM-SSC Prize in Statistics.

The dinner will be held on the final day of the conference, June 3rd, from 6:00-9:30 pm at the beautiful Thea Koerner Graduate Student Centre on UBC campus (<http://www.gss.ubc.ca/thea/rooms.html>). The cost of the event is \$33 and full dinner will be provided.

Le deuxième souper annuel des nouveaux chercheurs de la SSC cible les nouveaux chercheurs, incluant les étudiants au doctorat, les diplômés récents occupants des postes gouvernementaux ou en industrie, les chercheurs postdoctoraux, et les nouveaux professeurs. Il s'agit d'une occasion de rencontrer et de tisser des liens avec d'autres jeunes chercheurs et d'apprendre au sujet de la transition de la vie étudiante vers une carrière indépendante. De brèves présentations seront données par des chercheurs seniors, incluant le Dr. Paul Gustafson, le récipiendaire du prix en statistique CRM-SSC.

Le souper aura lieu le dernier jour de la conférence, le 3 juin, de 18h00 à 21h30 au magnifique Thea Koerner Graduate Student Centre sur le campus (<http://www.gss.ubc.ca/thea/rooms.html>). Le coût de l'événement est de 33\$, incluant le repas.

Internet Access • Accès internet

All conference participants will have wireless internet access across the UBC campus. See your registration package for details. Wireless access at UBC is very good inside and near academic buildings, and some other central facilities (like the Student Union Building). In addition, several public terminals are available in the Student Union Building.

Tous les participants à la conférence auront accès au réseau internet sans fil de l'université sur le campus. Veuillez consulter votre sac du congrès pour plus de détails. L'accès sans fil est de très bonne qualité dans et près des édifices académiques et certains autres édifices, comme le Student Union Building. De plus, plusieurs ordinateurs à accès public sont disponibles dans le « Student Union Building ».

Memorial Sessions • Les séances à la mémoire de Randy Sitter et Keith Worsley

Please note the following memorial sessions. A memorial session for Randy Sitter will be held on Monday June 1st, 3:30 to 5:00 pm in room IRC 5 (session 04C). There will be an informal reception afterwards at the nearby Mahoney's Pub, sponsored by Simon Fraser University. For further information please contact Charmaine Dean, Simon Fraser University.

Also a special session in honour of Keith Worsley will be held on Tuesday June 2nd, 3:30 to 5:00 pm in room MSL101 (session 09E). An informal reception will be held immediately after the session, partially supported by McGill University. For further information please contact George Styan, McGill University.

Veillez noter les séances à la mémoire des personnes suivantes. Une séance à la mémoire de Randy Sitter sera tenue lundi le 1er juin de 15h 30-17h dans la salle IRC 5 (séance 04C). Suite à la séance, il y aura une réception informelle au Pub Mahoney commandité par l'Université de Simon Fraser. Pour plus de renseignements, veuillez contacter Charmaine Dean, Université Simon Fraser.

De plus, une séance en honneur de Keith Worsley sera tenue mardi le 2 juin de 15h 30-17h dans la salle MSL 101 (séance 09E). Immédiatement après la séance il y aura une réception informelle en part subventionné par l'Université McGill. Pour plus de renseignements, veuillez contacter George Styan, Université McGill.

Committees and Meetings

Comités et réunions

WMAX rooms are part of PIMS (Pacific Institute of Mathematical Sciences) and are in West Mall Annex. IKB rooms = Irving K. Barber Learning Centre, IRC = Instructional Resources Centre

Les salles WMAX font partie du PIMS (Pacific Institute of Mathematical Sciences) et sont dans le West Mall Annex. Les salles IKB sont dans le Irving K. Barber Learning Centre. Les salles IRC sont dans le Instructional Resources Centre.

Calendar Listing • Horaire journalier

Saturday May 30 • Samedi 30 mai

18:00–22:00 Executive Committee • *Comité exécutif* WMAX 110

Sunday, May 31 • Dimanche 31 mai

9:00–11:00 Finance Committee • *Comité des finances* WMAX 110
 11:00–12:00 Publications Committee • *Comité des publications* WMAX 110
 10:00–12:00 Department Heads Meeting • *Rencontre des directeurs de départements* WMAX Library
 12:00–16:00 Board of Directors • *Conseil d'administration* WMAX 110

Monday, June 1st • Lundi 1er juin

12:00–13:30 Probability Section (Section Officers) • *Groupe de probabilité (directeurs du groupe)* IKB 181
 Biostatistics Section (Section Officers) • *Groupe de biostatistique (directeurs du groupe)* IKB 183
 Business and Industrial Statistics Section (Section Officers) • *Groupe de statistique industrielle et de gestion (directeurs du groupe)* IKB 184
 Survey Methods Section (Section Officers) • *Groupe des méthodes d'enquête (directeurs du groupe)* IKB 196
 Committee on Women in Statistics • *Comité sur les femmes en statistique* IKB 157
 17:15–18:15 Biostatistics Section (General Meeting) • *Groupe de biostatistique (assemblée générale)* IKB 260

	Probability Section (General Meeting) • <i>Groupe de probabilité (assemblée générale)</i>	IKB 158
	Survey Methods Section (General Meeting) • <i>Groupe des méthodes d'enquête (assemblée générale)</i>	IKB 185
	Business and Industrial Statistics Section (General Meeting) • <i>Groupe de statistique industrielle et de gestion (assemblée générale)</i>	IKB 155
18:15–19:15	Statistical Education Committee • <i>Comité d'éducation en statistique</i>	IKB 157
	Information on 2008/9 NSERC Competition and New Evaluation Model • <i>Information sur la compétition et le nouveau modèle d'évaluation du CRSNG 2008/9</i>	SCR 100

Tuesday, June 2 • Mardi 2 juin

12:00–13:30	Bilingualism Committee • <i>Comité du bilinguisme</i>	IKB 181
	Accreditation General Meeting • <i>Assemblée générale du comité d'accréditation</i>	IKB 155
	Public Relations Committee • <i>Comité des relations publiques</i>	IKB 157
	NICDS • <i>INSDC</i>	IKB 158
	CJS Editorial Board • <i>Comité de rédaction de la RCS</i>	IKB 191
17:15–18:30	SSC Annual General Meeting • <i>Assemblée générale annuelle de la SSC</i>	IRC 2

Wednesday, June 3 • Mercredi 3 juin

12:00–13:30	Accreditation Committee • <i>Comité sur l'accréditation</i>	IKB 156
	Liaison	IKB 157
12:15–13:30	Program Committee • <i>Comité du programme</i>	IKB 158
	Probability Section (Section Officers) • <i>Groupe de probabilité (directeurs du groupe)</i>	IKB 181
12:15–13:00	Biostatistics Section (Section Officers) • <i>Groupe de biostatistique (directeurs du groupe)</i>	IKB 183
	Business and Industrial Statistics Section (Section Officers) • <i>Groupe de statistique industrielle et de gestion (directeurs du groupe)</i>	IKB 184
	Awards Committee • <i>Comité des prix</i>	IKB 191
	Survey Methods Section (Section Officers) • <i>Groupe des méthodes d'enquête (directeurs du groupe)</i>	IKB 196
15:30–17:30	Board of Directors • <i>Conseil d'administration</i>	MSL 101
17:30	Executive Committee • <i>Comité exécutif</i>	MSL 101

Alphabetical Listing • Liste alphabétique

Accreditation Committee • *Comité sur l'accréditation*

Wednesday, June 3 • mercredi 3 juin, 12:00–13:30, IKB 156

Accreditation General Meeting • *Assemblée générale du comité d'accréditation*

Tuesday, June 2 • mardi 2 juin, 12:00–13:30, IKB 155

Awards Committee • Comité des prix

Wednesday, June 3 • mercredi 3 juin, 12:00–13:30, IKB 191

Bilingualism Committee • Comité du bilinguisme

Tuesday, June 2 • mardi 2 juin, 12:00–13:30, IKB 181

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

Monday, June 1st • lundi 1er juin, 17:15–18:15, IKB 260

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

Monday, June 1st 26 • lundi 1er juin, 12:00–13:30, IKB 183

Wednesday, June 3 • mercredi 3 juin, 12:00–13:00, IKB 183

Board of Directors • Conseil d'administration

Sunday, May 31 • dimanche 31 mai, 12:00–16:00, WMAX 110

Wednesday, June 3 • mercredi 3 juin, 15:30–17:30, MSL 101

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

Monday, June 1st • lundi 1er juin, 17:15–18:15, IKB 155

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

Monday, June 1st • lundi 1er juin, 12:00–13:30, IKB 184

Wednesday, June 3 • mercredi 3 juin, 12:00–13:00, IKB 184

CJS Editorial Board • Comité de rédaction de la RCS

Tuesday, June 2 • mardi 2 juin, 12:00–13:30, IKB 191

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

Monday, June 1st • lundi 1er juin, 12:00–13:30, IKB 157

Department Heads Meeting • Rencontre des directeurs de département

Sunday, May 31 • dimanche 31 mai, 10:00–12:00, WMAX Library

Executive Committee • Comité exécutif

Saturday May 30 • samedi 30 mai, 18:00–22:00, WMAX 110

Wednesday June 3 • mercredi 3 juin, 17:30, MSL 101

Finance Committee • Comité des finances

Sunday, May 31 • dimanche 31 mai, 9:00–11:00, WMAX 110

Information on 2008/9 NSERC Competition and New Evaluation Model • Information sur la compétition et le nouveau modèle d'évaluation du CRSNG 2008/9

Monday, June 1st • lundi 1er juin, 18:15–19:15, SCR 100

Liaison

Wednesday, June 3 • mercredi 3 juin, 12:00–13:30, IKB 157

NICDS • INSDC

Tuesday, June 2 • mardi 2 juin, 12:00–13:30, IKB 158

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

Monday, June 1st • lundi 1er juin, 17:15–18:15, IKB 158

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

Monday, June 1st • lundi 1er juin, 12:00–13:30, IKB 181

Thursday, May 29 • jeudi 29 mai, 12:00–13:00, IKB 181

Program Committee • *Comité du programme*

Wednesday, June 3 • mercredi 3 juin, 12:00–13:30, IKB 158

Public Relations Committee • *Comité des relations publiques*

Tuesday, June 2 • mardi 2 juin, 12:00–13:30, IKB 157

Publications Committee • *Comité des publications*

Sunday, May 31 • dimanche 31 mai, 11:00–12:00, WMAX 110

SSC General Meeting • *Assemblée générale de la SSC*

Tuesday, June 2 • mardi 2 juin, 17:15–18:30, IRC 2

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

Monday, June 1st • lundi 1er juin, 18:15–19:15, IKB 157

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

Monday, June 1st • lundi 1er juin, 17:15–18:15, IKB 185

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

Monday, June 1 • lundi 1er juin, 12:00–13:30, IKB 196

Wednesday, June 3 • mercredi 3 juin, 12:00–13:00, IKB 196

List of Sessions • Liste des sessions

Monday June 1st • Lundi 1er juin

08:15–08:30

IRC 2 Session 01A Inaugural Session
Séance inaugurale

08:30–09:45

IRC 2 Session 01B SSC Presidential Invited Address
Allocution de l'invité de la présidente de la SSC

10:30–12:00

IRC 6	Session 02A	Special Invited Session of the Survey Methods Section <i>Allocution de l'invité d'honneur du groupe des méthodes d'enquête</i>
IRC 1	Session 02B	Evaluation of Medical Screening and Diagnostic Tests <i>Évaluation des tests de dépistage et de diagnostic médicaux</i>
IRC 4	Session 02C	History of Statistics - Lifetime Modeling and Data <i>Histoire de la statistique - Modélisation et données de cycle de vie</i>
IRC 5	Session 02D	Functional Data Analysis <i>Analyse de données fonctionnelles</i>
IRC 3	Session 02E	Recent Development in Analysis of High Dimensional Data <i>Évolution récente de l'analyse des données de grande dimension</i>
MSL 102	Session 02F	Statistical Methods for Public Health Research <i>Méthodes statistiques pour la recherche en santé publique</i>
CHE 250	Session 02G	Probability and Finance <i>Probabilités et finance</i>
SCR 100	Session 02H	Design and Analysis for Biomedical Studies <i>Planification et analyse d'études biomédicales</i>
CHE 150	Session 02I	Methodology for Time-to-Event Data <i>Méthodologie pour données de temps d'événements</i>
IRC Lobby	Session 02J	Poster Session I <i>Séance d'affichage I</i>

12:00–13:30

IKB 185 NSERC Workshop: How to Prepare a Research Grant Application
Atelier du CRSNG : Préparation d'une demande de subvention à la découverte

13:30–15:00

IRC 6	Session 03A	Teaching Biostatistics in Public Health Programs: Struggles and Successes <i>Enseigner la biostatistique dans les programmes de santé publique : défis et succès</i>
IRC 1	Session 03B	Stochastic Spatial Models <i>Modèles spatiaux stochastiques</i>
IRC 5	Session 03C	Statistical Methods in Clinical Trials <i>Méthodes statistiques dans les essais cliniques</i>
IRC 3	Session 03D	Statistics and Science <i>La statistique et la science</i>
MSL 102	Session 03E	Simulator Assisted Inference for Environmental Systems <i>Inférence assistée par simulateur pour les systèmes environnementaux</i>
CHE 150	Session 03F	Some Recent Development in Analyzing Complex Survey Data <i>Évolution récente de l'analyse de données d'enquêtes complexes</i>
CHE 250	Session 03G	Estimation Methods in Business and Industrial Statistics <i>Méthodes d'estimation en statistique industrielle et de gestion</i>
SCR 100	Session 03H	Inference, Design, and Applications <i>Inférence, planification et applications</i>
IRC Lobby	Session 03I	Case Study 1: The Effects of Climate on the Growth of Lodgepole Pine <i>Étude de cas 1 : Les effets du climat sur la croissance du pin de Murray</i>
IRC Lobby	Session 03J	Case Study 2: Proteomic Biomarkers for Disease Status <i>Étude de cas 2 : Biomarqueurs protéomiques et états pathologiques</i>
15:30–17:00		
IRC 6	Session 04A	Isobel Loutit Invited Address on Business and Industrial Statistics <i>Allocution Isobel Loutit sur la statistique industrielle et de gestion</i>
IRC 1	Session 04B	Complex Methods for Longitudinal and Survival Data <i>Méthodes complexes pour les données longitudinales et de survie</i>
IRC 5	Session 04C	Memorial Session for Randy Sitter <i>Séance commémorative en la mémoire de Randy Sitter</i>
MSL 102	Session 04D	Runs, Patterns and the Finite Markov Chain Imbedding Technique <i>Courses, schémas et la technique d'insertion de la chaîne de Markov finie</i>
IRC 3	Session 04E	Some Recent Developments on Response Adaptive Designs of Clinical Trials <i>Évolution récente des plans d'essais cliniques adaptés à la réponse</i>
FNH 60	Session 04F	Statistical Challenges in Clinical Research: CANNeCTIN Research Program <i>Défis statistiques en recherche clinique : le programme de recherche CANNeCTIN</i>
CHE 150	Session 04G	Monte Carlo Markov Chains and Markov Models <i>Chaînes de Markov Monte-Carlo et modèles markoviens</i>
CHE 250	Session 04H	Statistical Genetics and Genomics I <i>Génétique et génomique statistiques I</i>

IRC Lobby Session 04I Greater Vancouver Regional Science Fair Student Projects
Projets de l'expo-sciences régionale de Vancouver

Tuesday June 2 • Mardi 2 juin

08:45–10:00

IRC 2 Session 05 Gold Medal Award Address
Allocution du récipiendaire de la médaille d'or

10:30–12:00

IRC 6 Session 06A Special Invited Session of the Probability Section
Allocution de l'invité d'honneur du groupe de probabilité

IRC 1 Session 06B Innovative Ideas in the Teaching of Statistics and Probability
Idées novatrices en enseignement de la statistique et des probabilités

IRC 4 Session 06C Statistical Inference Based on Data from Different Sources
Inférence statistique fondée sur des données de sources différentes

IRC 5 Session 06D Statistical Applications in the Study of Infectious Diseases
Modèles et méthodes statistiques dans le domaine de l'étude et la lutte contre les maladies infectieuses

IRC 3 Session 06E Nonparametric Multivariate Methods
Méthodes non paramétriques à variables multiples

MSL 102 Session 06F Emerging Issues in Graphical Models
Nouveaux problèmes dans les modèles graphiques

IRC G65/66 Session 06G Forests, Fires and Stochastic Modeling
Forêts, incendies et modélisation stochastique

MSL 101 Session 06H Models and Tests in Business and Industrial Statistics
Modèles et tests en statistique industrielle et de gestion

CHE 150 Session 06I Estimation and Models for Biomedical Data
Estimation et modèles pour données biomédicales

CHE 250 Session 06J Estimation Methods
Méthodes d'estimation

IRC Lobby Session 06K Poster Session II
Séance d'affichage II

13:30–14:30

IRC 6 Session 07 Canadian Journal of Statistics Award Address
Allocution du récipiendaire du prix de la Revue canadienne de statistique

13:30–15:00

IRC 1 Session 08A Quantitative and Qualitative Challenges in Climate Change and Ecosystem Sustainability Research
Problèmes quantitatifs et qualitatifs en changement climatique et recherches sur la durabilité des écosystèmes

IRC 4 Session 08B Quality Improvement in Healthcare
Amélioration de la qualité en soins de santé

MSL 102 Session 08C Stochastic Modeling of Disordered Spatial Patterns
Modélisation stochastique des structures spatiales désordonnées

IRC 3 Session 08D Analysis of Correlated Data with Excessive Zeros
Analyse de données corrélées à zéros excessifs

IRC G65/66	Session 08E	Analysis and Modelling of Marine Ecological Data <i>Analyse et modélisation de données écologiques marines</i>
MSL 101	Session 08F	Estimation Methods for Survey Data <i>Méthodes d'estimation pour données d'enquête</i>
CHE 150	Session 08G	Statistical Genetics and Genomics II <i>Génétique et génomique statistiques II</i>
CHE 250	Session 08H	Bayesian Statistics and Applications <i>Statistique et applications bayésiennes</i>
IRC Lobby	Session 08I	Poster Session III <i>Séance d'affichage III</i>
15:30–17:00		
IRC 6	Session 09A	Special Invited Session of the Biostatistics Section <i>Allocution de l'invité d'honneur du groupe de biostatistique</i>
IRC 1	Session 09B	Innovative Uses of Statistical Software in the (Remote) Classroom <i>Utilisations novatrices des logiciels statistiques dans la salle de classe (à distance)</i>
IRC 4	Session 09C	Statistics/Mathematics Finance 1 <i>Statistique/Mathématique Finance-1</i>
IRC 5	Session 09D	Small Area Estimation at Statistics Canada <i>Estimation relatives aux secteurs restreints chez Statistique Canada</i>
MSL 101	Session 09E	Special Session in Honour of Keith Worsley (1951-2009) <i>Séance en la mémoire de Keith Worsley (1951-2009)</i>
IRC G65/66	Session 09F	Design and Analysis of Computer Experiments for Complex Systems <i>Planification et analyse d'expériences informatiques pour systèmes complexes</i>
IRC 3	Session 09G	Analysis of Spatial and Temporal Data <i>Analyse de données spatiales et temporelles</i>
MSL 102	Session 09H	Bayesian Methods in Biostatistics <i>Méthodes bayésiennes en biostatistique</i>
CHE 150	Session 09I	Statistical Theory <i>Théorie statistique</i>
CHE 250	Session 09J	Computational Methods and Algorithms <i>Méthodes et algorithmes informatiques</i>
IRC Lobby	Session 09K	Poster Session IV <i>Séance d'affichage IV</i>

Wednesday June 3 • Mercredi 3 juin

8:30–09:30

IRC 6	Session 10	Pierre Robillard Award Addresses <i>Allocution du récipiendaire du prix Pierre-Robillard</i>
-------	------------	---

08:30–10:00

IRC 1	Session 11A	Statistics/Mathematics Finance <i>Statistique/Mathématique Finance-2</i>
IRC 4	Session 11B	Modelling and Measuring Pollution and Natural Systems <i>Modélisation et mesure de la pollution et des systèmes naturels</i>
IRC 5	Session 11C	Multivariate Survival Analysis <i>Analyse de survie à variables multiples</i>

IRC 3	Session 11D	Theory and Application of the Finite Mixture Models <i>Théorie et application des modèles de mélange fini</i>
CHE 150	Session 11E	Models for Count Data and Excessive Zeros <i>Modèles pour données de dénombrement et zéros excessifs</i>
CHE 250	Session 11F	Biostatistical Applications, Issues, and Interpretations <i>Applications, problèmes et interprétations biostatistiques</i>
SCR 100	Session 11G	Survey Design and Analysis <i>Planification et analyse d'enquêtes</i>
MSL 101	Session 11H	Statistical Education in Action <i>L'éducation statistique en action</i>
MSL 102	Session 11I	Stochastic Processes and Applications <i>Processus stochastiques et applications</i>

10:30–12:00

IRC 1	Session 12A	Ecological Processes, Mechanisms, and Statistical Evidence <i>Processus, mécanismes et preuves statistiques écologiques</i>
IRC 4	Session 12B	Empirical Processes and Applications to Finance <i>Processus empiriques et applications à la finance</i>
IRC 5	Session 12C	What do Mathematical Statisticians do at Statistics Canada <i>Travail des statisticiens mathématiques chez Statistique Canada</i>
IRC 3	Session 12D	Statistical Methods for Causal Inference <i>Méthodes statistiques pour l'inférence causale</i>
IRC 6	Session 12E	Session in Honor of Jack Kalbfleisch's 65th Birthday: Analysis of Life History Data Arising from Non-standard Selection and Observation Schemes <i>Séance en honneur du 65e anniversaire de Jack Kalbfleisch : analyse de données de cycles de vie provenant de schémas de sélection et d'observation non standard</i>
MSL 101	Session 12F	Design and Analysis for Business and Industrial Statistics <i>Planification et analyse pour statistique industrielle et de gestion</i>
CHE 150	Session 12G	Mixed Models, Multistate Modeling, and Measurement Error <i>Modèles mixtes, modélisation à plusieurs états et erreurs de mesure</i>
MSL 102	Session 12H	Modeling, Classification, and Algorithms <i>Modélisation, classification et algorithmes</i>
CHE 250	Session 12I	Statistical Inference <i>Inférence statistique</i>
SCR 100	Session 12J	Mixture Models <i>Modèles de mélanges</i>

13:30–14:30

IRC 6	Session 13	Address of the winner of the Centre de Recherches Mathématiques/Statistical Society of Canada Prize <i>Allocution du récipiendaire du prix Centre de Recherches Mathématiques/Société Statistique du Canada</i>
-------	------------	--

13:30–15:00

IRC 1	Session 14A	Statistics in Sports <i>La statistique dans le monde des sports</i>
-------	-------------	--

IRC 5	Session 14B	Statistical Inference in Studies with Large p and Small n <i>Inférence statistique dans les études à grand p et petit</i>
MSL 102	Session 14C	Recent Developments in Order-Restricted Inference <i>Évolution récente de l'inférence d'ordre restreint</i>
IRC 3	Session 14D	Matrices and Statistics <i>Matrices et statistique</i>

Program • Programme

Lundi 1 juin • Monday, June 1

08:15 – 08:30 Session 01A

IRC 2

Inaugural Session

Séance inaugurale

Special Session • Conférence spéciale

Special Guest Speakers

Chair • Présidente: Nancy HECKMAN, *University of British Columbia*Organizers • Responsables: Nancy HECKMAN, *University of British Columbia* and Román VIVEROS-AGUILERA, *McMaster University*08:15 Alejandro ADEM, *Director of Pacific Institute for the Mathematical Sciences*  

Welcoming Remarks • Remarques de bienvenue [MS-1]

08:25 Román VIVEROS-AGUILERA, *McMaster University*, Nancy HECKMAN, *University of British Columbia*

Announcements • Annonces [MS-2]



08:30 – 09:45 Session 01B

IRC 2

SSC Presidential Invited Address

Allocution de l'invité du président de la SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Roman VIVEROS-AGUILERA, *McMaster University*08:30 Chris WILD, *University of Auckland*  

Building the Pyramids • Construire les pyramides [MS-3]

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

IRC Lobby

10:30 – 12:00 Session 02A

IRC 6

Special Invited Session of the Survey Methods Section

Allocution de l'invité d'honneur du groupe des méthodes d'enquête

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Survey Methods Section

Organizer and Chair • Responsable et présidente: Julie TRÉPANIÉ, *Statistics Canada*

10:30 Jean-François BEAUMONT, *Statistique Canada / Statistics Canada*  

A New Approach to Weighting and Inference for Samples Drawn from a Finite Population • Une nouvelle approche de pondération et d'inférence pour des échantillons tirés d'une population finie [MS-4]

10:30 – 12:00 Session 02B

IRC 1

Evaluation of Medical Screening and Diagnostic Tests

Évaluation des tests de dépistage et de diagnostic médicaux

Invited Paper Session • Présentations sur invitation

Session Sponsored by the SSC Award for Impact of Applied and Collaborative Work

Organizer and Chair • Responsable et président: Stephen WALTER, *McMaster University*

10:30 James HANLEY, *McGill University*  

Assessing the Performance of Diagnostic and Prognostic Probability Functions in Medicine • Évaluer la performance des fonctions de probabilité diagnostique et pronostique en médecine [MS-5]

11:00 Eduardo FRANCO, *McGill University*  

Biases and Pitfalls in Interpreting the Efficacy of Cancer Screening Technologies: The Epidemiologists' Perspective • Biais et pièges dans l'interprétation de l'efficacité des technologies de dépistage du cancer : la perspective des épidémiologistes [MS-6]

11:30 Timo HAKULINEN, *Finnish Cancer Registry*  

Statistical Issues With Cancer Screening Programs - Experience From Finland • Problèmes statistiques avec les programmes de dépistage du cancer : l'expérience de la Finlande [MS-7]

10:30 – 12:00 Session 02C

IRC 4

History of Statistics - Lifetime Modeling and Data



Histoire de la statistique - Modélisation et données de cycle de vie

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Statistical Education Committee

Chair • Président: George STYAN, *McGill University*

Organizer • Responsable: David BELLHOUSE, *University of Western Ontario*

10:30 Elizabeth TURNER, *London School of Hygiene & Tropical Medicine*, James HANLEY, *McGill University*  

Cultural Imagery and Statistical Models of the Force of Mortality • Représentations culturelles et modèles statistiques de la force de mortalité [MS-8]

11:00 Antoine de FALGUEROLLES, *Université de Toulouse (III)*  

Cauchy's Method of Linear Regression • La méthode de régression linéaire de Cauchy [MS-9]

11:30 David BELLHOUSE, *University of Western Ontario*  

Edmond Halley and His Life Table • Edmond Halley et sa table de mortalité [MS-10]

10:30 – 12:00 Session 02D**IRC 5**



Functional Data Analysis

Analyse de données fonctionnelles

Invited Paper Session • Présentations sur invitation

Chair • Président: David CAMPBELL, *Simon Fraser University*Organizers • Responsables: Jiguo Cao, *Simon Fraser University*10:30 Hans-Georg MUELLER, *UC Davis*, Bitao LIU, *Affymetrix*  

Derivatives of Sparsely Sampled Random Trajectories • Dérivées de trajectoires aléatoires à échantillonnage éparé [MS-11]

11:00 Nancy HECKMAN, *University of British Columbia*, Richard LOCKHART, *Simon Fraser University*, Nielsen JASON, *Carleton University*  

Penalized Regression and Mixed Effects Models • Régression pénalisée et modèles à effets mixtes [MS-12]

11:30 Jiguo CAO, & David CAMPBELL, *Simon Fraser University*, Giles HOOKER, *Cornell University*, Jianhua HUANG, *A&M University*, Jim RAMSAY, *McGill University* 

Statistical Inference for Dynamic Models with the Generalized Profiling Method • Inférence statistique pour modèles dynamiques avec la méthode du profilage généralisé [MS-13]

10:30 – 12:00 Session 02E**IRC 3**

Recent Development in Analysis of High Dimensional Data

Évolution récente de l'analyse des données de grande dimension



Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Wenqing HE, *University of Western Ontario*10:30 Jian HUANG & Patrick BREHENY, *University of Iowa*  

Penalized Methods for Bi-level Variable Selection • Méthodes pénalisées pour sélection de variables à deux niveaux [MS-14]

11:00 Yi LI, *Harvard*  

Dantzig Selector for Censored Linear Regression Models With High Dimensional Covariates • Le Sélectionneur de Dantzig pour modèles de régression linéaires censurés avec covariables de dimension élevée [MS-15]

11:30 J. Sunil RAO, *Case Western Reserve University*, Simin HU, *Eli Lilly Corporation*  

Sparse Penalization with Censoring Constraints for Estimating AFT Models with Applications to Microarray Data Analysis • Pénalisation éparse avec contraintes de censure en estimation de modèles à temps de défaillance accéléré avec application à l'analyse de données de micropuces [MS-16]

10:30 – 12:00 Session 02F**MSL 102**

Statistical Methods for Public Health Research

Méthodes statistiques pour la recherche en santé publique

Topic Contributed Paper Session

National Institute for Complex Data Structures

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

10:30 Taraneh ABARIN & Liquin WANG, *University of Manitoba*  


Method of Moments Estimation in Logistic regression Models with Exposure Biomarkers • Estimation par la méthode des moments en modèles de régression logistique avec biomarqueurs d'exposition [MS-17]

10:45 Kyle RAYMOND & Richard COOK, *University of Waterloo*  



The Impact of Dependent Durations of Episodes in Recurrent Event Analysis • L'impact de durées dépendantes des épisodes en analyse d'événements récurrents [MS-18]

11:00 Chun-Po Steve FAN, *University of Toronto*, Patrick BROWN, *Cancer Care Ontario*, James STAFFORD, *University of Toronto*  

A Local Likelihood Method in Disease Mapping • Une méthode de vraisemblance locale en cartographie des maladies [MS-19]

11:15 Chel Hee LEE & Liquin WANG, *Department of Statistics, University of Manitoba*  

Efficient Monte Carlo Random Variates Generation Through Discretization • Génération de variables aléatoires efficace par Monte-Carlo par discrétisation [MS-20]

11:30 Ella HUSZTI & Michal ABRAHAMOWICZ, *McGill University*, Ahmadou ALIOUM, *ISPED-Université Victor Segalen Bordeaux 2, France*, Catherine QUANTIN, *Dijon University Hospital, France*  

Connecting Relative Survival with Multi-State Models • Connecter la survie relative aux modèles multi-états [MS-21]

10:30 – 12:00 Session 02G



CHE 250

Probability and Finance



Probabilités et finance

Contributed Paper Session • Séance de communications libres

Chair • Président: Murray BURKE, *University of Calgary*

10:30 Yung-Pin CHEN, *Lewis and Clark College*  



On Two Monotonic Properties of the Kendall Tau Coefficient of Order Statistics • Deux propriétés de monotonie du coefficient tau de Kendall de statistiques d'ordre [MS-22]

10:45 Leonard HEARNE, *University of Missouri at Columbia*  



On the Complexity of d-dimensional Delaunay Tessellations • De la complexité des mosaïques de Delaunay d-dimensionnelles [MS-23]

11:00 Chantal LABBÉ & Bruno RÉMILLARD, *HEC Montréal*, Jean-François RENAUD, *University of Waterloo*  

A Simple Discretization Scheme For Nonnegative Diffusion Processes, With Applications in Financial Mathematics • Un schéma de discrétisation simple pour des processus de diffusion non-négatifs, avec applications en finance mathématique [MS-24]

11:15 Natalia LYSENKO, *ETH Zurich*, Guus BALKEMA, *University of Amsterdam*, Paul EMBRECHTS, *ETH Zurich*  

Extremes from Meta Distributions and the Shape of Sample Clouds • Extrêmes des méta-distributions et forme des nuages échantillonnaires [MS-25]


11:30 François WATIER & René FERLAND, *Université du Québec à Montréal*  

Goal Achieving Probability of a Discrete-time Mean-variance Optimal Strategy • Probabilité d'atteinte d'objectif d'une stratégie moyenne-variance optimale à temps discret [MS-26]

10:30 – 12:00 Session 02H**SCR 100****Design and Analysis for Biomedical Studies****Planification et analyse d'études biomédicales**

Contributed Paper Session • Séance de communications libres



Chair • Président: Brajendra SUTRADHAR, *Memorial University*

- 10:30 Jeffrey BAKAL, Cynthia WESTERHOUT, Robert WELSH & Paul ARMSTRONG, *Canadian VIGOUR Centre*  



Weighted Composite Endpoint Models in the Analysis of Cardiovascular Clinical Trial Data • Évaluation des modèles à points limite composites pondérés dans l'analyse des essais cliniques cardio-vasculaire [MS-27]

- 10:45 Elizabeth JUAREZ & Charmaine DEAN, *Simon Fraser University*, Robert BALSHAW, *Syreon Corporation*  



Design Considerations for Recurrent Events Studies • Considérations pour la planification d'études d'événements récurrents [MS-28]

- 11:00 Elmabrok MASAPOUD & Henrik STRYHN, *University of Prince Edward Island*  

Statistical Modelling of Neighbour Vaccine Effects in Aquaculture Clinical Trials • Modélisation statistique de l'effet voisin des vaccins en essais cliniques en aquaculture [MS-29]

- 11:15 Reza RAMEZAN, Paul MARRIOTT & Shoja'eddin CHENOURI, *University of Waterloo*  

Analysis of Neural Spike Trains: a Review • Analyse de trains d'impulsion neuronaux : une revue [MS-30]

- 11:30 Lan-Yan YANG, *National Cheng Kung University, Tainan, Taiwan*, Shein-Chung CHOW, *Duke University*, Yunchan CHI, *National Cheng-Kung University, Tainan, Taiwan*  

Statistical Inference for Clinical Trials with Protocol Amendments • Inférence statistique pour essais cliniques avec amendements au protocole [MS-31]

- 11:45 Qiaohao ZHU, *University of Alberta*  

Power Calculation for Cluster Randomized Trial of Comparing Trends of Incidence Over Time • Calcul de la puissance pour la comparaison de tendances dans l'incidence en essais randomisés par grappes [MS-32]

10:30 – 12:00 Session 02I**CHE 150****Methodology for Time-to-Event Data****Méthodologie pour données de temps d'événements**

Contributed Paper Session • Séance de communications libres



Chair • Présidente: Eleanor PULLEYNAYEGUM, *McMaster University*

- 10:30 Pierre-Jérôme BERGERON, *University of Ottawa*  

Regression and Length-biased Sampling in Survival Analysis • Régression et biais de longueur en analyse de survie [MS-33]

- 10:45 Stephanie DIXON & Gerarda DARLINGTON, *University of Guelph*  

A Subdistribution Hazards Model for Clustered Data • Modèle de défaillance à sous-distributions pour données en grappes [MS-34]

- 11:00 Hyun LIM, *University of Saskatchewan*, Xu ZHANG, *Georgia State University*  



Additive Hazards Regression Model for Competing Risk Events • Modèle de régression à risques additifs pour événements à risque concurrents [MS-35]

11:15 Yan YUAN, *Alberta Cancer Board*, Jerald LAWLESS, *University of Waterloo*  

Estimation of the Prediction Error for Survival Models • Estimation de l'erreur de prévision pour modèles de survie [MS-36]

11:30 Jiajia ZHANG, *University of South Carolina*, Yingwei PENG, *Queen's University*  

Accelerated Hazards Mixture Cure Model • Modèle de mélange à défaillance accélérée avec fraction non à risque [MS-37]

11:45 Mohannad Hossein DEGHAN, *Laval university*, Thierry DUCHESNE, *Laval university*  

A Nonparametric Estimator of the Conditional Survivor Function with Interval Censored Data • Un estimateur non paramétrique de la fonction de survie conditionnelle avec données censurées par intervalles [MS-38]



10:30 – 12:00 Session 02J

IRC Lobby



Poster Session I

Séance d'affichage I

Contributed Poster Session

10:30 Li Hua YUE, Duncan MURDOCH & Wenqing HE, *University of Western Ontario*  

Cost-efficient Variable Selection with Non-additive Cost • Sélection de variables rentable sous coût non additif [MS-39]

10:30 Kelly BURKETT, Brad McNENEY & Jinko GRAHAM, *Simon Fraser University*  

Sampling Ancestries at a Genomic Location Conditional on Data from Surrounding Genetic Markers • Échantillonnage d'ascendance à une position génomique conditionnellement aux données des marqueurs génétiques environnants [MS-40]

10:30 Mateen SHAIKH, Paul McNICHOLAS & Tony DESMOND, *University of Guelph*  



Clustering Incomplete Longitudinal Data • Regroupement de données longitudinales incomplètes [MS-41]

10:30 Gerald LEBOVIC, *University of Toronto*, Patrick BROWN, *Cancer Care Ontario*, George TOMLINSON, *University Health Network*  

Estimating Non-Homogeneous Transition Intensities in Continuous Time Multi-State Markov Models: An Application to Ordinal Data. • Estimation d'intensités de transition non homogènes sous modèles markoviens multi-états à temps continu : une application à des données ordinales [MS-42]

10:30 Ruimin XU, Paul McNICHOLAS, Gerarda DARLINGTON & Anthony DESMOND, *University of Guelph*  



Inverse Gaussian Cure Rate Mixture Models • Modèles de mélanges gaussiens inverses avec fraction non à risque [MS-43]

10:30 Tolulope SAJOBI & Lisa LIX, *University of Saskatchewan*  

Testing Multiple Outcomes in the Presence of Non-Normality and Covariance Heterogeneity • Tester des réponses multiples en présence de non-normalité et d'hétérogénéité de la covariance [MS-44]

10:30 Hedy JIANG & Patrick BROWN, *Cancer Care Ontario*  

Sample Size and Design Effect for Multilevel Logistic Regression Models • Taille échantillonnale et effet du plan d'expérience pour modèles de régression logistique à plusieurs niveaux [MS-45]

10:30 Osama BATAINEH & Lisa LIX, *University of Saskatchewan*  

Ascertainment of Osteoporosis Cases Using Administrative Databases and Capture-Recapture Method • Découverte de cas d'ostéoporose au moyen de bases de données administratives et de la méthode capture-recapture [MS-46]

10:30 Daniel GILLIS, Tony DESMOND & John HOLT, *University of Guelph*  

Identification of Disease Source via Multivariate Spatial • Identification de la source d'une maladie au moyen de modèles spatiaux multivariés [MS-47]

12:00 – 13:30 Lunch Break • Dîner

12:00 – 13:30 NSERC Workshop: How to Prepare a Research Grant Application • Atelier du CRSNG : Comment préparer une demande de subvention à la découverte
IKB185

13:30 – 15:00 Session 03A

IRC 6


Teaching Biostatistics in Public Health Programs: Struggles and Successes

Enseigner la biostatistique dans les programmes de santé publique : défis et succès

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Statistical Education Committee

Organizer and Chair • Responsable et président: Rollan BRANT, *University of British Columbia*

13:30 Rollin BRANT & Kevin CRAIB, *University of British Columbia*, Joel DUBIN, *University of Waterloo*, Gian JHANGRI, *University of Alberta*, Punam PAHWA, *University of Saskatchewan*, Kevin THORPE, *University of Toronto*  

Round Table: Teaching Biostatistics in Public Health Programs • Table ronde : l'enseignement de la biostatistique en programmes de santé publique [MS-48]

13:30 – 15:00 Session 03B

IRC 1



Stochastic Spatial Models

Modèles spatiaux stochastiques

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Probability Section

Organizer and Chair • Responsable et président: Ed PERKINS, *University of British Columbia*

13:30 Ted Cox, *Syracuse University*  

Long Range Random Walk on the Two-dimensional Torus • Marche aléatoire à long terme sur un tore bidimensionnel [MS-49]

14:00 Frank DEN HOLLANDER, *University of Leiden, Netherlands*, Dimitris CHELIOTIS, *EURANDOM*  

Variational Approach to Pinning of Random Polymers • Approche variationnelle au blocage de polymères aléatoires [MS-50]

14:30 Amanda TURNER, *University of Lancaster*  



Planar Aggregation and the Coalescing Brownian Flow • L'agrégation planaire et le flot brownien coalescent [MS-51]

13:30 – 15:00 Session 03C**IRC 5**

Statistical Methods in Clinical Trials

Méthodes statistiques dans les essais cliniques



Invited Paper Session • Présentations sur invitation

Chair • Présidente: Ying ZHANG, *Acadia University*Organizer • Responsable: K.K. Gordon LAN, *Johnson & Johnson Inc.*13:30 Kuang-Kuo LAN, *Johnson & Johnson PRD*  

Flexibility and Optimality of Statistical Procedures in Drug Development • Flexibilité et optimalité des procédures statistiques en développement de médicaments [MS-52]

14:00 Michael PROSCHAN, *National Institute of Allergy and Infectious Diseases*  

Sample Size Re-estimation in Clinical Trials • Ré-estimation de taille échantillonnale en essais cliniques [MS-53]

14:30 Weichung SHIH & Yujun WU, *University of Medicine & Dentistry of New Jersey*  



Approaches to Handling Data When a Phase II Trial Deviates from the Pre-specified Simon's Two-Stage Design • Approches au traitement des données lorsqu'un essai de phase II dévie du plan d'expérience à deux étapes de Simon préspecifié [MS-54]

13:30 – 15:00 Session 03D**IRC 3**



Statistics and Science

La statistique et la science

Special Late Breaking Session • Conférence spéciale “dernière nouvelle”

Organizer and Chair • Responsable et président: Agnes M. HERZBERG, *Queen's University*13:30 John CAREY, *Water and Science Technology Directorate, Environment Canada*  

Will There Be Enough Water? • Y aura-t-il assez d'eau ? [MS-55]

14:00 John STONE, *formerly Environment Canada*  

Climate Change: The Role of Science and its Limitations • Changements climatiques : le rôle de la science et ses limites [MS-56]

14:30 Peter CALAMAI, *Columnist, Toronto Star*  

Engaging the Canadian Public with Sound Science • Communiquer avec le public canadien avec de la science bien fondée [MS-57]

13:30 – 15:00 Session 03E**MSL 102**

Simulator Assisted Inference for Environmental Systems

Inférence assistée par simulateur pour les systèmes environnementaux



Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Derek BINGHAM, *Simon Fraser University*13:30 Shane REESE, *Brigham Young University*, David HIGDON, David MOULTON & Jasper VRUGT, *Los Alamos National Laboratory*, Colin Fox, *University of Otago, New Zealand*  

Posterior Exploration for Computationally Intensive Forward Models • Exploration a posteriori pour modèles avant demandant une grande puissance de calcul [MS-58]

14:00 Stephan SAIN, *National Center for Atmospheric Research*  

A Spatial Analysis of Regional Climate Model Ensembles • Une analyse spatiale d'ensembles de modèles régionaux du climat [MS-59]

14:30 Cari KAUFMAN, *UC Berkeley*  

Calibration and Prediction Problems in Catchment Scale Hydrology • Problèmes de calibration et de prévision en hydrologie à l'échelle des bassins-versants [MS-60]

13:30 – 15:00 Session 03F

CHE 150

Some Recent Development in Analyzing Complex Survey Data


Évolution récente de l'analyse de données d'enquêtes complexes

Topic Contributed Paper Session



MITACS & National Institute for Complex Data Structures

Chair • Président: Changbao WU, *University of Waterloo*

Organizers • Responsables: Milorad KOVACEVIC, *Statistics Canada* and Changbao WU, *University of Waterloo*

13:30 Milorad KOVACEVIC, *Statistics Canada*, Changbao WU, *University of Waterloo*  

Statistics Canada's Collaboration with University Researchers: Statistical Methods for Complex Survey Data • La collaboration de Statistique Canada avec des chercheurs universitaires : méthodes pour données d'enquête complexes [MS-61]

13:52 Zhijian CHEN, *University of Waterloo*, Harold MANTEL, *Statistics Canada*  

Logistic Regression Analysis Using Complex Survey Data with Misclassification in an Ordinal Covariate • Analyse par régression logistique de données d'enquête complexes avec mauvaise classification dans une variable ordinale [MS-62]

14:15 Dagmar MARIACA-HAJDUCEK & Jerald LAWLESS, *University of Waterloo*  

Fitting Cox Models to Jobless Spell Durations in SLID • Ajuster des modèles de Cox aux durées de périodes sans emploi dans l'EDTR [MS-63]

14:38 Yan LIU, *University of British Columbia*, Milorad KOVACEVIC & Abdelnasser SAIDI, *Statistics Canada*, Bruno ZUMBO, *University of British Columbia*  

Challenges in Analyzing National Longitudinal Data: Cohort-Sequential Design and the Application of Sampling Weights When Using Structural Equation Modeling (SEM) • Défis pour l'analyse des données longitudinales nationales : Cohortes séquentielles par construction et application des poids de sondage dans les modèles à équations structurelles (MES) [MS-64]

13:30 – 15:00 Session 03G



CHE 250

Estimation Methods in Business and Industrial Statistics

Méthodes d'estimation en statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres



Chair • Président: Thierry DUCHESNE, *Université Laval*

13:30 Melody GHAHRAMANI, *University of Winnipeg*, Aerambamoorthy THAVANESWARAN, *University of Manitoba*  

Recursive Estimation Using Combined Optimal Estimating Functions • Estimation récursive utilisant des fonctions estimatrices optimales combinées [MS-65]

13:45 Kin Hung KAN & Mark REESOR, *The University of Western Ontario*  



Bias Reduction of the Least-Squares Monte Carlo Estimators of American Option Values • Réduction du biais des estimateurs des moindres carrés Monte-Carlo de la valeur d'options américaines [MS-66]

14:00 Asokan MULAYATH VARIYATH, *Memorial University of Newfoundland*, Shoja'eddin CHENOURI, *University of Waterloo*  

A Comparative Study of Robust Control Charts • Une étude comparative de cartes de contrôle robustes [MS-67]

14:15 Juan SORIANO, *University of Barcelona/Agroasemex*, F.J. SARRASI, *University of Barcelona*  

Estimation and Coverage of the Hurricane Risk with the use of Copulas • Estimation et couverture du risque d'ouragans avec copules [MS-68]

14:30 Yanqiao ZHANG, *University of Waterloo*  

Time-Dependent > Regime-switching Models: Estimation and Applications • Modèles à changement de régime dépendant du temps : Estimation et applications [MS-69]

13:30 – 15:00 Session 03H


SCR 100

Inference, Design, and Applications

Inférence, planification et applications

Contributed Paper Session • Séance de communications libres



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

13:30 François BELLAVANCE, Ahlem HAJJEM & Denis LAROCQUE, *HEC Montréal*  



Mixed Effects Regression Trees for Clustered Data • Arbres de régression à effets mixtes pour données en grappes [MS-70]

13:45 Maryse CHAMPAGNE, *Université du Québec à Trois-Rivières*, Anne-Catherine FAVRE, *INRS-ETE Québec*, Jean-François QUESSY, *Université du Québec à Trois-Rivières*  



Break-point Detection for Dependence Functions • Détection de points de rupture pour fonctions de dépendance [MS-71]

14:00 Lin ZHAO & Xiaojian XU, *Brock University*  

Robust Regression Designs for Wavelet Approximation Models • Plans de régression robuste pour modèles d'approximation par ondelettes [MS-72]

14:15 Aline TABET, Arnaud DOUCET, Paul GUSTAFSON & Kevin MURPHY, *University of British Columbia*  

Model selection and Inference in the Multivariate Probit Model • Sélection de modèle et inférence sous le modèle probit multivarié [MS-73]

14:30 Mehdi KIANI, *Athens University of Economics and Business*, Muhammad SALEEM, *GC University Faisalabad*  

A Note on the Modified Gram-Schmidt Orthogonalization Process in Design and Analysis • Une note sur le processus d'orthogonalisation de Gram-Schmidt modifié en planification et analyse [MS-74]

14:45 Liangliang WANG, Nancy HECKMAN & Matías SALIBIÁN-BARRERA, *University of British Columbia*  

Robust Functional Principal Components Analysis and its Application to Outlier Detection • L'analyse en composantes principales fonctionnelle robuste et ses applications en détection des données aberrantes [MS-75]

13:30 – 15:00 Session 03I**IRC Lobby**

Case Study 1: The Effects of Climate on the Growth of Lodgepole Pine

Étude de cas 1 : Les effets du climat sur la croissance du pin de Murray

Case Study • Étude de cas

Organizer and Chair • Responsable et présidente: Alison GIBBS, *University of Toronto*13:30 Cindy FENG, Carolyn HUSTON & Elizabeth JUAREZ, *Simon Fraser University*13:30 Eric CORMIER & Zheng SUN, *University of Victoria*13:30 Charlotte WICKHAM, *University of California, Berkeley*13:30 Janet SHAN & Jinning ZHANG, *University of Guelph*13:30 Mustafa Abdel-Aziz SALAMH, *University of Manitoba***13:30 – 15:00 Session 03J****IRC Lobby**

Case Study 2: Proteomic Biomarkers for Disease Status

Étude de cas 2 : Biomarqueurs protéomiques et états pathologiques

Case Study • Étude de cas

Organizer and Chair • Responsable et présidente: Alison GIBBS, *University of Toronto*13:30 Billy CHANG, Zhengfei CHEN, Nino DEMETRASHVILI & Matthew KOWGIER, *University of Toronto*13:30 Yalin CHEN, Qing GUO & Defen PENG, *McMaster University*13:30 Amy LIU & Rajat MALIK, *McMaster University*13:30 Steven WANG & Michelle C. XIA, *University of British Columbia*13:30 Linghong LU, Rena MANN, Rabih SAAB & Ryan STONE, *University of Victoria*13:30 Haimo CHEN, Dongdong LI, Chengchen SUN, Yuan WANG, Beilei WU & Yifan ZHU, *University of Calgary*13:30 Calvin JOHNSTON & Christopher MEANEY, *University of Toronto*, Jenna SYKES, *University of Waterloo***15:00 – 15:30 Coffee Break • Pause-café****IRC Lobby****15:30 – 17:00 Session 04A****IRC 6**

Isobel Loutit Invited Address on Business and Industrial Statistics

Allocution Isobel Loutit sur la statistique industrielle et de gestion

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Business and Industrial Statistics Section



Organizer and Chair • Responsable et président: Smiley CHENG, *University of Manitoba*15:30 William WOODALL, *Virginia Tech*  

The State of Statistical Process Control: An Update • L'état actuel du contrôle statistique du processus : une mise à jour [MS-78]


15:30 – 17:00 Session 04B**IRC 1****Complex Methods for Longitudinal and Survival Data****Méthodes complexes pour les données longitudinales et de survie**

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Biostatistics Section

Organizer and Chair • Responsable et président: Joel DUBIN, *University of Waterloo*15:30 Wesley JOHNSON, *UC Irvine*, Timothy HANSON, *University of Minnesota*, Adam BRANSCUM, *University of Kentucky*  

Bayesian Joint Semi-parametric Modeling of Longitudinal- Survival Data: Is Joint Modeling Really Necessary? • Modélisation bayésienne conjointe semi-paramétrique de données longitudinales de survie : la modélisation conjointe est-elle vraiment nécessaire ? [MS-79]

15:52 Robert STRAWDERMAN & David CLEMENT, *Cornell University*  

Conditional GEE for Recurrent Event Gap Time Data • ÉEG conditionnelles pour données de temps d'écarts entre événements récurrents [MS-80]

16:15 Grace YI, *University of Waterloo*, Wei LIU, *York University*, Lang WU, *University of British Columbia*  

Analysis of Longitudinal Data with Error in Covariates and Missingness in Responses under Generalized Linear Mixed Models • Analyse de données longitudinales avec erreurs dans les covariables et réponses manquantes sous modèles linéaires mixtes généralisés [MS-81]

16:37 Terry LEE, Charmaine DEAN & Leilei ZENG, *Simon Fraser University*  

Joint Analysis of Longitudinal Growth and Interval Censored Mortality Data • Analyse conjointe de croissance longitudinale et de données de mortalité censurées par intervalles [MS-82]

15:30 – 17:00 Session 04C**IRC 5****Memorial Session for Randy Sitter****Séance commémorative en la mémoire de Randy Sitter**

Invited Paper Session • Présentations sur invitation

Session Sponsored by Simon Fraser University, Informal reception immediately following session at Mahoney's - all welcome

Chair • Président: Charmaine DEAN, *Simon Fraser University*Organizers • Responsables: Charmaine DEAN & Derek BINGHAM, *Simon Fraser University*15:30 Arden MILLER, *University of Auckland*  

Using Folded Over Designs to Consider Interactions • Utiliser des plans d'expérience avec repli pour prendre en considération les interactions [MS-83]

16:00 J.N.K. RAO, *Carleton University*  

Impact of Randy Sitter's Contributions to Sample Survey Theory and Practice • Impact des contributions de Randy Sitter sur la théorie et la pratique des enquêtes par sondage [MS-84]

16:30 Richard LOCKHART, Derek BINGHAM & Charmaine DEAN, *Simon Fraser University*  

Randy Sitter: His Career, His Humour and His Many Accomplishments • Randy Sitter : sa carrière, son humour et ses nombreuses réalisations [MS-85]

15:30 – 17:00 Session 04D**MSL 102****Runs, Patterns and the Finite Markov Chain Imbedding Technique****Courses, schémas et la technique d'insertion de la chaîne de Markov finie**



Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Brad JOHNSON, *University of Manitoba*15:30 Markos KOUTRAS, *University of Piraeus, Greece*, N. BALAKRISHNAN, *McMaster University, I TRI-ANTAFYLLOU, University of Piraeus, Greece*  

Nonparametric Control Charts Based on Run Statistics • Cartes de contrôle non paramétriques basée sur des statistique de séries [MS-86]

16:00 Gregory NUEL, *MAP5, CNRS, University Paris Descartes*  



Dealing With Patterns in Degenerated Sequences • Méthodes pour l'étude de motifs dans des séquences dégénérées [MS-87]

16:30 John ASTON, *CRiSM, Warwick University, UK*, Jyh-Ying PENG, *Academia Sinica*  

Some Applications of FMCI in Segmentation • Quelques applications de l'ECMF en segmentation [MS-88]

15:30 – 17:00 Session 04E**IRC 3****Some Recent Developments on Response Adaptive Designs of Clinical Trials****Évolution récente des plans d'essais cliniques adaptés à la réponse**

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Xikui WANG, *University of Manitoba*15:30 Nancy FLOURNOY, Ping YAO & Seung Won HYUN, *University of Missouri*  

Adaptive Treatment Allocation Procedures With Application to Phase I/II Clinical Trials and Toxicology Studies. • Procédures adaptatives d'assignation de traitement avec application à des essais cliniques de phase I/II et à des études de toxicologie [MS-89]

15:52 Yi CHENG, *Indiana University South Bend*  

Bayesian Adaptive Clinical Trial Designs • Planification bayésienne d'essais cliniques adaptatifs [MS-90]

16:15 Janis HARDWICK & Quentin STOUT, *University of Michigan*  

Optimizing Tradeoffs Between Randomization and Classical Objectives • Optimisation de compromis entre objectifs classiques et de randomisation [MS-91]











16:37 Yanqing YI, *Memorial University of Newfoundland*  

Asymptotic inference of Response Adaptive Clinical Trials • Inférence asymptotique pour essais cliniques adaptatifs en fonction de la réponse [MS-92]

15:30 – 17:00 Session 04F**FNH 60****Statistical Challenges in Clinical Research: CANNeCTIN Research Program****Défis statistiques en recherche clinique : le programme de recherche CANNeCTIN**

Topic Contributed Paper Session

Chair • Président: David MATTHEWS, *University of Waterloo*Organizers • Responsables: Richard Cook, *University of Waterloo*

- 15:30 Meaghan CUERDEN & Amit GARG, *University of Western Ontario*, Richard COOK, *University of Waterloo*  
 Estimating Mediation Effects in Randomized Controlled Trials with a Failure Time Response • Évaluer les effets de médiation en essais comparatifs randomisés avec données de temps de défaillance [MS-93]
- 15:45 Rong CHU, Jinhui MA, Lehana THABANE, Anne HOLBROOK & Eleanor PULLENAYEGUM, *McMaster University*, COMPETE GROUP  
 Dealing with Stratification and Clustering in Randomized Controlled Trials with a Continuous Outcome • Traiter la stratification et la segmentation en essais comparatifs randomisés avec réponse continue [MS-94]
- 16:00 Audrey BORUVKA & Richard COOK, *University of Waterloo*  
 Analysis of Multi-state Data with Missing Covariates • Analyse de données à multi-états avec covariables manquantes [MS-95]
- 16:15 Jinhui MA & Lehana THABANE, *McMaster University*  
 Comparing the Efficiency of Different Statistical Methods in the Analysis of Binary Outcomes from Cluster Randomized Controlled Trials: A Simulation Study • Comparer l'efficacité de différentes méthodes statistiques pour l'analyse de réponses binaires d'essais contrôlés randomisés par grappes : une étude de simulation [MS-96]
- 16:30 Janice POGUE, ON BEHALF OF THE PHRI STATISTICS GROUP & CANNeCTIN (CANADIAN NETWORK AND CENTRE FOR TRIALS INTERNATIONALLY), *McMaster University*  
 Statistical Challenges Faced by the Population Health Research Institute (PHRI) in Conducting Large Simple Trials and Epidemiologic Studies • Défis statistiques rencontrés par l'Institut de recherche en santé de la population (PHRI) en essais simples à grande échelle et en études épidémiologiques [MS-97]

15:30 – 17:00 Session 04G

CHE 150

Monte Carlo Markov Chains and Markov Models

Chaînes de Markov Monte-Carlo et modèles markoviens

Contributed Paper Session • Séance de communications libres

Chair • Président: Brian ALLEN, *University of Guelph*

- 15:30 Yan BAI, *University of Toronto*, G.O. ROBERTS, *University of Warwick*, J.S. ROSENTHAL, *University of Toronto*  
 On the Containment Condition for Adaptive Markov Chain Monte Carlo Algorithms • Sur la condition de confinement pour algorithmes adaptatifs à chaînes de Markov Monte-Carlo [MS-98]
- 15:45 Roman HOLENSTEIN, *University of British Columbia*, Arnaud DOUCET, *University of British Columbia / The Institute of Statistical Mathematics, Tokyo, Japan*, Christophe ANDRIEU, *University of Bristol, UK*  
 Particle Markov Chain Monte Carlo • Chaînes de Markov Monte-Carlo à particules [MS-99]
- 16:00 Reza HOSSEINI, James ZIDEK & Nhu LE, *University of British Columbia*  
 r^{th} Order Categorical Markov Chains • Chaînes de Markov catégoriques du r^{e} ordre [MS-100]
- 16:15 Wai Kong YUEN, *Brock University*  
 Reptation Quantum Monte Carlo Algorithm • Algorithme Monte-Carlo quantique à reptation [MS-101]
- 16:30 Amaan MEHRABIAN & David STEPHENS, *McGill University*  

On Markov Chain Monte Carlo Simulation for High-Frequency Financial Data • Simulation par chaînes de Markov Monte-Carlo pour données financières de haute fréquence [MS-102]

16:45 Hsing-Ming CHANG, *University of Manitoba*  

Priority Modelling of Emergency Department Wait Times by Finite Markov Chain Imbedding • Modélisation de priorité des temps d'attente aux urgences par emboîtement de chaînes de Markov finies [MS-103]

15:30 – 17:00 Session 04H



CHE 250

Statistical Genetics and Genomics I

Génétique et génomique statistiques I

Contributed Paper Session • Séance de communications libres


Chair • Présidente: Jinko GRAHAM, *Simon Fraser University*

15:30 Kenneth Lo, *University of British Columbia*, Raphael GOTTARDO, *Institut de recherches cliniques de Montreal*  

Flexible Empirical Bayes Models for Differential Gene Expression • Modélisation bayésienne empirique flexible de données d'expression génétique différentielle [MS-104]

15:45 Juli ATHERTON, *McGill University*, Nathan BOLEY, Ben BROWN & Peter BICKEL, *UC Berkeley*, Mark BIGGIN, *Lawrence Berkeley Labs*  

Using SELEX Data to Model the Affinity of DNA Sequences to the Transcription Factor Bicoid • Utilisation de données ESLEE pour modéliser l'affinité de séquences d'ADN pour le facteur de transcription bicoïde [MS-105]

16:00 Yun-Hee CHOI, *University of Western Ontario*, Karen KOPCIUK, *Alberta Cancer Board*, Jane GREEN & Pat PARFREY, *Memorial University of Newfoundland*, Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital*  

Evaluating Disease Risk for Successive Events in Carriers of a Major Gene using a Progressive Multi-state Modelling Approach. • Évaluer le risque de maladie pour événements successifs chez les porteurs d'un gène majeur par modélisation progressive à plusieurs états [MS-106]

16:15 Juan XIONG & Wenqing HE, *University of Western Ontario*  

Survival Relevant Gene Selection in Microarray Data Analysis with Gene Expression Subject to Measurement Error • Sélection de gènes liés à la survie en analyse de données de micropuces avec expression génétique sujette à erreurs de mesure [MS-107]

16:30 Xuekui ZHANG & Raphael GOTTARDO & Kaida NING *University of British Columbia*, Gordon ROBERTSON, *British Columbia Cancer Agency Genome Sciences Centre*  

Probabilistic Inference for ChIP-Seq data: a Model Based Approach • Inférences probabilistes pour données ChIP-Seq : une approche basée sur un modèle [MS-108]

15:00 – 16:30 Session 04I

IRC Lobby

Greater Vancouver Regional Science Fair Student Projects

Projets de l'expo-sciences régionale de Vancouver

Topic Contributed Poster Session • Séance d'affichage

Organizer and Chair • Responsable et présidente: Nancy HECKMAN, *University of British Columbia*

15:00 Michelle NG & Meiyang ZHUANG, *University Transitions Program - Early Entrance to UBC*

Effects of Light on Algae • Effets de la lumière sur l'algue [MS-109]

15:00 Jonathan ZHOU, *Pinetree Secondary*, Andrew J. JUNG, *Walnut Grove Secondary*

Probabilistic Evolution in Dynamic Equilibrium • Évolution probabiliste dans l'équilibre dynamique [MS-110]

15:00 Jieqing XU, *University Transitions Program - Early Entrance to UBC*

The Effect of Traditional Chinese Herbal Medicines on Human Cancer Cells • L'effet des herbes médicinales chinoises traditionnelles sur les cellules humaines cancéreuses [MS-111]

15:00 Jessica ZHANG, *University Transitions Program - Early Entrance to UBC*

Effects of Common Fruit Extracts on Cell Division in Human Cancers • Effets des essences communes de fruits sur la division cellulaire dans les cancers humains [MS-112]

Mardi 2 juin • Tuesday, June 2

08:45 – 10:00 Session 05

IRC 2

Gold Medal Award Address

Allocution du récipiendaire de la médaille d'or

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: David BINDER, *Statistics Canada*08:45 Luc DEVROYE, *McGill University*  

Randomness is Not a Bad Word • Aléatoire n'est pas un mauvais mot [MS-113]

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

IRC Lobby

10:30 – 12:00 Session 06A



IRC 6

Special Invited Session of the Probability Section

Allocution de l'invité d'honneur du groupe de probabilité

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Probability Section

Organizer and Chair • Responsable et président: Reg KULPERGER, *University of Western Ontario*10:30 Peter GUTTORP, *University of Washington*  

Extremes in Air Pollution and Climate • Extrêmes en pollution atmosphérique et en climat [MS-114]

10:30 – 12:00 Session 06B

IRC 1

Innovative Ideas in the Teaching of Statistics and Probability

Idées novatrices en enseignement de la statistique et des probabilités

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Statistical Education Committee

Organizer and Chair • Responsable et président: Alison GIBBS, *University of Toronto*10:30 Larry WELDON, *SSC*  

Authentic Statistics Education - Updating the Apprenticeship Model • Une éducation authentique en statistique - mettre à jour le modèle de l'apprentissage [MS-115]

11:00 Christine ANDERSON-COOK, *Los Alamos National Laboratory*  

A Case Study about Statistical Communication with Practicing Engineers • Une étude de cas à propos de la communication statistique avec des ingénieurs praticiens [MS-116]



11:30 Wayne OLDFORD, *University of Waterloo*  

Eikosograms: The Picture of Probability • Eikosogrammes : l'image des probabilités [MS-117]


10:30 – 12:00 Session 06C**IRC 4****Statistical Inference Based on Data from Different Sources****Inférence statistique fondée sur des données de sources différentes**

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Survey Methods Section

Chair • Président: David BINDER, *Statistics Canada*Organizer • Responsable: Milorad KOVACEVIC, *Statistics Canada*10:30 Ray CHAMBERS, *University of Wollongong*  

Regression with Probability-Linked Data • Régression avec données couplées de façon probabiliste [MS-118]

11:00 Milorad KOVACEVIC, *Statistics Canada*  

Regression Analysis of Record-Linked Survey Data • Analyse de régression pour données d'enquêtes liées à des enregistrements [MS-119]



10:30 – 12:00 Session 06D**IRC 5****Statistical Applications in the Study of Infectious Diseases****Modèles et méthodes statistiques dans le domaine de l'étude et la lutte contre les maladies infectieuses**

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Public Health Agency of Canada

Organizer and Chair • Responsable et président: Ping YAN, *Public Health Agency of Canada*10:30 Nicholas JEWELL, *UC Berkeley*  

Alternatives to Intention to Treat—the MIRA Trial • Alternatives à l'intention de traitement-l'étude MIRA [MS-120]

11:00 Christl DONNELLY & Helen JENKINS, *MRC Centre for Outbreak Analysis and Modelling, Imperial College London*, Bruce AYLWARD & Alex GASASIRA, *World Health Organization*, Emmanuel ABANIDA & Titi KOLEOSHO-ADELEKAN, *National Primary Health Care Development Agency, Abuja, Nigeria*, Nicholas GRASSLY, *MRC Centre for Outbreak Analysis and Modelling, Imperial College London*  

Estimating the Effectiveness of Immunization against Paralytic Poliomyelitis in Nigeria • Estimation de l'efficacité de l'immunisation contre la poliomyélite paralytique au Nigéria [MS-121]



11:30 Rob DEARDON, *University of Guelph*, TJ MCKINLEY, *University of Cambridge*, Alex COOK, *National University of Singapore*  

Likelihood-free Inference for Infectious Disease Models • Inférence sans vraisemblance pour modèles de maladies infectieuses [MS-122]



10:30 – 12:00 Session 06E**IRC 3****Nonparametric Multivariate Methods****Méthodes non paramétriques à variables multiples**

Invited Paper Session • Présentations sur invitation



Organizer and Chair • Responsable et président: Shojaeddin CHENOURI & Christopher SMALL, *University of Waterloo*

10:30 Ivan MIZERA, *University of Alberta*  

How to Find the Shape of a Banana • Comment trouver la forme d'une banane [MS-123]

11:00 Hannu OJA, *University of Tampere*, Davy PAINDAVEINE, *Université libre de Bruxelles*, Klaus NORDHAUSEN, *University of Tampere*, Sara TASKINEN, *University of Jyväskylä*  

Nonparametric Tests and the Independent Components Model • Tests non paramétriques et le modèle à composantes indépendantes [MS-124]

11:30 Yijun ZUO, *Michigan State University*  

On Multivariate Medians • À propos de médianes multivariées [MS-125]

10:30 – 12:00 Session 06F



MSL 102

Emerging Issues in Graphical Models

Nouveaux problèmes dans les modèles graphiques

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Lang WU, *University of British Columbia*

10:30 Helene MASSAM, *York University*, Jinnan LIU, *Samuel Lunenfeld Research Institute*, Adrian DOBRA, *University of Washington*  

A Conjugate Prior for Discrete Hierarchical Loglinear Models • Une loi a priori conjuguée pour les modèles hiérarchiques discrets loglinéaires [MS-126]

11:00 Bo THIESSON, *Microsoft Research*, Chong WANG, *Princeton University*, Christopher MEEK, *Microsoft Research*, David BLEI, *Princeton University*  

Probabilistic Inference for CART Networks • Inférence probabiliste pour réseaux CART [MS-127]

11:30 Kevin MURPHY & Ben MARLIN, *UBC*  

Learning Graphs With unknown Group Sparsity • Graphes d'apprentissage avec éparsité de groupe inconnue [MS-128]

10:30 – 12:00 Session 06G

IRC G65/66

Forests, Fires and Stochastic Modeling



Forêts, incendies et modélisation stochastique

Topic Contributed Poster Session

National Institute for Complex Data Structures

Chair • Présidente: Charmaine DEAN, *Simon Fraser University*

Organizers • Responsables: Charmaine DEAN, *Simon Fraser University* and W. John BRAUN, *University of Western Ontario*

10:30 Lengyi HAN, W. John BRAUN & Xiaoming LIU, *University of Western Ontario*  


Spatial Bootstrap Block Size Selection • Sélection de la taille des blocs pour bootstrap spatial [MS-129]

10:30 Douglas WOOLFORD, Jiguo CAO & Charmaine DEAN, *Simon Fraser University*, David MARTELL, *University of Toronto*  

Looking for Climate Change Signals in the Canadian Forest Fire Ignition Record • Chercher des signaux des changements climatiques dans le déclenchement de feux de forêts au Canada [MS-130]

10:30 Katherine DAVIES, *University of Manitoba*, W. John BRAUN, *University of Western Ontario*  

Modelling Point Patterns with the Ivanoff-Merzbach Renewal Process • Modélisation de motifs ponctuels avec le processus de renouvellement Ivanoff-Merzbach [MS-131]

10:30 Sean MICHALETZ & Edward A. JOHNSON, *University of Calgary*  

Wind Effects on Bark Char Patterns in Forest Fires: a Heat Transfer Analysis • Effets du vent sur la répartition de résidus d'écorce charbonneux en feux de forêts : une analyse de transfert de chaleur [MS-132]

10:30 Yolanda LI, *British Columbia Ministry of Health Services*, Steve W. TAYLOR, *Pacific Forestry Centre*, Farouk NATHOO, *University of Victoria*  

Bayesian Hierarchical Models for Spatial Count Data with Application to Fire Frequency in British Columbia • Modèles hiérarchiques bayésiens pour données de dénombrement spatiales avec application à la fréquence des feux de forêts en Colombie-Britannique [MS-133]

10:30 Jim GOULD, *Natural Resources Canada- Canadian Forest Service*  

Predicting Fire Behaviour: Field Experiments for Fire Behaviour Modelling • Prévoir le comportement du feu : expériences sur le terrain en modélisation du comportement du feu [MS-134]

10:30 – 12:00 Session 06H

MSL 101

Models and Tests in Business and Industrial Statistics

Modèles et tests en statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

Chair • Président: Kevin CRAIB, *University of British Columbia*

10:30 Uditha BALASOORIYA & Jackie LI, *Nanyang Technological University*  



On the Significance of the Limited Value Averages Test • De la signification du test des moyennes de valeur limitée [MS-135]

10:45 Amanda LAFONTAINE & Dayan YEHUDA, *Ipsos MORI*  

Gaging Public Opinion Through Prediction Markets • Mesurer l'opinion publique à travers les marchés de prévision [MS-136]

11:00 Jasper TEY, Daniel GILLIS, Neil ROONEY, Greg KLOTZ, Amanda CASKENETTE & Steve CRAWFORD *University of Guelph*, Bill HARFORD, *University of Miami*  

A Review of the Bence Statistical Catch-at-Age Population for Lake Whitefish Stock Management • Une revue du modèle statistique d'âge au moment de la capture de Bence pour la gestion du stock de grand corégone [MS-137]

11:15 Vineetha WARRIYAR. K.V, *Memorial University*, Abraham BOVAS, *University of Waterloo*, Mulayath Variyath ASOKAN, *Memorial University*  

Performance of Some Diagnostic Tests in Financial Time Series • Performance de certains tests diagnostics en séries chronologiques financières [MS-138]

10:30 – 12:00 Session 06I

CHE 150

Estimation and Models for Biomedical Data

Estimation et modèles pour données biomédicales

Contributed Paper Session • Séance de communications libres



Chair • Président: John BRAUN, *University of Western Ontario*

10:30 Shahedul KHAN, Grace CHIU & Joel DUBIN, *University of Waterloo*  



Random Changepoint Model for Longitudinal Data: An Extension of Bent-Cable Regression • Modèle à point de changement aléatoire pour données longitudinales : une extension de la régression à cordes pliées [MS-139]

10:45 Wendell CHALLENGER & Carl SCHWARZ, *Simon Fraser University*  

Occupancy modeling with multiple states and multiple seasons • Modélisation de l'occupation avec états multiples et saisons multiples [MS-140]

11:00 Jun LI & Yogendra CHAUBEY, *Concordia University*, Pranab SEN, *University of North Carolina*  

Selection of Smoothing Parameter • Sélection du paramètre de lissage [MS-141]

11:15 Cheng CHENG & Jianrong WU, *St. Jude Children's Research Hospital*  

Interval Estimation of Quantile Ratios Applied to Anti-Cancer Drug Screening by Xenograft Experiments • Estimation par intervalle de rapports de quantiles appliquée à des expériences de dépistage de médicaments contre le cancer par xénogreffe [MS-142]

11:30 Bin DONG & David E. MATTHEWS, *University of Waterloo*  

Empirical likelihood estimator for diagnostic test likelihood ratio • Estimateur par vraisemblance empirique pour rapport de vraisemblance d'un test diagnostique [MS-143]

10:30 – 12:00 Session 06J



CHE 250

Estimation Methods



Méthodes d'estimation

Contributed Paper Session • Séance de communications libres


Chair • Président: Saumen MANDAL, *University of Manitoba*

10:30 Simon BONNER, *University of British Columbia*, Carl SCHWARZ, *Simon Fraser University*  

Capture-Recapture, Jolly-Seber, and the Horvitz-Thompson Estimator: ML Estimation of Abundance for Populations with Variable Capture Probabilities • Capture-recapture, Jolly-Seber et l'estimateur Horvitz-Thompson : estimation par maximum de vraisemblance de l'abondance de populations avec probabilités variables de capture [MS-144]

10:45 Min JIANG & Serge B. PROVOST, *University of Western Ontario*  



Optimal Orthogonal Series Density Estimates • Estimation optimale de densités par séries orthogonales [MS-145]

11:00 Mark SCHMIDT, EWOUT VAN DEN BERG, Friedlander MICHAEL & Murphy KEVIN, *University of British Columbia*  

Optimizing Costly Functions with Simple Constraints: A Limited-Memory Projected Quasi-Newton Algorithm • Optimisation de fonctions coûteuses sous contraintes simples : un algorithme de projection quasi-Newton à mémoire limitée [MS-146]

11:15 Zheng ZHANG & Mei Ling HUANG, *Brock University*  

On High Quantile Estimation for Heavy Tailed Distributions • Estimation des quantiles supérieurs pour distributions à ailes relevées [MS-147]

11:30 Yildiz YILMAZ, *University of Waterloo*  

Semiparametric Estimation for Sequential Survival Times • Estimation semi-paramétrique pour temps de survie séquentiels [MS-148]

11:45 Ivan CARRILLO-GARCIA & Milorad KOVACEVIC, *Statistics Canada*  

Some Issues in Estimation for Population Projections using Microsimulations • Quelques problèmes en estimation de projections de populations au moyen de microsimulations [MS-149]

10:30 – 12:00 Session 06K

IRC Lobby

Poster Session II

Séance d'affichage II

Contributed Poster Session

10:30 Tingting GOU & Duncan MURDOCH, *University of Western Ontario*  



BARS: An R Package for Binary Adaptive Rejection Sampling • BARS : une librairie R pour échantillonnage par rejet adaptatif binaire [MS-150]

10:30 Gun Ho JANG & Michael EVANS, *University of Toronto*  

The Information in One Prior Relative to Another • L'information dans une loi a priori relative à une autre [MS-151]

10:30 John GIBSON, *University of Waikato*, Xiangzheng DENG, *Chinese Academy of Sciences*, Scott ROZELLE, *Stanford University*, Jikun HUANG, *Chinese Academy of Sciences*  



Does Spatial Sampling Remove Spatial Autocorrelation while Preserving Inferences from Environmental Data? The Case of Roads and Forested Areas in China • L'échantillonnage spatial enlève-t-il la corrélation spatiale tout en conservant les inférences obtenues à partir de données environnementales ? Le cas des routes et forêts en Chine [MS-152]

10:30 Luke BORN, *University of British Columbia*, Ingo STEINWART & Marian ANGHEL, *Los Alamos National Labs*  

Filtering Vs. Support Vector Machine Forecasters: A Simulation Study • Filtrage contre prédicteurs par machines à vecteurs de support : une étude de simulation [MS-153]

10:30 Xiaoming LIU & Liqun WANG, *University of Manitoba*  

A Class of Generalized Shrunk Least Squares Estimators in Linear Models • Une classe d'estimateurs par rétrécissement des moindres carrés généralisés pour modèles linéaires [MS-154]

10:30 Mohammad CHOWDHURY & Saumen MANDAL, *University of Manitoba*, Ben TORSNEY, *University of Glasgow*  

Designing to Estimate Certain Parameters or Parametric Functions Independently of Others • Planification afin d'estimer certains paramètres ou certaines fonctions paramétriques indépendamment des autres [MS-155]

10:30 Kenneth Lo, *University of British Columbia*, Ryan BRINKMAN, *BC Cancer Research Centre*, Raphael GOTTARDO, *Institut de recherches cliniques de Montreal*  

flowClust: A Clustering Software for Automated Gating of Flow Cytometry Data • flowClust : une librairie de classification pour automatiser le portillonnage de données de cytométrie en flux [MS-156]

10:30 Sévérien NKURUNZIZA & Ejaz AHMED, *University of Windsor*  

Estimating and Pretesting in Multi-factors Ornstein-Uhlenbeck Processes • Estimation et pré-test dans les processus multivariés d'Ornstein-Uhlenbeck [MS-157]



12:00 – 13:30 Lunch Break • Dîner

13:30 – 14:30 Session 07**IRC 6**

Canadian Journal of Statistics Award Address

Allocution du récipiendaire du prix de la Revue canadienne de statistique

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Mary LESPERANCE, *University of Victoria*13:30 Qing PAN, *George Washington University*, Douglas SCHAUBEL, *University of Michigan*  Proportional Hazards Models based on Biased Samples and Estimated Selection Probabilities •
Modèles à risques proportionnels basés sur des échantillons biaisés et les probabilités de sélection
estimées [MS-158]**13:30 – 15:00 Session 08A****IRC 1**

Quantitative and Qualitative Challenges in Climate Change and Ecosystem Sustainability Research

Problèmes quantitatifs et qualitatifs en changement climatique et recherches sur la durabilité des écosystèmes

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Biostatistics Section

Chair • Présidente: Lisa LIX, *University of Saskatchewan*Organizer • Responsable: Colin SOSKOLNE, *University of Alberta*13:30 Colin SOSKOLNE, *University of Alberta*, Maud HUYNEN, *Maastricht University*, Brian LADD, *Alberta Health Services*, Pim MARTENS, *Maastricht University*  Eco-epidemiological Enquiry under Global Ecological Change • Enquête éco-épidémiologique
sous changements écologiques globaux [MS-159]13:50 Charmaine DEAN, Douglas WOOLFORD & Jiguo CAO, *Simon Fraser University*, David MARTELL, *University of Toronto*  Testing for Possible Climate Change-Caused Shifts in Forest Fire Ignitions • Tester l'hypothèse
d'une évolution due aux changements climatiques dans le déclenchement de feux de forêts [MS-160]14:10 Rick ROUTLEDGE, *Simon Fraser University*  March Winds and April Showers: Impacts on Coastal Sockeye Salmon • Vents de mars et pluies
d'avril : impacts sur le saumon rouge côtier [MS-161]14:30 Peter GUTTORP, *University of Washington* 

Discussant • Modérateur [MS-162]

13:30 – 15:00 Session 08B**IRC 4**

Quality Improvement in Healthcare

Amélioration de la qualité en soins de santé

Invited Paper Session • Présentations sur invitation



Session Sponsored by the Business and Industrial Statistics Section

Chair • Président: Bill WOODALL, *Virginia Tech*Organizer • Responsable: Charlie H GOLDSMITH, *McMaster University*13:30 Paul CASCAGNETTE, *Health Quality Council*  

Overcoming Barriers to the Use of SPC by Front Line Quality Improvement Teams • Surmonter les barrières à l'utilisation de la MSP par des équipes sur le terrain d'amélioration de la qualité [MS-163]

14:00 Charlie GOLDSMITH, *McMaster University*  

Experiences With Quality Improvement in a New Not-for-profit Nursing home • Expériences en amélioration de la qualité dans une nouvelle maison de soins infirmiers sans but lucratif [MS-164]

14:30 Malcolm MACLURE, *University of BC and BC Ministry of Health Services*  

Case-crossover Designs for Studies of Medical Errors and 'Near Misses' • Plans cas-croisés pour études d'erreurs médicales et de quasi-erreurs [MS-165]

13:30 – 15:00 Session 08C

MSL 102



Stochastic Modeling of Disordered Spatial Patterns

Modélisation stochastique des structures spatiales désordonnées

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Probability Section

Organizer and Chair • Responsable et président: Jeff PICKA, *University of New Brunswick*

13:30 Christine SOTEROS, *University of Saskatchewan*  

The Statistics of Self-avoiding Walk and Polygon Models of Polymers • La statistique des modèles de marches et polygones autoévitants pour polymères [MS-166]

14:00 Balint VIRAG, *University of Toronto*  

Determinantal Processes – Models Point Repulsion • Processus déterminantaux – modèles ponctuels avec répulsion [MS-167]

13:30 – 15:00 Session 08D



IRC 3

Analysis of Correlated Data with Excessive Zeros



Analyse de données corrélées à zéros excessifs

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Renjun MA, *University of New Brunswick*

13:30 Odd AALEN, *University of Oslo, Norway*, Tron A MOGER, *University of Oslo*  

Compound Poisson Distributions for Multivariate Frailty • Distributions de Poisson composées pour fragilité multivariée [MS-168]

14:00 M. Tariqul HASAN & Renjun MA, *University of New Brunswick*, Gary SNEDDON, *Memorial University of Newfoundland*  

Modeling Heterogeneity in Clustered Count Data With Extra Zeros Using Compound Poisson Random Effect • Modélisation de l'hétérogénéité dans des données de dénombrement en grappes avec surplus de zéros au moyen d'un effet aléatoire de Poisson composé [MS-169]

14:30 Dulal BHAUMIK & Kush KAPUR, *University of Illinois at Chicago*, Anindya ROY, *University of Maryland Baltimore County*, Robert GIBBONS, *University of Illinois at Chicago*  

Sample Size Determination for Mixed-Effects Longitudinal Logistic Regression Models • Détermination de la taille échantillonnale pour modèles de régression logistique à effets mixtes longitudinaux [MS-170]



13:30 – 15:00 Session 08E**IRC G65/66****Analysis and Modelling of Marine Ecological Data****Analyse et modélisation de données écologiques marines**

Topic Contributed Poster Session

National Institute for Complex Data Structures

Organizer and Chair • Responsable et président: Michael Dowd, *Dalhousie University*13:30 Jann Paul MATTERN, Katja FENNEL & Mike DOWD, *Dalhousie University*  

Application of Sequential Data Assimilation to a Biogeochemical Ocean Model • Application de l'assimilation de données séquentielle à un modèle biogéochimique de l'océan [MS-171]

13:30 Désirée TOMMASI, *University of British Columbia*, Rick ROUTLEDGE, *Simon Fraser University*, Brian HUNT & Evgeny PAKHOMOV, *University of British Columbia*  

Modelling Spring Phytoplankton Bloom Dynamics in a British Columbia Fjord • Modélisation de l'efflorescence dynamique du phytoplancton dans un fjord de la Colombie-Britannique [MS-172]

13:30 Greg BREED, Joanna MILLS FLEMMING, Mike DOWD & Christopher FIELD, *Dalhousie University*, Dan COSTA, *University of California Santa Cruz*  

Parameter Augmented Particle Filter/Smother Methods for Animal Tracking Data • Méthodes de filtres à particules/lisseuses augmentées de paramètres pour données de pistage d'animaux [MS-173]

13:30 Laurie AINSWORTH, Rick ROUTLEDGE & Jiguo CAO, *SFU*  

Functional Data Analysis in Ecosystem Research: the Decline of Oweekeno Lake Sockeye • Analyse de données fonctionnelles en recherche en écosystèmes : le déclin du saumon rouge du lac Oweekeno [MS-174]

13:30 – 15:00 Session 08F**MSL 101****Estimation Methods for Survey Data****Méthodes d'estimation pour données d'enquête**

Contributed Paper Session • Séance de communications libres

Chair • Président: Edward CHEN, *Statistics Canada*13:30 Jessica ANDREWS, *Statistics Canada*  

Changing the Methodology from Generic to Detail Allocation for Incorporated Business Tax Data • Changer la méthodologie de la désagrégation des données génériques aux détails pour les données fiscales des entreprises incorporées [MS-175]

13:45 Chang-Tai CHAO, *National Cheng Kung University, Taiwan*  

Ratio Estimators Improved by Rao-Blackwellization under Adaptive Cluster Sampling • Estimateurs par le quotient améliorés par Rao-Blackwellisation dans le cadre de l'échantillonnage en grappes adaptatif [MS-176]

14:00 Abdellatif DEMNATI, *Statistique Canada*, J.N.K. RAO, *Carleton University*  

Linearization Variance Estimation for Two-phase Sampling under Mass Imputation • Estimation de la variance de linéarisation pour échantillonnage à deux phases avec imputation massive [MS-177]

14:15 Stanislav KOLENIKOV, *University of Missouri, Columbia, USA*  

Applications of Quasi-Monte Carlo Methods in Inference for Complex Survey Data • Applications de méthodes quasi Monte-Carlo en inférence pour données d'enquête complexes [MS-178]

14:30 Alfredo NAVARRO & Michael STARSINIC, *US Census Bureau*  

Alternative Criteria for Data Quality Filtering in the ACS • Critères alternatifs en filtrage des données au sein de l'ACS [MS-179]

14:45 Kyle VINCENT, & Steve THOMPSON, *Simon Fraser University*  

Design Variations in Adaptive Web Sampling • Variations de plans d'expérience en échantillonnage en toile adaptatif [MS-180]

13:30 – 15:00 Session 08G

CHE 150

Statistical Genetics and Genomics II



Génétique et génomique statistiques II

Contributed Paper Session • Séance de communications libres

Chair • Président: Joseph BEYENE, *University of Toronto*

13:30 Zeny FENG, *University of Guelph*, William WONG, *University of Waterloo*, Gao XIN, *York University*, Flavio SCHENKEL, *University of Guelph*  

Generalized Association Study for Subjects Sampled from a Complex Pedigree • Étude d'association généralisée pour sujets à généalogie complexe [MS-181]

13:45 Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute & University of Toronto*, Jinnan LIU, *Samuel Lunenfeld Research Institute*, Adrian DOBRA, *University of Washington*, Helene MASSAM, *York University*  

Impact of Informative Prior in Discrete Bayesian Graphical Models: Application to Genome-Wide Association Studies • Impact de lois a priori informatives dans les modèles graphiques discrets bayésiens : application aux études sur l'ensemble du génome [MS-182]

14:00 Xuekui ZHANG & Kevin MURPHY, *University of British Columbia*, Peter PARÉ & Andrew SANDFORD, *The James Hogg iCAPTURE Center, University of British Columbia*  

Analysis of Family-based Genetic Association Study Data, a Bayesian Approach • Une approche bayésienne à l'analyse de données d'une étude d'association génétique basée sur les familles [MS-183]

14:15 Laura FAYE, *University of Toronto*, Shelley B BULL, *University of Toronto & Samuel Lunenfeld Research Institute, Mount Sinai Hospital*, Lei SUN, *University of Toronto*  

Effect Estimates in 2-stage Design: from Genome-wide Association to Sequencing • Estimateurs des effets en plans à deux étapes : de l'association sur tout le génome au séquençage [MS-184]

14:30 Ji-Hyung SHIN, Brad McNENEY & Jinko GRAHAM, *Simon Fraser University*  

Can Gene-by-environment Interaction be Inferred from Parent-case Transmission Rates? • Peut-on inférer l'interaction gène-environnement à partir de taux de transmission parent-cas ? [MS-185]

13:30 – 15:00 Session 08H

CHE 250

Bayesian Statistics and Applications

Statistique et applications bayésiennes

Contributed Paper Session • Séance de communications libres



Chair • Président: Gary SNEDDON, *Memorial University*

13:30 Gun Ho JANG, *University of Toronto*, Jaeyong LEE & Sangyeol LEE, *Seoul National University*  

Posterior Consistency of Species Sampling Priors • Cohérence de la loi a posteriori pour lois a priori d'échantillonnage des espèces [MS-186]

13:45 Anne-Sophie CHAREST, *Carnegie-Mellon University*  

Analysis of Tropical Cyclone Trends in the Atlantic • Analyse de la tendance des cyclones tropicaux dans l'Atlantique [MS-187]

14:00 Longhai LI, *University of Saskatchewan*  

Calibrating Predictions Based on a Selected Subset of Features from Bayesian Gaussian Classification Models • Calibration de prévisions basée sur un sous-ensemble de caractéristiques de modèles de classification bayésiens gaussiens [MS-188]

14:15 Saman MUTHUKUMARANA & Tim SWARTZ, *Simon Fraser University*  

Bayesian Analysis of Ordinal Response Survey Data using Dirichlet Process Priors • Analyse bayésienne de données d'enquête ordinales avec processus de Dirichlet comme loi a priori [MS-189]

14:30 Changjiang XU & A. Ian McLEOD, *University of Western Ontario*  

Adjustable Bayesian Information Criterion for Model Selection • Critère d'information bayésien ajustable pour sélection de modèles [MS-190]

14:45 Azizur RAHMAN, *University of Canberra*  

Effects of Different Prior Distributions on the Bayesian Predictive Inference • Effets de différentes distributions a priori sur l'inférence bayésienne [MS-191]

13:30 – 15:00 Session 08I

IRC Lobby

Poster Session III

Séance d'affichage III



Contributed Poster Session

13:30 Ye LI, *University of Toronto*  

Spatio-Temporal Modelling for Lupus Incidence in Toronto, Canada since 196 • Modélisation spatio-temporelle de l'incidence de Lupus à Toronto, Canada depuis 196 [MS-192]

13:30 Michelle STEANE, Paul McNICHOLAS & Rickey YADA, *University of Guelph*  



Model-Based Classification in Food Authenticity Studies • Classification basée sur un modèle dans les études d'authenticité des aliments [MS-193]

13:30 Ying QU, *AQL Management Consulting Inc.*, Keumhee CARRIERE, *University of Alberta*  


Direct Parametric Regression Models on Cumulative Incidence Function for Competing Risk Data • Modèles paramétriques de régression directe sur la fonction d'incidence cumulative pour données de risques concurrents [MS-194]

13:30 Mark WOLTERS & W. John BRAUN, *University of Western Ontario*  

Optimization Heuristics for Unimodal Kernel Density Estimation by Data Sharpening • Heuristiques d'optimisation pour estimation de densités unimodales par noyaux avec affinage de données [MS-195]

13:30 Jeong Eun MIN, Brad McNENEY & Jinko GRAHAM, *Simon Fraser University*  



A Statistical Method for High-throughput Screening of Predicted Orthologs • Une méthode statistique pour le tri à haut débit d'orthologues prédits [MS-196]

13:30 Xiuli KANG, & W. John BRAUN, *University of Western Ontario*  

Interval-censored Regression • Régression censurée par intervalles [MS-197]

13:30 Carrie PAQUETTE, *University of Manitoba*  

Statistical Analysis of Trends in Red River Water Quality at the International Boundary Over a 45-year Period • Analyse statistique des tendances de la qualité d'eau de la rivière Rouge à la frontière internationale pendant une période de 45 ans [MS-198]

- 13:30 Song CAI, *University of British Columbia*, Denise NEILSEN & Nathaniel NEWLANDS, *Agriculture and Agri-Food Canada*, James ZIDEK, *University of British Columbia*  

A Bayesian Survival Model of Phenology of High-value Perennial Agricultural Crops in the Okanagan, British Columbia • Un modèle de survie bayésien pour la phénologie de cultures vivaces de grande valeur dans l'Okanagan, Colombie-Britannique [MS-199]

- 13:30 Bingshu CHEN, *Queen's University*, Barry GRAUBARD, *NIH/NCI*, Katherine FLEGAL, *Centers for Disease Control*, Mitchell GAIL, *NIH/NCI*  

Evaluating Strategies to Estimate the Association of Obesity with Mortality via a Markov Model • Évaluer des stratégies d'estimation de l'association entre obésité et mortalité au moyen d'un modèle markovien [MS-200]

15:00 – 15:30 Coffee Break • Pause-café

IRC Lobby

15:30 – 17:00 Session 09A

IRC 6

Special Invited Session of the Biostatistics Section

Allocution de l'invité d'honneur du groupe de biostatistique

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Biostatistics Section

Organizer and Chair • Responsable et président: Subhash LELE, *University of Alberta*

- 15:30 Charles McCULLOCH & John NEUHAUS, *UCSF*  

Misspecifying a Random Effects Distribution: Why Getting it Wrong Probably Doesn't Matter • Pourquoi une mauvaise spécification d'une distribution à effets aléatoires n'est probablement pas très problématique [MS-201]

15:30 – 17:00 Session 09B

IRC 1

Innovative Uses of Statistical Software in the (Remote) Classroom

Utilisations novatrices des logiciels statistiques dans la salle de classe (à distance)


Invited Paper Session • Présentations sur invitation

Session Sponsored by the Statistical Education Committee

Organizer and Chair • Responsable et président: John AMRHEIN, *McDougall Scientific Ltd.*

- 15:30 Mark BAILEY, *SAS Institute Inc.*  

Interactive Demonstrations of Statistics with JMP • Démonstrations statistiques interactives avec JMP [MS-202]

- 16:00 Tim DACIUK, *SPSS Inc.*  

Using SPSS in Teaching Predictive Analytics • Utiliser SPSS dans l'enseignement de l'analytique prédictive [MS-203]

- 16:30 Roberto GUTIERREZ, *StataCorp LP*  

Teaching Stata in the Remote Classroom • Enseigner l'utilisation de Stata à distance [MS-204]

15:30 – 17:00 Session 09C**IRC 4**

Statistics/Mathematics Finance 1

Statistique/Mathématique Finance-1

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Business and Industrial Statistics Section

Organizer and Chair • Responsable et président: Mark REESOR, *University of Western Ontario*15:30 Adam METZLER, *University of Western Ontario*, Don L. McLEISH, *University of Waterloo*  

A Multiname First Passage Model for Credit Risk • Un modèle de premier passage à plusieurs noms pour le risque de crédit [MS-205]

16:00 Xikui WANG, *University of Manitoba*  

The Use of Statistical Models in Dynamic Pricing Problems • L'utilisation de modèles statistiques pour la fixation dynamique des prix [MS-206]

16:30 Yonggan ZHAO, *Dalhousie University*  

Equity Risk Premium and Volatility: A Correlation Structure • Prime de risque sur capitaux propres et volatilité : une structure de corrélation [MS-207]

15:30 – 17:00 Session 09D**IRC 5**

Small Area Estimation at Statistics Canada

Estimation relatives aux secteurs restreints chez Statistique Canada

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Survey Methods Section

Chair • Président: Jean-François BEAUMONT, *Statistics Canada*Organizer • Responsable: Mike HIDIROGLOU, *Statistics Canada*15:30 Emily BERG & Wayne A. FULLER, *Iowa State University*  

A SPREE Small Area Procedure for Estimating Population Counts • Une procédure SPREE pour petits domaines afin d'estimer les comptes spécifiques de la population [MS-208]

15:52 Mike HIDIROGLOU & Zdenek PATAK, *Statistics Canada*  

Estimation for Occupations Counts Using the Labour Force Survey • Estimations des comptes d'occupation en se servant de l'Enquête sur la population active du Canada [MS-209]

16:15 Yong You, *Statistics Canada*  

Area Level Modeling Approaches to Small Area Estimation with Applications • Estimation pour petits domaines au moyen de modèles régionaux avec exemples [MS-210]

16:37 Wesley YUNG & Susana RUBIN-BLEUER, *Statistics Canada*  

Small Area Estimation: A Business Survey Application • L'estimation de petits domaines appliquée à une enquête auprès des entreprises [MS-211]



15:30 – 17:00 Session 09E**MSL 101**

Special Session in Honour of Keith Worsley (1951-2009)

Séance en la mémoire de Keith Worsley (1951-2009)

Invited Paper Session • Présentations sur invitation

Session Sponsored partially by McGill University, Reception immediately following the session

Chair • Président: Román VIVEROS-AGUILERA, *McMaster University*Organizers • Responsables: George STYAN, *McGill University* & Jonathan TAYLOR, *Stanford University*15:30 Alastair SCOTT, *University of Auckland*  

Keith Worsley: his Early Years • Keith Worsley : ses jeunes années [MS-212]

16:00 Jonathan TAYLOR, *Stanford University*  

Keith Worsley's Contributions to Brain Mapping • Les contributions de Keith Worsley à l'imagerie cérébrale [MS-213]

16:30 George STYAN, *McGill University*  

A Photo Album for Keith Worsley • Un album de photos pour Keith Worsley [MS-214]



15:30 – 17:00 Session 09F**IRC G65/66**

Design and Analysis of Computer Experiments for Complex Systems



Planification et analyse d'expériences informatiques pour systèmes complexes

Topic Contributed Poster Session

National Institute for Complex Data Structures

Organizer and Chair • Responsable et président: Derek BINGHAM, *Simon Fraser University*15:30 Matthew PRATOLA, *Simon Fraser University*, Stephan SAIN, *National Center for Atmospheric Research*, Derek BINGHAM, *Simon Fraser University*  

Fast Calibration of Complex Computer Models • Calibration rapide de modèles informatiques complexes [MS-215]

15:30 Ryan LEKIVETZ, Randy SITTER & Derek BINGHAM, *Simon Fraser University*, M. HAMADA, L. MOORE & J. WENDELBERGER, *Los Alamos National Laboratory*  

A New Algorithm for the Construction of Computer Experiments with Mixed Levels • Un nouvel algorithme pour la construction d'expériences informatiques à niveaux mixtes [MS-216]

15:30 Pritam RANJAN, *Acadia University*, Derek BINGHAM, *Simon Fraser University*, George MICHAILIDIS, *University of Michigan*  

Sequential Design for Solving Inverse Problem for Expensive Deterministic Computer Simulators • Plan séquentiel pour résoudre le problème inverse pour simulateurs informatiques déterministes complexes [MS-217]












15:30 – 17:00 Session 09G**IRC 3**

Analysis of Spatial and Temporal Data

Analyse de données spatiales et temporelles

Contributed Paper Session • Séance de communications libres









Chair • Présidente: Grace CHIU, *University of Waterloo*

- 15:30 Patrick BROWN, *Cancer Care Ontario*, Peter A HENRYS, *JBA Consulting*  
Longitudinal Spatial Point Processes and Residential Histories • Processus ponctuel spatial longitudinal et histoires résidentielles [MS-218]
- 15:45 Carolyn HUSTON & Carl SCHWARZ, *Simon Fraser University*  
Spatial CAR Models And Compositional Data • Modèles ARC spatiaux et données compositionnelles [MS-219]
- 16:00 Xin FENG, *Simon Fraser University*, Charmaine DEAN, *Simon Fraser University*  
Hot Spot Detection in Disease Mapping • Détection de points chauds en cartographie des maladies [MS-220]
- 16:15 Olaf BERKE, *University of Guelph*  
Influence of Imperfect Diagnostic Tests on Spatial Patterns in Regional Epidemiological Data • L'influence de tests diagnostics imparfaits sur les motifs spatiaux dans des données épidémiologiques régionales [MS-221]
- 16:30 Mahmoud TORABI & Rhonda J. ROSYCHUK, *University of Alberta*  
Spatio-temporal Modeling Using Spline for Disease Mapping: Analysis of Childhood Cancer Trends • Modélisation spatio-temporelle au moyen de splines en cartographie des maladies : une analyse des tendances du cancer pédiatrique [MS-222]
- 16:45 Glen TAKAHARA, Azadeh MOGHADERI & David THOMSON, *Queen's University*  
Unaliasing of Line Components in Bivariate Time Series • Dépliement de composantes lignes en séries chronologiques bivariées [MS-223]



15:30 – 17:00 Session 09H**MSL 102****Bayesian Methods in Biostatistics****Méthodes bayésiennes en biostatistique**

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Therese STUKEL, *Institute for Clinical Evaluative Sciences*

- 15:30 Ji CHENG, *McMaster University and St. Joseph's Healthcare Hamilton*, Shelley SALPETER, *Santa Clara Valley Medical Center and Stanford University*, Lehana THABANE, *McMaster University and St. Joseph's Healthcare Hamilton*, Nicholas BUCKLEY, *California Institute of Technology*, Edwin SALPETER, *Cornell University*  
Hormone Therapy and Mortality in Younger Postmenopausal Women: Synthesizing the Evidence using Bayesian Meta-analysis • Thérapie hormonale et mortalité chez les plus jeunes femmes post-ménopause : synthétiser l'information par méta-analyse bayésienne [MS-224]
- 15:45 Lawrence McCANDLESS, *Simon Fraser University*  
Hierarchical Priors for Bias Parameters in Bayesian Adjustment for Unmeasured Confounding • Lois a priori hiérarchiques pour paramètres de biais en ajustement bayésien pour confusion non mesurée [MS-225]
- 16:00 Eleanor PULLENAYEGUM, *McMaster University*, Kevin WONG, *University of Toronto*, Savio YU & Anne HOLBROOK, *McMaster University*  
Informed Priors for Between-study Variability in Bayesian Meta-analyses • Lois a priori informées pour la variabilité inter-études en méta-analyse bayésienne [MS-226]
- 16:15 Guohua YAN & Renjun MA, *University of New Brunswick*  
Bayesian Hierarchical Mixture Modeling for Clustered Count Data with Extra Zeros • Modélisation

bayésienne hiérarchique de mélanges pour données de dénombrement en grappes avec excès de zéros [MS-227]

16:30 Shahadut HOSSAIN, *BC Cancer Research Centre*, Paul GUSTAFSON, *University of British Columbia*  

Bayesian Adjustment for Covariate Measurement Errors: A Flexible Parametric Approach • Ajustement bayésien pour erreurs de mesures dans les covariables : une approche paramétrique flexible [MS-228]

15:30 – 17:00 Session 09I



CHE 150

Statistical Theory

Théorie statistique

Contributed Paper Session • Séance de communications libres

Chair • Président: Mikelis BICKIS, *University of Saskatchewan*

15:30 Noomen El Beji BEN GHORBAL & Christian GENEST, *Université Laval*, Johanna NEÖLEHOV, *ETH Zürich*  

On the Test of Ghoudi, Khoudraji and Rivest • À propos du test de Ghoudi, Khoudraji et Rivest [MS-229]

15:45 Katherine DAVIES, *University of Manitoba*, N. BALAKRISHNAN, *McMaster University*, Jerome P. KEATING, *University of Texas at San Antonio*  

Pitman Closeness of Order Statistics to Population Quantiles • Proximité de Pitman des statistiques d'ordre par rapport aux quantiles de population [MS-230]

16:00 Qifeng JIANG, Michael STEPHENS & Richard LOCKHART, *Simon Fraser University*  



On Fitting a Mixture of two von Mises Distributions • Ajustement d'un mélange de deux distributions von Mises [MS-231]

16:15 Taoufik BOUEZMARNI & Roch ROY, *Université de Montréal*, Abderrahim TAAMOUTI, *Universidad Carlos III de Madrid Calle Madrid, Spain*  

Nonparametric Tests for Conditional Independence Using Conditional Distribution and Quantile Regression • Tests non paramétriques d'indépendance conditionnelle basés sur la distribution conditionnelle et la régression de quantile [MS-232]

16:30 Zheng SUN, *University of Victoria*, Michael STEPHENS, *Simon Fraser University*  

Estimation of Testing fit for the von Mises Distribution with the Grouped Data • Estimation et test d'adéquation pour distribution von Mises avec données groupées [MS-233]

16:45 Qian ZHOU, *University of Waterloo*, Peter X.-K.SONG, *University of Michigan*, Mary THOMPSON, *University of Waterloo*  

Information Ratio Test for Model Misspecification • Test du rapport des informations pour mauvaise spécification de modèle [MS-234]

15:30 – 17:00 Session 09J**CHE 250****Computational Methods and Algorithms****Méthodes et algorithmes informatiques**



Contributed Paper Session • Séance de communications libres

Chair • Présidente: Shelley BULL, *University of Toronto*15:30 Iain CURRIE, *Heriot-Watt University, UK*  



GLAM: An Introduction to Array Methods in Statistics • MLGT : une introduction aux méthodes tabulaires en statistique [MS-235]

15:45 Olena FROLOVA, Jean-Marc FILLION & Jean-Louis TAMBAY, *Statistics Canada*  



Confid2: Statistics Canada's New Tabular Data Confidentiality Software • Confid2 : nouveau logiciel de confidentialité des données tabulaires de Statistique Canada [MS-236]

16:00 Hui SHEN & William J. WELCH, *University of British Columbia*, Jacqueline M. HUGHES-OLIVER, *North Carolina State University*  

Efficient, Adaptive Cross Validation for Tuning and Comparing Models • Validation croisée adaptative efficace pour mise au point et comparaison de modèles [MS-237]

16:15 Yasaman HOSSEINKASHI, Christopher SMALL & Shoja'eddin CHENOURI, *University of Waterloo*  

Analyzing the 2001 FMD Outbreak Using Dynamic Graph Models • Analyse de la poussée de fièvre aphteuse de 2001 au moyen de modèles de graphes dynamiques [MS-238]

16:30 Colin BABYAK, Asma ALAVI, Krista COLLINS, Amanda HALLADAY & Dawn TAPPER, *Statistics Canada*  



The Crime Severity Index - Finally, the One, True Measure of Crime...? • L'indice de sévérité des crimes - finalement la vraie mesure du crime... ? [MS-239]

16:45 Ali GHODSI, Stephen VAVASIS & Michael BIGGS, *University of Waterloo*  

A Modified Power Iteration for Nonnegative Matrix Factorization • Une méthode des puissances modifiée pour la factorisation de matrices non négatives [MS-240]

15:30 – 17:00 Session 09K**IRC Lobby****Poster Session IV****Séance d'affichage IV**

Contributed Poster Session

15:30 William HARPER, *Otterbein College*, Ted ESCHENBACH, *TEG Consulting*, Thomas JAMES, *Otterbein College*  



Reliability Confidence Intervals for Oil Spills in the Gulf of Mexico • Intervalles de confiance de la fiabilité pour déversements d'hydrocarbures dans le golfe du Mexique [MS-241]

15:30 Yi LIN & Ying MACNAB, *University of British Columbia*  

Bayesian Ecological Modeling of Small Area Suicide Rates • Modélisation écologique bayésienne de taux de suicide dans de petits domaines [MS-242]

15:30 Conghui QU & Jinko GRAHAM, *Simon Fraser University*, John SPINELLI, *B.C. Cancer Agency*  



Multiple Hypothesis Testing Procedures with applications to Epidemiologic Studies of Genetic Susceptibility • Procédures pour tester de multiples hypothèses avec applications à des études épidémiologiques de susceptibilité génétique [MS-243]

15:30 Paul NGUYEN & Duncan MURDOCH, *The University of Western Ontario*  

Confidence Intervals for Binomial Probability Predictions • Intervalles de confiance pour prévisions de probabilités binomiales [MS-244]

15:30 John GIBSON, *University of Waikato*, Kathleen BEEGLE, *World Bank*, Joachim DEWEERDT, *EDI*, Jed FRIEDMAN, *World Bank*  

Measuring Household Consumption: A Survey Methods Experiment in Tanzania • Mesurer la consommation des ménages : une expérience en méthodes d'enquête en Tanzanie [MS-245]

15:30 Camila PEDROSO ESTEVAM DE SOUZA, *University of British Columbia*, Ronaldo DIAS, *University of Campinas*  

Hypothesis Tests for Functional Data Based on Distances: a Study Using Splines • Tests d'hypothèses pour données fonctionnelles basés sur les distances : une étude utilisant des splines [MS-246]

Mercredi 3 juin • Wednesday, June 3

08:30 – 09:30 Session 10

IRC 6

Pierre Robillard Award Address

Allocution du récipiendaire du prix Pierre-Robillard

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Sudhir R. PAUL, *University of Windsor*

08:30 Baojiang CHEN, *University of Washington*, Grace Yi & Richard COOK, *University of Waterloo*  

Statistical Methods for Multi-State Analysis of Incomplete Longitudinal Data • Méthodes statistiques pour l'analyse multi-états de données longitudinales incomplètes [MS-247]

08:30 – 10:00 Session 11A

IRC 1

Statistics/Mathematics Finance 2



Statistique/Mathématique Finance-2

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Business and Industrial Statistics Section

Chair • Président: Adam METZLER, *University of Western Ontario*

Organizers • Responsables: Mark REESOR, *University of Western Ontario*

08:30 Alex BADESCU & Robert ELLIOT, *University of Calgary*, Tak Ken SIU, *Curtin University of Technology*  

Bond Valuation Under Regime-switching models • Évaluation d'obligations sous modèles à changement de régimes [MS-248]

09:00 Adam KOLKIEWICZ, *University of Waterloo*, Joonghee HUH, *Prudential Financial*  

Pricing Basket Options With Barriers • Fixation des prix d'options sur panier en présence de barrières [MS-249]

09:30 Mark REESOR, *University of Western Ontario*, Shudan LIU, *TD*  

Swinging in Liquidity: The Use of Swing Options to Manage Liquidity Risk • L'utilisation d'options swing pour gérer le risque d'illiquidité [MS-250]

08:30 – 10:00 Session 11B

IRC 4



Modelling and Measuring Pollution and Natural Systems

Modélisation et mesure de la pollution et des systèmes naturels



Invited Paper Session • Présentations sur invitation

Session Sponsored by the Committee on Women in Statistics



Organizer and Chair • Responsable et présidente: Sylvia ESTERBY, *University of British Columbia, Okanagan*

08:30 Yulia GEL & Bei CHEN, *University of Waterloo*  

Probabilistic Weather Prediction for Local Weather Users based on Statistical Ensembles of Forecasts • Prédiction probabiliste météorologique pour usagers locaux basée sur des ensembles de prévisions statistiques [MS-251]

09:00 Hyang Mi KIM, *University of Calgary*, Igor BURSTYN, Ambikaipakan SENTHILSELVAN & Nicola CHERRY, *University of Alberta*, Elise PIETRONIRO & Cheryl WALDNER, *University of Saskatchewan*  

Industrial Sources Influence Air Concentrations of Hydrogen Sulfide and Sulfur Dioxide in Rural Areas of Western Canada • Des sources industrielles influencent la concentration atmosphérique de sulfure d'hydrogène et de dioxyde de soufre des zones rurales de l'ouest canadien [MS-252]

09:30 Laura COWEN, *University of Victoria*  

A Bayesian Capture-recapture Model With Premature Radio-tag Failure • Un modèle bayésien de capture-recapture avec défaillance prématurée des radio-étiquettes [MS-253]

08:30 – 10:00 Session 11C

IRC 5

Multivariate Survival Analysis

Analyse de survie à variables multiples

Invited Paper Session • Présentations sur invitation



Organizer and Chair • Responsable et président: Lajmi LAKHAL-CHAIEB, *Université Laval*

08:30 Yingwei PENG, *Queen's University*, Jeremy TAYLOR, *University of Michigan*  

Cure Models for Clustered Survival Data • Modèles avec fraction non à risque pour données de survie en grappes [MS-254]

09:00 Richard COOK & Jerald LAWLESS, *University of Waterloo*  

A Copula-based Mixed Poisson Model for Bivariate Recurrent Events under Event-Dependent Censoring • Un modèle mixte de Poisson basé sur les copules pour événements récurrents bivariés avec censure dépendante des événements [MS-255]

09:30 Arusharka SEN, *Concordia University*, Winfried STUTE, *Justus-Liebig University, Giessen, Germany*  

The Multivariate Kaplan-Meier Estimator via an Eigenfunction Problem • L'estimateur de Kaplan-Meier multivarié au moyen d'un problème de fonctions propres [MS-256]

08:30 – 10:00 Session 11D

IRC 3

Theory and Application of the Finite Mixture Models



Théorie et application des modèles de mélange fini

Invited Paper Session • Présentations sur invitation


Organizer and Chair • Responsable et président: Jiahua CHEN, *University of British Columbia*

08:30 Leonard MACLEAN & Yonggan ZHAO, *Dalhousie University*  

Financial Market Regimes and Mixture Models • Régimes de marchés financiers et modèles de mélanges [MS-257]

09:00 Jiayang SUN, *Case Western Reserve University*, Peng LIU, *Eli Lilley, USA*, Chen JIAHUA, *University of British Columbia*  

Efficient Adaptive Mixture Estimation • Estimation adaptative efficace de mélanges [MS-258]

09:30 Pengfei LI, *University of Alberta*, Jiahua CHEN, *University of British Columbia*  

Testing the Order of a Finite Mixture • Tester l'ordre d'un mélange fini [MS-259]

08:30 – 10:00 Session 11E



CHE 150

Models for Count Data and Excessive Zeros



Modèles pour données de dénombrement et zéros excessifs

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Rhonda ROSYCHUK, *University of Alberta*

08:30 Nabil CHANNOUF, *Group for Research in Decision Analysis (GERAD)*, Marc FREDETTE, *HEC Montreal*, Brenda MACGIBBON, *UQAM*  



Sample Size Calculation For Hierarchical Poisson Regression Models • Détermination de la taille échantillonnale dans un modèle hiérarchique de régression de Poisson [MS-260]

08:45 Jeffrey DOWDEN, *NL Centre for Health Information*, Khokan SIKDAR & Reza ALAGHEHBANDAN, *NL Centre for Health Information, Memorial University*  

Statistical Inferences of Childhood Injury Data using Mixed-Effects Zero-Inflated Poisson Models • Inférences statistiques sur des données de blessures chez les enfants au moyen de modèles à effets mixtes de Poisson avec surreprésentation de zéros [MS-261]

09:00 Nader FALLAH, Kenneth ROCKWOOD & Arnold MITNITSKI, *Dalhousie University*  



Non-linear Poisson Regression Using Neural Networks: Application to Cognitive Score Changes in Elderly • Régression de Poisson non linéaire utilisant des réseaux de neurones : application aux changements de scores cognitifs chez les personnes âgées [MS-262]

09:15 Khokan SIKDAR, Jeffrey DOWDEN & Reza ALAGHEHBANDAN, *Newfoundland and Labrador Centre for Health Information, Memorial University of Newfoundland*, Don MACDONALD, *Newfoundland and Labrador Centre for Health Information*, Veeresh GADAG, *Memorial University of Newfoundland*  

Modeling Count Data with an Application to Adverse Drug Reaction in Hospitalized Patients • Modélisation de données de dénombrement avec une application à la réaction adverse aux médicaments chez les patients hospitalisés [MS-263]

09:30 Connie STEWART, *University of New Brunswick Saint John*, Christopher FIELD, *Dalhousie University*  

Managing the Essential Zeros in Quantitative Fatty Acid Signature Analysis • Gérer les zéros essentiels en analyse quantitative de signature d'acides gras [MS-264]

09:45 Sadia MAHMUD, *Aga Khan University, Karachi, Pakistan*, Neil JOHNSTON, *McMaster University*  

Probit- Log- Skew-Normal Mixture Model for longitudinal outcomes with excess zeros • Modèle de mélanges probit/log-normaux-assymétrique pour réponses longitudinales avec excès de zéros [MS-265]

08:30 – 10:00 Session 11F

CHE 250

Biostatistical Applications, Issues, and Interpretations



Applications, problèmes et interprétations biostatistiques

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Karen KOPCIUK, *Alberta Cancer Board*

08:30 Gordon FICK, *University of Calgary*  

Modifying a Modifier, Confounding a Modifier, Confounding a Confounder • Modifier un modificateur, confondre un modificateur et confondre un confondant [MS-266]

08:45 Jennifer LIU & Lehana THABANE, *McMaster University*, Karen MILLIGAN, *Credit Valley Hospital*, Alison NICCOLS & Wendy SWORD, *McMaster University*, Joanna HENDERSON, *Centre for Addiction and Mental Health*, Ainsley SMITH, *McMaster University*  



Statistical Issues in a Meta-analysis of Studies of Integrated Treatment Programs for Women with Substance Use Problems and Their Children • Difficultés statistiques en méta-analyses d'études de programmes de traitements intégrés pour femmes ayant des problèmes d'abus de substances et leurs enfants [MS-267]

09:00 Binod NEUPANE, Mark LOEB, Stephen WALTER & Paul KRUEGER, *McMaster University*  

Analysis and Interpretation Issues When There are Multiple Control Groups in a Case-control Study • Problèmes d'analyse et d'interprétation lorsqu'il y a plusieurs groupes contrôle dans une étude cas-témoins [MS-268]

09:15 Lilia RAMIREZ RAMIREZ & Mary THOMPSON, *University of Waterloo*  

Effectiveness of Control Measures for Diseases That Spread in Populations With Contact Network Structure • Efficacité des mesures de contrôle de maladies transmises par structure de réseaux de contacts [MS-269]

09:30 Julia TALEBAN & GuangYong ZOU, *University of Western Ontario*  

Confidence Intervals for Cost Data in Cluster Randomization Trials: Does it Make a Difference? • Intervalles de confiance pour données de coûts en essais randomisés par grappe : y a-t-il une différence ? [MS-270]

09:45 Swarna WEERASINGHE, *Dalhousie University*.  

Development and Evaluation of an Efficient and Valid Missing Values Imputation Method to Investigate the Health Effects of Air Pollution Levels • Développement et évaluation d'une méthode efficace et valide pour l'imputation de valeurs manquantes dans l'étude des effets sur la santé de niveaux de pollution atmosphérique [MS-271]

08:30 – 10:00 Session 11G



SCR 100

Survey Design and Analysis

Planification et analyse d'enquêtes

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Patricia WHITRIDGE, *Canada Revenue Agency*

08:30 Nora BOHOSSIAN, Pierre DAOUST & Sylvie GAUTHIER, *Statistics Canada*  

The Sample Design of the Quarterly Survey of Financial Statistics: Review And Improvements • Plan de sondage de l'Enquête trimestrielle sur les statistiques financières : revue et améliorations [MS-272]

08:45 Jeannine CLAVEAU, Jack LOTHIAN & Sylvie GAUTHIER, *Statistique Canada*  

Rent Component of the Consumer Price Index • Méthodologie de la composante loyer de l'Indice des prix à la consommation [MS-273]

09:00 José GAUDET, *Statistique Canada*  

Recent Improvements to Statistics Canada's Business Register Industry Autocoding Process • Améliorations récentes au processus de codage automatisé de l'activité industrielle pour le Registre des entreprises de Statistique Canada [MS-274]

09:15 Javier OYARZUN, *Statistique Canada*  

New Methodology for the T1 tax data processing • Nouvelle méthodologie pour le traitement des données fiscales T1 [MS-275]

08:30 – 10:00 Session 11H

MSL 101

Statistical Education in Action

L'éducation statistique en action

Contributed Paper Session • Séance de communications libres

Chair • Président: Kevin THORPE, *University of Toronto*

08:30 Mikelis BICKIS, *University of Saskatchewan*  

Teaching Regression using Interactive Visualization • L'enseignement de la régression en utilisant la visualisation interactive. [MS-276]

08:45 Anne Michele MILLAR, *Mount Saint Vincent University*  

Choosing a Text for a Service Course in Introductory Statistics • Choisir un livre pour un cours de service d'introduction à la statistique [MS-277]

08:30 – 10:00 Session 11I

MSL 102

Stochastic Processes and Applications

Processus stochastiques et applications

Contributed Paper Session • Séance de communications libres

Chair • Président: John ASTON, *University of Warwick*

08:30 David BRILLINGER, *UC Berkeley*  

Modelling Trajectories via Stochastic Functional Differential Equations • Modélisation de trajectoires au moyen d'équations différentielles stochastiques fonctionnelles [MS-278]

08:45 Sirlei CAVASSIN, Xiaoming LIU & Hao YU, *University of Western Ontario*  

Fractional Integrated Process for Mortality Index k_t in Lee-Carter Model • Processus intégré fractionnel pour l'indice de mortalité k_t dans le modèle Lee-Carter [MS-279]

09:00 Jisoo JANG & Xiaoming LIU, *University of Western Ontario*  

Stochastic Life Annuity • Rente viagère stochastique [MS-280]

09:15 Mahmoud ZAREPOUR, *University of Ottawa*, Hemant ISHWARAN, *Cleveland clinic foundation*  

Series Representations for Multivariate Generalized Gamma Processes via a Scale Invariance Principle • Représentations en série pour processus gamma généralisés multivariés au moyen d'un principe d'invariance aux changements d'échelle [MS-281]

09:30 Candemir CIGSAR, *University of Waterloo*  

Testing for Carryover Effects in Recurrent Event Processes • Tester pour les effets résiduels dans des processus d'événements récurrents [MS-282]

09:45 Lihui ZHAO & X. Joan HU, *Simon Fraser University*  

Estimation with Incompletely Observed Semi-Markov Processes • Estimation avec processus semi-markoviens incomplètement observés [MS-283]

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

IRC Lobby

10:30 – 12:00 Session 12A**IRC 1****Ecological Processes, Mechanisms, and Statistical Evidence****Processus, mécanismes et preuves statistiques écologiques**

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Biostatistics Section

Organizer and Chair • Responsable et président: Christopher JERDE, *University of Notre Dame*

10:30 Christopher JERDE, *University of Notre Dame*, David FINNOFF, *University of Wyoming*, Katherine SMITH, *Brown University*, Peter DASZAK, *Consortium for Conservation Medicine*, David LODGE, *University of Notre Dame*  

The Economic Uncertainty of Emerging Infectious Diseases and Biological Invasions • L'incertitude économique des maladies infectieuses et invasions biologiques émergentes [MS-284]

11:00 Eric WARD & Eli HOLMES, *NOAA*, Haridas CHIRAKKAL & Leah GERBER, *ASU*  

Revealing Spatial Population Structure With Time Series Data • Révéler des structures spatiales des populations au moyen de données de séries chronologiques [MS-285]

11:30 Andrew EDWARDS, *Pacific Biological Station, DFO*  

How a Lack of Statistical Analysis Led to Claims that Many Animals Forage Using Similar Random Walks Described by Power-law Distributions • Comment un manque d'analyse statistique mène à l'affirmation que plusieurs animaux paissent utilisant des marches aléatoires similaires décrites par des distributions de loi de puissance [MS-286]

10:30 – 12:00 Session 12B**IRC 4****Empirical Processes and Applications to Finance****Processus empiriques et applications à la finance**

Invited Paper Session • Présentations sur invitation

Session Sponsored by the Probability Section



Organizer and Chair • Responsable et président: Hao YU, *University of Western Ontario*

10:30 Piotr KOKOSZKA, *Utah State University*, Lajos HORVATH, *University of Utah*, Robertas GABRYS, *Utah State University*  

Residuals in the Functional Linear Model With Application to Goodness-of-fit testing • Résidus dans le modèle fonctionnel linéaire avec application aux tests d'ajustement [MS-287]

11:00 Bruno RÉMILLARD, *HEC Montreal*  

Monte Carlo Methods for Testing Goodness-of-fit • Méthodes Monte Carlo pour tests d'adéquation [MS-288]

11:30 Hao YU, *University of Western Ontario*  

Empirical Processes of Standardized Residuals in GARCH Models • Processus empiriques de résidus standardisés en modèles GARCH [MS-289]

10:30 – 12:00 Session 12C**IRC 5****What do Mathematical Statisticians do at Statistics Canada****Travail des statisticiens mathématiques chez Statistique Canada**

Invited Paper Session • Présentations sur invitation


Session Sponsored by the Survey Methods Section

Chair • Président: Julie TRÉPANIÉ, *Statistics Canada*

Organizers • Responsables: Lenka MACH & Michelle SIMARD, *Statistics Canada*

10:30 Cindy UBARTAS, *Statistique Canada*  

Synthetic Data Files and PUMFs, Two Applications Taken from Post-Censal Surveys • Les fichiers de données synthétiques et les FMGD, deux applications tirées d'enquêtes post-censitaires [MS-290]

11:00 Chi Wai YEUNG, *Statistics Canada*  

Challenges and Gains in Using Administrative Data in the Unified Enterprise Survey • Défis et avantages reliés à l'utilisation de données administratives dans le cadre de l'Enquête unifiée auprès des entreprises [MS-291]

11:30 Krista COLLINS, *Statistics Canada*  

Spatial Modelling of Geocoded Crime Data • Modélisation spatiale de données géocodées sur la criminalité [MS-292]

10:30 – 12:00 Session 12D



IRC 3

Statistical Methods for Causal Inference

Méthodes statistiques pour l'inférence causale

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Lawrence McCANDLESS, *Simon Fraser University*

10:30 Ashkan ERTEFAIE, *McGill University*  

Comparing Approaches to Causal Inference for Longitudinal Data: GPS vs IPTW • Comparaison de méthodes d'inférence causale de données longitudinales : SGP vs PPIT [MS-293]

11:00 Zhiqiang TAN, *Rutgers University*, Zhiqiang TAN,  

Outcome Regression, Propensity Scores, and Instrumental Variables • Régression des résultats thérapeutiques, scores de propension et variables instrumentales [MS-294]

11:30 Robert PLATT, *McGill University*  

Flexible Marginal Structural Models • Modèles structuraux marginaux flexibles [MS-295]

10:30 – 12:00 Session 12E

IRC 6

Session in Honor of Jack Kalbfleisch's 65th Birthday: Analysis of Life History Data Arising from Non-standard Selection and Observation Schemes

Séance en honneur du 65^e anniversaire de Jack Kalbfleisch : analyse de données de cycles de vie provenant de schémas de sélection et d'observation non standards

Invited Paper Session • Présentations sur invitation

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

10:30 Kajsa KVIST, Per KRAGH ANDERSEN, Jules ANGST & Lars Vedel KESSING, *University of Copenhagen*  

Effect of Event Dependent Sampling in the Analysis of Recurrent Events • Effet de l'échantillonnage dépendant des événements en analyse d'événements récurrents [MS-296]

11:00 Jerald LAWLESS & Dagmar MARIACA-HAJDUCEK, *University of Waterloo*  

Analysis of Duration Data From Longitudinal Surveys • Analyse de données de durée pour études longitudinales [MS-297]

11:30 Douglas SCHAUBEL & Jack KALBFLEISCH, *University of Michigan*  

Estimating the Effect of a Time-Dependent Therapy on Restricted Mean Lifetime using Observational Data • Estimation de l'effet d'une thérapie dépendant du temps sur l'espérance de vie restreinte au moyen de données d'observations [MS-298]

10:30 – 12:00 Session 12F

MSL 101

Design and Analysis for Business and Industrial Statistics

Planification et analyse pour statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

Chair • Président: Jason NIELSEN, *Carleton University*

10:30 Sheau-Chiann CHEN & Jeh-Nan PAN, *National Cheng Kung University, Taiwan*  



Determining Optimal Number of Samples for Constructing Multivariate Control Chart • Détermination du nombre optimal d'échantillons pour la construction d'une carte de contrôle multivariée [MS-299]

10:45 Oana DANILA & R. Jock MacKAY & Stefan H. STEINER, *Univeristy of Waterloo*  



Assessment of Binary Measurement Systems • Évaluation de systèmes de mesure binaires [MS-300]

11:00 Amaan MEHRABIAN, Tigran ATOYAN & David STEPHENS, *McGill University*  

On informational Content of Model-Based Volatility Forecasts and Implied Volatility • Contenu en information des prévisions de volatilité basées sur des modèles et de la volatilité impliquée [MS-301]

11:15 Nathaniel NEWLANDS, *Environmental Health, Agriculture and Agri-Food Canada*, Mohua PODDER, *Joint with University of British Columbia and Agriculture and Agri-Food Canada*, James ZIDEK, *University of British Columbia*, Lawrence TOWNLEY-SMITH & Brian McCONKEY, *Agriculture and Agri-Food Canada*  

Bayesian Network Model of Regional Biofuel Supply Chains • Modélisation de chaînes logistiques de biocarburants au moyen de réseaux bayésiens [MS-302]

11:30 Jenna TICHON, *University of Manitoba*, Robert McLEOD, *University of Winnipeg*  

Semifolding Two-Level Fractional Factorial Split-Plot Designs • Plans semi-repliés à deux niveaux fractionnels factoriels avec subdivision des parcelles [MS-303]

10:30 – 12:00 Session 12G

CHE 150

Mixed Models, Multistate Modeling, and Measurement Error

Modèles mixtes, modélisation à plusieurs états et erreurs de mesure

Contributed Paper Session • Séance de communications libres



Chair • Président: Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute*

10:30 Depeng JIANG, *St Michael's Hospital*, Dorcas BEATON & Kenneth TANG, *St. Michael's Hospital*, Sheilah HOGG-JOHNSON, *Institute for Work and Health*  

Two-Part Mixed and Mixture Models for Longitudinal Data with both Random and Non-Random Missing Data • Modèle aléatoire et modèle de croissance en deux parties pour données longitudinales avec données manquantes aléatoires et non-aléatoires [MS-304]

10:45 Wei LIU, *York University*, Lang WU, *University of British Columbia*  



Two-Step and Likelihood Methods for Semiparametric Nonlinear Mixed-effects Models with Measurement Errors and Missing Covariates • Méthodes par vraisemblance et à deux étapes pour modèles à effets mixtes non linéaires semi-paramétriques avec erreurs dans les mesures et covariables manquantes [MS-305]

11:00 Arnold MITNITSKI & Nader FALLAH, *Dalhousie University*, Charmaine DEAN, *Simon Fraser University*, Kenneth ROCKWOOD, *Dalhousie University*  

A Multi-state Modeling Approach to Changes in Cognitive Scores Over a Fixed Time Interval • Une approche multi-états à la modélisation de changements dans les scores cognitifs sur un intervalle de temps fixe [MS-306]

11:15 Jesse RAFFA & Joel DUBIN, *University of Waterloo*  

Multistate Models for Hepatitis C Virus Testing, Seroconversion and Death • Modèles à plusieurs états pour le dépistage, la séroconversion et la mortalité associés au virus de l'hépatite C [MS-307]

11:30 Michael REGIER, *University of British Columbia*  

Generalized Imperfect Variables for Generalized Linear Models • Variables imparfaites généralisées pour modèles linéaires généralisés [MS-308]

11:45 Peng ZHANG, *University of Alberta*, Peter X.-K. SONG, *University of Michigan*  

Joint Mean-covariance Modeling With Nonlinear Random Mean Models for Longitudinal Data Analysis • Modélisation conjointe de la moyenne et de la covariance avec modèles à moyennes aléatoires non linéaires en analyse de données longitudinales [MS-309]

10:30 – 12:00 Session 12H

MSL 102

Modeling, Classification, and Algorithms

Modélisation, classification et algorithmes

Contributed Paper Session • Séance de communications libres



Chair • Présidente: Aurélie LABBE, *McGill University*

10:30 Chris WILD, Alan LEE & Alastair SCOTT, *University of Auckland*  



Estimating Equations and Two-phase Sampling • Équations d'estimation et échantillonnage à deux phases [MS-310]

10:45 Rod WALKER, *Group Health Center for Health Studies, Seattle, WA*, Sebastien HANEUSE, *Group Health Center for Health Studies; University of Washington*  



Adjustment for Selection Bias in Constructing and Evaluating Prediction Models • Ajustement pour le biais de sélection en construction et évaluation de modèles de prévision [MS-311]

11:00 Mark BEASLEY, *University of Alabama at Birmingham*  



Issues with Rank-Based Inverse Normal Transformations • Problèmes avec les transformations normales inverses basées sur les rangs [MS-312]

11:15 Keyue DING, *Queen's University*, Keramat NOURIJELYANI & Philip TWUMASI-ANKRAH, *NCIC CTG, Queen's University*  

Prognostic and Predictive Biomarker Model Building for Time to Event Outcomes • Construction de pronostics et de modèles prédictifs basés sur biomarqueurs pour données de temps d'évènements [MS-313]

11:30 Jiguo CAO, *Simon Fraser University*, Guangzhe FAN, *University of Waterloo*  

Kernel-Induced Random Forest Classification for functional Gene Analysis • Classification aléatoire de forêts induite par noyaux en analyse de gènes fonctionnelle [MS-314]

11:45 Sam STEWART & Mohamed ABDOLELL, *Dalhousie University*, Michael LeBLANC, *Fred Hutchinson Cancer Research Center*  

Customizing the rpart library for multivariate gaussian outcomes: the longRPart library • Customiser la librairie rpart pour réponses gaussiennes multivariées : la librairie longRPart [MS-315]

10:30 – 12:00 Session 12I

CHE 250

Statistical Inference



Inférence statistique

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Min GAO, *Cardiac Services BC*

10:30 Jeffrey PICKA, *University of New Brunswick*  

Random Set Modeling of Forest Fires • Modélisation par ensembles aléatoires de feux de forêts [MS-316]

10:45 Michael STEPHENS & Richard LOCKHART, *Simon Fraser University*, Federico O'REILLY, *UNAM*  

A Surprising Comparison of Tests of Fit • Une comparaison surprenante de tests d'adéquation [MS-317]

11:00 Chul PARK, *Carleton University*, Pranab SEN, *University of North Carolina*  

Testing Marginal Homogeneity Against Uniform Stochastic Ordering in a Contingency Table • Tester l'homogénéité marginale contre un classement stochastique uniforme dans une table de contingence [MS-318]

11:15 Juxin LIU, *University of Saskatchewan*, Paul GUSTAFSON, *University of British Columbia*  

On Average Predictive Comparisons and Interactions • Comparaison et interactions prédictives moyennes [MS-319]

11:30 Wanhua SU & Peng ZHANG, *University of Alberta*  

Statistical Inference on Recall, Precision and Average Precision • Inférence statistique sur le rappel, la précision et la précision moyenne [MS-320]

11:45 Kai YU, *National Institutes of Health*  

On the Moments and the Moment-generating Function of the Maximum of a Negatively Drifting Random Walk • Moments et fonction génératrice des moments du maximum d'une marche aléatoire à dérive négative [MS-321]

10:30 – 12:00 Session 12J

SCR 100

Mixture Models



Modèles de mélanges

Contributed Paper Session • Séance de communications libres

Chair • Président: Patrick BROWN, *Cancer Care Ontario*

10:30 Shaheena BASHIR, *The Aga Khan University Karachi, Pakistan*, Shagufta SULTAN, *University of Ottawa*, E CARTER, *University of Guelph*  

Robust Mixture Regression Models • Mélanges de modèles de régression robustes [MS-322]

10:45 Guillaume ÉVIN, *Université du Québec - Institut national de la recherche scientifique - Eau, Terre et Environnement (INRS-ETE)*, James MERLEAU & LUC PERREAULT, *Institut de recherche d'Hydro-Québec (IREQ)*  



Mixtures of Normal, Gamma, and Gumbel Distributions • Mélanges de lois normales, gamma et Gumbel [MS-323]

11:00 Jeffrey ANDREWS & Paul McNICHOLAS, *University of Guelph*  



Extending Mixtures of t-Factors for Model-Based Clustering • Étendue des mélanges de t-facteurs pour la classification à base de modèle [MS-324]

11:15 Abbas KHALILI, *University of British Columbia*  

Feature Selection in Mixture-of-Experts Models • Sélection de caractéristiques en modèles de mélanges d'experts [MS-325]

11:30 Peiming WANG, *Auckland University of Technology*, Joseph ALBA, *Nanyang Technological University*  

Analysis of Bivariate FDI Panel Count Data: a Markov Bivariate Zero-inflated Negative Binomial Regression Approach • Analyse de données de dénombrement bivariées d'IÉD : une approche bivariée markovienne par régression négative binomiale à surreprésentation de zéros [MS-326]

11:45 Gary SNEDDON, *Memorial University*, M. Tariqul HASAN & Renjun MA, *University of New Brunswick*  

Modelling Longitudinal Count Data With Excess Zeros Using a Compound Poisson Random Effect • Modélisation de données de dénombrement longitudinales avec surreprésentation de zéros au moyen d'un effet aléatoire de Poisson composé [MS-327]

12:00 – 13:30 Lunch Break • Dîner

13:30 – 14:30 Session 13

IRC 6

Address of the winner of the Centre de Recherches Mathématiques/Statistical Society of Canada Prize

Allocution du récipiendaire du prix Centre de Recherches Mathématiques/Société Statistique du Canada

Special Session • Conférence spéciale

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

13:30 Hugh CHIPMAN, *Acadia University*  

Statistical Learning With Trees • Apprentissage statistique avec arbres [MS-328]

13:30 – 15:00 Session 14A

IRC 1



Statistics in Sports

La statistique dans le monde des sports



Invited Paper Session • Présentations sur invitation

Chair • Président: Shane REESE, *Brigham Young University*

Organizers • Responsables: Paramijit GILL, *University of British Columbia, Okanagan*

13:30 Hal STERN, *University of California, Irvine*  


Small Sample Statistics in Baseball - the Batter-pitcher Matchup • Statistique en présence de petits échantillons au baseball - la rencontre batteur-lanceur [MS-329]

13:55 Tim SWARTZ, *Simon Fraser University*  

A Graduate Course in Statistic in Sport • Un cours de cycle supérieur de statistique dans le sport [MS-330]

14:20 Paramjit GILL, *UBC Okanagan*  

NHL Scoring and Fighting: Before and After 2005 • Compter et cogner dans la LNH : avant et après 2005 [MS-331]

14:45 Shane REESE, *Brigham Young University* 

Discussant • Modérateur [MS-332]

13:30 – 15:00 Session 14B

IRC 5

Statistical Inference in Studies with Large p and Small n

Inférence statistique dans les études à grand p et petit n

Invited Paper Session • Présentations sur invitation



Organizer and Chair • Responsable et présidente: Zeny FENG, *Guelph University*

13:30 Yuedong WANG, *University of California - Santa Barbara*, Tiejun TONG, *University of Colorado*  

Optimal Shrinkage Estimation of Variances With Applications to Microarray Data Analysis • Estimation optimale par rétrécissement de variances avec applications à l'analyse de données de micropuces [MS-333]

13:52 Yinglei LAI, *The George Washington University*  

Improving Differential Expression Analysis With the Consideration of Genome-wide Co-expression Information • Améliorer l'analyse de l'expression différentielle en considérant l'information de co-expression sur tout le génome [MS-334]

14:15 Tiejun TONG, *University of Colorado at Boulder*, Liang CHEN, *University of Southern California*, Hongyu ZHAO, *Yale University*  

Shrinkage Estimation of Means for High-dimensional Data • Estimation par rétrécissement de moyennes pour données de haute dimension [MS-335]

14:37 Paul McNICHOLAS, *University of Guelph*  

Clustering Strategies for Large p, Small n Problems • Stratégies de classification pour problèmes à grand p et petit n [MS-336]

13:30 – 15:00 Session 14C


MSL 102

Recent Developments in Order-Restricted Inference



Évolution récente de l'inférence d'ordre restreint

Invited Paper Session • Présentations sur invitation



Organizer and Chair • Responsable et président: Lang WU, *University of British Columbia*

13:30 Mervyn SILVAPULLE, *Monash University, Australia*  

Recent Advances in Constrained Statistical Inference in Parametric Models • Développements récents en inférence statistique sous contraintes pour modèles paramétriques [MS-337]

14:00 Hidetoshi SHIMODAIRA, *Tokyo Institute of Technology*  

Approximately Unbiased Tests for Cone Shaped Regions via Multiscale Bootstrap • Tests approximativement sans biais pour régions en forme de cône au moyen de bootstrap multi-échelles [MS-338]

14:30 Jianan PENG, *Acadia University*, Lang WU, *University of British Columbia*, Weizhen WANG, *Wright State University*  

Stepwise Confidence Intervals for Partial Monotone Dose-Response Studies • Intervalles de confiances pas-à-pas pour études dose-réponse partiellement monotones [MS-339]

13:30 – 15:00 Session 14D

IRC 3

Matrices and Statistics


Matrices et statistique

Invited Paper Session • Présentations sur invitation


Organizer and Chair • Responsable et président: George STYAN, *McGill University*

13:30 Simo PUNTANEN, *University of Tampere, Finland*, Stephen J. HASLETT, *Massey University, New Zealand* 

Equality of the BLUPs for New Observations Under two Linear Models • Égalité des MPLSB pour de nouvelles observations sous deux modèles linéaires [MS-340]

14:00 Goetz TRENKLER, *Dortmund University of Technology, Germany*, Oskar M. BAKSALARY, *Adam Mickiewicz University, Poland* 

Classical Rank Formulas From Projectors' Point of View • Formules classiques pour le rang du point de vue des projecteurs [MS-341]

14:30 Oskar M. BAKSALARY, *Adam Mickiewicz University, Poland*, Goetz TRENKLER, *Dortmund University of Technology, Germany* 

Core Inverse of Matrices • Inverse nucléaire de matrices [MS-342]

Abstracts • Résumés

Session 01A **Lundi 1 juin • Monday, June 1, 08:15 – 08:30** **IRC 2**
 Inaugural Session
 Séance inaugurale

[MS-1]

Welcoming Remarks

Remarques de bienvenue

Alejandro ADEM, *Director of Pacific Institute for the Mathematical Sciences*

[MS-2]

Announcements

Annonces

Román VIVEROS-AGUILERA, *McMaster University*, Nancy HECKMAN, *University of British Columbia*

Session 01B **Lundi 1 juin • Monday, June 1, 08:30 – 09:45** **IRC 2**
 SSC Presidential Invited Address
 Allocution de l'invité du président de la SSC

[MS-3]

Building the Pyramids

Construire les pyramides

Chris WILD, *University of Auckland*

This wide ranging talk will explore principles and strategies for expanding the penetration, relevance and effectiveness of statistics education programmes at all levels. We will travel through many dimensions including identity, market appeal and market positioning, buyers' markets and sellers' markets, communities of interest, allies and competitors, supply and demand, why the models that work well at the most elite universities endanger others, applied statistics and service led growth, quality improvement and working smarter not harder, and how technology really has changed everything but we haven't woken up to that yet.

Cet exposé explorera différents principes et stratégies pour améliorer la pénétration, la pertinence et l'efficacité des programmes d'éducation statistique à tous les niveaux. Nous allons voyager à travers plusieurs dimensions incluant l'identité, l'évaluation du marché, le positionnement, les marchés d'acheteurs et de vendeurs, les communautés d'intérêt, les alliés et compétiteurs, l'offre et la demande, pourquoi les modèles qui fonctionnent bien aux universités d'élite mettent les autres en danger, la statistique appliquée et la croissance basée sur les services, l'amélioration de la qualité et travailler plus intelligemment et non plus fort, et comment la technologie a réellement tout changé sans que nous en prenions encore réellement conscience.

Session 02A**Lundi 1 juin • Monday, June 1, 10:30 – 12:00****IRC 6**

Special Invited Session of the Survey Methods Section

Allocution de l'invité d'honneur du groupe des méthodes d'enquête

[MS-4]**A New Approach to Weighting and Inference for Samples Drawn from a Finite Population****Une nouvelle approche de pondération et d'inférence pour des échantillons tirés d'une population finie**Jean-François BEAUMONT, *Statistique Canada / Statistics Canada*

The validity of design-based inference is not dependent on model assumptions. However, it is well known that estimators derived through this theory may be inefficient when the design weights are weakly related to the variables of interest and have widely dispersed values. We propose estimators that have the potential of improving the efficiency of any standard estimator derived under the design-based theory. The new estimators are obtained by smoothing survey weights using an appropriate model. Our approach to inference requires the modelling of only one variable, the weight, and leads to a single set of smoothed weights in multipurpose surveys.

La validité de l'inférence fondée sur le plan de sondage ne dépend d'aucune hypothèse modéliste. Cependant, il est bien connu que les estimateurs dérivés sous cette théorie peuvent être inefficaces quand les poids de sondage sont faiblement corrélés aux variables d'intérêt et ont des valeurs fortement répandues. Nous proposons des estimateurs qui ont le potentiel d'améliorer l'efficacité de n'importe quel estimateur classique fondé sur le plan. Ces nouveaux estimateurs sont obtenus en lissant les poids d'enquête au moyen d'un modèle approprié. Notre approche d'inférence requiert la modélisation d'une seule variable, le poids, et résulte en un seul ensemble de poids lissés dans les enquêtes à objectifs multiples.

Session 02B**Lundi 1 juin • Monday, June 1, 10:30 – 12:00****IRC 1**

Evaluation of Medical Screening and Diagnostic Tests

Évaluation des tests de dépistage et de diagnostic médicaux

[MS-5]**Assessing the Performance of Diagnostic and Prognostic Probability Functions in Medicine****Évaluer la performance des fonctions de probabilité diagnostique et pronostique en médecine**James HANLEY, *McGill University*

This overview will revisit the origins of receiver-operating-characteristic analysis in psychophysics, describe some of the early applications in medical diagnosis, and highlight the substantial statistical developments over the last three decades. I will also comment on the growing use of "risk prediction" models for prognosis, on the differences between diagnostic and prognostic probabilities, and on the criteria used to assess the performance of these probability functions.

Cette revue va survoler les origines de l'analyse de fonction d'efficacité de l'observateur en psychophysique, décrire quelques-unes de ses premières applications en diagnostics médicaux, et souligner les développements statistiques substantiels des trois dernières décennies. Je vais aussi commenter l'utilisation croissante de modèles de prévision du risque pour pronostics, les différences entre probabilités diagnostiques et pronostiques, et les critères utilisés pour évaluer la performance de ces fonctions de probabilité.

[MS-6]**Biases and Pitfalls in Interpreting the Efficacy of Cancer Screening Technologies: The Epidemiologists' Perspective****Biais et pièges dans l'interprétation de l'efficacité des technologies de dépistage du cancer : la perspective des épidémiologistes**Eduardo FRANCO, *McGill University*

Interpreting the efficacy of cancer screening is affected by important study design and analysis issues. Some are easy to communicate, such as self-selection and length-biased sampling, lead time bias, and overdiagnosis bias. Others are more insidious and complex, such as verification bias, false sensitivity gains, and sticky diagnosis and slippery linkage biases. Overcoming these biases poses challenges that incur increased research costs, ethical concerns, and statistical methodology that is difficult to communicate to clinicians and policy-makers. Some of these biases have led to inappropriate claims by the technology industry that adjunct screening can minimize risk of malpractice litigation.

L'interprétation de l'efficacité du dépistage du cancer est affectée par d'importantes questions de planification et d'analyse d'enquêtes. Certaines sont faciles à communiquer, tels l'échantillonnage par autosélection et biaisé en longueur, le biais de délai de mise en oeuvre et le biais de surdiagnostic. D'autres sont plus insidieuses et complexes, comme le biais de vérification, les faux gains en sensibilité, et les biais de diagnostic absorbant et de couplage dissipant. Surmonter ces biais représente des défis entraînant une augmentation des coûts de la recherche, des considérations éthiques et une méthodologie statistique difficile à communiquer aux cliniciens et aux preneurs de décision. Certains de ces biais ont mené à l'affirmation inappropriée que le dépistage adjoint peut minimiser le risque de poursuites pour faute professionnelle.

[MS-7]**Statistical Issues With Cancer Screening Programs - Experience From Finland****Problèmes statistiques avec les programmes de dépistage du cancer : l'expérience de la Finlande**Timo HAKULINEN, *Finnish Cancer Registry*

Statistical issues with cancer screening programs - experience from Finland

Finland has three population-based screening programs – for cervical, breast and colorectal cancers – and a new large trial of prostate cancer screening. The cervical cancer program was launched without a study design, the breast cancer program targets alternate birth-year cohorts, and the colorectal cancer program is a prevention trial based on randomized individual invitations. Evaluation of effectiveness for these programs is at an early stage, and poses statistical challenges.

The challenges and solutions will be reviewed. The importance of a good design is emphasized for successful application of methods.

Problèmes statistiques avec les programmes de dépistage du cancer : l'expérience de la Finlande

La Finlande a trois programmes de dépistage en population générale - pour les cancers cervicaux, colorectaux et du sein - et un nouvel essai à grande échelle de dépistage du cancer de la prostate.

Le programme pour le cancer cervical a été lancé sans plan d'expérience, le programme pour le cancer du sein alterne entre différentes cohortes basées sur l'année de naissance, et le programme pour le cancer colorectal est un essai en prévention basé sur des invitations individuelles randomisées. L'évaluation de l'efficacité de ces programmes est à ses débuts et pose des défis statistiques.

Nous ferons la revue des défis et de leurs solutions, en insistant sur l'importance d'un bon plan d'expérience pour une bonne application des méthodes.

Session 02C**Lundi 1 juin • Monday, June 1, 10:30 – 12:00****IRC 4**

History of Statistics - Lifetime Modeling and Data

Histoire de la statistique - Modélisation et données de cycle de vie

[MS-8]**Cultural Imagery and Statistical Models of the Force of Mortality****Représentations culturelles et modèles statistiques de la force de mortalité**Elizabeth TURNER, *London School of Hygiene & Tropical Medicine*, James HANLEY, *McGill University*

We describe selected artistic and scientific depictions of the force of mortality, a concept that has long pre-occupied actuaries and statisticians. The "Bridge of Human Life" (Addison, 1711) provides a particularly vivid image. The Chances of Death (Pearson, 1897) mathematically "modernizes" the medieval conception of the relation between Death and Chance by decomposing the full mortality curve into five distributions along the age axis, the results of five "marksmen" aiming at the human mass crossing this bridge. We describe the imagery, critique Pearson's statistical model, and give the Bridge of Life a modern form, illustrating it via statistical animation.

Nous décrivons des représentations artistiques et scientifiques choisies de la force de mortalité, un concept ayant longtemps préoccupé les actuaires et statisticiens. Le «Bridge of Human Life» (Addison, 1711) en fournit une image particulièrement vive. «The Chances of Death» (Pearson, 1897) modernise mathématiquement la conception médiévale de la relation entre Mort et Chance en décomposant l'ensemble de la courbe de mortalité en cinq distributions le long de l'axe des âges, résultats de cinq «tireurs» visant la masse humaine traversant ce pont. Nous décrivons l'imagerie, critiquons le modèle statistique de Pearson, et donnons une forme moderne au Pont de la Vie, l'illustrant au moyen d'une animation statistique.

[MS-9]**Cauchy's Method of Linear Regression****La méthode de régression linéaire de Cauchy**Antoine de FALGUEROLLES, *Université de Toulouse (III)*

The method of least-squares for linear fits has been published by Legendre in 1805 and by Gauss in 1809. Circa 1835, Cauchy introduced another method of linear regression. Pareto (1897) and March (1898) consider both methods in papers presented at the Société de Statistique de Paris. In this presentation, Cauchy's method is introduced in the present-day framework for regression and the papers of Pareto and March are revisited. It turns out that the estimators of the coefficients in Cauchy's regression are also linear and unbiased.

La méthode des moindres carrés pour des ajustements linéaires a été publiée par Legendre en 1805 et par Gauss vers 1809. Autour de 1835, Cauchy proposa une autre méthode d'ajustement. Pareto (1897) et March (1898) considèrent les deux méthodes dans des communications faites à la Société de Statistique de Paris. Dans cet exposé, la méthode de Cauchy est présentée dans le cadre actuel de la régression et les communications de Pareto et March sont revues. Il apparaît que les estimateurs des coefficients de la régression selon Cauchy sont aussi linéaires et sans biais.

[MS-10]**Edmond Halley and His Life Table****Edmond Halley et sa table de mortalité**David BELLHOUSE, *University of Western Ontario*

In 1693 data on births and deaths from the City of Breslau were presented to the Royal Society. As clerk to the society, Edmond Halley took these data

En 1693 des données de naissances et de décès dans la cité de Breslau furent présentées à la Royal Society. Edmond Halley établit une table de mortalité à partir de ces données. Je vais

and computed a life table. I will examine Halley's method of table construction as well as his life annuity calculations based on the table. I will show that Halley made some numerical errors in his annuity calculations and fudged some of his data for convenience. I will also examine Halley's method of smoothing life table entries, a technique that received much technical attention from actuaries in the nineteenth century.

examiner la méthode de construction de la table de Halley et ses calculs de rentes viagères basées sur la table. Je vais montrer que Halley a commis certaines erreurs numériques dans ces calculs de rentes et a manipulé certaines données à des fins de commodité. Je vais aussi examiner la méthode de Halley pour lisser les entrées de la table de mortalité, une technique très étudiée par les actuaires du dix-neuvième siècle.

Session 02D

Lundi 1 juin • Monday, June 1, 10:30 – 12:00

IRC 5

Functional Data Analysis

Analyse de données fonctionnelles

[MS-11]

Derivatives of Sparsely Sampled Random Trajectories

Dérivées de trajectoires aléatoires à échantillonnage éparé

Hans-Georg MUELLER, *UC Davis*, Bitao LIU, *Affymetrix*

Derivatives of functions play an important role in assessing underlying dynamics. Estimating derivatives from sparse, irregular and noisy measurements, as typically encountered in longitudinal studies, poses challenges. It is demonstrated how these can be overcome under minimal assumptions if one has a sparsely measured sample of random functions. For the case of Gaussian processes, an application of derivative estimation is to identify a simple linear empirical differential equation that describes overall trends in the derivatives of the processes. We apply this approach to study the dynamics of on-line auctions.

Les dérivées d'une fonction jouent un rôle important dans l'étude de la dynamique sous-jacente à celle-ci. L'estimation de dérivées à partir de mesures éparées, irrégulières et comportant du bruit représente un défi important souvent rencontré en études longitudinales. Nous montrons comment il peut être surmonté avec un minimum d'hypothèses si l'on dispose d'un échantillon éparé de fonctions aléatoires. Dans le cas des processus gaussiens, une application de l'estimation des dérivées est l'identification d'une équation différentielle empirique linéaire simple décrivant la tendance générale des dérivées du processus. Nous appliquons cette approche à l'étude de la dynamique des enchères en ligne.

[MS-12]

Penalized Regression and Mixed Effects Models

Régression pénalisée et modèles à effets mixtes

Nancy HECKMAN, *University of British Columbia*, Richard LOCKHART, *Simon Fraser University*, Nielsen JASON, *Carleton University*

Linear mixed effects models for the analysis of longitudinal data are convenient for modelling within-individual correlation. Using spline functions in a mixed effects model allows for flexible modelling of response functions. The use of spline functions in longitudinal data analysis is even more appealing because of a computational connection between linear mixed effects modelling and spline smoothing. However, resulting estimates of the underlying

Les modèles linéaires à effets mixtes en analyse de données longitudinales sont commodes pour modéliser la corrélation intra-individu. L'utilisation de fonctions splines dans un modèle à effets mixtes permet une modélisation flexible des fonctions réponse. L'utilisation de fonctions splines en modélisation de données longitudinales est encore plus attrayante grâce à un lien computationnel entre la modélisation linéaire à effets mixtes et le lissage par splines. Cependant, les estimés résultants pour la moyenne de la population

population mean, although fast to compute, might not track the data well and associated standard errors might be unreasonably large. We discuss these shortcomings and suggest some easy-to-compute methods to eliminate them.

sous-jacente, bien que rapides à calculer, peuvent ne pas bien suivre les données et les erreurs standard associées peuvent être très grandes. Nous discutons de ces lacunes et suggérons quelques méthodes faciles à calculer pour les combler.

[MS-13]

Statistical Inference for Dynamic Models with the Generalized Profiling Method

Inférence statistique pour modèles dynamiques avec la méthode du profilage généralisé

Jiguo CAO, & David CAMPBELL, *Simon Fraser University*, Giles HOOKER, *Cornell University*, Jianhua HUANG, *A&M University*, Jim RAMSAY, *McGill University*

Differential equations (DEs) model the rate of change of a process. One central and difficult problem is how to estimate DE parameters from noisy data. We have developed the generalized profiling method to solve this problem. DE solutions are approximated by nonparametric functions, which are estimated by penalized smoothing with DE-defined penalty. The computation is much faster than other methods. A modified delta method is proposed to estimate variances of DE parameters, which include all the uncertainty of the smoothing process. I will demonstrate our method with estimating a predator-prey dynamic model and gene regulatory networks.

Les équations différentielles (ÉD) modélisent le taux de changement d'un processus. Un problème difficile est l'estimation de paramètres d'ÉD à partir de données avec bruit. Nous avons développé la méthode du profilage généralisé pour résoudre ce problème. Les solutions des ÉD sont approximées par des fonctions non paramétriques qui sont estimées par lissage pénalisé avec pénalité définie par les ÉD. Les calculs sont beaucoup plus rapides que sous les autres méthodes. Nous proposons une méthode delta modifiée pour l'estimation de la variance des paramètres d'ÉD, qui comprend toute l'incertitude liée au processus de lissage. Notre méthode est appliquée à l'estimation d'un modèle dynamique prédateur-proie et de réseaux de gènes régulateurs.

Session 02E

Lundi 1 juin • Monday, June 1, 10:30 – 12:00

IRC 3

Recent Development in Analysis of High Dimensional Data

Évolution récente de l'analyse des données de grande dimension

[MS-14]

Penalized Methods for Bi-level Variable Selection

Méthodes pénalisées pour sélection de variables à deux niveaux

Jian HUANG & Patrick BREHENY, *University of Iowa*

In many applications, covariates possess a grouping structure that can be incorporated into the analysis to select important groups and important members of those groups. This work focuses on the incorporation of grouping structure into penalized regression. We investigate the group lasso, group bridge and a novel method, group MCP, introducing a framework and conducting simulation studies that shed light on the behavior of these methods. To fit these models, we use a locally coordinate descent algorithm which are fast and stable even

Dans plusieurs applications, les covariables ont une structure de groupe pouvant être incorporée à l'analyse de façon à sélectionner les groupes importants et les membres importants de ces groupes. Ce travail porte sur l'incorporation de structures de groupe à la régression pénalisée. Nous étudions la régression lasso en groupe, les ponts en groupe et une nouvelle méthode, le MCP en groupe. Nous effectuons des études de simulation éclairant le comportement de ces méthodes. Afin d'ajuster ces modèles, nous utilisons un algorithme de descente locale rapide et stable même lorsque p est plus grand que n . Ces méthodes sont appliquées à une

when p is larger than n . These methods are applied to a genetic association study.

étude d'association génétique.

[MS-15]

Dantzig Selector for Censored Linear Regression Models With High Dimensional Covariates **Le Sélectionneur de Dantzig pour modèles de régression linéaires censurés avec covariables de dimension élevée**

Yi LI, *Harvard*

This project stems from a cancer study that is to detect predictive genes out of thousands of candidates for myeloma patients' phenotype, namely, overall survival. We propose a new weighted Dantzig variable selector for linear regression models when the response variable is subject to right censoring. We establish consistency in model selection and the oracle property of the estimation. Extensive numerical studies show that the Dantzig selector is a strong competitor with the other competing methods. We apply the new method to the myeloma study and identify important predictive genes for patients' overall survival.

Ce projet découle d'une étude sur le cancer portant sur la détection de gènes prédictifs à partir de milliers de candidats pour le phénotype de patients de myélome, c'est-à-dire la survie globale. Nous proposons un nouveau sélectionneur de variables de Dantzig pondéré pour modèles de régression linéaire lorsque la variable réponse est sujette à une censure à droite. Nous établissons la cohérence en sélection de modèle et la propriété oracle de l'estimation. Des études numériques étendues montrent que le sélectionneur de Dantzig est hautement compétitif par rapport aux méthodes concurrentes. Nous appliquons la nouvelle méthode à l'étude du myélome et identifions les gènes prédictifs importants pour la survie globale des patients.

[MS-16]

Sparse Penalization with Censoring Constraints for Estimating AFT Models with Applications to Microarray Data Analysis **Pénalisation éparse avec contraintes de censure en estimation de modèles à temps de défaillance accéléré avec application à l'analyse de données de micropuces**

J. Sunil RAO, *Case Western Reserve University*, Simin HU, *Eli Lilly Corporation*

We use a weighted least squares (WLS) method with sparse penalization and censoring constraints to estimate accelerated failure time models with high-dimensional covariates. Our interest is microarray analysis and identifying genes that are predictive of survival times. In this approach, we add a LASSO penalty to a WLS objective function to produce sparse coefficients, and use right censored observations as censoring constraints when optimizing the objective function. We also include a ridge penalty to induce a grouping effect amongst related variables. Studies demonstrate that adding censoring constraints improves model estimation and variable selection especially in the situation of heavy censoring.

Nous utilisons une méthode des moindres carrés pondérés (MCP) avec pénalité éparse et contraintes de censure pour estimer des modèles à défaillance accélérée avec variables de dimension élevée. Notre intérêt est l'analyse de micropuces et l'identification de gènes prédictifs de temps de survie élevés. Dans cette approche, nous ajoutons une pénalité LASSO à une fonction objectif MCP pour produire les coefficients épars, et utilisons les observations censurées à droite comme contrainte de censure lors de l'optimisation de la fonction objectif. Nous incluons également une pénalité ridge pour induire un effet de regroupement des variables similaires. Des études démontrent que l'ajout des contraintes de censure améliore l'estimation du modèle et la sélection de variables, particulièrement en présence de forte censure.

Session 02F**Lundi 1 juin • Monday, June 1, 10:30 – 12:00****MSL 102**

Statistical Methods for Public Health Research

Méthodes statistiques pour la recherche en santé publique

[MS-17]**Method of Moments Estimation in Logistic regression Models with Exposure Biomarkers****Estimation par la méthode des moments en modèles de régression logistique avec biomarqueurs d'exposition**Taraneh ABARIN & Liqun WANG, *University of Manitoba*

Motivated by problems in nutritional and physical activity epidemiology, an extended method of moments estimation is applied to a logistic regression measurement error model. Our new approach is an extension of the method of Abarin and Wang (2008) and it allows us to combine a random effect component with the instrumental variable method. Simulation studies based on a Biomarker data set are presented to evaluate and contrast the method and more common procedures such as regression calibration and conditional score.

Nous utilisons une méthode des moments étendue pour l'ajustement d'un modèle de régression logistique d'erreurs de mesure. Cette approche est motivée par des problèmes en épidémiologie nutritionnelle et de l'activité physique. Notre approche est une extension de la méthode d'Abarin et Wang (2008) et elle nous permet de combiner une composante à effets aléatoires à la méthode de la variable instrumentale. Des études de simulations basées sur un jeu de données de biomarqueurs sont présentées afin d'évaluer la méthode et de la comparer à des procédures plus usuelles tels la calibration de régression et les scores conditionnels.

[MS-18]**The Impact of Dependent Durations of Episodes in Recurrent Event Analysis****L'impact de durées dépendantes des épisodes en analyse d'événements récurrents**Kyle RAYMOND & Richard COOK, *University of Waterloo*

Clinical trials often aim to assess the effect of treatments on preventing recurrent outbreaks of symptoms such as exacerbations in patients with asthma or chronic bronchitis. Standard marginal estimates of rate and mean functions or associated regression coefficients require the observation process to yield consistent estimates. This is not satisfied when an association exists between the duration and risk of the exacerbations. We discuss the bias arising from marginal analysis of processes exhibiting the above association and demonstrate it through simulation. Data from a clinical trial of chronic bronchitis are analyzed for illustration.

Les essais cliniques visent souvent à évaluer la capacité de traitement à prévenir l'apparition récurrente de symptômes tels les exacerbations chez les patients souffrant d'asthme ou de bronchite chronique. Les estimateurs standards marginaux de fonctions de la moyenne et de la fréquence, tout comme les coefficients de régression associés, nécessitent que le processus d'observation soit complètement indépendant du processus des événements pour assurer la cohérence des estimateurs. Cette condition n'est pas satisfaite lorsqu'il y a une association entre la durée et le risque des exacerbations. Nous discutons du biais résultant de l'analyse marginale de processus contenant une telle association et le démontrons au moyen d'une étude de simulation. Des données d'essai clinique pour bronchite chronique sont analysées à fins illustratives.

[MS-19]**A Local Likelihood Method in Disease Mapping****Une méthode de vraisemblance locale en cartographie des maladies**

Chun-Po Steve FAN, *University of Toronto*, Patrick BROWN, *Cancer Care Ontario*, James STAFFORD, *University of Toronto*

Disease incidence data are often area-censored with only cases at the regional level being available. The regions may vary in size. While estimating the risk of rare diseases often requires data collected over a long period of time, the boundaries of these regions may not be the same throughout the entire collection period. We extend the EMS algorithm proposed by Fan and Stafford (2008) to combine multiple disease maps with the longitudinal census and estimate a spatially smooth risk surface. We then apply the proposed method to analyze newly diagnosed lupus cases in Greater Toronto Area between 1966 and 2005.

Les données d'incidence de maladies sont souvent censurées par zones et seuls les cas au niveau régional sont disponibles. Les régions peuvent varier en taille. Alors que l'estimation du risque de maladies rares nécessite souvent des données recueillies sur une longue période de temps, les frontières des régions peuvent ne pas être les mêmes durant toute cette période. Nous étendons l'algorithme EMS proposé par Fan et Stafford (2008) pour combiner plusieurs cartes de maladies à des recensements longitudinaux et estimons une surface de risque spatialement lisse. Nous appliquons alors la méthode proposée à l'analyse de cas de lupus nouvellement diagnostiqués dans la région du Toronto métropolitain entre 1966 et 2005.

[MS-20]**Efficient Monte Carlo Random Variates Generation Through Discretization****Génération de variables aléatoires efficace par Monte-Carlo par discrétisation**

Chel Hee LEE & Liqun WANG, *Department of Statistics, University of Manitoba*

We propose an efficient Monte Carlo method for random variates generation from high dimensional distributions of complex structures. This method is based on the discretization of the sample space and efficient numerical inversion of a multivariate cumulative distribution function. It is non-iterative, easy to implement and overcomes many typical drawbacks of the Markov chain Monte Carlo (MCMC) methods. This method improves the discretization-based algorithm of Fu and Wang (2002). Numerical examples will be presented to illustrate the new method and algorithm.

Nous proposons une méthode de Monte-Carlo efficace pour générer des variables aléatoires de distributions de dimension élevée et de structure complexe. Cette méthode est basée sur la discrétisation de l'espace échantillonnal et l'inversion numérique efficace d'une fonction de répartition multivariée. Elle est non- itérative, facile à mettre en oeuvre et surmonte plusieurs désavantages typiques des méthodes MCMC. Cette méthode améliore l'algorithme à base de discrétisation de Fu et Wang (2002). Des exemples numériques seront présentés afin d'illustrer le nouvel algorithme.

[MS-21]**Connecting Relative Survival with Multi-State Models****Connecter la survie relative aux modèles multi-états**

Ella HUSZTI & Michal ABRAHAMOWICZ, *McGill University*, Ahmadou ALIOUM, *ISPED-Université Victor Segalen Bordeaux 2, France*, Catherine QUANTIN, *Dijon University Hospital, France*

We propose an extension of the traditional Markov model to incorporate Relative Survival (RS) methods. We redefine the transition intensity leading to death in the Markov model through an additive RS model, by splitting the intensity into an estimable

Nous proposons une étendue du modèle de Markov traditionnel pour incorporer des méthodes de survie relative (SR). Nous redéfinissons l'intensité de transition menant à la mort dans le modèle de Markov au moyen d'un modèle SR additif, en séparant l'intensité en un risque excessif dû à la maladie

excess hazard due to the disease of interest and an 'offset': the expected hazard of all-cause mortality obtained from life-tables. This "offset" modifies the MLE of intensities of transitions to death but also affects the estimated intensities of transitions to other states and the estimated effects of prognostic factors. We evaluate the new Markov Relative Survival model through simulations.

d'intérêt et un «décalage» : le risque espéré pour la mortalité due à toutes les causes obtenu de tables de mortalité. Ce «décalage» modifie l'EMV des intensités de transition vers la mort, mais affecte également l'estimation des intensités de transition vers d'autres états et des effets de facteurs pronostics. Nous évaluons le nouveau modèle de survie relative de Markov au moyen de simulations.

Session 02G

Lundi 1 juin • Monday, June 1, 10:30 – 12:00

CHE 250

Probability and Finance Probabilités et finance

[MS-22]

On Two Monotonic Properties of the Kendall Tau Coefficient of Order Statistics

Deux propriétés de monotonie du coefficient tau de Kendall de statistiques d'ordre

Yung-Pin CHEN, *Lewis and Clark College*

We prove two monotonic properties of the Kendall tau coefficient of order statistics of a random sample from a continuous distribution. We first show that the Kendall tau coefficient of the j^{th} and k^{th} order statistics increases as the two indexes j and k concurrently shift toward the middle. The second monotonic property involves the sample size: As the sample size increases, the Kendall tau coefficient of the j^{th} and k^{th} order statistics increases as well.

Nous prouvons deux propriétés de monotonie du coefficient tau de Kendall des statistiques d'ordre d'un échantillon aléatoire de distribution continue. Nous montrons d'abord que le coefficient tau de Kendall de la j^{e} et k^{e} statistiques d'ordre augmente alors que les deux indices j et k se déplacent vers le milieu. La seconde propriété de monotonie implique la taille échantillonnale : alors que la taille augmente, le coefficient tau de Kendall de la j^{e} et k^{e} statistiques d'ordre augmente aussi.

[MS-23]

On the Complexity of d-dimensional Delaunay Tessellations

De la complexité des mosaïques de Delaunay d-dimensionnelles

Leonard HEARNE, *University of Missouri at Columbia*

Given a set of n points in a d -dimensional space, a Delaunay tessellation can be constructed. If the generating points are representative random instances from a single generating density function then the function can be estimated using the tessellation. The accuracy of the estimate is a function of the cardinality of the set of tiles in the tessellation. In 1980 V. Klee established bounds on the cardinality of tiles in the tessellation, given n and d . By applying constructive geometric methods, tighter bounds have been established.

Une mosaïque de Delaunay peut être construite à partir d'un ensemble de n points dans un espace d dimensionnel. Si les points générateurs sont des instances aléatoires d'une unique fonction de densité génératrice, alors cette fonction peut être estimée à l'aide de la mosaïque. La précision de l'estimation est fonction de la cardinalité de l'ensemble des tuiles dans la mosaïque. En 1980 V. Klee a établi des bornes pour la cardinalité des tuiles dans la mosaïque étant donnés n et d . Des bornes plus strictes sont obtenues en appliquant des méthodes constructives géométriques.

[MS-24]**A Simple Discretization Scheme For Nonnegative Diffusion Processes, With Applications in Financial Mathematics****Un schéma de discrétisation simple pour des processus de diffusion non-négatifs, avec applications en finance mathématique**Chantal LABBÉ & BRUNO RÉMILLARD, *HEC Montréal*, Jean-François RENAUD, *University of Waterloo*

The Cox-Ingersoll-Ross (CIR) process is broadly used in financial engineering, mostly since it assumes nonnegative values. Applying a standard Euler discretization scheme for simulations is problematic, as nothing prevents the simulated values to be negative. Several fixes are found in the literature, most involving the normal distribution. We propose a scheme producing nonnegative discrete approximations for nonnegative diffusions (e.g. CIR), in which nonnegativity is secured by sampling from a nonnegative distribution. We use the martingale problem to establish convergence of the approximating processes to the target diffusion. The proposed scheme is suitable for several applications in pricing derivative securities.

L'usage du processus de Cox-Ingersoll-Ross (CIR) est répandu en ingénierie financière, principalement parce qu'il est à valeurs non négatives. L'utilisation d'un schéma de discrétisation standard d'Euler pour le simuler est problématique, puisque rien n'empêche alors les valeurs simulées d'être négatives. Nous proposons un schéma produisant, pour des diffusions non négatives (telles le CIR), des processus d'approximation discrets non-négatifs, la non-négativité de ces derniers résultant de l'emploi d'une distribution à valeurs non négatives, plutôt que la loi normale sur laquelle repose la plupart des schémas présentés dans la littérature pour pallier aux problèmes de non-négativité. Nous utilisons le problème de martingale pour établir la convergence des processus d'approximation vers la diffusion visée. Le schéma proposé est approprié pour la tarification de plusieurs produits dérivés.

[MS-25]**Extremes from Meta Distributions and the Shape of Sample Clouds****Extrêmes des méta-distributions et forme des nuages échantillonnaires**Natalia LYSENKO, *ETH Zurich*, GUUS BALKEMA, *University of Amsterdam*, Paul EMBRECHTS, *ETH Zurich*

Given a multivariate light-tailed density g with the copula of a heavy-tailed distribution, we show that under our assumptions the level sets $\{g > c\}$ may be scaled to converge to a limit set as $c \downarrow 0$. With the same scaling factors, this limit set is also the set onto which normalized random samples from g converge. We derive an analytic expression for the limit set and examine some of its properties. Limit sets are of interest in the study of extremes and in risk analysis since they determine how extremal observations in different directions relate.

Considérons une densité à ailes plates multivariée g avec la copule d'une distribution à ailes relevées. Sous certaines conditions, nous montrons que les ensembles de niveaux $\{g > c\}$ peuvent être normalisés pour converger vers un ensemble limite alors que $c \downarrow 0$. Avec les mêmes facteurs de normalisation, cet ensemble limite est aussi l'ensemble vers lequel des échantillons aléatoires normalisés tirés de g convergent. Nous obtenons une expression analytique pour l'ensemble limite et examinons certaines de ses propriétés. Les ensembles limites sont d'intérêt pour l'étude des extrêmes et en analyse de risque puisqu'ils déterminent comment sont liées les observations extrêmes dans différentes directions.

[MS-26]**Goal Achieving Probability of a Discrete-time Mean-variance Optimal Strategy****Probabilité d'atteinte d'objectif d'une stratégie moyenne-variance optimale à temps discret**Francois WATIER & Rene FERLAND, *Université du Québec à Montréal*

First, consider a discrete-time market model where an investor follows a multiperiod mean-variance efficient strategy. Then, let us introduce a stopping-time rule as follows: safely reallocate the total wealth accumulated into a bank account as soon as it hits the discounted value of the financial target (predetermined mean return), otherwise just follow the mean-variance optimal strategy until you reach the terminal time. We will show that the hitting-time probability is independent of the targeted goal and obtain explicit analytical lower bounds for this probability in the case of binomial- tree stock market models.

Considérons d'abord un modèle de marchés à temps discret où un investisseur suit une stratégie moyenne-variance efficace à multiples périodes. Nous introduisons ensuite une règle de temps d'arrêt consistant à réallouer la richesse totale accumulée dans un compte bancaire dès qu'elle atteint la valeur actualisée de la cible financière (rendement moyen prédéterminé) ou à suivre la stratégie moyenne-variance optimale jusqu'à ce que le temps terminal soit atteint. Nous allons montrer que la probabilité d'atteinte du temps terminal est indépendante du but ciblé et obtenir des bornes inférieures analytiques explicites pour cette probabilité dans le cas de modèles de marchés financiers à arbres binomiaux.

Session 02H**Lundi 1 juin • Monday, June 1, 10:30 – 12:00****SCR 100****Design and Analysis for Biomedical Studies****Planification et analyse d'études biomédicales****[MS-27]****Weighted Composite Endpoint Models in the Analysis of Cardiovascular Clinical Trial Data****Évaluation des modèles à points limite composites pondérés dans l'analyse des essais cliniques cardio-vasculaire**Jeffrey BAKAL, Cynthia WESTERHOUT, Robert WELSH & Paul ARMSTRONG, *Canadian VIGOUR Centre*

Due to the decreasing numbers of death events in cardiovascular clinical trials, combined fatal and non-fatal endpoints (e.g., heart failure, cardiogenic shock) are frequently used to assess the treatment efficacy. Typically both fatal and non-fatal events are given equal weight in the analysis. As an alternative we consider the implications on the creation of survival models based on assigning weights to the non-fatal outcomes determined by their relative severity. These weights were established using a survey tool in which clinical investigators allocated 20 points among 4 events. An example from a recent acute myocardial infarction trial is presented for discussion.

Avec la réduction des événements fatals en essais cliniques cardio-vasculaires, les investigateurs utilisent souvent les points terminaux composites pour l'évaluation de l'efficacité des traitements. Traditionnellement les points terminaux fatals et non fatals sont à poids égaux. Nous considérerons l'implication sur l'analyse des modèles de survie à poids différent pour les événements non fatals. Ces poids sont établis par les cliniciens qui allouent 20 points entre 4 événements. Un exemple d'un essai clinique récent d'infarctus aigu du myocarde est présenté pour discussion.

[MS-28]**Design Considerations for Recurrent Events Studies****Considérations pour la planification d'études d'évènements récurrents**

Elizabeth JUAREZ & Charmaine DEAN, *Simon Fraser University*, Robert BALSHAW, *Syreon Corporation*

In longitudinal recurrent event studies, sometimes partial information on the counting process, such as the number of events occurring in specific intervals, provides nearly the same precision for estimation of treatment effects as full information based on data from continuous observation. This talk will compare the analysis of such panel data with the analysis of data recorded as times of recurrences. The main purpose is inference concerning a treatment effect where the rate of the process is modelled as a proportional intensity function, with the baseline intensity being a spline. We discuss efficient panel designs, and conditions for optimal designs.

En études longitudinales avec évènements récurrents, de l'information partielle sur le processus de dénombrement, tel le nombre d'évènements survenant durant ces intervalles spécifiques, apporte presque la même précision pour estimation des effets des traitements que l'information complète basée sur des données d'observations continues. Cet exposé va comparer l'analyse de telles données de dénombrement avec l'analyse de données enregistrées comme temps de récurrences. Nous nous intéressons principalement à l'inférence relative à un effet de traitement lorsque le taux du processus est modélisé comme une fonction d'intensité proportionnelle, avec une spline comme intensité de base. Nous discutons de plans efficaces pour données de dénombrement, et de conditions pour l'optimalité des plans.

[MS-29]**Statistical Modelling of Neighbour Vaccine Effects in Aquaculture Clinical Trials****Modélisation statistique de l'effet voisin des vaccins en essais cliniques en aquaculture**

Elmabrok MASAPOUD & Henrik STRYHN, *University of Prince Edward Island*

Aquaculture trials sometimes hold multiple treatment groups in the same tanks. By neighbour treatment effects in an incomplete block design setting, we mean that treatments present in the same tank may affect each other in their performance. Multiple membership and non-linear mixed models were proposed and illustrated on simulated data and data from an ISAV clinical trial carried out at the Atlantic Veterinary College.

Results demonstrated that both models showed promise in capturing neighbour effects. Analysis of the ISAV trial data by both models did not provide any evidence of substantial neighbour effects.

En études en aquaculture, plusieurs groupes traitements se retrouvent parfois dans les mêmes bassins. Par effets de traitements voisins dans le cadre d'un plan en blocs incomplets, nous entendons que les traitements présents dans le même réservoir peuvent affecter leurs performances mutuelles. Nous proposons l'utilisation de modèles mixtes non linéaires et à appartenances multiples et les illustrons à l'aide de données simulées et de données d'un essai clinique du virus de l'anémie infectieuse du saumon mené à l'Atlantic Veterinary College.

Les résultats montrent que les deux modèles sont prometteurs pour le traitement d'effets voisins. L'analyse des données du VAIS par les deux modèles ne permet pas de conclure qu'il existe des effets voisins significatifs.

[MS-30]**Analysis of Neural Spike Trains: a Review****Analyse de trains d'impulsion neuronaux : une revue**

Reza RAMEZAN, Paul MARRIOTT & Shoja'eddin CHENOURI, *University of Waterloo*

Neural action potentials are often called spikes because they tend to be strongly localized in time. A sequence of such spikes from a single neuron is called a spike train. The modeling and analy-

Les potentiels d'action neuronaux sont souvent appelés impulsions, car ils sont souvent très localisés dans le temps. Une série de telles impulsions provenant d'un seul neurone est appelée train d'impulsions. La modélisation et

sis of spike train data is of interest in mathematics, statistics, neuroscience and biology. One method is to model the intensity function of the corresponding inhomogeneous point process based upon the spike train data. Two well-known approaches in intensity function approximation are the spline-based and multi-resolution-based methods. This talk reviews these methods of spike train analysis.

Keywords: BARS, density estimation, intensity function, multi-resolution analysis, spike train data

l'analyse de données de trains d'impulsions sont d'intérêt en mathématiques, statistiques, neuroscience et biologie. Une méthode est de modéliser la fonction d'intensité du processus ponctuel non homogène correspondant basé sur les données de trains d'impulsion. Deux approches bien connues en approximation de fonctions d'intensité sont les méthodes de splines et les méthodes à résolutions multiples. Cet exposé fait la revue de ces méthodes d'analyse de trains d'impulsions.

Mots-clés : BARS, estimation de densité, fonction d'intensité, analyse à résolutions multiples, données de trains d'impulsions

[MS-31]

Statistical Inference for Clinical Trials with Protocol Amendments

Inférence statistique pour essais cliniques avec amendements au protocole

Lan-Yan YANG, *National Cheng Kung University, Tainan, Taiwan*, Shein-Chung CHOW, *Duke University, Yunchan CHI, National Cheng-Kung University, Tainan, Taiwan*

It is not uncommon to modify trial procedure and/or statistical methods of on-going clinical trials through protocol amendments. A major modification could result in a shift in target patient population. In addition, frequent and significant modifications could lead to a totally different study, which is unable to address the medical questions that the original study is intended to answer. Chow and Shao (2005) proposed a covariate-adjusted model with continuous study endpoint. In this research, following similar ideas of Chow and Shao (2005), statistical inference and sample size adjustment based on binary study endpoint for trials with protocol amendments are derived.

Il n'est pas rare que la procédure d'essai ou les méthodes statistiques d'un essai clinique en cours soient modifiées par amendement de protocole. Une modification majeure pourrait entraîner un changement dans la population de patients cible. De plus, des modifications fréquentes et importantes pourraient mener à une étude complètement différente, incapable de répondre aux questions médicales auxquelles devait répondre l'étude originale. Chow et Shao (2005) ont proposé un modèle ajusté par covariables avec point terminal continu. Dans cette recherche, suivant des idées similaires à Chow et Shao (2005), nous obtenons une inférence statistique et des ajustements de taille échantillonnale basés sur un point terminal binaire pour essais avec amendements de protocole.

[MS-32]

Power Calculation for Cluster Randomized Trial of Comparing Trends of Incidence Over Time

Calcul de la puissance pour la comparaison de tendances dans l'incidence en essais randomisés par grappes

Qiaohao ZHU, *University of Alberta*

This research deals with statistical issues of power calculation for a cluster randomized trial to compare trends of incidence over time between the control and intervention groups, for a binary outcome measured from individuals within a cluster. We used a logistic regression model with random intercepts and different time slopes, aiming at detecting the difference of the slopes. We constructed a data set based on specified incidence rates and sample size, and fitted the postulated model to the data. Following Stroup (2002), we obtained

Cette recherche traite du calcul de la puissance d'une étude randomisée par grappes pour la comparaison de tendances d'incidence entre les groupes contrôle et intervention pour une réponse binaire mesurée chez les individus dans chaque grappe. Nous utilisons un modèle de régression logistique avec ordonnées à l'origine aléatoires et pentes différentes visant à détecter la différence entre les pentes. Nous construisons un jeu de données basé sur les taux d'incidence et la taille échantillonnale spécifiés, et ajustons le modèle proposé aux données. Nous obtenons la valeur critique et le paramètre de décentralité et déterminons la puissance à partir de

the critical value and the non-central parameter and determined the power from the non-central F-distribution.

la distribution F non centrée en suivant Stroup (2002).

Session 02I

Lundi 1 juin • Monday, June 1, 10:30 – 12:00

CHE 150

Methodology for Time-to-Event Data

Méthodologie pour données de temps d'événements

[MS-33]

Regression and Length-biased Sampling in Survival Analysis

Régression et biais de longueur en analyse de survie

Pierre-Jérôme BERGERON, *University of Ottawa*

In standard regression, though one samples from the joint distribution of the variable of interest and covariates, the analysis is carried out conditionally because the covariate distribution is considered ancillary to the parameters of interest. Length-biased sampling of the response variable results in biased sampling of the covariates. Their marginal distribution holds information about the parameters, thus one should adapt the usual methods of analysis to account for it. We present an adjusted (joint) likelihood approach for length-biased left-truncated right-censored data that yields more efficient estimates than a conditional approach which ignores the information in the covariate distribution.

Dans la plupart des analyses de régression, bien que l'échantillonnage se fasse à partir de la loi jointe de la variable d'intérêt et des covariables, l'analyse est effectuée en conditionnant sur les valeurs des covariables, parce que la loi marginale des covariables ne contient aucune information sur les paramètres étudiés lorsque l'échantillonnage se fait sans biais. Lorsqu'il y a échantillonnage avec biais de longueur pour la variable d'intérêt, les covariables souffrent également d'un biais et leur distribution dépend des paramètres de régression. Il faut en tenir compte dans l'analyse, en adaptant les méthodes existantes. En utilisant des méthodes basées sur la vraisemblance pour des données tronquées à gauche et censurées à droite, on démontre que l'on peut obtenir des estimateurs de moindre variance par l'approche jointe (tenant compte des covariables) en comparaison à l'approche conditionnelle.

[MS-34]

A Subdistribution Hazards Model for Clustered Data

Modèle de défaillance à sous-distributions pour données en grappes

Stephanie DIXON & Gerarda DARLINGTON, *University of Guelph*

While extensive work has been published in the area of competing risks, there is a notable gap in the literature in terms of the subdistribution hazards (SDHs) model for clustered subjects. For example, when multiple diseases aggregate within families, it may be of interest to follow a set of families over time to investigate their ages of onset for these multiple endpoints. A SDH model is proposed to account for multiple endpoints and intra-family correlation. This model is illustrated using data from a prospective cohort study of hereditary breast and ovarian cancer.

Bien que plusieurs travaux ont été publiés dans le domaine des risques concurrents, il y a un manque notable dans la littérature concernant les modèles de défaillance à sous-distributions (DSD) pour sujets en grappes. Par exemple, lorsque plusieurs maladies s'agrègent parmi des familles, il peut être intéressant de suivre un ensemble de familles à travers le temps pour étudier l'âge d'apparition de ces multiples maladies. Un modèle DSD est proposé permettant de tenir compte de ces multiples points terminaux et de la corrélation intra-famille. Ce modèle est illustré avec des données d'une étude cohorte prospective des cancers héréditaires du sein et des ovaires.

[MS-35]**Additive Hazards Regression Model for Competing Risk Events****Modèle de régression à risques additifs pour événements à risque concurrents**Hyun LIM, *University of Saskatchewan*, Xu ZHANG, *Georgia State University*

Competing risks arise commonly in the analysis of cancer studies. Adopting widely used Cox models to competing risk data causes the problem of internal inconsistency. Additive models for either the hazard rates or the cumulative incidence functions are more natural. We advocate the cause-specific hazards with the additive hazards model that includes both constant and time-varying covariate effects. We illustrate that the covariate effect on all-cause failure can be estimated by sum of the effects on all competing risks and that interpretation of additive models for the hazard rate with a real dataset.

Les risques concurrents se retrouvent fréquemment en analyse d'études sur le cancer. Adopter les modèles de Cox fréquemment utilisés pour modéliser des données à risques concurrents peut mener à des problèmes d'incohérence interne. Les modèles additifs pour les taux de défaillance ou les fonctions d'incidence cumulative sont plus naturels. Nous recommandons des défaillances par cause avec un modèle de défaillances additives incluant des covariables à effets constant et variant avec le temps. Nous montrons que l'effet des covariables sur la mortalité totale peut être estimé par la somme des effets sur les risques concurrents et illustrons l'interprétation de modèles additifs pour les défaillances avec un jeu de données réelles.

[MS-36]**Estimation of the Prediction Error for Survival Models****Estimation de l'erreur de prévision pour modèles de survie**Yan YUAN, *Alberta Cancer Board*, Jerald LAWLESS, *University of Waterloo*

Statistical models are often used for the prediction of future random variables. Their prediction accuracy is typically quantified using loss functions. We examine three estimators of prediction error for survival models. They are the model based, apparent loss and cross-validation loss based estimators. We use the inverse probability censoring weighting method to accommodate censoring and handle arbitrary loss functions. Confidence interval procedures for these estimators are developed. The performance of the estimators are compared under model misspecification and variable selection. Simulation results indicate that cross-validation produces reliable point estimates and confidence intervals, whereas model-based estimates are sensitive to model misspecification.

Des modèles statistiques sont souvent utilisés pour la prévision de variables aléatoires futures. La précision de leurs prévisions est habituellement quantifiée par des fonctions de perte. Nous examinons trois estimateurs de l'erreur de prévision pour modèles de survie. Il s'agit des estimateurs à base de modèles, par perte apparente et par validation croisée. Nous utilisons la méthode de pondération de censure des probabilités inverses pour tenir compte de la censure et traiter des fonctions de perte arbitraires. Nous développons des intervalles de confiance pour ces estimateurs. La performance des estimateurs est comparée dans le contexte d'une mauvaise spécification de modèle et de sélection de variables. Les résultats de simulation indiquent que la validation croisée produit des estimateurs ponctuels et intervalles de confiance fiables, alors que les estimateurs à base de modèles sont sensibles à une mauvaise spécification du modèle.

[MS-37]**Accelerated Hazards Mixture Cure Model****Modèle de mélange à défaillance accélérée avec fraction non à risque**Jiajia ZHANG, *University of South Carolina*, Yingwei PENG, *Queen's University*

Accelerated hazards mixture cure model allows covariates to have effects on the failure time distribution of uncured patients to be negligible at time

Les modèles de mélanges à défaillance accélérée avec fraction non à risque permettent aux covariables d'avoir un effet sur la distribution du temps de défaillance de patients

zero and to increase as time goes by. It is particularly useful in some cancer treatments when the treatment effect increases gradually from zero, and the existing models usually cannot handle this situation properly. We develop a rank based semi-parametric estimation method to obtain the maximum likelihood estimates of the parameters in the model. The numerical studies show that the new model provides a useful addition to the cure model literature.

à risques négligeable au temps zéro et croissant avec le temps. Ils sont particulièrement utiles en étude de traitements du cancer où l'effet du traitement augmente graduellement en s'éloignant du temps zéro, et les modèles existants ne peuvent habituellement pas traiter convenablement cette situation. Nous développons une méthode d'estimation semi-paramétrique basée sur les rangs pour obtenir les estimateurs du maximum de vraisemblance des paramètres du modèle. Des études numériques montrent que le nouveau modèle constitue un apport utile à la littérature sur les modèles avec fraction non à risque.

[MS-38]

A Nonparametric Estimator of the Conditional Survivor Function with Interval Censored Data **Un estimateur non paramétrique de la fonction de survie conditionnelle avec données censurées par intervalles**

Mohannad Hossein DEGHAN, *Laval university*, Thierry DUCHESNE, *Laval university*

In this paper we extend the survival function estimator for interval censored data proposed by Turnbull (1976) to obtain a nonparametric estimator of the conditional survival function for an interval censored response given a covariate. We get this generalized Turnbull estimator (GTE) by adding kernel weights that depend on the covariate values in Turnbull's self-consistency equation. We show that the GTE reduces to Turnbull's estimator when the kernel weights are equal and to the generalized Kaplan-Meier (Beran, 1981) under right-censoring. We also illustrate how appropriate bandwidth values can be set and show how to estimate the GTE's variance by bootstrap.

Dans cet article, nous étendons l'estimateur de la fonction de survie pour données censurées par intervalles proposé par Turnbull (1976) pour obtenir un estimateur non paramétrique de la fonction de survie conditionnelle pour réponse censurée par intervalles, étant donnée une covariable. Nous obtenons cet estimateur de Turnbull généralisé (ETG) en ajoutant des poids aux noyaux dépendants des valeurs de la covariable dans l'équation d'autocohérence de Turnbull. Nous montrons que l'ETG se ramène à l'estimateur de Turnbull lorsque les poids sont égaux, et à l'estimateur de Kaplan-Meier généralisé (1981) sous censure à droite. Nous illustrons aussi comment des valeurs appropriées du paramètre de lissage peuvent être déterminées et montrons comment estimer la variance de l'ETG par bootstrap.

Session 02J

Lundi 1 juin • Monday, June 1, 10:30 – 12:00

IRC Lobby

Poster Session I

Séance d'affichage I

[MS-39]

Cost-efficient Variable Selection with Non-additive Cost **Sélection de variables rentable sous coût non additif**

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

Non-additive costs of variable selection occur when obtaining one variable leads to lower costs for some other variables, e.g. higher order or interaction terms add no data collection costs. Non-additive cost is very common in health research, and identification of cost efficient diagnostic fac-

Des coûts non additifs de sélection de variables surviennent lorsque l'obtention d'une variable entraîne une diminution du coût de certaines autres variables. Par exemple, les termes d'interaction ou d'ordre supérieur n'ajoutent pas de coûts de collecte de données. Les coûts non additifs sont très fréquents en recherche en santé, et l'identification de facteurs

tors is of great interest for health researchers. A branching LARS procedure has been developed to select the important predictors to build a model not only good at prediction but also cost efficient. We present some approaches for addressing non-additive cost in cost efficient variable selection using Branching LARS.

diagnostics rentables est d'un grand intérêt pour les chercheurs en santé. Une procédure LARS à branchements a été développée pour sélectionner les prédicteurs importants afin de construire un modèle faisant de bonnes prévisions tout en étant rentable. Nous présentons quelques approches pour traiter les coûts non additifs en sélection de variables rentable au moyen de LARS à branchements.

[MS-40]

Sampling Ancestries at a Genomic Location Conditional on Data from Surrounding Genetic Markers

Échantillonnage d'ascendance à une position génomique conditionnellement aux données des marqueurs génétiques environnants

Kelly BURKETT, Brad McNENEY & Jinko GRAHAM, *Simon Fraser University*

For understanding genetic associations with disease outcomes, it can be useful to model the latent ancestries that give rise to the sample's genetic variability. Incorporating ancestry into genetic association requires Monte Carlo methods to sample from the ancestry space as this multidimensional space is too large to enumerate. We describe our implementation of a Markov Chain Monte Carlo sampler for the topology, node times and data at internal nodes of the ancestral tree at a genomic location, conditional on observed genotype data at surrounding markers. We apply the sampler to real data and suggest some potential applications.

Il peut être utile de modéliser les ascendances latentes à la variabilité génétique de l'échantillon afin de comprendre les associations génétiques avec l'évolution de la maladie. Incorporer les ascendances aux associations génétiques nécessite des méthodes de Monte-Carlo pour échantillonner dans l'espace des ascendances, car cet espace multidimensionnel est trop grand pour être énuméré. Nous décrivons un échantillonneur MCMC pour la topologie, les temps des noeuds et les données aux noeuds intérieurs de l'arbre généalogique à un emplacement génomique donné, conditionnellement aux données de génotype observées aux marqueurs environnants. Nous appliquons l'échantillonneur à des données réelles et suggérons quelques applications possibles.

[MS-41]

Clustering Incomplete Longitudinal Data

Regroupement de données longitudinales incomplètes

Mateen SHAIKH, Paul McNICHOLAS & Tony DESMOND, *University of Guelph*

Clustering longitudinal data is an important emerging research area with applications in a wide variety of disciplines. One issue that complicates the clustering of longitudinal data is that, in many applications, the data is often incomplete. A technique for clustering incomplete longitudinal data is outlined and applied to real data.

Le regroupement de données longitudinales est un domaine de recherche émergent important avec des applications dans une large gamme de disciplines. Un problème compliquant le regroupement de données longitudinales est que, dans plusieurs applications, les données sont souvent incomplètes. Une technique de classification de données longitudinales incomplètes est présentée et appliquée à des données réelles.

[MS-42]**Estimating Non-Homogeneous Transition Intensities in Continuous Time Multi-State Markov Models: An Application to Ordinal Data.****Estimation d'intensités de transition non homogènes sous modèles markoviens multi-états à temps continu : une application à des données ordinales**

Gerald LEBOVIC, *University of Toronto*, Patrick BROWN, *Cancer Care Ontario*, George TOMLINSON, *University Health Network*

Often, longitudinal studies examine ordinal outcomes (i.e. disease severity). In some cases, a Multi-State-Markov model may adequately characterize the behaviour of these outcomes. Kalbfleisch and Lawless (1985) and Gentleman et al. (1994) examine this model under the assumption of time homogeneous transition intensities. Currently, though less than ideal, the most common method to examine time varying intensities uses piecewise constant intensities. We propose a local likelihood method, (Tibshirani and Hastie 1987, Loader 1996), to estimate the transition intensities as continuous functions of time. A simulation study showing that this method enables more precise estimates of transition intensities will be presented.

Souvent, les études longitudinales observent des réponses ordinales (par exemple, des degrés de sévérité d'une maladie). Dans certains cas, un modèle multi-états markoviens peut adéquatement caractériser le comportement de ces réponses. Kalbfleisch et Lawless (1985) et Gentleman et al. (1994) ont examiné ce modèle sous l'hypothèse d'intensité de transitions homogènes à travers le temps. Présentement, la façon la plus habituelle d'étudier des intensités variant avec le temps est de passer par des intensités constantes par morceaux, ce qui n'est pas idéal. Nous proposons une méthode par vraisemblance locale (Tibshirani et Hastie 1987, Loader 1996) pour estimer les intensités de transition comme fonctions continues du temps. Une étude de simulation montrant que cette méthode permet des estimés plus précis des intensités de transition sera présentée.

[MS-43]**Inverse Gaussian Cure Rate Mixture Models****Modèles de mélanges gaussiens inverses avec fraction non à risque**

Ruimin XU, Paul McNICHOLAS, Gerarda DARLINGTON & Anthony DESMOND, *University of Guelph*

Cure rate mixture models have been widely used to model failure time data with the existence of long-term survivors. The problems of interest in these models include the proportion of cured patients, known as cure rate, and the failure time distribution of uncured patients. The inverse Gaussian cure rate mixture model, along with a modified long-term survival component in the presence of competing risks, will be introduced. The EM algorithm is employed to estimate the parameters of interest in the model. The model will be applied to real data to compare with other popular failure time distributions of uncured patients.

Les modèles de mélanges à fraction non à risque ont été largement utilisés pour modéliser des données de temps de défaillance avec survivants à long terme. Les problèmes d'intérêt dans ces modèles incluent la proportion de patients non à risque et la distribution du temps de défaillance des patients à risque. Le modèle de mélange inverse gaussien à fraction non à risque, avec une composante de survie à long terme modifiée en présence de risques concurrents, sera présenté. L'algorithme EM est utilisé pour estimer les paramètres d'intérêt dans le modèle. Le modèle sera ensuite appliqué à des données réelles pour le comparer à d'autres distributions de la défaillance à long terme pour patients non à risque.

[MS-44]**Testing Multiple Outcomes in the Presence of Non-Normality and Covariance Heterogeneity
Tester des réponses multiples en présence de non-normalité et d'hétérogénéité de la covariance**Tolulope SAJOBI & Lisa LIX, *University of Saskatchewan*

This study investigates procedures for testing mean differences between treatment and control groups on multiple, correlated outcomes. Resampling-based procedures with least-squares or robust estimators and an approximate degrees of freedom test statistic are considered. A simulation study investigates the effects of population shape, covariance heterogeneity, correlation structure, sample size, and mean configuration on the familywise error rate (FWER), all-variable power, and any-variable power. Procedures based on the bootstrap with robust estimators controlled the FWER better than procedures based on least squares estimators. Power differences between bootstrap step-up and step-down procedures were less than 10% even for non-normal data.

Cette étude porte sur les procédures pour tester les différences en moyennes entre groupes traitement et contrôle pour de multiples réponses corrélées. Des procédures basées sur le rééchantillonnage avec estimateurs des moindres carrés ou robustes et une statistique de test de type degrés de liberté approximatifs sont considérées. Une étude de simulation sur les effets de la forme de la population, l'hétérogénéité de covariance, la structure de corrélation, la taille d'échantillon, la configuration moyenne d'erreur de type global, la puissance pour toutes les variables, et la puissance pour n'importe quelle variable. Les procédures basées sur le bootstrap avec estimateurs robustes contrôlaient mieux l'erreur de type global que les procédures basées sur des estimateurs des moindres carrés. Les différences en puissance entre procédures bootstrap «step-up» et procédures «step-down» étaient de moins de 10% même pour données non normales.

[MS-45]**Sample Size and Design Effect for Multilevel Logistic Regression Models****Taille échantillonnale et effet du plan d'expérience pour modèles de régression logistique à plusieurs niveaux**Hedy JIANG & Patrick BROWN, *Cancer Care Ontario*

Many studies conducted in health-care, screening or educational interventions collect both individual and group level data as outcome measures. Outcomes of interest in those fields often reflect binary measures such as presence or absence of a symptom. In this paper, bootstrap method is used to assess the effect of varying sample size at both the individual and group level on the accuracy of the estimates. In addition, a simulation based method for calculating the intra-class correlation coefficient is introduced and compared with the existing methods only depending on the variance of the random intercept in an unconditional multilevel logistic model.

Plusieurs études en soins de santé, en dépistage ou en interventions en éducation recueillent des données au niveau des individus et au niveau des groupes comme réponses. Les réponses d'intérêt dans ces domaines sont souvent des mesures binaires telles la présence ou l'absence d'un symptôme. Dans cet article, la méthode du bootstrap est utilisée pour évaluer l'effet de différentes tailles d'échantillon au niveau des groupes et au niveau des individus sur la précision des estimateurs. De plus, une méthode par simulation pour le calcul du coefficient de corrélation intraclasse est présentée et comparée aux méthodes existantes dépendant uniquement de la variance de l'ordonnée à l'origine dans un modèle logistique multi-niveaux inconditionnel.

[MS-46]

Ascertainment of Osteoporosis Cases Using Administrative Databases and Capture-Recapture Method
Découverte de cas d'ostéoporose au moyen de bases de données administratives et de la méthode capture-recapture

Osama BATAINEH & Lisa LIX, *University of Saskatchewan*

Osteoporosis prevalence estimates from population-based physician and hospital administrative data are known to be conservative. Capture-recapture (CR) methods were applied to hospital, physician, pharmaceutical, and bone mineral density test data from Manitoba to improve case ascertainment. Three- and four- source log- linear CR models that assume independence and homogeneity of capture probability were compared to models that adjust for dependence and heterogeneity. A four-source model resulted in an estimated prevalence of 14.1% which is comparable to national estimates from primary data. Using three sources without controlling for heterogeneity resulted in substantially lower estimates (i.e., 8.1%).

Les estimations de la prévalence de l'ostéoporose à partir de données administratives d'hôpitaux et de médecins sont conservatrices. Des méthodes capture-recapture ont été appliquées à des données pharmaceutiques et d'hôpitaux, de médecins et de tests de densité minérale osseuse du Manitoba pour améliorer l'évaluation des cas. Des modèles CR log-linéaires à trois ou quatre sources supposant l'indépendance et l'homogénéité des probabilités de capture ont été comparés à des modèles ajustant pour la dépendance et l'hétérogénéité. Un modèle à quatre sources donne une prévalence estimée de 14.1%, ce qui est comparable aux estimations nationales à partir de données primaires. En utilisant trois sources sans contrôler pour l'hétérogénéité, des estimations substantiellement plus basses (par exemple, 8.1%) sont obtenues.

[MS-47]

Identification of Disease Source via Multivariate Spatial
Identification de la source d'une maladie au moyen de modèles spatiaux multivariés

Daniel GILLIS, Tony DESMOND & John HOLT, *University of Guelph*

Gastrointestinal (GI) disease data obtained from the Canadian Institute for Health Information provides motivation to extend Mixture Model (MM) literature in order to classify disease based on infection source. Specifically, MMs are employed to classify GI data as foodborne or waterborne. This presentation will account for spatially indexed disease using two methods. In the first case, independent conditionally autoregressive spatial priors will be assigned to the log linear term of each of the mixture components (one per disease source). The second case investigates a non-independence assumption. Models are compared via simulation, with application to Alberta GI data (1992-1998).

Les données de maladies gastro-intestinales (GI) obtenues de l'Institut canadien d'informations sur la santé apportent une motivation pour étendre la littérature sur les modèles de mélanges (MM) à la classification de maladies basée sur la source d'infection. Spécifiquement, les MM sont utilisés pour classer les données GI comme étant d'origine hydrique ou alimentaire. Cette présentation va tenir compte de maladies spatialement indexées en utilisant deux méthodes. Dans le premier cas, des lois a priori indépendantes conditionnellement autorégressives spatiales seront assignées au terme log linéaire de chaque composante du mélange (une par source de maladie). Le second cas étudie une hypothèse de non-indépendance. Les modèles sont comparés au moyen de simulations, avec application à des données GI d'Alberta (1992-1998).

Session 03A**Lundi 1 juin • Monday, June 1, 13:30 – 15:00****IRC 6**

Teaching Biostatistics in Public Health Programs: Struggles and Successes

Enseigner la biostatistique dans les programmes de santé publique : défis et succès

[MS-48]**Round Table: Teaching Biostatistics in Public Health Programs****Table ronde : l'enseignement de la biostatistique en programmes de santé publique**

Rollin BRANT & Kevin CRAIB, *University of British Columbia*, Joel DUBIN, *University of Waterloo*, Gian JHANGRI, *University of Alberta*, Punam PAHWA, *University of Saskatchewan*, Kevin THORPE, *University of Toronto*

Recent challenges in the health system have revealed a shortage of appropriately trained public health professionals in Canada. This has fostered a re-investment in existing educational programs and to the creation of new schools of public health at universities across Canada. Correspondingly, there is growing demand for teaching in the core discipline of Biostatistics. The aim of this round-table discussion is to provide a discussion forum for meeting this challenge. Among the issues to be covered are not only what needs to be taught, but how best to teach it, taking advantage of current technology and recent research in pedagogy.

De récents défis dans le système de santé ont révélé un manque de professionnels qualifiés en santé publique au Canada. Ceci a entraîné un ré-investissement dans les programmes éducatifs existant et la création de nouvelles écoles de santé publique dans plusieurs universités à travers le Canada, ce qui a entraîné un accroissement de la demande d'enseignement en biostatistique. Le but de cette table ronde est de fournir un forum de discussion pour surmonter ces défis. Les questions à aborder comprennent ce qu'il faut enseigner, mais aussi comme l'enseigner le mieux possible, en prenant avantage des technologies actuelles et de récentes recherches en pédagogie.

Session 03B**Lundi 1 juin • Monday, June 1, 13:30 – 15:00****IRC 1**

Stochastic Spatial Models

Modèles spatiaux stochastiques

[MS-49]**Long Range Random Walk on the Two-dimensional Torus****Marche aléatoire à long terme sur un tore bidimensionnel**Ted Cox, *Syracuse University*

We consider random walk on the two-dimensional torus of side N with step-distribution uniform on the square of side M_N . The problem is to determine the asymptotic hitting time of the origin starting at different sites x . We show that this time is of order N^2 , uniformly in x , if and only if M_N is of larger order than $(\log N)^{1/2}$. This result complements one of Matsen and Wakeley, who considered the case that M_N is of order N , and can be used to determine when the corresponding stepping stone model is homogeneously mixing or not.

Nous considérons une marche aléatoire sur le tore bidimensionnel de côté N avec distribution des pas uniforme sur le carré de côté M_N . Le problème consiste à déterminer le temps d'atteinte asymptotique de l'origine à partir de différents sites x . Nous montrons que ce temps est d'ordre N^2 uniformément en x si et seulement si M_N est plus grand que $(\log N)^{1/2}$. Ce résultat complète un de ceux de Matsen et Wakeley, qui ont considéré le cas où M_N est d'ordre N , et peut être utilisé pour déterminer quand le modèle tremplin correspondant est homogènement mélangeant ou non.

[MS-50]**Variational Approach to Pinning of Random Polymers****Approche variationnelle au blocage de polymères aléatoires**

Frank DEN HOLLANDER, *University of Leiden, Netherlands*, Dimitris CHELIOTIS, *EURANDOM*

We look at the pinning of a directed polymer by a one-dimensional linear interface carrying random charges. There are two phases – localized and delocalized – depending on the temperature and the disorder bias. We derive a variational formula for the quenched critical curve separating the two phases. Applications of this variational formula include a necessary and sufficient criterion for the quenched critical curve to differ from its annealed counterpart, a property referred to as relevance of the disorder.

Nous considérons le blocage d'un polymère dirigé par une interface linéaire unidimensionnelle comportant des changements aléatoires. Il y a deux phases - localisée et délocalisée - dépendant de la température et du biais de désordre. Nous obtenons une formule variationnelle pour la courbe critique trempée séparant les deux phases. Une application de cette formule variationnelle est un critère nécessaire et suffisant pour que la courbe critique trempée diffère de sa contrepartie recuite, une propriété dite pertinence du désordre.

[MS-51]**Planar Aggregation and the Coalescing Brownian Flow****L'agrégation planaire et le flot brownien coalescent**

Amanda TURNER, *University of Lancaster*

Diffusion limited aggregation (DLA) is a random growth model which was originally introduced in 1981 by Witten and Sander. This model is prevalent in nature and has many applications in the physical sciences as well as industrial processes. Unfortunately it is notoriously difficult to understand, and only one rigorous result has been proved in the last 25 years. We consider a simplified version of the Hastings-Levitov model of planar aggregation which is obtained by composing certain independent random conformal maps and show that the evolution of harmonic measure on the boundary of the cluster converges to the coalescing Brownian flow.

L'agrégation limitée par diffusion (ALD) est un modèle de croissance aléatoire originellement introduit par Witten et Sander en 1981. Ce modèle est prévalent dans la nature et a plusieurs applications en sciences physiques ainsi qu'en processus industriels. Malheureusement, il est difficile à comprendre, et un seul résultat rigoureux à son sujet a été prouvé au cours des 25 dernières années. Nous considérons une version simplifiée du modèle Hastings-Levitov d'agrégation planaire obtenu en composant certaines cartes conformelles aléatoires indépendantes et montrons que l'évolution de la mesure harmonique sur la frontière de l'agrégat converge vers le flot brownien coalescent.

Session 03C**Lundi 1 juin • Monday, June 1, 13:30 – 15:00****IRC 5****Statistical Methods in Clinical Trials****Méthodes statistiques dans les essais cliniques****[MS-52]****Flexibility and Optimality of Statistical Procedures in Drug Development****Flexibilité et optimalité des procédures statistiques en développement de médicaments**

Kuang-Kuo LAN, *Johnson & Johnson PRD*

Many statistical procedures used in drug development were derived from simple mathematical frameworks so that optimal solutions can be ob-

Plusieurs procédures statistiques utilisées en développement de médicaments ont été dérivées de cadres mathématiques simples permettant l'obtention de solutions opti-

tained. In practice, we often have to trade flexibility for optimality. Several examples will be given, including: the choice of a primary endpoint in a clinical trial when multiple endpoints are under consideration; the over-ruling of a group sequential boundary; and the potential misinterpretation of survival data when the proportional hazards assumption is violated.

males. En pratique, nous avons souvent à équilibrer optimalité et flexibilité. Plusieurs exemples seront donnés, incluant le choix d'un point d'arrêt principal dans un essai clinique lorsque plusieurs points d'arrêts sont considérés, passer outre la frontière dans un plan séquentiel en groupes, et la mésinterprétation potentielle de données de survie lorsque l'hypothèse des défaillances proportionnelles n'est pas respectée.

[MS-53]

Sample Size Re-estimation in Clinical Trials

Ré-estimation de taille échantillonnale en essais cliniques

Michael PROSCHAN, *National Institute of Allergy and Infectious Diseases*

Sample size calculations for clinical trials depend on both nuisance parameters and the treatment effect. If we change the sample size after peeking at interim data, under what circumstances can we ignore that aspect and pretend that the sample size had been fixed in advance? Will this lead to the correct type I error rate averaged over all possible sample sizes? How about conditioned on the sample size actually selected? These are some of the questions answered in this talk.

La détermination de la taille échantillonnale en essais cliniques dépend à la fois de paramètres de nuisance et de l'effet de traitement. Si nous changeons la taille échantillonnale après avoir jeté un coup d'oeil à des données partielles, sous quelles circonstances est-il possible d'ignorer cet aspect et prétendre que la taille échantillonnale était fixée d'avance ? Ceci nous mènera-t-il à un taux d'erreur de type I correct en moyenne pour toutes les tailles échantillonnales possibles ? Que se passe-t-il si nous conditionnons sur la taille échantillonnale finalement choisie ? Ces questions sont parmi celles abordées dans cet exposé.

[MS-54]

Approaches to Handling Data When a Phase II Trial Deviates from the Pre-specified Simon's Two-Stage Design

Approches au traitement des données lorsqu'un essai de phase II dévie du plan d'expérience à deux étapes de Simon préspecifié

Weichung SHIH & Yujun WU, *University of Medicine & Dentistry of New Jersey*

Simon's "optimal" and "minimax" two-stage designs are common methods for conducting phase IIA studies investigating new cancer therapies. In practice we often are unable to adhere to the original event number and sample size of the design. In this paper/talk, we consider some approaches in handling situations where deviations or interruptions from the original Simon's two-stage design occur because recruitment of patients is slower than expected. We consider four scenarios and use conditional probabilities to address the issues commonly inquired by the scientific review board. We also discuss how to report p-values in these situations.

Les plans à deux phases "optimal" et "minimax" de Simon sont des méthodes communes pour mener des études de phase IIA concernant de nouvelles thérapies pour le cancer. En pratique, il est souvent impossible d'adhérer au nombre d'événements et à la taille des échantillons prévus par le plan. Dans cet article/exposé, nous considérons certaines approches pour traiter les situations où une déviation ou une interruption du plan à deux phases de Simon se produit parce que le recrutement de patients est plus lent que prévu. Nous considérons quatre scénarios et utilisons les probabilités conditionnelles pour répondre aux questions fréquemment posées par les comités de lecture. Nous discutons aussi des valeurs-p à rapporter dans ces situations.

Session 03D**Lundi 1 juin • Monday, June 1, 13:30 – 15:00****IRC 3**

Statistics and Science

La statistique et la science

[MS-55]**Will There Be Enough Water?****Y aura-t-il assez d'eau ?**John CAREY, *Water and Science Technology Directorate, Environment Canada*

The question, "Will There Be Enough Water?", will be discussed.

Nous discuterons de la question «y aura-t-il assez d'eau ?»

[MS-56]**Climate Change: The Role of Science and its Limitations****Changements climatiques : le rôle de la science et ses limites**John STONE, *formerly Environment Canada*

Science, most powerfully through the Intergovernmental Panel on Climate Change, has been effective in putting the issue of climate change on the international and national policy agendas and in raising the awareness of the general public. However, there are limitations both physical limitations, as a result of the incompleteness of our understanding or because of the complexity of the system, and institutional limitations, because decisions have to be based in part on values and this requires a political process. In this talk we will explore some of these limitations and how they may be overcome.

La science, principalement grâce aux travaux du Groupe d'experts intergouvernemental sur l'évolution du climat, a réussi à mettre les changements climatiques aux ordres du jour nationaux et internationaux et à sensibiliser le grand public. Cependant, il y a des limites tant physiques (dus à l'incomplétude de notre compréhension ou à la complexité du système étudié) qu'institutionnelles (parce que les décisions à prendre seront basées sur des valeurs, ce qui requiert un processus politique) à ce que la science peut faire. Dans cet exposé, nous allons explorer certaines de ces limites et des façons de les surmonter.

[MS-57]**Engaging the Canadian Public with Sound Science****Communiquer avec le public canadien avec de la science bien fondée**Peter CALAMAI, *Columnist, Toronto Star*

The Science Media Centre of Canada is a new, non-profit body that will probably emerge in the second half of 2009. Its purpose is to help journalists cover stories in which science plays an important part. This means everything from stories where science is the story, such as the discovery of a new Earth-like planet, to stories where science provides the crucial factual underpinning, such as citizen opposition to cell phone towers. The goal is better journalism through coverage of science that is more informed, more accurate and more incisive. Scientists, journalists, policy makers and the public will benefit.

Le centre canadien sciences et médias est un nouvel organisme à but non lucratif qui émergera probablement durant la deuxième moitié de 2009. Son but est d'aider les journalistes à couvrir des questions où la science joue un rôle important. Ces questions comprennent les découvertes scientifiques elles-mêmes, par exemple la découverte d'une planète semblable à la Terre, mais aussi des questions dont notre compréhension est facilitée par la science, comme l'opposition citoyenne aux tours de téléphonie cellulaire. Le but est d'améliorer le journalisme par une couverture plus informée, plus précise et plus incisive de la science. Les scientifiques, les journalistes, les preneurs de décisions et le grand public en bénéficieront.

Session 03E**Lundi 1 juin • Monday, June 1, 13:30 – 15:00****MSL 102****Simulator Assisted Inference for Environmental Systems****Inférence assistée par simulateur pour les systèmes environnementaux****[MS-58]****Posterior Exploration for Computationally Intensive Forward Models****Exploration a posteriori pour modèles avant demandant une grande puissance de calcul**

Shane REESE, *Brigham Young University*, David HIGDON, David MOULTON & Jasper VRUGT, *Los Alamos National Laboratory*, Colin FOX, *University of Otago, New Zealand*

While standard single-site Metropolis updating proves effective in a variety of applications, it has the drawback of requiring many calls to the simulation model. Here we compare two MCMC schemes that use highly multivariate updates to sample from the posterior: the multivariate random walk Metropolis algorithm and the distributed evolution-MCMC sampler. Such schemes are alluring for computationally demanding inverse problems since they have the potential to update many components at once, while requiring only a single evaluation of the simulator. We consider new formulations based on faster, approximate simulators created by altering the multi-grid solver used in the simulator.

Bien que la mise à jour habituelle de Metropolis en un seul site soit efficace dans une variété d'applications, elle a le désavantage de nécessiter beaucoup d'appels au modèle de simulation. Nous comparons ici deux schémas MCMC utilisant des mises à jour hautement multivariées pour échantillonner de la loi a posteriori : l'algorithme Metropolis à marche aléatoire multivariée et l'échantillonneur MCMC à évolution distribuée. De tels schémas sont attrayants pour des problèmes inverses demandant une grande puissance de calcul, car ils ont le potentiel de mettre à jour simultanément beaucoup de composantes en nécessitant une seule évaluation du simulateur. Nous considérons de nouvelles formulations basées sur des simulateurs approximatés plus rapides créés en altérant le résolveur multigrilles utilisé dans le simulateur.

[MS-59]**A Spatial Analysis of Regional Climate Model Ensembles****Une analyse spatiale d'ensembles de modèles régionaux du climat**

Stephan SAIN, *National Center for Atmospheric Research*

The North American Regional Climate Change Assessment Program (NARCCAP) examines uncertainty in the output of regional climate models and projections of climate. The program seeks to use several regional models with boundary conditions supplied by different atmosphere-ocean general circulation models to produce a variety of output. We aim to develop statistical methodology to analyze this output and assess sources of uncertainty. We present a Bayesian hierarchical framework, based upon a multivariate spatial model, that captures the complex distribution of the spatial fields. Case studies will be presented based as well as an initial analysis of model output associated with NARCCAP.

Le North American Regional Climate Change Assessment Program (NARCCAP, programme d'évaluation des changements climatiques régionaux nord-américain) examine l'incertitude dans les sorties des modèles régionaux du climat et dans les projections climatiques. Le programme cherche à utiliser plusieurs modèles régionaux dont les conditions aux limites sont fournies par différents modèles généraux de circulation atmosphère-océan afin de produire une variété de sorties. Nous visons à développer une méthodologie statistique afin d'analyser ces sorties et d'évaluer les sources d'incertitude. Nous présentons un cadre de modélisation hiérarchique bayésien, basé sur un modèle spatial multivarié, qui représente la distribution complexe des champs spatiaux. Des études de cas et une analyse initiale des sorties des modèles associés au NARCCAP seront présentées.

[MS-60]**Calibration and Prediction Problems in Catchment Scale Hydrology****Problèmes de calibration et de prévision en hydrologie à l'échelle des bassins-versants**Cari KAUFMAN, *UC Berkeley*

Statistical models in environmental applications are making increasing use of the scientific knowledge represented in systems of partial differential equations describing the evolution of environmental processes over time. I will discuss statistical methodology for simultaneously 1) inferring unknown model parameters, and 2) predicting the behavior of the process. I will focus on modelling soil moisture fields under various climatic conditions, using a hierarchical model in which soil moisture is the process of interest. Its behavior is governed by a catchment-level hydrology model, and the observations are modeled according to a stochastic redistribution of total water.

Les modèles statistiques en applications environnementales font un usage croissant des connaissances scientifiques que représentent les systèmes d'équations différentielles partielles décrivant l'évolution des processus environnementaux à travers le temps. Je vais discuter d'une méthodologie statistique pour simultanément 1) inférer les paramètres inconnus d'un modèle, et 2) prévoir le comportement du processus. Je vais me concentrer sur la modélisation de champs d'humidité du sol sous différentes conditions climatiques au moyen d'un modèle hiérarchique où l'humidité du sol est le processus d'intérêt. Son comportement est dirigé par un modèle hydrologique au niveau des bassins-versants et les observations sont modélisées d'après une redistribution stochastique de l'eau entière.

Session 03F**Lundi 1 juin • Monday, June 1, 13:30 – 15:00****CHE 150****Some Recent Development in Analyzing Complex Survey Data****Évolution récente de l'analyse de données d'enquêtes complexes****[MS-61]****Statistics Canada's Collaboration with University Researchers: Statistical Methods for Complex Survey Data****La collaboration de Statistique Canada avec des chercheurs universitaires : méthodes pour données d'enquête complexes**Milorad KOVACEVIC, *Statistics Canada*, Changbao WU, *University of Waterloo*

The internship program for graduate students in statistics co-sponsored by the MITACS, NPCDS and Statistics Canada is entering its sixth year. This program is part of the larger project on Statistical methods for complex survey data envisioned as a collaboration among Canadian researchers, with an objective to develop methods and to improve methods for analysis of complex survey data. In this overview we look at the program from the initial idea up to today's presentations. While emphasizing its success, we'll also talk about its drawbacks, and discuss its future.

Le programme de stages pour étudiants de cycles supérieurs en statistique, co-parrainé par le MITACS, le NPCDS et Statistique Canada amorce sa sixième année. Ce programme fait partie d'un plus grand projet de collaboration entre les chercheurs canadiens sur les méthodes statistiques pour données d'enquête complexes, visant à développer et améliorer les méthodes d'analyse de données d'enquête complexes. Nous faisons la revue du programme, de sa conception jusqu'aux présentations d'aujourd'hui. Bien que nous mettions l'accent sur ses réussites, ses désavantages et son avenir seront aussi abordés.

[MS-62]**Logistic Regression Analysis Using Complex Survey Data with Misclassification in an Ordinal Covariate
Analyse par régression logistique de données d'enquête complexes avec mauvaise classification dans une variable ordinale**Zhijian CHEN, *University of Waterloo*, Harold MANTEL, *Statistics Canada*

Measurement errors are common in complex surveys where variables are often collected through non-standard procedures. We consider estimation of regression coefficients in logistic regression analysis using survey data, where misclassification of an ordinal covariate is present and dependent on other variables. We propose to use the expected score method which employs a parametric assumption for the measurement-error process. The method is applied to a dataset from Canadian Community Health Survey, where self-reported body mass index is believed to be a risk factor for several chronic conditions. A limited simulation study is carried out to investigate the proposed method's performance.

Les erreurs de mesures sont fréquentes en enquêtes complexes où les variables sont souvent recueillies au moyen de procédures inhabituelles. Nous considérons l'estimation de coefficients de régression en analyse de régression logistique avec données d'enquête avec mauvaise classification d'une covariable ordinale dépendante de d'autres variables. Nous proposons d'utiliser la méthode des scores espérés employant une hypothèse paramétrique sur le processus d'erreurs de mesures. La méthode est appliquée à un jeu de données de l'Enquête sur la santé dans les collectivités canadiennes, où l'indice de masse corporelle déclaré par les sujets est peut-être un facteur de risque pour plusieurs conditions chroniques. Une étude restreinte de simulation est effectuée afin d'étudier la performance de la méthode proposée.

[MS-63]**Fitting Cox Models to Jobless Spell Durations in SLID****Ajuster des modèles de Cox aux durées de périodes sans emploi dans l'EDTR**Dagmar MARIACA-HAJDUCEK & Jerald LAWLESS, *University of Waterloo*

This talk will discuss fitting Cox PH models to jobless spell durations for individuals from the Survey of Labour and Income Dynamics (SLID), over a six-year period. Features like within-individual and within-cluster association in spell durations, dependent loss to follow-up (LTF) and non-ignorable sampling design are considered. Within-individual dependence is taken into account by including previous jobless history in the form of covariates. Dependent loss to follow-up and non-ignorable sampling are accounted for by using combined sampling and LTF inverse probability weights in the estimation procedure.

Cet exposé discutera de l'ajustement de modèles de risque proportionnel de Cox pour la durée de périodes sans emploi pour des individus de l'Enquête sur la dynamique du travail et du revenu (EDTR), sur une période de six ans. Des caractéristiques comme des associations intra-individu et intra-grappe dans les durées de périodes, la perte dépendante au suivi et le plan d'échantillonnage non ignorable sont considérés. La dépendance intra-individu est traitée en incluant l'historiques des périodes de sans emploi précédentes des individus comme covariables. La perte dépendante au suivi et le plan d'échantillonnage non ignorable sont traités en utilisant l'échantillonnage combiné et des poids de probabilités inverses de perte au suivi dans la procédure d'estimation.

[MS-64]

Challenges in Analyzing National Longitudinal Data: Cohort-Sequential Design and the Application of Sampling Weights When Using Structural Equation Modeling (SEM)

Défis pour l'analyse des données longitudinales nationales : Cohortes séquentielles par construction et application des poids de sondage dans les modèles à équations structurelles (MES)

Yan LIU, *University of British Columbia*, Milorad KOVACEVIC & Abdelnasser SAIDI, *Statistics Canada*, Bruno ZUMBO, *University of British Columbia*

Cohort-sequential design has been introduced into longitudinal studies in order to deal with time constraints and maturation of population groups of interest, as well as with sample attrition. The analysis of longitudinal survey data using this research design brings several challenges for researchers. This paper aims to demonstrate and compare two analytical methods adopting SEM approach for analyzing cohort-sequential design of data. These two methods are illustrated by the NLSCY data. Because sampling weights have often been neglected in using SEM, which often biases the results, we also demonstrate how to apply sampling weights in this kind of modeling.

Les cohortes séquentielles par construction ont été introduites dans les études longitudinales pour tenir compte des contraintes du temps, de la maturation des groupes de population d'intérêt et aussi de l'érosion de l'échantillon. L'analyse des données d'enquêtes longitudinales utilisant cette recherche selon le plan pose plusieurs défis aux chercheurs. Le but de cet article est d'étudier et de comparer deux méthodes analytiques adoptant l'approche MES pour analyser des données de cohortes séquentielles par construction. Ces deux méthodes sont illustrées à l'aide des données de l'ELNEJ. Comme les poids de sondage ont été souvent négligés lors de la modélisation MES, ce qui engendre un biais dans les résultats, nous démontrons aussi comment appliquer les poids de sondage pour cette sorte de modèles.

Session 03G

Lundi 1 juin • Monday, June 1, 13:30 – 15:00

CHE 250

Estimation Methods in Business and Industrial Statistics

Méthodes d'estimation en statistique industrielle et de gestion

[MS-65]

Recursive Estimation Using Combined Optimal Estimating Functions

Estimation réursive utilisant des fonctions estimatrices optimales combinées

Melody GHAHRAMANI, *University of Winnipeg*, Aerambamoorthy THAVANESWARAN, *University of Manitoba*

In this talk, we discuss combining estimating functions for the parameter of discrete time stochastic processes where the conditional mean and the conditional variance depend on the same parameter with applications to volatility models. The optimal combined estimating function is shown to contain more information than each of the component estimating functions. As an application, nonlinear recursive estimation of the parameter in nonlinear time series models, based on the optimal combined estimating function is derived. Pre-filtered optimal estimation of the parameter of doubly stochastic time series based on the optimal combined estimating function will also be discussed.

Dans cet exposé, nous nous intéressons à combiner les fonctions d'estimation pour le paramètre de processus stochastiques à temps discret où la moyenne et la variance conditionnelles dépendent du même paramètre, avec applications aux modèles de volatilité. Nous montrons que la fonction d'estimation combinée optimale contient plus d'information que chacune des fonctions d'estimation composantes. Comme application, nous effectuons l'estimation non linéaire réursive du paramètre en séries chronologiques non linéaires en nous basant sur la fonction d'estimation combinée optimale. Nous discuterons aussi de l'estimation préfiltrée optimale du paramètre de séries chronologiques doublement stochastiques basée sur la fonction d'estimation combinée optimale.

[MS-66]**Bias Reduction of the Least-Squares Monte Carlo Estimators of American Option Values****Réduction du biais des estimateurs des moindres carrés Monte-Carlo de la valeur d'options américaines**Kin Hung KAN & Mark REESOR, *The University of Western Ontario*

Whitehead, Davison and Reesor (2008) introduce a general bias reduction technique for pricing American options by Monte Carlo methods based on large sample theory that corrects the stochastic tree estimators. We apply this technique to the popular least-squares Monte Carlo estimators using well-known results about the large sample properties of least-squares estimators. We derive an expression for bias-corrected estimators. Numerical results show the effectiveness of this technique.

Whitehead, Davison et Reesor (2008) présentent une technique générale de réduction de biais dans la fixation de prix d'options américaines au moyen de méthodes de Monte-Carlo basée sur la théorie asymptotique et qui corrigé les estimateurs d'arbres stochastiques. Nous appliquons cette technique aux estimateurs des moindres carrés Monte-Carlo en utilisant des résultats bien connus sur les propriétés asymptotiques des estimateurs des moindres carrés. Nous obtenons une expression pour les estimateurs avec correction pour le biais. Des résultats numériques démontrent l'efficacité de cette technique.

[MS-67]**A Comparative Study of Robust Control Charts****Une étude comparative de cartes de contrôle robustes**ASOKAN MULAYATH VARIYATH, *Memorial University of Newfoundland*, Shoja'eddin CHENOURI, *University of Waterloo*

For monitoring multivariate individual observations, robust control charts are often used. Recently, Chenouri et al. (2008) proposed re-weighted MCD based Hotelling's T₂ control charts for monitoring Phase II multivariate observations. This approach is computationally fast and has better performance compared to T₂ charts when outlier rejection is done in phase I with ordinary MCD and MVE estimates. In this paper, we construct the phase II control limits of the Hotelling's T₂ charts based on the re-weighted MVE, and S- estimators empirically and assess the performance of these charts using Monte Carlo simulations. We implement our proposed methods in a real example.

Pour surveiller des observations individuelles multivariées, des cartes de contrôle robustes sont souvent utilisées. Récemment, Chenouri et al (2008) ont proposé des cartes de contrôle T₂ de Hotelling basées sur une repondération MCD pour la surveillance d'observations multivariées de phase II. Les calculs sont rapides sous cette approche, et la performance est meilleure comparativement aux cartes T₂ lorsque les observations aberrantes sont rejetées en phase I avec estimateurs MCD et MVE ordinaires. Dans cet article, nous construisons les limites de contrôle de phase II des cartes T₂ de Hotelling à partir des MVE repondérés et des S-estimateurs empiriques et évaluons la performance de ces cartes par simulations de Monte- Carlo. Nous implémentons notre méthode dans un exemple réel.

[MS-68]**Estimation and Coverage of the Hurricane Risk with the use of Copulas****Estimation et couverture du risque d'ouragans avec copules**Juan SORIANO, *University of Barcelona/Agroasemex*, F.J. SARRASI, *University of Barcelona*

The hurricanes are classified as the principal natural hazard that can originate catastrophic losses. This characteristic, together with the great uncertainty associated with them, has generated a lack of coverage for their derived economical losses. A solution for this problem is the implementation of alternative instruments for risk transfer, such as

Les ouragans sont classés comme étant le principal risque naturel pouvant causer des pertes catastrophiques. Cette caractéristique et la grande incertitude leur étant associée ont généré une absence de couverture pour les pertes économiques qu'ils entraînent.

Une solution à ce problème est la mise en oeuvre d'instruments alternatifs pour le transfert de risques, comme les titres liés

Insurance Link Securities.

In order to define the instrument's rate a deeper evaluation based in the interaction of variables such as: exposition, vulnerability and location, is needed. In this work we present the use the theory of copulas to describe the interaction of these variables.

aux taux d'assurances.

Pour définir le taux d'un instrument, une évaluation plus approfondie basée sur l'interaction de variables telles l'exposition, la vulnérabilité et l'emplacement, sera nécessaire. Dans ce travail, nous présentons l'utilisation de la théorie des copules pour décrire l'interaction de ces variables.

[MS-69]

Time-Dependent > Regime-switching Models: Estimation and Applications

Modèles à changement de régime dépendant du temps : Estimation et applications

Yanqiao ZHANG, *University of Waterloo*

Linearity of the drift coefficient in the diffusion process for interest rates has been studied by many authors and is yet inconclusive. Fan, et al (2003) introduced a time-dependent model and found, using non-parametric estimation methods, that there is no evidence of time variation in the drift coefficient. On the other hand, models used for pricing of interest rates securities typically are time dependent. We attempt to fill this gap by proposing a continuous time-dependent regime-switching model. We verify stationarity of the proposed process and develop algorithms to estimate a discrete time analogue of our model. Applications are also conducted.

La linéarité du coefficient de dérive dans le processus de diffusion pour taux d'intérêt a été étudiée par plusieurs auteurs, mais reste non concluante. Fan et al (2003) ont présenté un modèle dépendant du temps et ont trouvé, au moyen de méthodes d'estimation non paramétrique, qu'il n'y a pas d'évidence de variation avec le temps dans le coefficient de dérive. D'un autre côté, les modèles utilisés pour la fixation des prix de titres sur taux d'intérêt sont habituellement dépendants du temps. Nous tentons de combler cet écart en proposant un modèle à changement de régime dépendant du temps à temps continu. Nous vérifions la stationnarité du processus proposé et développons des algorithmes pour estimer un analogue à temps discret de notre modèle. Des applications sont également présentées.

Session 03H

Lundi 1 juin • Monday, June 1, 13:30 – 15:00

SCR 100

Inference, Design, and Applications

Inférence, planification et applications

[MS-70]

Mixed Effects Regression Trees for Clustered Data

Arbres de régression à effets mixtes pour données en grappes

François BELLAVANCE, Ahlem HAJJEM & Denis LAROCQUE, *HEC Montréal*

Previous works extending tree based methods to accommodate correlated data are mainly based on the multivariate repeated measures approach. We propose a mixed effects regression tree method which is more flexible because the observations are viewed as nested within clusters rather than as multivariate vectors. It can handle unbalanced clusters, allows observations within a cluster to be in different nodes, and can incorporate observation-level covariates and random effects. The method is implemented using a standard tree algorithm within

Les travaux de recherche antérieurs qui ont adapté les méthodes d'arbres de régression pour traiter les données corrélées sont principalement basés sur l'approche multivariée des mesures répétées. Nous proposons une méthode d'arbres de régression plus flexible qui s'inspire des modèles à effets mixtes. Les observations sont vues comme étant nichées à l'intérieur d'un groupe. Elle peut traiter un nombre inégal d'observations par groupe, permet à des observations d'un même groupe de se retrouver dans différents noeuds, et peut inclure des variables explicatives de différents niveaux pour les observations et des effets aléatoires. Elle est implémentée

the framework of the EM algorithm. Simulation results show that the proposed method can provide substantial improvements over standard tree algorithms.

tée en utilisant un algorithme standard d'arbres à l'intérieur du cadre bien connu de l'algorithme EM. Les résultats d'une étude de simulation montrent que la méthode proposée améliore substantiellement la performance des algorithmes standards d'arbre de régression.

[MS-71]

Break-point Detection for Dependence Functions

Détection de points de rupture pour fonctions de dépendance

Maryse CHAMPAGNE, *Université du Québec à Trois-Rivières*, Anne-Catherine FAVRE, *INRS-ETE Quebec*, Jean-François QUESSY, *Université du Québec à Trois-Rivières*

For many phenomena which imply several random variables, the underlying multivariate distribution can change over time. In this talk, a new method to detect such variations is presented. The approach suggested makes it possible to identify modifications in the dependence structure, i.e. the copula, independently of changes that could affect the margins. An asymptotically valid technique to compute p-values will be discussed and simulation results will show its accuracy in small samples. The conclusions of a power study and an illustration with hydrological data will be offered.

Pour certains phénomènes mettant en scène plusieurs variables aléatoires, la loi multivariée sous-jacente risque de changer au cours du temps. Dans cet exposé, il sera question d'une nouvelle méthode pour détecter de telles variations. L'approche proposée permet d'identifier des modifications dans la structure de dépendance, c'est-à-dire la copule, indépendamment de changements pouvant affecter les lois marginales. Une technique asymptotiquement valide pour l'obtention de valeurs-p sera introduite et des résultats de simulation vont montrer sa fiabilité pour de petits échantillons. Les conclusions d'une étude de puissance et une illustration sur des données hydrologiques seront offertes.

[MS-72]

Robust Regression Designs for Wavelet Approximation Models

Plans de régression robuste pour modèles d'approximation par ondelettes

Lin ZHAO & Xiaojian XU, *Brock University*

We discuss the construction of robust designs for wavelet regression models when the assumed models are possibly contaminated over two different neighbourhoods: G1, and G2. Our main findings are: (1) A recursive formula for constructing D-optimal designs under G1; (2) Equivalency of Q-optimal and A-optimal designs under G1 and G2; and (3) Analytic forms for A- and Q-optimal robust design densities under G2. Additionally, we have extended the results of others' study on homoscedastic wavelet models to their corresponding heteroscedastic models, under G2. Several examples are given for the comparison, and the results demonstrate that our designs are efficient.

Nous discutons de la construction de plans d'expérience robustes pour modèles de régression à ondelettes lorsqu'il y a possible contamination du modèle sur deux voisinages possibles G1 et G2. Nos principaux résultats sont : (1) une formule récursive pour construire des plans D-optimaux sous G1, (2) l'équivalence des plans Q-optimaux et A-optimaux sous G1 et G2 et (3) une forme analytique pour les densités de plans d'expériences robustes A- et Q-optimaux sous G2. De plus, nous étendons les résultats d'études antérieures sur les modèles à ondelettes homoscédastiques aux modèles hétéroscédastiques correspondants sous G2. Plusieurs exemples sont donnés à des fins de comparaisons et les résultats montrent que nos plans d'expérience sont efficaces.

[MS-73]**Model selection and Inference in the Multivariate Probit Model****Sélection de modèle et inférence sous le modèle probit multivarié**Aline TABET, Arnaud DOUCET, Paul GUSTAFSON & Kevin MURPHY, *University of British Columbia*

We propose an efficient parameter expansion for a data augmentation sampling algorithm to facilitate estimation in the multivariate probit class of model. Particularly, we show that given our choice of marginally uniform prior on the correlation matrix, we are able to accurately estimate the structure of the inverse correlation among the Gaussian latent variables, along with the other parameters. We consider the mathematical and computational challenges in extending the algorithm to accommodate multivariate longitudinal binary data, whereby one wishes to model two or more processes evolving together in time.

Nous proposons une expansion efficace des paramètres pour un algorithme d'augmentation de données facilitant l'estimation de modèles probit multivariés. En particulier, nous montrons que notre choix de lois a priori pour la matrice de corrélation marginalement uniformes nous permet d'estimer précisément la structure de la corrélation inverse parmi les variables latentes gaussiennes, ainsi que les autres paramètres. Nous considérons les défis mathématiques et informatiques dans l'extension de l'algorithme pour accommoder des données binaires longitudinales multivariées, où l'on souhaite modéliser au moins deux processus évoluant ensemble dans le temps.

[MS-74]**A Note on the Modified Gram-Schmidt Orthogonalization Process in Design and Analysis****Une note sur le processus d'orthogonalisation de Gram-Schmidt modifié en planification et analyse**Mehdi KIANI, *Athens University of Economics and Business*, Muhammad SALEEM, *GC University Faisalabad*

The orthogonalization of non-orthogonal experiments is presented to increase the statistical precision of the estimated regression coefficients. The purpose is to minimize the covariance and the bias of least squares estimators for estimating the path of steepest ascent that leads the users toward the neighbourhood of the optimum response. An orthogonal design is established to decrease the inverse determinant of $X'X$ and the angle between the true and the estimated steepest ascent paths. For orthogonalization of an un-designed matrix, our proposed solution is constructed on the modified Gram-Schmidt strategy relevant to the process of the Gaussian elimination.

Nous présentons l'orthogonalisation d'expériences non orthogonales dans le but d'augmenter la précision statistique des coefficients de régression estimés. Le but est de minimiser la covariance et le biais des estimateurs des moindres carrés pour l'estimation du chemin de plus grande ascension menant les utilisateurs vers le voisinage de la réponse optimale. Nous établissons un plan orthogonal diminuant le déterminant inverse de $X'X$ et l'angle entre les chemins d'ascension réel et estimé. Pour l'orthogonalisation d'une matrice non planifiée, la solution que nous proposons est construite à partir d'une modification de la stratégie de Gram-Schmidt relative au processus d'élimination gaussienne.

[MS-75]**Robust Functional Principal Components Analysis and its Application to Outlier Detection****L'analyse en composantes principales fonctionnelle robuste et ses applications en détection des données aberrantes**Liangliang WANG, Nancy HECKMAN & Matías SALIBIÁN-BARRERA, *University of British Columbia*

We propose robust functional principal components analyses for irregularly spaced functional data when the conditional distribution of the dependent variable, given the independent variable,

Nous proposons des analyses en composantes principales fonctionnelles et robustes pour données fonctionnelles à espacement irrégulier lorsque la distribution de la variable dépendante, conditionnelle à la variable indépendante, est

is skewed. Further, we propose several ways of detecting outliers, where an outlier could be an atypical trajectory or an unusual observation on an otherwise typical trajectory. The methods are motivated and illustrated by applications to detect outliers for meteorological data which are paired measurements of pressure and temperature, recorded while radiosondes rise through the earth's atmosphere. In our analysis, we view temperature as a function of pressure.

asymétrique. Nous proposons aussi plusieurs façons de détecter les observations aberrantes, où une observation aberrante pourrait être une trajectoire atypique ou une observation aberrante sur une trajectoire autrement typique. Ces méthodes sont motivées et illustrées par des applications à la détection d'observations aberrantes en données météorologiques qui sont des mesures paires de pression et de température, enregistrées alors qu'une radiosonde s'élève à travers l'atmosphère. Dans notre analyse, nous traitons la température comme une fonction de la pression.

Session 03I

Lundi 1 juin • Monday, June 1, 13:30 – 15:00

IRC Lobby

Case Study 1: The Effects of Climate on the Growth of Lodgepole Pine

Étude de cas 1 : Les effets du climat sur la croissance du pin de Murray

Case Study 1: The Effects of Climate on the Growth of Lodgepole Pine

Étude de cas 1 : Les effets du climat sur la croissance du pin de Murray

Alison GIBBS, *University of Toronto*

Lodgepole pine is an important commercial species in British Columbia, used for lumber, poles, railroad ties, furniture, cabinetry, and construction timbers. In order to predict how lodgepole pine trees will grow and accumulate wood over time, the use of high resolution satellite images is under consideration. However, for these images to be useful, the relationships between the crown properties and the wood properties must be understood. In this case study, students will examine the effects of climate variables, particularly temperature and precipitation, on these relationships.

Utilisé pour le bois d'œuvre, les poteaux, les traverses de chemin de fer, les meubles, l'ébénisterie et le bois de charpente, le pin de Murray est une essence commerciale importante en Colombie-Britannique. Il est envisagé d'utiliser des images satellite à haute résolution afin de prédire comment ces arbres pousseront et accumuleront du bois au fil du temps. Toutefois, pour que ces images soient utiles, il faut comprendre les relations entre les propriétés de la couronne et celles du bois. Dans cette étude de cas, les étudiants examineront les effets de variables climatiques, notamment la température et les précipitations, sur ces relations.

Session 03J

Lundi 1 juin • Monday, June 1, 13:30 – 15:00

IRC Lobby

Case Study 2: Proteomic Biomarkers for Disease Status

Étude de cas 2 : Biomarqueurs protéomiques et états pathologiques

Case Study 2: Proteomic Biomarkers for Disease Status

Étude de cas 2 : Biomarqueurs protéomiques et états pathologiques

Alison GIBBS, *University of Toronto*

Patients with a particular disease require urgent and aggressive treatment when their disease is in its active state, so it is important to know if their disease is active or inactive. The current method for determining the status of the disease is both invasive and expensive. Proteomics is the large-scale study of

Les patients souffrant d'une maladie particulière ont besoin d'un traitement urgent et agressif lorsque leur maladie est à l'état actif, si bien qu'il est important de savoir si leur maladie est active ou inactive. La méthode employée actuellement pour déterminer le statut de la maladie est à la fois agressive et coûteuse. La protéomique est l'étude à grande

proteins in blood. In this case study, students will examine whether the information from proteomic analysis of blood samples can be used to identify whether the samples are from active or inactive patients.

échelle des protéines sanguines. Dans cette étude de cas, les étudiants chercheront à savoir si les résultats d'une analyse protéomique d'échantillons sanguins permettent de déterminer si ces échantillons proviennent de patients dont la maladie est à l'état actif ou inactif.

Session 04A

Lundi 1 juin • Monday, June 1, 15:30 – 17:00

IRC 6

Isobel Loutit Invited Address on Business and Industrial Statistics

Allocution Isobel Loutit sur la statistique industrielle et de gestion

[MS-78]

The State of Statistical Process Control: An Update

L'état actuel du contrôle statistique du processus : une mise à jour

William WOODALL, *Virginia Tech*

In this non-mathematical overview presentation, applications of profile monitoring and health-related surveillance will be described. These are two of the most exciting applications of control charts. After a brief overview of process monitoring in general, some profile monitoring applications will be given. In profile monitoring, changes over time in a function are to be detected. A number of applications of health related quality monitoring will be presented. Some of the differences between the industrial process monitoring and health-related surveillance environments are to be discussed. Other areas of research on control charting will be briefly summarized with some key references provided.

Dans cette présentation non mathématique, des applications de la surveillance des profils et de la surveillance liée à la santé seront décrites. Ce sont deux des plus intéressantes des applications des cartes de contrôle. Après une brève vue d'ensemble de la surveillance des profils en général, quelques-unes de leurs applications seront présentées. En surveillance des profils, la modification d'une fonction avec le temps doit être détectée. Certaines applications de la surveillance de la qualité liée à la santé seront présentées. Nous discuterons de certaines des différences entre la surveillance des procédés industriels et la surveillance liée à la santé. D'autres domaines de recherche en cartes de contrôle seront brièvement résumés à l'aide de quelques références-clefs.

Session 04B

Lundi 1 juin • Monday, June 1, 15:30 – 17:00

IRC 1

Complex Methods for Longitudinal and Survival Data

Méthodes complexes pour les données longitudinales et de survie

[MS-79]

Bayesian Joint Semi-parametric Modeling of Longitudinal- Survival Data: Is Joint Modeling Really Necessary?

Modélisation bayésienne conjointe semi-paramétrique de données longitudinales de survie : la modélisation conjointe est-elle vraiment nécessaire ?

Wesley JOHNSON, *UC Irvine*, Timothy HANSON, *University of Minnesota*, Adam BRANSCUM, *University of Kentucky*

We discuss joint modeling of survival data with time dependent covariates (TDC). Historically, TDCs were regarded as fixed rather than modeled. We discuss the problem assuming the need for joint modeling and allow baseline distributions

Nous discutons de la modélisation conjointe de données de survie avec covariables dépendantes du temps (CDT). Historiquement, les CDT étaient considérées comme fixes et non modélisées. Nous discutons du problème en supposant la nécessité d'une modélisation conjointe et en permettant aux

to be drawn from three well-known broad families of survival distributions. We compare models using Mediterranean fruit fly data that were analyzed in the literature using a Cox and Oakes model. Our results indicate that the proportional odds and Cox TDC models were preferable based on a prediction based model fit criterion, and our results indicate that joint modeling is overkill for these particular data.

distributions de base d'être tirées parmi trois grandes classes bien connue de distributions de survie. Nous comparons les modèles en utilisant des données de mouches à fruits méditerranéennes analysées dans la littérature avec le modèle Cox et Oakes. Nos résultats indiquent que le modèle à cotes proportionnelles et le modèle CDT de Cox étaient préférables d'après un critère d'ajustement de modèle basé sur la prévision, et nos résultats indiquent que la modélisation conjointe est excessive pour ces données particulières.

[MS-80]

Conditional GEE for Recurrent Event Gap Time Data

ÉEG conditionnelles pour données de temps d'écarts entre événements récurrents

Robert STRAWDERMAN & David CLEMENT, *Cornell University*

This talk considers the analysis of recurrent event data subject to censored observation. Using a suitable adaptation of generalized estimating equations for longitudinal data, we propose a useful methodology for estimating the parameters indexing the conditional means and variances of the process inter-event (i.e. gap) times that accomodates censoring, transformations of gap times, and both time-fixed and time-varying covariates. The methods are used to analyze data from a randomized trial of asthma prevention in young children.

Cet exposé considère l'analyse de données d'événements récurrents sujettes à censure. Utilisant une adaptation adéquate d'équations d'estimation généralisées pour données longitudinales, nous proposons une méthodologie utile pour l'estimation des paramètres indexant les moyennes et variances conditionnelles de temps d'écarts entre événements tenant compte de la censure, de transformations des temps d'écarts et de covariables variant et ne variant pas avec le temps. Ces méthodes sont utilisées pour analyser des données d'essai randomisé en prévention de l'asthme chez les jeunes enfants.

[MS-81]

Analysis of Longitudinal Data with Error in Covariates and Missingness in Responses under Generalized Linear Mixed Models

Aanalyse de données longitudinales avec erreurs dans les covariables et réponses manquantes sous modèles linéaires mixtes généralisés

Grace YI, *University of Waterloo*, Wei LIU, *York University*, Lang WU, *University of British Columbia*

Longitudinal data arise commonly in medical studies and it is common practice to analyze such data with generalized linear mixed models. It is known that results based on standard methods can be seriously biased if incompleteness or measurement error is present. There has been considerable interest in accommodating either incompleteness or measurement error. However, relatively little attention has been directed to concerning both features simultaneously. It is not clear how measurement error and incompleteness may interact and influence inference. In this talk we will investigate the asymptotic biases and develop valid methods to handle longitudinal data with both features.

Les données longitudinales sont fréquentes en études médicales et il est d'usage courant d'analyser de telles données au moyen de modèles linéaires généralisés mixtes. En présence de mesures manquantes ou comportant des erreurs, les résultats basés sur les méthodes habituelles peuvent être fortement biaisés. Il y a eu un intérêt considérable dans les façons de traiter les mesures manquantes ou erronées. Cependant, relativement peu d'attention a été portée au traitement simultané des deux cas. La façon dont les mesures erronées et manquantes peuvent interagir et influencer l'inférence n'est pas claire. Dans cet exposé, nous allons étudier les biais asymptotiques et développer des méthodes valides pouvant traiter des données longitudinales présentant ces deux cas.

[MS-82]**Joint Analysis of Longitudinal Growth and Interval Censored Mortality Data****Analyse conjointe de croissance longitudinale et de données de mortalité censurées par intervalles**Terry LEE, Charmaine DEAN & Leilei ZENG, *Simon Fraser University*

Joint analysis of longitudinal and survival data has received considerable attention. We develop a joint model for the analysis of longitudinal data monitoring the growth and survival of trees. The challenge is to handle features of the data which are not common in considerations of joint analyses. Two main features are interval censoring of the survival response and the need to account for spatial effects jointly in the survival and longitudinal outcomes. We adopt linkages in random effects over multiple outcomes as in Chi and Ibrahim (2006), and propose to use functional approach to smooth the data to handle interval censoring.

L'analyse conjointe de données longitudinales et de survie a reçu beaucoup d'attention. Nous développons un modèle conjoint pour l'analyse de données longitudinales de croissance et de survie des arbres. Le défi est de traiter certaines caractéristiques des données dont on ne tient pas souvent compte en analyse conjointe. Deux principales caractéristiques sont la censure par intervalles de la survie et le besoin de tenir compte d'effets spatiaux conjointement dans les réponses longitudinales et de survie. Nous utilisons des liens dans les effets aléatoires sur de multiples réponses comme dans Chi et Ibrahim (2006), et proposons l'utilisation de l'approche fonctionnelle pour lisser les données afin de traiter la censure par intervalle.

Session 04C**Lundi 1 juin • Monday, June 1, 15:30 – 17:00****IRC 5****Memorial Session for Randy Sitter****Séance commémorative en la mémoire de Randy Sitter****[MS-83]****Using Folded Over Designs to Consider Interactions****Utiliser des plans d'expérience avec repli pour prendre en considération les interactions**Arden MILLER, *University of Auckland*

Randy Sitter and I collaborated on research into screening experiments where the primary goal is the identification of important main effects and a secondary goal is to consider the possibility that a small number of two-factor interactions are also important. This led us to consider the use of non-standard folded over designs which involve complex aliasing. This talk gives an overview of this work and summarizes the key findings.

Randy Sitter et moi avons collaboré sur des travaux en expériences de dépistage où le but principal est l'identification d'effets principaux importants et où un but secondaire est de considérer la possibilité qu'un petit nombre d'interactions à deux facteurs soient également importantes. Ceci nous a menés à considérer l'utilisation de plans repliés inhabituels à crénelage complexe. Cet exposé fait la revue de ces travaux et en résume les résultats clés.

[MS-84]**Impact of Randy Sitter's Contributions to Sample Survey Theory and Practice****Impact des contributions de Randy Sitter sur la théorie et la pratique des enquêtes par sondage**J.N.K. RAO, *Carleton University*

Randy Sitter has made fundamental contributions to the theory and practice of sample surveys. In particular, he has developed ingenious re-sampling methods for variance estimation from complex survey data including imputation for missing data,

Randy Sitter a apporté des contributions fondamentales à la théorie et à la pratique d'enquêtes par sondage. En particulier, il a développé des méthodes de rééchantillonnage ingénieuses pour estimer la variance de données d'enquêtes complexes, incluant l'imputation de données manquantes,

pseudo-empirical likelihood methods that take account of survey design, model-assisted inference and minimizing disclosure risk associated with a replicate weights data file. In this talk, I will try to highlight some of Randy Sitter's major contributions to those areas in survey sampling.

des méthodes de pseudo-vraisemblance empirique tenant compte du plan d'enquête, l'inférence assistée par un modèle et la minimisation de risques de divulgation associés à un fichier des données contenant les poids de rééchantillonnage. Dans cet exposé, je vais tenter de souligner certaines des principales contributions de Randy Sitter à ces domaines des enquêtes par sondage.

[MS-85]

Randy Sitter: His Career, His Humour and His Many Accomplishments

Randy Sitter : sa carrière, son humour et ses nombreuses réalisations

Richard LOCKHART, Derek BINGHAM & Charmaine DEAN, *Simon Fraser University*

Randy Sitter was a Canadian statistical star and an international leader in sampling and industrial statistics. Randy was a pioneer in the development of re-sampling methods for variance estimation in complex surveys. An expert in design of experiments, he developed theory and algorithms to find and analyze many important designs. Randy supervised over 20 graduate students with 6 PhD students taking academic positions at prestigious institutions. We will highlight Randy's research interests and talk about his superb career and generous, passionate and fun-loving nature. SFU's Statistics and Actuarial Science Department welcomes conference participants to an informal reception following the talk.

Randy Sitter était une star dans le monde statistique canadien et un leader international en échantillonnage et en statistique industrielle. Randy était un pionnier dans le développement de méthodes de rééchantillonnage pour l'estimation de la variance en enquêtes complexes. Un expert en planification d'expériences, il a développé une théorie et des algorithmes pour trouver et analyser plusieurs plans expérimentaux importants. Randy a supervisé plus de 20 étudiants de cycles supérieurs, dont 6 étudiants au doctorat qui occuperont par la suite des positions académiques dans des institutions prestigieuses. Nous soulignerons les intérêts de recherche de Randy et parlerons de sa superbe carrière et de sa nature généreuse, passionnée et enjouée. Le département de statistique et d'actuariat de l'Université Simon Fraser invite les participants à la conférence à une réception informelle après la présentation.

Session 04D

Lundi 1 juin • Monday, June 1, 15:30 – 17:00

MSL 102

Runs, Patterns and the Finite Markov Chain Imbedding Technique

Courses, schémas et la technique d'insertion de la chaîne de Markov finie

[MS-86]

Nonparametric Control Charts Based on Run Statistics

Cartes de contrôle non paramétriques basée sur des statistiques de séries

Markos KOUTRAS, *University of Piraeus, Greece*, N. BALAKRISHNAN, *McMaster University*, I TRIANTAFYLLOU, *University of Piraeus, Greece*

In the present work, we introduce three new distribution-free Shewhart-type control charts. They use order statistics of a reference sample to establish appropriate control limits. They then exploit run-based statistics defined through the test sample observations that lie between the control limits to decide whether the process is in-control or not. Exact formulas are developed for the in-control distri-

Dans le présent travail, nous présentons trois nouvelles cartes de contrôle libres de distribution de type Shewhart. Elles utilisent les statistiques d'ordre d'un échantillon de référence afin d'établir des limites de contrôle appropriées. Elles exploitent ensuite des statistiques basées sur des séries définies grâce aux observations de l'échantillon test entre les limites de contrôle afin de décider si le processus est sous contrôle ou non. Les formules exactes de la distribution sous

bution, the false alarm rate and the exact run length distribution. An explicit formula is provided for the alarm rate under Lehmann-type out-of-control alternatives. Several numerical computations are presented which display the efficacy of the new charts and their robustness features.

contrôle, des taux de fausse alerte et de la distribution de la longueur exacte des séries sont obtenues. Une formule explicite est fournie pour le taux d'alerte sous les alternatives hors contrôle de type Lehmann. Plusieurs calculs numériques démontrant l'efficacité et la robustesse des nouvelles cartes sont présentés.

[MS-87]

Dealing With Patterns in Degenerated Sequences

Méthodes pour l'étude de motifs dans des séquences dégénérées

Gregory NUEL, *MAP5, CNRS, University Paris Descartes*

It is quite common to encounter sequences defined on a degenerated alphabet: some positions may be ambiguous, allowing a specific set of letters to appear instead of a single one. A example of this kind is provided by the DNA sequences and the well-known degenerated IUPAC alphabet. Assuming that the (unambiguous) underlying sequence is generated according to a Markov model, we show that the posterior distribution of the sequence given the observation is a heterogeneous Markov chain which can then be used to study patterns occurrences.

Il est fréquent de rencontrer des séquences définies sur un alphabet dégénéré : certaines positions peuvent être ambiguës, permettant ainsi à un sous-ensemble de lettres d'apparaître, au lieu d'une seule. Un exemple typique est celui des séquences d'ADN et de l'alphabet dégénéré bien connu IUPAC. En supposant que la séquence (non ambiguë) sous-jacente est générée selon un modèle Markovien, nous montrons que la distribution a posteriori de la séquence, conditionnellement à l'observation, est une chaîne de Markov hétérogène qui peut alors être utilisée pour étudier des occurrences de motifs.

[MS-88]

Some Applications of FMCI in Segmentation

Quelques applications de l'ECMF en segmentation

John ASTON, *CRISM, Warwick University, UK*, Jyh-Ying PENG, *Academia Sinica*

Finite Markov Chain Imbedding is a powerful technique for investigating run and pattern distributions. However, its use in data analysis is usually restricted to estimating distributions against which a dataset is to be evaluated. Here, a different approach is taken where we investigate the distributions within the dataset itself. This will allow the quantification of distributions associated with segments in fairly general time series models. Using FMCI techniques and posterior Markov chains, it will be possible to quantify uncertainty in segment starts, the number of segments, and the length of segments. Applications in biology and medicine will be considered.

L'emboîtement de chaînes de Markov finies est une technique utile pour étudier la distribution de séquences et de motifs. Cependant, son utilisation en analyse de données est généralement restreinte à l'estimation de distributions auxquelles un jeu de données doit être confronté. Ici, une approche différente est adoptée : nous étudions les distributions à l'intérieur du jeu de données. Ceci permettra la quantification de distributions associées à des segments dans des modèles de séries chronologiques assez généraux. Au moyen d'ECMF et de chaînes de Markov a posteriori, il sera possible de quantifier l'incertitude par rapport aux points de départ des segments, au nombre de segments et à leur longueur. Des applications en médecine et en biologie seront considérées.

Session 04E**Lundi 1 juin • Monday, June 1, 15:30 – 17:00****IRC 3****Some Recent Developments on Response Adaptive Designs of Clinical Trials****Évolution récente des plans d'essais cliniques adaptés à la réponse****[MS-89]****Adaptive Treatment Allocation Procedures With Application to Phase I/II Clinical Trials and Toxicology Studies.****Procédures adaptatives d'assignation de traitement avec application à des essais cliniques de phase I/II et à des études de toxicologie**Nancy FLOURNOY, Ping YAO & Seung Won HYUN, *University of Missouri*

Assume dose-response functions are unimodal. An intuitively appealing procedure is to treat next, with best intentions, to maximize estimated patient gain. We discuss this procedure and alternatives in the context of phase I/II trials. In the context of toxicology studies, researchers experiment to estimate the peak dose. Considering this estimate fixed, they run additional trials to estimate the EC50. We find the variance of the EC50 recognizing peak dose is random and obtain designs that minimize this variance. These designs cannot be implemented because they are functions of unknown parameters. However, they are approximated by sequentially estimating optimal designs.

Supposons l'unimodalité des fonctions dose-réponse. Une procédure intuitivement attirante est de traiter ensuite, avec les meilleures intentions, de façon à maximiser le gain estimé des patients. Nous discutons de cette procédure et d'alternatives dans le contexte d'essais de phase I/II. Dans le contexte d'études de toxicologie, les chercheurs expérimentent afin de réduire la dose maximale. En considérant l'estimé fixe, ils effectuent des tests supplémentaires afin d'estimer la CE50. Nous trouvons la variance de la CE50 en reconnaissant que la dose maximale est aléatoire et obtenons des plans d'expérience minimisant cette variance. Ces plans ne peuvent être mis en oeuvre, car ils sont fonctions de paramètres inconnus. Cependant, ils sont approximatés par estimation séquentielle de plans optimaux.

[MS-90]**Bayesian Adaptive Clinical Trial Designs****Planification bayésienne d'essais cliniques adaptatifs**Yi CHENG, *Indiana University South Bend*

A clinical trial requires decisions about how an experiment will be conducted. The process is often dynamic and a Bayesian formulation provides a mechanism for running such a process. The decision theoretical approach considers the consequences of possible design parameter values and selects one that maximizes the utility. Decisions depend on the goals of the experiment and choices may be restricted by available resources and ethical considerations. Bayesian designs also have the flexibility of allowing interpretation of the final results along frequentist lines. The frequentist error rates can be established so that the desired properties can be maintained for regulatory settings.

Une étude clinique requiert une prise de décision relative à la façon dont l'expérience sera menée. Ce processus est souvent dynamique et une formulation bayésienne apporte un mécanisme pour gérer un tel processus. L'approche de la théorie de la décision considère les conséquences des différentes valeurs possibles pour les paramètres de planification et choisit celle maximisant l'utilité. Les décisions dépendent des buts de l'expérience et peuvent être restreintes par les ressources disponibles et par des considérations éthiques. La planification bayésienne a aussi la flexibilité de permettre l'interprétation des résultats finaux dans une optique fréquentiste. Il est possible de fixer les taux d'erreurs fréquentistes de façon à ce que les propriétés désirées puissent être conservées dans un cadre réglementaire.

[MS-91]**Optimizing Tradeoffs Between Randomization and Classical Objectives****Optimisation de compromis entre objectifs classiques et de randomisation**Janis HARDWICK & Quentin STOUT, *University of Michigan*

In many response adaptive clinical trials designed to optimize a given objective, each treatment decision is completely determined once one is given the design parameters and the observations to date. Some have suggested that this deterministic behavior renders optimal designs unsuitable for certain applications.

Here we introduce controlled randomization that allows researchers to specify a precise tradeoff between their randomization and optimization goals for a wide variety of settings. We present examples of formally defined randomization goals and designs that incur relatively little loss of the optimization objective. We also compare to randomized optimal designs and various suboptimal designs.

Dans plusieurs essais cliniques adaptatifs planifiés afin d'optimiser un objectif donné, chaque décision de traitement est complètement déterminée une fois obtenus les paramètres du plan d'expérience et les observations à jour. Il a été suggéré que ce comportement déterministe rend les plans optimaux non convenables pour certaines applications.

Ici, nous présentons une randomisation contrôlée permettant aux chercheurs de spécifier un compromis précis entre leurs objectifs de randomisation et d'optimalité pour une grande variété de conditions. Nous présentons des exemples d'objectifs et de plans de randomisation formellement définis entraînant une perte relativement faible au niveau de l'objectif d'optimalité. Nous faisons une comparaison aux plans randomisés optimaux et à divers plans suboptimaux.

[MS-92]**Asymptotic inference of Response Adaptive Clinical Trials****Inférence asymptotique pour essais cliniques adaptatifs en fonction de la réponse**Yanqing YI, *Memorial University of Newfoundland*

We explore the issue of asymptotic inference of response adaptive clinical trials. Asymptotic properties of the likelihood ratio and goodness-of-fit tests are discussed, together with the consistency, asymptotic normality and asymptotic efficiency of the maximum likelihood estimators.

Nous explorons le problème de l'inférence asymptotique pour essais cliniques adaptatifs en fonction de la réponse. Nous discutons des propriétés asymptotiques des tests du rapport de vraisemblance et d'adéquation, ainsi que de la cohérence, normalité asymptotique et efficacité asymptotique des estimateurs du maximum de vraisemblance.

Session 04F**Lundi 1 juin • Monday, June 1, 15:30 – 17:00****FNH 60****Statistical Challenges in Clinical Research: CANNeCTIN Research Program****Défis statistiques en recherche clinique : le programme de recherche CANNeCTIN****[MS-93]****Estimating Mediation Effects in Randomized Controlled Trials with a Failure Time Response****Évaluer les effets de médiation en essais comparatifs randomisés avec données de temps de défaillance**Meaghan CUERDEN & Amit GARG, *University of Western Ontario*, Richard COOK, *University of Waterloo*

Mediation occurs when an independent variable is associated with an intermediate response, or mediator, which in turn is predictive of the outcome. Estimating the mediation effect when the outcome is continuous or dichotomous is well described; es-

De la médiation se produit lorsqu'une variable indépendante est associée à une réponse intermédiaire, ou médiatrice, qui est elle-même un prédicteur de la réponse. L'estimation de l'effet de médiation lorsque la réponse est continue ou dichotomique est bien décrite ; l'estimation de cet effet a été peu

timation of this effect has been less studied for the case in which the outcome is a failure time. We propose a statistical test to assess mediation effects for failure time responses in randomized controlled trials. We apply this test in a trial of perioperative medicine, estimating the effect of the mediator "acute kidney injury" on the outcome of time to chronic kidney disease.

étudiée pour le cas où la réponse est un temps de défaillance. Nous proposons un test statistique pour évaluer les effets de médiation pour temps de défaillance en essais comparatifs randomisés. Nous appliquons ce test à une étude en médecine périopératoire, estimant l'effet du médiateur «blessure aiguë au rein» sur la réponse «temps de maladie rénale chronique».

[MS-94]

Dealing with Stratification and Clustering in Randomized Controlled Trials with a Continuous Outcome Traiter la stratification et la segmentation en essais comparatifs randomisés avec réponse continue

Rong CHU, Jinhui MA, Lehana THABANE, Anne HOLBROOK & Eleanor PULLENAYEGUM, *McMaster University*,
COMPETE GROUP

Stratified randomization is used in clinical trials to achieve balance of important baseline covariates between the experimental and control groups. Hospital, centre, family physician or practice are common stratification variables to reduce outcome variability, accounting for differences in environment, social, key demographic and prognostic factors in the study sample. Correlation among observations within stratum sometimes arises and may increase outcome variability among strata. We compare several statistical approaches to analyzing continuous outcomes in trials where there is high clustering within strata. Performance of cluster level analysis, individual level analysis (fixed or random effects) will be investigated based on simulated datasets.

La randomisation stratifiée est utilisée en essais cliniques pour atteindre un équilibre dans certaines covariables de base importantes entre les groupes contrôle et expérimental. L'hôpital, le centre, le médecin ou la pratique de famille sont des variables de stratification fréquentes pour réduire la variabilité des réponses, tenir compte des différences dans les facteurs environnementaux, sociaux, démographiques et pronostics dans l'échantillon. Une corrélation parmi les observations à l'intérieur des strates survient parfois et peut alors augmenter la variabilité de la réponse dans les strates. La performance d'analyse au niveau des grappes et des individus (effets fixes ou aléatoires) sera étudiée à partir de jeux de données simulés.

[MS-95]

Analysis of Multi-state Data with Missing Covariates Analyse de données à multi-états avec covariables manquantes

Audrey BORUVKA & Richard COOK, *University of Waterloo*

Missing covariates is a common problem that can hinder efficient use of data and generate biased estimators. Methods to address missing covariates in event history data have been primarily limited to survival analysis. In this paper we consider the use of likelihood-based approaches and weighted estimating functions to handle missing covariates in multi-state models. Performance of our proposed estimators is assessed using a simulation study involving a simple illness death model. Possible extensions to more general state structures are discussed.

La présence de covariables manquantes est un problème fréquent nuisant à l'utilisation efficace de donnée et générant des estimateurs biaisés. Les méthodes traitant des données manquantes en données d'histoire événementielle ont été surtout limitées à l'analyse de survie. Dans cet article nous considérons l'utilisation d'approches basées sur la vraisemblance et les fonctions d'estimation pondérées pour traiter les covariables manquantes dans des modèles multi-états. La performance des estimateurs proposés est évaluée par une étude de simulation impliquant un simple modèle de mortalité causée par la maladie. De possibles extensions à des structures d'états plus générales sont aussi abordées.

[MS-96]**Comparing the Efficiency of Different Statistical Methods in the Analysis of Binary Outcomes from Cluster Randomized Controlled Trials: A Simulation Study****Comparer l'efficacité de différentes méthodes statistiques pour l'analyse de réponses binaires d'essais contrôlés randomisés par grappes : une étude de simulation**Jinhui MA & Lehana THABANE, *McMaster University*

Cluster randomized trials (CRTs) are increasingly used in the assessment of the effectiveness of interventions to improve health outcomes or prevent diseases. The units of randomization for such trials are groups or clusters such as family practices, families, hospitals, or entire communities rather than individuals themselves to minimize the potential contaminations. CRTs may lead to substantially reduced statistical efficiency compared to trials that randomize the same number of individuals. We will discuss several cluster and individual level statistical methods to analyze binary outcomes from CRTs and compare their efficiency based on the simulation study.

Les essais randomisés par grappes (ERG) sont de plus en plus utilisés dans l'évaluation de l'efficacité d'interventions pour améliorer les résultats de santé ou prévenir les maladies. Les unités de randomisation pour de tels essais sont des groupes ou grappes tels les médecins de famille, les familles, hôpitaux ou communautés entières plutôt que les individus eux-mêmes, afin de minimiser les contaminations potentielles. Les ERG peuvent mener à une efficacité statistique substantiellement plus faible que des essais randomisant le même nombre d'individus. Nous allons discuter de plusieurs méthodes statistiques au niveau des grappes et des individus pour analyse des réponses binaires d'ERG et comparer leur efficacité en nous basant sur une étude de simulation.

[MS-97]**Statistical Challenges Faced by the Population Health Research Institute (PHRI) in Conducting Large Simple Trials and Epidemiologic Studies****Défis statistiques rencontrés par l'Institut de recherche en santé de la population (PHRI) en essais simples à grande échelle et en études épidémiologiques**Janice POGUE, ON BEHALF OF THE PHRI STATISTICS GROUP & CANNeCTIN (CANADIAN NETWORK AND CENTRE FOR TRIALS INTERNATIONALLY), *McMaster University*

Innovations in biostatistical methods have played a major role in advancing health research, but challenges to remain. PHRI has faced a number of such challenges over its 16 years conducting research in cardiovascular and related diseases. Issues include: struggling to integrate multiple outcomes/comparisons within a single study, in order to understand the totality of the data, without inflating type I error rate; non-inferiority trials complexities and their margins, interim analyses, interpretation, and direct comparisons vs. placebo; and others. CANNeCTIN seeks to bring together biostatisticians and researchers from across Canada to address complex biostatistical issues and provide practical and interactive training.

Les innovations en méthodes biostatistiques ont joué un rôle majeur en recherche en santé, mais plusieurs défis restent à résoudre. Le PHRI a fait face à certains de ces défis pendant ses 16 ans de recherches en maladies cardiovasculaires. Les défis comprennent : l'intégration de plusieurs réponses ou comparaisons dans une même étude, de façon à comprendre la totalité des données sans gonfler le taux d'erreur de type I ; les complexités des tests de non-infériorité de leurs marges, analyses intérimaires, interprétation et comparaisons directes avec placebo ; et autres. Le CANNeCTIN vise à rapprocher biostatisticiens et chercheurs du Canada pour répondre à ces questions biostatistiques complexes et fournir un apprentissage pratique et interactif.

Session 04G**Lundi 1 juin • Monday, June 1, 15:30 – 17:00****CHE 150**

Monte Carlo Markov Chains and Markov Models

Chaînes de Markov Monte-Carlo et modèles markoviens

[MS-98]**On the Containment Condition for Adaptive Markov Chain Monte Carlo Algorithms****Sur la condition de confinement pour algorithmes adaptatifs à chaînes de Markov Monte-Carlo**Yan BAI, *University of Toronto*, G.O. ROBERTS, *University of Warwick*, J.S. ROSENTHAL, *University of Toronto*

This paper considers ergodicity properties of certain adaptive Markov chain Monte Carlo (MCMC) algorithms for multidimensional target distributions, in particular Adaptive Metropolis and Adaptive Metropolis-within-Gibbs. It was previously shown (Roberts and Rosenthal) that Diminishing Adaptation and Containment imply ergodicity of adaptive MCMC. We derive various sufficient conditions to ensure Containment, and connect the convergence rates of algorithms with the tail properties of the corresponding target distributions. An example is given to show that Diminishing Adaptation alone does not imply ergodicity. We also present a Summable Adaptive Condition which, when satisfied, proves ergodicity more easily.

Cet article considère les propriétés d'ergodicité de certains algorithmes MCMC adaptatifs pour distributions cibles multidimensionnelles, en particulier l'algorithme de Metropolis adaptatif et l'algorithme de Metropolis-dans-Gibbs adaptatif. Il a déjà été démontré (Roberts et Rosenthal) que l'adaptation décroissante et le confinement impliquent l'ergodicité de MCMC adaptatif. Nous obtenons des conditions suffisantes variées pour garantir le confinement, et lions le taux de convergence d'algorithmes aux propriétés des queues des distributions cibles correspondantes. Un exemple est donné afin de montrer que l'adaptation décroissante à elle seule n'implique pas l'ergodicité. Nous présentons aussi une condition adaptative sommable qui, lorsque satisfaite, permet de prouver plus aisément l'ergodicité.

[MS-99]**Particle Markov Chain Monte Carlo****Chaînes de Markov Monte-Carlo à particules**Roman HOLENSTEIN, *University of British Columbia*, Arnaud DOUCET, *University of British Columbia / The Institute of Statistical Mathematics, Tokyo, Japan*, Christophe ANDRIEU, *University of Bristol, UK*

MCMC and SMC methods have emerged as the two main tools to sample from high-dimensional probability distributions. Although asymptotic convergence of MCMC algorithms is ensured under weak assumptions, the performance of these latter is unreliable when the proposal distributions used to explore the space are poorly chosen and/or if highly correlated variables are updated independently. We propose a new Monte Carlo framework in which we build efficient high-dimensional proposal distributions using SMC methods. This allows us to design effective MCMC algorithms in complex scenarios where standard strategies fail.

Les méthodes MCMC et SMC sont devenues les deux outils principaux pour échantillonner de distributions de dimension élevée. Bien que la convergence asymptotique des algorithmes MCMC soit assurée sous des hypothèses faibles, leur performance est peu fiable lorsque les distributions instrumentales utilisées pour explorer l'espace sont mal choisies, ou si des variables fortement corrélées sont mises à jour de façon indépendante. Nous proposons un nouveau cadre Monte-Carlo dans lequel nous construisons des distributions instrumentales de dimension élevée efficaces au moyen de méthodes SMC. Ceci nous permet d'élaborer des algorithmes MCMC dans des scénarios complexes où les stratégies habituelles ne peuvent être appliquées.

[MS-100] **r^{th} Order Categorical Markov Chains****Chaînes de Markov catégoriques du r^{e} ordre**Reza HOSSEINI, James ZIDEK & Nhu LE, *University of British Columbia*

We prove a representation theorem for a r -th order categorical Markov chain $\{X_t\}$, where t denotes time. We express the conditional probability $P(X_t|X_{t-1}, \dots, X_{t-r})$ as a linear combination of the monomials of past processes X_{t-1}, \dots, X_{t-r} . Simulations show that the partial likelihood estimation and the representation together give us satisfactory results. We also check the performance of "BIC" to pick optimal models. An advantage of this model to the pre-existing models is the capacity to accommodate other continuous variables as linear terms. For example, we can add some seasonal processes to get a non-stationary chain.

Nous prouvons un théorème de représentation d'une chaîne de Markov catégorique du r -ième ordre $\{X_t\}$, où t désigne le temps. Nous exprimons la probabilité conditionnelle $P(X_t|X_{t-1}, \dots, X_{t-r})$ comme une combinaison linéaire de monômes de processus passés X_{t-1}, \dots, X_{t-r} . Les simulations montrent que l'estimation de la vraisemblance partielle et la représentation suggérée produisent des résultats satisfaisants. Nous vérifions aussi la performance du «BIC» pour choisir des modèles optimaux. L'un des avantages de ce modèle comparé aux modèles existants est la capacité d'ajouter d'autres variables continues comme termes linéaires. Par exemple, nous pouvons ajouter certains processus saisonniers pour obtenir une chaîne non stationnaire.

[MS-101]**Reptation Quantum Monte Carlo Algorithm****Algorithme Monte-Carlo quantique à reptation**Wai Kong YUEN, *Brock University*

The recent development of the reptation quantum Monte Carlo algorithm (RQMCA) by Baroni and Moroni (1999) has been promising in quantum chemistry. Mathematically, RQMCA simulates a Markov chain that converges to a high dimensional target distribution, which models the electron configurations of a quantum system, with a transition kernel motivated by its dynamical properties. In this presentation, we review the mathematical foundations of RQMCA by comparing it with a Metropolis-Hastings algorithm and discuss how to improve its efficiency through the relaxation of some assumptions. We report significant improvements in the estimations of some important physical quantities in various systems.

Le développement récent de l'algorithme Monte-Carlo quantique à reptation (AMCQR) par Baroni et Moroni (1999) fut prometteur en chimie quantique. Mathématiquement, l'AMCQR simule une chaîne de Markov convergeant vers une distribution cible de dimension élevée, modélisant la configuration des électrons d'un système quantique, avec un noyau de transition motivé par ses propriétés dynamiques. Dans cette présentation, nous faisons la revue des fondations mathématiques de l'AMCQR en le comparant à un algorithme de type Metropolis-Hastings et discutons de possibles améliorations à son efficacité par un assouplissement de certaines hypothèses. Nous obtenons des améliorations significatives dans l'estimation de certaines quantités physiques importantes dans divers systèmes.

[MS-102]**On Markov Chain Monte Carlo Simulation for High-Frequency Financial Data****Simulation par chaînes de Markov Monte-Carlo pour données financières de haute fréquence**Amaan MEHRABIAN & David STEPHENS, *McGill University*

This paper presents a Bayesian analysis approach to state-space models. In particular, the problem of calibrating Stochastic Volatility (SV) models to high-frequency financial data is studied.

Cet article présente une approche bayésienne pour l'analyse de modèles à espaces d'états. En particulier, le problème de la calibration de modèles de volatilité stochastique (VS) pour des données financières de haute fréquence est étudié. Nous

We demonstrate a comprehensive Markov Chain Monte Carlo (MCMC) method for this purpose and address different issues raised by high-dimensional state-space model. Specifically, we show how changing the observation frequency through data aggregation can affect the model parameters. We then study the expected auxiliary variable method for MCMC and illustrate how this method helps to construct Markov chains which mix faster and are easier to simulate in complex scenarios.

montrons une méthode MCMC à cette fin et répondons à plusieurs questions soulevées par le modèle à espaces d'états de dimension élevée. Plus particulièrement, nous montrons comment un changement des fréquences d'observation par agrégation des données peut affecter les paramètres du modèle. Nous étudions alors la méthode de la variable auxiliaire espérée pour MCMC et illustrons comment cette méthode aide à la construction de chaînes de Markov se mélangeant plus rapidement et plus faciles à simuler sous scénarios complexes.

[MS-103]

Priority Modelling of Emergency Department Wait Times by Finite Markov Chain Imbedding

Modélisation de priorité des temps d'attente aux urgences par emboîtement de chaînes de Markov finies

Hsing-Ming CHANG, *University of Manitoba*

Knowing only the average rate of emergency department (ED) volume per time interval, the average time spent in ED from initial physician assessment or treatment until the patient's visit is complete, the maximum number of patients admissible to a single ED, and assuming at any given time only a single doctor attending multiple emergency patients is necessary, Finite Markov Chain Imbedding technique can be applied to obtain wait-time distributions and estimating expected wait-times of each of the Canadian Triage & Acuity Scale level. An analysis of a wait-time dataset which was provided by a hospital in Taiwan will be illustrated.

Sachant seulement le taux moyen de volume des urgences par intervalle de temps, le temps moyen passé aux urgences de l'évaluation initiale ou du traitement jusqu'à la fin de la visite du patient, le nombre maximal des patients admissibles à un service des urgences simple, et supposant qu'à un moment donné un seul docteur s'occupe de patients multiples si nécessaire, la technique d'emboîtement de chaînes de Markov finies peut être appliquée pour obtenir des distributions de temps d'attente et estimer les temps d'attente prévus pour chacun des niveaux de l'Échelle canadienne de triage et de gravité. Une analyse d'un jeu de données de temps d'attente fourni par un hôpital de Taiwan sera présentée.

Session 04H

Lundi 1 juin • Monday, June 1, 15:30 – 17:00

CHE 250

Statistical Genetics and Genomics I

Génétique et génomique statistiques I

[MS-104]

Flexible Empirical Bayes Models for Differential Gene Expression

Modélisation bayésienne empirique flexible de données d'expression génétique différentielle

Kenneth Lo, *University of British Columbia*, Raphael GOTTARDO, *Institut de recherches cliniques de Montreal*

Inference about differential expression is a typical objective when analyzing gene expression data. The application of Bayesian hierarchical models with the Gamma-Gamma or Lognormal-Normal framework is a popular approach for this type of problem. However, the unrealistic assumption of a common coefficient of variation across genes has been implicitly made. In this talk, we propose an extension of these models to allow for gene-

L'inférence à partir de données d'expression différentielle est un objectif typique en analyse de données d'expression génétique. L'application de modèles hiérarchiques bayésiens dans le cadre gamma-gamma ou lognormal-normal est une approche populaire pour ce type de problème. Cependant, l'hypothèse peu réaliste d'un coefficient de variation commun pour tous les gènes a été implicitement fait. Dans cet exposé, nous proposons une étendue de ces modèles permettant des coefficients de variation spécifiques aux gènes, et

specific coefficients of variation, and EM-based algorithms for parameter estimation. The proposed methodology is evaluated on some experimental datasets. It shows that the two extended models significantly reduce the false discovery rate while maintaining a high sensitivity.

des algorithmes à base EM pour l'estimation des paramètres. La méthodologie proposée est évaluée sur des jeux de données expérimentales. L'évaluation montre que les deux modèles étendus réduisent significativement le taux de fausses découvertes tout en maintenant une sensibilité élevée.

[MS-105]

Using SELEX Data to Model the Affinity of DNA Sequences to the Transcription Factor Bicoid

Utilisation de données ESLEE pour modéliser l'affinité de séquences d'ADN pour le facteur de transcription bicoïde

Juli ATHERTON, *McGill University*, Nathan BOLEY, Ben BROWN & Peter BICKEL, *UC Berkeley*, Mark BIGGIN, *Lawrence Berkeley Labs*

In genomic research, determining locations on the genome to which a transcription factor binds with medium to high affinity might help identify possible transcription factor binding sites. Interest then lies in developing models that predict the affinity of a transcription factor based on nucleotide sequence. Data from in-vitro experiments are often used to build these models. We present new methodology to analyze data from the in-vitro experiment called systematic evolution of ligands by exponential enrichment (SELEX).

En recherche génomique, déterminer l'emplacement dans le génome auquel un facteur de transcription se lie avec affinité moyenne ou élevée peut aider à identifier de possibles sites de liaison de facteurs de transcription. Il est donc d'intérêt de développer des modèles afin de prédire l'affinité d'un facteur de transcription à partir de la séquence de nucléotides. Des données d'expériences in vitro sont souvent utilisées pour construire de tels modèles. Nous présentons une nouvelle méthodologie pour analyser des données d'expériences in vitro appelée évolution systématique des ligands par enrichissement exponentiel (ÉSLEE).

[MS-106]

Evaluating Disease Risk for Successive Events in Carriers of a Major Gene using a Progressive Multi-state Modelling Approach.

Évaluer le risque de maladie pour événements successifs chez les porteurs d'un gène majeur par modélisation progressive à plusieurs états

Yun-Hee CHOI, *University of Western Ontario*, Karen KOPCIUK, *Alberta Cancer Board*, Jane GREEN & Pat PARFREY, *Memorial University of Newfoundland*, Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital*

Carriers of a major disease-causing gene are often at risk for more than one phenotype. To estimate transition intensity functions for risk of developing successive phenotypes, a progressive multi-state model approach is adopted using both parametric and non-parametric baseline hazards. We also develop an Expectation-Maximization algorithm to estimate the parameters of the model which can infer missing genotypes for family members. Data from a cohort of 12 independently ascertained HNPCC families harbouring a founder MSH2 mutation and experiencing more than one HNPCC-related cancer from Newfoundland are used to illustrate these methods. We provide age-specific risk estimates associated with subsequent cancers.

Les porteurs d'un gène majeur causant la maladie sont souvent à risque pour plus d'un phénotype. Afin d'estimer les fonctions d'intensité de transition pour le risque de développer des phénotypes successifs, une approche de modélisation multi-états progressive est adoptée, utilisant des taux de panne de base paramétriques et non paramétriques. Nous développons aussi un algorithme EM pour estimer les paramètres du modèle pouvant inférer les génotypes manquant pour membres d'une famille. Les données d'une cohorte de 12 familles de Terre-Neuve à cancer colorectal héréditaire non-polypose (HNPCC) dont le fondateur a la mutation MSH2 et chez lesquelles plus d'un cancer lié au HNPCC est présent sont utilisées pour illustrer ces méthodes. Nous fournissons des estimations du risque spécifique à l'âge associé à des cancers subséquents.

[MS-107]**Survival Relevant Gene Selection in Microarray Data Analysis with Gene Expression Subject to Measurement Error****Sélection de gènes liés à la survie en analyse de données de micropuces avec expression génétique sujette à erreurs de mesure**Juan XIONG & Wenqing HE, *University of Western Ontario*

Microarray technology is essentially a measurement tool for measuring biological features such as gene expression. Gene expression could be employed as predictors for patient survival, and the measurement error involved in the expression data is often ignored. We studied the effect of the measurement error on survival relevant gene selection under the AFT model setting by regularizing weighted least square estimator with adaptive LASSO penalty. The simulation results show that ignoring measurement error will affect survival relevant gene selection and lead to erroneous predictive model. SIMEX method is investigated to adjust the impact of gene selection and survival prediction.

Les micropuces sont essentiellement un outil permettant de mesurer des caractéristiques biologiques telles l'expression génétique. L'expression génétique pourrait servir de prédicteurs pour la survie de patient, et l'erreur de mesure dans les données d'expression génétique est souvent ignorée. Nous avons étudié l'effet de l'erreur de mesure sur la sélection de gènes liés à la survie sous le modèle à défaillance accélérée (AFT) en régularisant l'estimateur des moindres carrés pondérés avec une pénalité LASSO adaptative. Les résultats de simulation montrent qu'ignorer l'erreur de mesure affecte la sélection de gènes liés à la survie et mène à des modèles prédictifs erronés. L'utilisation de la méthode SIMEX pour ajuster l'impact de la sélection de gènes et la prévision de survie est étudiée.

[MS-108]**Probabilistic Inference for ChIP-Seq data: a Model Based Approach****Inférences probabilistes pour données ChIP-Seq : une approche basée sur un modèle**Xuekui ZHANG & Raphael GOTTARDO & Kaida NING *University of British Columbia, Gordon ROBERTSON, British Columbia Cancer Agency Genome Sciences Centre*

ChIP-Seq is a massively parallel short-read sequencing system that can characterize genome-wide profiles for vivo transcription factor binding sites in mammals. This high-throughput sequencing technology has been available commercially since 2007, and become the new standard tool of research. We propose a method called PICS to extract information from the ChIP-Seq data and detect Transcription Factor bound regions across the genome. PICS identifies enriched regions by modeling local concentrations of directional reads adjusted by mappability profiles to discriminate adjacent binding events using a truncated t -mixture model. Using two published data sets we showed PICS out-performs its competitors.

ChIP-Seq est un système de séquençage massivement parallèle pouvant caractériser les profils sur tout le génome pour sites de liaison des facteurs de liaison chez les mammifères. Cette technologie de séquençage à haut débit est commercialement disponible depuis 2007, et est devenue le nouvel outil standard en recherche. Nous proposons une méthode appelée PICS pour extraire l'information de données ChIP-Seq et détecter les régions de liaison de facteurs de transcription sur le génome. PICS identifie les régions enrichies en modélisant les concentrations locales de lectures directionnelles afin de discriminer entre liaisons adjacentes en utilisant un modèle tronqué de mélanges de lois t . À l'aide de deux jeux de données publiés, nous montrons que PICS produit de meilleurs résultats que ses compétiteurs.

Session 04I Lundi 1 juin • Monday, June 1, 15:00 – 16:30 IRC Lobby
Greater Vancouver Regional Science Fair Student Projects
Projets de l'expo-sciences régionale de Vancouver

[MS-109]

Effects of Light on Algae
Effets de la lumière sur l'algue

Michelle NG & Meiyang ZHUANG, *University Transitions Program - Early Entrance to UBC*

[MS-110]

Probabilistic Evolution in Dynamic Equilibrium
Évolution probabiliste dans l'équilibre dynamique

Jonathan ZHOU, *Pinetree Secondary*, Andrew J. JUNG, *Walnut Grove Secondary*

[MS-111]

The Effect of Traditional Chinese Herbal Medicines on Human Cancer Cells
L'effet des herbes médicinales chinoises traditionnelles sur les cellules humaines cancéreuses

Jieqing XU, *University Transitions Program - Early Entrance to UBC*

[MS-112]

Effects of Common Fruit Extracts on Cell Division in Human Cancers
Effets des essences communes de fruits sur la division cellulaire dans les cancers humains

Jessica ZHANG, *University Transitions Program - Early Entrance to UBC*

Session 05 Mardi 2 juin • Tuesday, June 2, 08:45 – 10:00 IRC 2
Gold Medal Award Address
Allocution du récipiendaire de la médaille d'or

[MS-113]

Randomness is Not a Bad Word
Aléatoire n'est pas un mauvais mot

LUC DEVROYE, *McGill University*

"Randomness" has been jargonized in the last decade. Before it becomes a politically incorrect word, I will attempt to restore its decline by showing how one can create more randomness, some-

Le terme «aléatoire» a été jargonisé au cours de la dernière décennie. Avant qu'il ne devienne un mot politiquement incorrect, je vais tenter de renverser son déclin en montrant comment on peut créer plus de hasard, ce que j'ai fait durant

thing that I have been practicing for most of my life. More precisely, I will discuss the subdiscipline of random variate generation, starting from the pioneering fifties, and ending with the challenges that lie ahead.

la majeure partie de ma vie. Plus précisément, je vais discuter de la génération de variables aléatoires, de ses débuts dans les années 50 aux défis qui nous attendent.

Session 06A

Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00

IRC 6

Special Invited Session of the Probability Section

Allocution de l'invité d'honneur du groupe de probabilité

[MS-114]

Extremes in Air Pollution and Climate

Extrêmes en pollution atmosphérique et en climat

Peter GUTTORP, *University of Washington*

We outline some of the current issues regarding the use of extreme value theory in environmental sciences. In particular, we look at the conditional distribution of spatial maxima given observations at a network of stations and apply this to the operational evaluation of air quality standards. Another issue is looking for expected phenomena (based on climate simulations) in long time series of temperature and/or pressure, and relating this to extremes of long memory processes.

Nous présentons dans les grandes lignes certaines questions actuelles relatives à l'utilisation de la théorie des valeurs extrêmes en sciences de l'environnement. En particulier, nous regardons la distribution conditionnelle de maximums spatiaux étant données des observations d'un réseau de stations de surveillance et appliquons cela à l'évaluation opérationnelle de standards de qualité de l'air. Une autre question est de chercher des phénomènes espérés (basés sur des simulations climatiques) dans les séries chronologiques longues de température et/ou de pression, et de mettre cette recherche en relation avec les extrêmes de processus à mémoire longue.

Session 06B

Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00

IRC 1

Innovative Ideas in the Teaching of Statistics and Probability

Idées novatrices en enseignement de la statistique et des probabilités

[MS-115]

Authentic Statistics Education - Updating the Apprenticeship Model

Une éducation authentique en statistique - mettre à jour le modèle de l'apprentissage

Larry WELDON, *SSC*

In most disciplines, the apprenticeship model of education has been abandoned in favor of the current course-requirement system. In statistics, textbook-based courses have evolved to cover a narrow range of techniques, and the needs of professionals for statistical practice are not well-served. Instructor-guided experiential learning has the potential to provide a way for courses to fix this content shortfall. Textbooks in experiential learning serve as a reference, rather than as a sequential guide. Moreover, guided experiential learning is

Dans la plupart des disciplines, le modèle d'éducation par apprentissage a été abandonné pour le système actuel de cours exigés. En statistique, les cours basés sur des manuels ont évolué pour couvrir une gamme limitée de techniques, et les besoins des professionnels en matière de pratique statistique ne sont pas comblés. L'apprentissage expérimental guidé par un instructeur a le potentiel de permettre aux cours de combler ce déficit en contenu. Les manuels en apprentissage expérimental servent de référence plutôt que de guides séquentiels. En outre, l'apprentissage expérimental dirigé est plus facile à adapter aux besoins des étudiants

more easily adapted to suit student-group needs than in textbook-based courses. This presentation discusses the implementation of this "experiential" approach.

que l'enseignement basé sur des manuels. Cette présentation discute de la mise en oeuvre de cette approche «expérimentale».

[MS-116]

A Case Study about Statistical Communication with Practicing Engineers

Une étude de cas à propos de la communication statistique avec des ingénieurs praticiens

Christine ANDERSON-COOK, *Los Alamos National Laboratory*

Developing appropriate statistical models for system reliability of complex systems involve many diverse aspects and expertise from a variety of different experts. Understanding the system structure typically requires the expertise of several engineers, while building a model that can incorporate multiple types of data involves advanced statistical thinking. Through a case study, the talk describes the process and some of the potential pitfalls for eliciting information from collaborators with a variety of backgrounds, synthesizing different types of engineering and statistical knowledge, and verifying that the final statistical model reflects actual system characteristics.

Le développement de modèles statistiques appropriés pour la fiabilité de systèmes complexes implique plusieurs aspects et l'expertise d'une variété d'experts. La compréhension de la structure du système nécessite habituellement l'expertise de plusieurs ingénieurs, alors que la construction d'un modèle pouvant incorporer différents types de données demande une pensée statistique avancée. Au moyen d'une étude de cas, cet exposé décrit le processus d'obtention d'information de collaborateurs de différents domaines, de synthèse de différents types de connaissances en génie et en statistique, et de vérification que le modèle statistique final reflète les caractéristiques réelles du système, ainsi que certaines embûches potentielles du processus.

[MS-117]

Eikosograms: The Picture of Probability

Eikosogrammes : l'image des probabilités

Wayne OLDFORD, *University of Waterloo*

Good diagrams clarify. Very good diagrams force the ideas upon the viewer. The best diagrams embody the ideas themselves. An eikosogram, or probability picture, is a simple and easily constructed figure that ostensibly defines probability. It embeds the rules of probability – conditional, joint and marginal – within its construction. As such, the eikosogram is a natural tool to formally introduce probability theory to students. This will be demonstrated in the talk, along with how the eikosogram can be used to visually ground seemingly abstract concepts such as conditional independence. Some probability puzzles/paradoxes will be used as illustration.

De bons diagrammes clarifient. De très bons diagrammes imposent des idées chez le spectateur. Les meilleurs diagrammes incarnent les idées elles-mêmes. Un eikosogramme, ou image de probabilité, est une figure simple et facile à construire définissant ostensiblement une probabilité. Les règles des probabilités - conditionnelles, marginales et conjointes - apparaissent intrinsèquement dans sa construction. L'eikosogramme est donc un outil naturel pour formellement présenter la théorie des probabilités aux étudiants. Ceci sera démontré durant cet exposé, ainsi qu'une façon d'utiliser l'eikosogramme pour visuellement fixer des concepts abstraits telle l'indépendance conditionnelle. Certains casse-têtes et paradoxes en probabilités seront utilisés à des fins illustratives.

Session 06C**Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00****IRC 4**

Statistical Inference Based on Data from Different Sources

Inférence statistique fondée sur des données de sources différentes

[MS-118]**Regression with Probability-Linked Data****Régression avec données couplées de façon probabiliste**Ray CHAMBERS, *University of Wollongong*

Data obtained after probability linkage of administrative registers will include errors due to the fact that some linked records contain data items sourced from different individuals. Such errors can induce bias in standard statistical analyses if ignored. I describe some approaches to eliminating this bias in the case of linear regression analysis and, more generally when inference is based on an estimating equation, with an emphasis on logistic regression. Simulation results that illustrate the gains from allowing for linkage error in linear and logistic regression analysis are presented, as are extensions of the approach to more complex linkage situations.

Les données obtenues après le couplage probabiliste de registres administratifs vont comprendre des erreurs dues au fait que certains fichiers couplés contiennent des éléments de données provenant d'individus différents. De telles erreurs peuvent induire un biais dans les analyses statistiques standard si ignorées. Je décris quelques approches pour éliminer ce biais dans le cas de l'analyse de régression linéaire et, plus généralement, lorsque l'inférence est basée sur une équation d'estimation, avec une emphase sur la régression logistique. Des résultats de simulation illustrant les gains obtenus en tenant compte de l'erreur de couplage en régression linéaire et logistique sont présentés. Nous abordons également des extensions de cette approche à des situations de couplage plus complexes.

[MS-119]**Regression Analysis of Record-Linked Survey Data****Analyse de régression pour données d'enquêtes liées à des enregistrements**Milorad KOVACEVIC, *Statistics Canada*

Data from different sources are often combined using computerized record linkage methodology. However, due to a possible mismatch of records from two data sets the linkage errors are introduced. Scheuren and Winkler (1993, 1997), Lahiri and Larsen (2005), and more recently Chambers (2008), pointed out that the modeling based on such data can be subject to bias and additional variability in the presence of linkage errors. We extend their methods to joint accounting for linkage errors and for sample design when analyzing the linked survey data. In particular, we deal with the linear and the generalized linear models.

Des données de différentes sources sont souvent combinées par des méthodes de couplage des fichiers. Cependant, à cause d'un possible mauvais jumelage de deux jeux de données des erreurs de couplage sont introduites. Scheuren et Winkler (1993, 1997), Lahiri et Larsen (2005), et, plus récemment, Chambers (2008) ont souligné que la modélisation basée sur de tels modèles peut être sujette à un biais et à une variabilité additionnelle causés par la présence d'erreurs de couplage. Nous étendons leurs méthodes pour tenir compte des erreurs de couplage et du plan d'échantillonnage en analyse de données d'enquêtes jumelées. En particulier, nous traitons de modèles linéaires et linéaires généralisés.

Session 06D**Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00****IRC 5****Statistical Applications in the Study of Infectious Diseases****Modèles et méthodes statistiques dans le domaine de l'étude et la lutte contre les maladies infectieuses****[MS-120]****Alternatives to Intention to Treat—the MIRA Trial****Alternatives à l'intention de traitement-l'étude MIRA**Nicholas JEWELL, *UC Berkeley*

Clinical trial methodology applied to intervention trials presents challenges in design, analysis and interpretation. Consider The Methods for Improving Reproductive Health in Africa trial that investigated diaphragm and lubricant gel use in reducing HIV infection among women, randomly assigned to either a treatment or control arm. All subjects received intensive condom counseling and provision. There was lower reported condom use in the intervention arm, complicating the interpretation of a standard intention-to-treat analysis. I will describe an analysis technique designed to estimate the "direct effects" of assignment to the diaphragm arm, adjusting for condom use in an appropriate sense.

L'application de la méthodologie des essais cliniques à des études d'intervention pose des défis au niveau de la planification, de l'analyse et de l'interprétation. Considérons l'étude Méthodes pour améliorer la santé reproductive en Afrique qui portait sur l'utilisation du diaphragme et de gel lubrifiant dans la réduction de l'infection au VIH chez les femmes, assignées aléatoirement à un groupe traitement ou au groupe contrôle. Tous les sujets ont reçu une formation au sujet de l'utilisation du condom et une réserve de condoms. Le taux d'utilisation du condom rapporté dans le groupe d'intervention étant inférieur complique l'interprétation d'une analyse d'intention de traitement standard. Je vais décrire une technique d'analyse pour estimer les effets directs de l'assignation au groupe diaphragme, ajustant de façon appropriée pour l'utilisation du condom.

[MS-121]**Estimating the Effectiveness of Immunization against Paralytic Poliomyelitis in Nigeria****Estimation de l'efficacité de l'immunisation contre la poliomyélite paralytique au Nigéria**

Christl DONNELLY & Helen JENKINS, *MRC Centre for Outbreak Analysis and Modelling, Imperial College London*, Bruce AYLWARD & Alex GASASIRA, *World Health Organization*, Emmanuel ABANIDA & Titi KOLEOSHO-ADELEKAN, *National Primary Health Care Development Agency, Abuja, Nigeria*, Nicholas GRASSLY, *MRC Centre for Outbreak Analysis and Modelling, Imperial College London*

The incidence of poliomyelitis in Nigeria declined in 2007 following the introduction of monovalent oral poliovirus vaccines (OPVs) and new techniques of vaccine delivery. Estimated per-dose efficacies of monovalent type 1 OPV and trivalent OPV against type 1 paralytic poliomyelitis were 67% and 16% respectively, with that of trivalent OPV against type 3 paralytic poliomyelitis being 18%. Between 2005 and 2007, vaccine-induced immunity levels in high-incidence areas more than doubled, to 56%. Further increases in coverage in affected states are required to achieve the levels of vaccine-induced immunity associated with the sustained elimination achieved in other parts of Nigeria.

L'incidence de la poliomyélite au Nigéria a diminué en 2007 suite à l'introduction des vaccins antipoliomyélitiques oraux (VPOs) monovalents et à la mise en place de nouvelles techniques de distribution du vaccin. L'efficacité par dose contre la poliomyélite paralytique de type 1 a été estimée à 67% pour le vaccin VPO monovalent de type 1 et à 16% pour le vaccin VPO trivalent, l'efficacité de ce dernier contre la poliomyélite paralytique de type 3 étant estimée à 18%. Entre 2005 et 2007, la fraction d'enfants entre 0 et 4 ans directement immunisés par le vaccin a plus que doublé atteignant 56%. L'augmentation de la couverture vaccinale dans les zones affectées est nécessaire pour obtenir un niveau d'immunité induit par le vaccin similaire à celui des zones du Nigéria où l'on observe une élimination durable de la maladie.

[MS-122]**Likelihood-free Inference for Infectious Disease Models****Inférence sans vraisemblance pour modèles de maladies infectieuses**

Rob DEARDON, *University of Guelph*, TJ MCKINLEY, *University of Cambridge*, Alex COOK, *National University of Singapore*

Likelihood-based inference for epidemic models has always proved difficult, mainly due to difficulties associated with calculating or evaluating the likelihood, particularly in large scale models. Unobserved or partially observed data often further complicates this process. Here we investigate the performance of Markov chain Monte Carlo routines based on approximate likelihoods generated from model simulations as a means of estimating parameters in epidemic models for both complete and incomplete data. We illustrate our techniques using common cold data from the island of Tristan da Cunha.

L'inférence basée sur la vraisemblance pour modèles épidémiques s'est toujours avérée difficile, principalement à cause des difficultés associées au calcul ou à l'évaluation de la vraisemblance, en particulier dans le cas de modèles à grande échelle. Des données manquantes ou partiellement observées apportent souvent des complications supplémentaires. Nous étudions ici la performance de sous-programmes MCMC basés sur une vraisemblance approximée générée à partir de simulations du modèle comme méthode d'estimation des paramètres de modèles épidémiques pour données complètes et incomplètes. Nous illustrons nos techniques avec des données sur le rhume dans l'île Tristan da Cunha.

Session 06E**Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00****IRC 3****Nonparametric Multivariate Methods****Méthodes non paramétriques à variables multiples****[MS-123]****How to Find the Shape of a Banana****Comment trouver la forme d'une banane**

Ivan MIZERA, *University of Alberta*

The problem of finding a center and analogs of quantile contours for multivariate data will be addressed. The existing proposals, based on distance principles or data depth notions, adapt well rather to convex-shaped data clouds; the talk will focus on methods suitable for data clouds exhibiting non-convex, "curved" shapes, the prototypical example being a crescent-like arrangement.

Nous discuterons du problème de l'identification d'un centre et d'analogues aux contours des quantiles pour des données multivariées. Les propositions existantes, basées sur des principes de distance ou des notions de profondeurs des données, s'adaptent plutôt bien aux nuages de données de forme convexe. L'exposé sera centré sur des méthodes convenant à des nuages de données de forme «courbe» non convexes, l'exemple type étant une forme de croissant.

[MS-124]**Nonparametric Tests and the Independent Components Model****Tests non paramétriques et le modèle à composantes indépendantes**

Hannu OJA, *University of Tampere*, Davy PAINDAVEINE, *Université libre de Bruxelles*, Klaus NORDHAUSEN, *University of Tampere*, Sara TASKINEN, *University of Jyväskylä*

Several extensions of the standard multivariate normal model have been introduced in the literature and have been shown to be useful in practical multivariate data analysis. The extensions include the

Plusieurs extensions au modèle normal multivarié standard ont été présentées dans la littérature et se sont avérées utiles en analyse de données multivariées. Ces extensions comprennent, entre autres, les modèles de distributions ellipti-

model of elliptically symmetric distributions, the mixture models and the independent components (IC) model among others. The IC model assumes that the p -variate observations are linear combinations of p independent latent variables. In the independent component analysis the idea is to estimate the unknown linear transformation matrix. In this talk we introduce and discuss multivariate rank tests for multivariate location and independence problems in the IC model.

quement symétriques, les modèles de mélanges et le modèle à composantes indépendantes (CI). Le modèle CI suppose que des observations p -variées sont des combinaisons linéaires de p variables latentes indépendantes. En analyse de composantes indépendantes, l'idée est d'estimer la matrice de transformation linéaire inconnue. Dans cet exposé, nous présentons et discutons des tests de rangs multivariés pour problèmes de position et d'indépendance multivariée dans le modèle CI.

[MS-125]

On Multivariate Medians

À propos de médianes multivariées

Yijun Zuo, *Michigan State University*

The best breakdown point is one of the most outstanding features of the univariate median. For this property, it, however, has to pay the price of a low efficiency at normal and light-tailed models. Affine equivariant multivariate analogues in \mathbb{R}^d ($d > 1$) were constructed in the past two decades. Most of them also have to sacrifice their efficiency at normal and other models, nevertheless. In this talk, we review the various multi-dimensional medians and recommend a promising maximum depth median. It possesses a highest breakdown point among all its multivariate competitors and furthermore it is highly efficient at normal and other distributions.

Le meilleur point de rupture est une des caractéristiques les plus importantes de la médiane univariée. Elle paye cependant le prix de cette propriété sous la forme d'une faible efficacité pour modèles à ailes normales ou aplaties. Des analogues affines multivariés équivariants dans \mathbb{R}^d ($d > 1$) ont été construits au cours des deux dernières décennies. La plupart doivent sacrifier leur efficacité pour modèles à ailes normales ou aplaties. Dans cet exposé, nous faisons la revue de diverses médianes multidimensionnelles et recommandons une médiane à profondeur maximale prometteuse. Elle possède le point de rupture le plus élevé parmi toutes ses concurrentes multivariées et est en outre hautement efficace sous distributions normales et autres.

Session 06F

Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00

MSL 102

Emerging Issues in Graphical Models

Nouveaux problèmes dans les modèles graphiques

[MS-126]

A Conjugate Prior for Discrete Hierarchical Loglinear Models

Une loi a priori conjuguée pour les modèles hiérarchiques discrets loglinéaires

Helene MASSAM, *York University*, Jinnan LIU, *Samuel Lunenfeld Research Institute*, Adrian DOBRA, *University of Washington*

In Bayesian analysis of multi-way contingency tables, the selection of a prior distribution for either the loglinear parameters or the cell probabilities parameters is a major challenge. In this paper we define a flexible family of conjugate priors for the wide class of discrete hierarchical loglinear models which includes the class of graphical models. These priors are defined as the Diaconis-

En analyse bayésienne de tableaux de contingence, le choix de la distribution a priori pour les paramètres est une des difficultés principales. Nous définissons une famille d'a priori pour les paramètres loglinéaires valables pour la classe de modèles hiérarchiques discrets (qui inclut la classe des modèles graphiques discrets). Nous obtenons aussi la loi a priori correspondante pour les probabilités des cellules du tableau et nous montrons que cette nouvelle famille de lois a priori

Ylvisaker conjugate priors on the loglinear parameters subject to "baseline constraints" under multinomial sampling. We also derive the induced prior on the cell probabilities and show that the induced prior is a generalization of the hyper Dirichlet prior. We show that this prior has several desirable properties and illustrate its usefulness by identifying the most probable decomposable, graphical and hierarchical loglinear models for a six-way contingency table.

est une généralisation de l'hyper Dirichlet. Cette famille a de nombreuses propriétés utiles. Nous donnons un exemple d'application pour un tableau de contingence de dimension six.

[MS-127]

Probabilistic Inference for CART Networks

Inférence probabiliste pour réseaux CART

BO THIESSEN, *Microsoft Research*, CHONG WANG, *Princeton University*, CHRISTOPHER MEEK, *Microsoft Research*, DAVID BLEI, *Princeton University*

We describe CART networks—Bayesian networks in which the local structure associated with each variable is either a classification tree or a regression tree. We develop a sum-product junction tree inference algorithm for CART networks which utilizes a novel region partitioned potential. Finally, we provide an analysis of the computational complexity of inference for CART networks using our junction tree algorithm.

Dans cet article, nous décrivons les réseaux CART (Classification And Regression Tree). Dans un réseau CART, chaque variable est un arbre de classification ou un arbre de régression. Nous avons développé un algorithme d'inférence d'arbre basé sur les jonctions somme-produits et un nouveau potentiel de partitionnement. Nous présentons une analyse de la complexité computationnelle d'un moteur d'inférence utilisant notre algorithme d'arbre de jonction.

[MS-128]

Learning Graphs With unknown Group Sparsity

Graphes d'apprentissage avec éparsité de groupe inconnue

KEVIN MURPHY & BEN MARLIN, *UBC*

We propose a method for learning the structure of Gaussian graphical models from observational data based on group L1 regularization. In contrast to previous work, we estimate the group structure (using a variational EM algorithm), rather than assuming it is known. We present results on motion capture and other data sets.

Nous proposons une méthode pour l'apprentissage de la structure de modèles graphiques gaussiens à partir de données d'observation basées sur la régularisation du groupe L1. Contrairement aux travaux précédents, nous estimons la structure du groupe (en utilisant un algorithme EM variationnel) plutôt que de la supposer connue. Nous présentons des résultats de capture de mouvement et de d'autres jeux de données.

Session 06G Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00 IRC G65/66

Forests, Fires and Stochastic Modeling Forêts, incendies et modélisation stochastique

[MS-129]

Spatial Bootstrap Block Size Selection

Sélection de la taille des blocs pour bootstrap spatial

Lengyi HAN, W. John BRAUN & Xiaoming LIU, *University of Western Ontario*

The stationary block bootstrap was proposed for time series 20 years ago (Kunsch, 1989), and a spatial version has been proposed more recently. Block size choices, based on asymptotic arguments, have been proposed in both situations. We study these choices by simulation and compare with empirical choices based on matching autocorrelations or correlograms, for data of a type arising in mortality studies and in fire spread modelling.

Le bootstrap en blocs stationnaire a été proposé il y a 20 ans (Kunsch, 1989) pour l'étude de séries chronologiques et une version spatiale a été proposée plus récemment. Des choix pour la taille des blocs, basés sur des arguments asymptotiques, ont été proposés dans les deux situations. Nous étudions ces choix par simulation et les comparons à des choix empiriques, basés sur le jumelage d'autocorrélations ou de corrélogrammes, pour des données d'un type survenant en études de mortalité et en modélisation de propagation des incendies.

[MS-130]

Looking for Climate Change Signals in the Canadian Forest Fire Ignition Record

Chercher des signaux des changements climatiques dans le déclenchement de feux de forêts au Canada

Douglas WOOLFORD, Jiguo CAO & Charmaine DEAN, *Simon Fraser University*, David MARTELL, *University of Toronto*

We propose a mixture model framework to investigate for climate change signals in forest fire ignitions. A logistic generalized additive model with a mixture of normal and extreme value distributions in the residual structure is necessary, since the smooth terms do not adequately describe the extreme events. The general modelling framework, a discussion of model selection, and a simple estimation technique for a preliminary analysis of the extreme residuals is presented. We illustrate its application using data for lightning-caused forest fire ignitions over a period of 42 years in a remote region of boreal forest in Ontario, Canada.

Nous proposons un modèle de mélange pour étudier les signaux de changements climatiques dans le déclenchement de feux de forêts. Un modèle additif généralisé logistique avec un mélange de distributions normales et valeurs extrêmes dans la structure résiduelle est nécessaire, puisque les termes lisses ne décrivent pas adéquatement les événements extrêmes. Le cadre de modélisation général, une discussion de la sélection de modèle et une technique d'estimation simple pour l'analyse préliminaire des résidus extrêmes sont présentés. Nous illustrons son application à des données de feux de forêts déclenchés par la foudre sur une période de 42 ans dans une région reculée de la forêt boréale en Ontario, Canada.

[MS-131]

Modelling Point Patterns with the Ivanoff-Merzbach Renewal Process

Modélisation de motifs ponctuels avec le processus de renouvellement Ivanoff-Merzbach

Katherine DAVIES, *University of Manitoba*, W. John BRAUN, *University of Western Ontario*

This poster demonstrates how to apply the Ivanoff-Merzbach renewal process (Ivanoff and Merzbach, 2006 and Ivanoff, 2008) to the modelling of environmental and other forms of point process data.

Cette affiche démontre comment appliquer le processus de renouvellement d'Ivanoff-Merzbach (Ivanoff et Merzbach, 2006 et Ivanoff, 2008) à la modélisation de données de processus ponctuels, dont les données environnementales. Le

The model is explored via simulation, focussing on two special cases. Maximum likelihood estimation of model parameters is described and demonstrated for these cases. The methods are applied to two point patterns, one arising from history and one from ecology. It is also envisioned this model will be useful for the modelling of forest fires and the spread of infections through a forest.

modèle est exploré par simulation, et nous nous concentrons sur deux cas particuliers. L'estimation par le maximum de vraisemblance des paramètres du modèle est décrite et montrée pour ces cas. Les méthodes sont appliquées à deux nuages de points, provenant de données historiques et écologiques. Nous croyons également que ce modèle sera utile en modélisation de la propagation de feux et d'infections à travers la forêt.

[MS-132]

Wind Effects on Bark Char Patterns in Forest Fires: a Heat Transfer Analysis

Effets du vent sur la répartition de résidus d'écorce charbonneux en feux de forêts : une analyse de transfert de chaleur

Sean MICHALETZ & Edward A. JOHNSON, *University of Calgary*

In forest fires, bark char patterns vary with fire spread direction, wind direction and wind velocity. In low velocity winds, charring occurs on the tree side from which the fire approaches; as velocity increases, charring shifts to the leeward side. We model bark surface temperature rise to pyrolysis as one-dimensional conduction with convection boundary conditions. Convection coefficients are estimated from local Nu data at typical fire wind velocities. Results agree with observations and show that leeward char shift results from increased convection rates due to turbulent boundary layer formation and mixing associated with leeward vortex formation.

En feux de forêts, les motifs de résidus charbonneux d'écorce varient selon la direction de la propagation de l'incendie, la direction et la vitesse du vent. En présence de vents de vitesse peu élevée, la carbonisation se produit sur le côté de l'arbre duquel s'approche le feu ; alors que la vitesse du vent augmente, la carbonisation se déplace vers le côté sous le vent. Nous modélisons l'augmentation de la température de l'écorce jusqu'à pyrolyse par une conduction unidimensionnelle avec conditions limites de convection. Les coefficients de convection sont estimés à partir de données Nu locales sous vitesses habituelles du vent et de l'incendie. Nos résultats sont en accord avec les observations et montrent que le déplacement de la carbonisation sous le vent résulte de taux de convection accrus résultants de la formation d'une couche frontière turbulente associée à la formation d'un tourbillon à l'ombre du vent.

[MS-133]

Bayesian Hierarchical Models for Spatial Count Data with Application to Fire Frequency in British Columbia

Modèles hiérarchiques bayésiens pour données de dénombrement spatiales avec application à la fréquence des feux de forêts en Colombie-Britannique

Yolanda LI, *British Columbia Ministry of Health Services*, Steve W. TAYLOR, *Pacific Forestry Centre*, Farouk NATHOO, *University of Victoria*

In this article we develop spatial models for fire frequency based on the negative binomial distribution. Model development is motivated by a large scale study of fire frequency in British Columbia, conducted by the Canadian Forest Service. Specific to our analysis, the main focus lies in examining the interaction between wildfire and mountain pine beetle (MPB) outbreaks. Understanding this interaction is critically important as British Columbia is

Dans cet article, nous développons des modèles spatiaux pour la fréquence des feux de forêts basés sur la distribution binomiale négative. Le développement des modèles est motivé par une étude à grande échelle de la fréquence des feux en Colombie-Britannique menée par le Service canadien des forêts. Notre analyse porte principalement sur l'interaction entre feux de friches et poussées de dendroctone du pin ponderosa. Comprendre cette interaction est critique, car la Colombie-Britannique vit présentement une poussée sans

currently experiencing a historically unprecedented MPB outbreak. We specify hierarchical regression models for fire frequency incorporating spatially correlated random effects, and conduct model selection and inference within the Bayesian framework.

précédent de dendroctone du pin ponderosa. Nous spécifions des modèles de régression hiérarchiques pour la fréquence des feux incorporant des effets aléatoires spatialement corrélés et effectuons la sélection de modèle et l'inférence dans un cadre bayésien.

[MS-134]

Predicting Fire Behaviour: Field Experiments for Fire Behaviour Modelling

Prévoir le comportement du feu : expériences sur le terrain en modélisation du comportement du feu

Jim GOULD, *Natural Resources Canada- Canadian Forest Service*

Conducting experimental fires in a field environment, involving natural fuel complexes, terrain and range of fire weather conditions, is an integral component in either validating a theoretical and/or new fire behaviour concept or constructing empirical fire behaviour models. The real benefit of conducting experimental fires is the opportunity to view and collate the natural variation of fire behaviour under different burning conditions. This poster presents a collective review of experiences of numerous experimental burning projects over the past 20 years using different experimental fires as illustrated examples of conducting fire behaviour field experiments for fire behaviour modelling.

L'observation de feux expérimentaux sur le terrain, impliquant des combustibles naturels, le terrain et une gamme de conditions météorologiques, est une composante intégrante de la validation d'un nouveau concept sur le comportement du feu ou de la construction de modèles empiriques de comportement du feu. L'avantage réel des feux expérimentaux est l'opportunité qu'ils offrent d'observer la variation naturelle du comportement du feu sous différentes conditions. Cette affiche présente une revue de plusieurs projets d'incendies expérimentaux au cours des 20 dernières années et utilise différents feux expérimentaux comme exemples d'expériences sur le terrain pour modéliser le comportement des incendies.

Session 06H

Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00

MSL 101

Models and Tests in Business and Industrial Statistics

Modèles et tests en statistique industrielle et de gestion

[MS-135]

On the Significance of the Limited Value Averages Test

De la signification du test des moyennes de valeur limitée

Uditha BALASOORIYA & Jackie LI, *Nanyang Technological University*

In the literature, various tests have been proposed for selecting the best fitting model from several competing models, that explain a given set of data. In actuarial studies, in addition to the commonly used goodness-of-fit tests, often limited value averages are also examined in choosing the best model. In the present work, we investigate the significance of the limited value averages test as a goodness-of-fit test. Asymptotic results and some applications to real data are discussed.

Dans la littérature, plusieurs tests ont été proposés pour sélectionner le modèle ajustant le mieux les données parmi un ensemble de modèles concurrents. En études actuarielles, en plus des tests d'adéquations habituels, les moyennes de valeur limitée sont aussi examinées en sélection de modèles. Dans le présent travail, nous étudions la signification du test des moyennes à valeur limitée comme test d'ajustement. Des résultats asymptotiques et quelques applications à des données réelles sont abordés.

[MS-136]**Gaging Public Opinion Through Prediction Markets****Mesurer l'opinion publique à travers les marchés de prévision**Amanda LAFONTAINE & Dayan YEHUDA, *Ipsos MORI*

Political polling is an important part of the market research industry. Polls have the unique ability to gauge the precision of estimates of electoral results. In this light, Ipsos MORI is testing a virtual stock exchange to help predict poll results. This procedure allows respondents to be registered traders who use virtual money to trade on the likely outcome of election results. We hypothesize that this methodology will circumvent some undesired biases associated with online surveys (coverage bias and self selection). This presentation will review the results of a pilot study, as well as present key findings regarding behaviour patterns.

Les sondages politiques représentent une part importante de l'industrie des études de marchés. Ils sont en effet l'unique moyen de mesurer la précision des estimations électorales. Vu sous cet angle, Ipsos MORI est en train de tester une bourse virtuelle dont le but est de prédire les résultats des sondages. Cette expérience permet aux répondants de devenir 'traders' et, en utilisant une monnaie virtuelle, de spéculer sur les résultats probables des élections. Nous faisons l'hypothèse que cette méthodologie puisse permettre d'éviter certains biais, non désirés, mais pourtant associés aux enquêtes en ligne (biais de couverture, auto-sélection). Cette présentation va couvrir les résultats d'une étude pilote, ainsi que présenter des conclusions clés à propos du mode de comportement des participants.

[MS-137]**A Review of the Bence Statistical Catch-at-Age Population for Lake Whitefish Stock Management****Une revue du modèle statistique d'âge au moment de la capture de Bence pour la gestion du stock de grand corégone**Jasper TEY, Daniel GILLIS, Neil ROONEY, Greg KLOTZ, Amanda CASKENETTE & Steve CRAWFORD *University of Guelph*, Bill HARFORD, *University of Miami*

Our research aims to decompose and rebuild the Bence Statistical Catch-at-Age (SCAA) population model utilized in Great Lakes fisheries management. Simulated data sets and sensitivity analyses are used to identify SCAA parameters that significantly impact the ability to predict biomass abundance. The primary objective is to rebuild and implement a more modular and scalable SCAA modeling framework that incorporates appropriate population level dynamics. Ultimately, the outcome of this study will be used to improve population modeling for a wider variety of research and management needs, including fish habitat risk assessment and evaluating the effects of electrical power generating facilities.

Notre recherche vise à décomposer et reconstruire le modèle statistique de population d'âge au moment de la capture de Bence (SACB) utilisé en gestion des pêcheries des Grands Lacs. Des jeux de données simulés et des analyses de sensibilité sont utilisés pour identifier les paramètres SACB qui ont un impact significatif sur notre capacité prédire l'abondance de biomasse. L'objectif principal est de reconstruire et de mettre en oeuvre un SACB plus modulaire et échelonnable incorporant la dynamique du niveau de population approprié. Ultimement, le produit de cette étude sera utilisé pour améliorer la modélisation de la population pour une grande variété de besoins de recherche et de gestion, incluant l'évaluation du risque pour habitats de poissons et des effets de centrales électriques.

[MS-138]**Performance of Some Diagnostic Tests in Financial Time Series****Performance de certains tests diagnostics en séries chronologiques financières**

Vineetha WARRIYAR. K .V, *Memorial University*, Abraham BOVAS, *University of Waterloo*, Mulayath Variyath ASOKAN, *Memorial University*

Box-Pierce test and Ljung-Box test are portman-teau tests generally used to test the independence of the residuals in time series modeling. Since most financial time series data show heavy tailed behavior, these tests may reject the null hypothesis of independence when there is volatility clustering in the data. Diebold (1988) modified these tests to capture the heavy tailedness. We perform a series of Monte Carlo Simulations to compare the performance of the Box-Pierce and Ljung-Box tests and its modified versions in different situations where data exhibit ARCH, GARCH and SV effects and find that modified tests perform better.

Les tests de Box-Pierce et de Ljung-Box sont des tests de type portemanteau généralement utilisés pour tester l'indépendance des résidus en modélisation de séries chronologiques. Puisque la plupart des données de séries chronologiques financières sont à ailes relevées, ces tests peuvent rejeter l'hypothèse nulle d'indépendance lorsqu'il y a regroupement de la volatilité dans les données. Diebold (1988) a modifié ces tests pour prendre en considération l'épaisseur des ailes. Nous effectuons une série de simulations de Monte-Carlo afin de comparer la performance des tests de Box-Pierce et de Ljung-Box et de leur version modifiée dans différentes situations où les données présentent des effets ARCH, GARCH et SV et trouvons que les tests modifiés performant mieux.

Session 06I**Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00****CHE 150****Estimation and Models for Biomedical Data****Estimation et modèles pour données biomédicales****[MS-139]****Random Changepoint Model for Longitudinal Data: An Extension of Bent-Cable Regression****Modèle à point de changement aléatoire pour données longitudinales : une extension de la régression à cordes pliées**

Shahedul KHAN, Grace CHIU & Joel DUBIN, *University of Waterloo*

The bent-cable is a parametric regression model to study data that exhibits change. In this talk, we consider longitudinal data, where measurements taken over time are nested within observational units (individuals). A random-effects model extension of bent-cable regression is useful here. We develop a special type of bent-cable methodology in the framework of a random-effects model with within-individual serial correlation. The methodology is illustrated with an application where serial correlations are demonstrated by an AR(p) process. We also discuss extensions of the model to account for changepoint data that may be characterized by special cases of the bent-cable structure.

Les cordes pliées sont un modèle de régression paramétrique pour étudier les données comprenant des changements. Dans cet exposé, nous considérons des données longitudinales, où les mesures prises à travers le temps sont emboîtées dans des unités observationnelles (des individus). Une extension par effets aléatoires de la régression à cordes pliées est utile ici. Nous développons un type particulier de méthodologie à cordes pliées dans le cadre des modèles à effets aléatoires avec corrélation sérielle inter-individus. La méthodologie est illustrée avec une application où les corrélations sérielles sont démontrées par un processus AR(p). Nous abordons aussi des extensions du modèle pour tenir compte des données de point de changement pouvant être caractérisées par des cas particuliers de la structure à cordes pliées.

[MS-140]**Occupancy modeling with multiple states and multiple seasons****Modélisation de l'occupation avec états multiples et saisons multiples**Wendell CHALLENGER & Carl SCHWARZ, *Simon Fraser University*

The distribution of species over space is a central goal of ecology. However, the state of interest (e.g. species occurrences) tends not to be directly observable. Recently, various forms of occupancy models have been purposed to handle this shortcoming. We further expand this class of models by presenting a general framework for modeling multi-state occupancy over multiple seasons. We also provide examples and consider optimal design criterion for differing study objectives.

L'évaluation de la distribution d'espèces dans l'espace est un des buts centraux de l'écologie. Cependant, l'état d'intérêt (la présence d'espèces) n'est généralement pas directement observable. Récemment, différentes formes de modèles d'occupation ont été proposées pour allier à ce problème. Nous élargissons cette classe de modèles en présentant un cadre général pour modéliser plusieurs états d'occupation sur des saisons multiples. Nous présentons aussi des exemples et considérons des critères de planification optimale pour différents objectifs d'étude.

[MS-141]**Slection of Smoothing Parameter****Sélection du paramètre de lissage**Jun LI & Yogendra CHAUBEY, *Concordia University*, Pranab SEN, *University of North Carolina*

A smooth estimator of density function for Length-biased Data is found. In this estimator, there is a parameter which controls the smoothness of the estimator. In order to make a good choice of the parameter, we investigate two choices of cross validation methods and obtain some conclusions.

Un estimateur lisse de la fonction de densité pour données biaisées en longueur est obtenu. Un paramètre permet de contrôler le lissage de cet estimateur. Nous étudions deux méthodes de validation croisée pour permettre un bon choix du paramètre de lissage et obtenons quelques conclusions.

[MS-142]**Interval Estimation of Qantile Ratios Applied to Anti-Cancer Drug Screening by Xenograft Experiments****Estimation par intervalle de rapports de quantiles appliquée à des expériences de dépistage de médicaments contre le cancer par xénogreffe**Cheng CHENG & Jianrong WU, *St. Jude Children's Research Hospital*

Currently the analysis of xenograft experiments lacks statistical consideration to account for experimental noise. A novel confidence bound and interval procedure for estimating quantile ratios is developed to fill this void. Justified by large-sample theory and a simulation study of small-sample performance, this method performs well in a wide range of scenarios. By providing rigorous inference accounting for experimental noise, the proposed method greatly improves the current practice of analyzing drug activity data in xenograft experiments. This method is fully nonparametric and applicable to any inference of a "fold change" that can be formulated as a quantile ratio.

Présentement, l'analyse d'expérience xénogreffes manque de considération statistique pour tenir compte du bruit expérimental. Une nouvelle procédure de bornes et intervalles de confiance en estimation de rapports de quantile est développée pour combler ce vide. Justifiée par la théorie asymptotique et une étude de simulation de sa performance sous petits échantillons, la méthode performe bien dans une variété de scénarios. En apportant une inférence rigoureuse tenant compte du bruit expérimental, la méthode proposée améliore grandement la pratique actuelle consistant à analyser les données d'activité de médicaments en expériences xénogreffes. Cette méthode est complètement non paramétrique et applicable à n'importe quelle inférence sur un «changement de repli» pouvant être formulé comme un rapport de quantiles.

[MS-143]**Empirical likelihood estimator for diagnostic test likelihood ratio****Estimateur par vraisemblance empirique pour rapport de vraisemblance d'un test diagnostique**Bin DONG & David E. MATTHEWS, *University of Waterloo*

Diagnostic test likelihood ratios (DLRs) have been advocated in the clinical and epidemiologic literature for several decades. In this paper we derive smooth empirical likelihood-based estimators, involving kernel methods, for these ratios of continuous-scale diagnostic test measurements in diseased and disease-free populations. To avoid the resulting bandwidth selection problem, we also consider the use of non-smooth estimating equations to obtain estimators of these same DLRs. In a simulation study that we describe, interval estimates derived from the empirical likelihood estimators retain a high degree of coverage accuracy compared to the usual estimators based on normal approximations.

Les rapports de vraisemblance de test diagnostics (RVDs) ont été recommandés dans la littérature clinique et épidémiologique depuis plusieurs décennies. Dans cet article nous obtenons des estimateurs empiriques lisses basés sur la vraisemblance, impliquant des méthodes à noyaux, pour ces rapports de mesures de tests à échelle continue parmi les populations saine et malade. Pour éviter le problème de sélection du paramètre de lissage, nous considérons aussi l'utilisation d'équations d'estimation non lisses pour obtenir des estimateurs de ces mêmes RVDs. Dans une étude de simulation que nous décrivons, les estimateurs par intervalles obtenus à partir des estimateurs par vraisemblance empirique conservaient une grande précision de couverture comparativement aux estimateurs usuels basés sur l'approximation normale.

Session 06J**Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00****CHE 250****Estimation Methods****Méthodes d'estimation****[MS-144]****Capture-Recapture, Jolly-Seber, and the Horvitz-Thompson Estimator: ML Estimation of Abundance for Populations with Variable Capture Probabilities****Capture-recapture, Jolly-Seber et l'estimateur Horvitz-Thompson : estimation par maximum de vraisemblance de l'abondance de populations avec probabilités variables de capture**Simon BONNER, *University of British Columbia*, Carl SCHWARZ, *Simon Fraser University*

Capture-recapture estimation of abundance for populations with variable capture probabilities has remained a problem for 40 years. Recently, McDonald and Amstrup described an ad hoc approach to estimate abundance when capture probabilities are modelled as functions of covariates. Each individual's capture probability is first estimated from a conditional likelihood and then substituted into the Horvitz-Thompson estimator of abundance. We present a new formulation of the Jolly-Seber likelihood producing a similar estimator of abundance with a modification that accounts for extra knowledge from the survival of marked individuals. This provides a unified, likelihood framework for incorporating covariate effects in abundance estimation.

L'estimation par capture-recapture de l'abondance pour populations avec probabilités de capture variables est restée problématique pendant 40 ans. Récemment, McDonald et Amstrup ont décrit une approche ad hoc pour estimer l'abondance lorsque les probabilités de capture sont modélisées comme fonctions de covariables. La probabilité de capture de chaque individu est d'abord estimée à partir d'une vraisemblance conditionnelle et ensuite substituée dans l'estimateur d'abondance Horvitz-Thompson. Nous présentons une nouvelle formulation de la vraisemblance de Jolly-Seber produisant un estimateur d'abondance similaire avec une modification tenant compte de l'information additionnelle apportée par la survie d'individus marqués. Ceci constitue un cadre unifié et basé sur la vraisemblance pour incorporer l'effet de covariables dans l'estimation d'abondance.

[MS-145]**Optimal Orthogonal Series Density Estimates
Estimation optimale de densités par séries orthogonales**Min JIANG & Serge B. PROVOST, *University of Western Ontario*

An improved methodology for obtaining orthogonal series density estimates is presented. Semiparametric density estimates that are expressible as the product of base density function and a linear combination of orthogonal polynomials are being considered. Various approaches are proposed for selecting the base density function and the degree of the polynomial adjustment so that the estimates become 'optimal' in their class. A simulation study indicates that the integrated squared error determined from the resulting density estimates is generally smaller than that associated with other estimates currently in use. The proposed density estimation methodology is applied to several univariate and bivariate datasets.

Une méthode améliorée pour obtenir des estimateurs de densités par séries orthogonales est présentée. Les estimateurs de densité semi-paramétriques pouvant être exprimés comme le produit d'une fonction de densité de base et d'une combinaison linéaire de polynômes orthogonaux sont considérés. Plusieurs approches sont proposées pour sélectionner la densité de base et le degré de l'ajustement polynomial de façon à ce que les estimateurs deviennent optimaux dans leur classe. Une étude de simulation indique que l'erreur quadratique intégrée obtenue à partir des estimateurs de densité résultants est généralement plus petite que celle associée aux autres estimateurs présentement utilisés. La méthode d'estimation de densité proposée est appliquée à plusieurs jeux de données univariés et bivariés.

[MS-146]**Optimizing Costly Functions with Simple Constraints: A Limited-Memory Projected Quasi-Newton Algorithm****Optimisation de fonctions coûteuses sous contraintes simples : un algorithme de projection quasi-Newton à mémoire limitée**Mark SCHMIDT, EWOUT VAN DEN BERG, Friedlander MICHAEL & Murphy KEVIN, *University of British Columbia*

We can formulate parameter estimation in many statistical models as constrained optimization problems. However, standard constrained optimization methods are not particularly suited for solving these problems; They have a large number of variables and evaluating the likelihood can be very complex, while the constraints are typically simple. We give a projected quasi-Newton method designed for these types of problems that can estimate millions of parameters with only a small number of likelihood evaluations. The method substantially improves on state-of-the-art methods for problems such as learning the structure of graphical models with group-Lasso constraints.

Beaucoup de problèmes d'estimation de paramètres pour des modèles statistiques peuvent être reformulés en termes de problèmes d'optimisation avec contraintes. Malgré les contraintes souvent simples, les méthodes standard d'optimisation ne sont généralement pas appropriées pour ce type de problèmes. Ces derniers comprennent typiquement une grande quantité de variables et nécessitent plusieurs évaluations coûteuses de la vraisemblance des paramètres. Nous présentons une méthode de projection quasi-Newton conçue pour ce type de problèmes pouvant estimer des millions de paramètres avec très peu d'évaluations de la vraisemblance. Nous démontrons comment cette méthode est plus performante que les techniques de pointe pour résoudre des problèmes tels que la découverte de structure dans les modèles graphiques avec les contraintes de Lasso de groupe.

[MS-147]**On High Quantile Estimation for Heavy Tailed Distributions****Estimation des quantiles supérieurs pour distributions à ailes relevées**Zheng ZHANG & Mei Ling HUANG, *Brock University*

The paper studies high quantile estimation for heavy tailed distributions. High quantile estimation has many applications, such as insurance claims and financial markets. There are theoretical difficulties to reduce bias and variance of estimating high quantiles since the moments don't always exist in heavy-tailed distributions. This paper proposes a nonparametric weighted estimation method and other methods to avoid those difficulties. We derive the theoretical efficiency functions of the new estimators relative to existing Weissman estimators. Monte Carlo simulations confirm the theoretical results. A real world example shows that the proposed estimators are more efficient and fit data well.

Cet article étudie l'estimation des quantiles supérieurs pour distributions à ailes relevées. L'estimation des quantiles supérieurs a plusieurs applications, notamment en réclamations d'assurances et en marchés financiers. Il y a des difficultés théoriques liées à la réduction du biais et de la variance dans l'estimation de quantiles supérieurs puisque les moments n'existent pas toujours en présence de distributions à ailes relevées. Cet article présente une méthode d'estimation pondérée non paramétrique et d'autres méthodes pour éviter ces difficultés. Nous obtenons les fonctions d'efficacité théorique des nouveaux estimateurs relativement aux estimateurs de Weissman existants. Nos résultats théoriques sont confirmés par simulations de Monte-Carlo. Un exemple avec des données réelles montre que les estimateurs proposés ajustent mieux les données.

[MS-148]**Semiparametric Estimation for Sequential Survival Times****Estimation semi-paramétrique pour temps de survie séquentiels**Yildiz YILMAZ, *University of Waterloo*

Sequentially ordered survival times are of interest in many studies. We introduce an approach overcoming some well-known difficulties in modeling and analyzing sequential data such as induced dependent censoring, non-identifiability of the marginal distributions and some individuals never experiencing the first failure. We model the joint distribution of the successive survival times by using copula functions. We propose a semiparametric estimation method in which the copula parameter is estimated without assumptions on the marginal distributions. The methodology developed is applied to a motivating example involving relapse and survival following colon cancer treatment.

Les temps de survie séquentiellement ordonnés sont d'intérêt dans plusieurs études. Nous présentons une approche surmontant plusieurs difficultés bien connues en modélisation et analyse de données séquentielles telles la censure dépendante induite, la non-identifiabilité des distributions marginales et le fait que certains individus ne subissent jamais la première défaillance. Nous modélisons la distribution conjointe des temps de survie successifs par des fonctions copules. Nous proposons une méthode d'estimation semi-paramétrique où les paramètres des copules sont estimés sans hypothèses sur les distributions marginales. La méthode développée est appliquée à un exemple de survie et rechute suivant un traitement pour le cancer du côlon.

[MS-149]**Some Issues in Estimation for Population Projections using Microsimulations****Quelques problèmes en estimation de projections de populations au moyen de microsimulations**IVAN CARRILLO-GARCIA & Milorad KOVACEVIC, *Statistics Canada*

In microsimulations for population projections one is often interested in estimating an event probability. A request for microsimulation flexibility leads

En microsimulations pour projections de populations il est souvent d'intérêt d'estimer une probabilité d'évènement. Une demande pour plus de flexibilité en microsimulations

to a two-step method in which estimated base-risks are adjusted by factors based on relative risks. The concepts of odds ratios and relative risks are contrasted under these settings. To calculate the relative risks, we want to set the whole population (i.e. population average) as the reference category. Besides the obvious risk ratio with the denominator being an average risk for the population, we examine some alternative ways of duplicating the population for estimation of relative risks.

mène à une méthode à deux étapes où les risques de bases estimés sont ajustés par des facteurs basés sur les risques relatifs. Les concepts de rapports de cotes et de risques relatifs sont comparés dans ce contexte. Pour calculer les risques relatifs, nous voulons prendre la population entière (c'est-à-dire la moyenne de la population) comme catégorie de référence. Outre le rapport de risques évident ayant le risque moyen pour la population comme dénominateur, nous examinons certaines façons alternatives de dupliquer la population pour l'estimation de risques relatifs.

Session 06K

Mardi 2 juin • Tuesday, June 2, 10:30 – 12:00

IRC Lobby

Poster Session II

Séance d'affichage II

[MS-150]

BARS: An R Package for Binary Adaptive Rejection Sampling

BARS : une librairie R pour échantillonnage par rejet adaptatif binaire

Tingting GOU & Duncan MURDOCH, *University of Western Ontario*

Binary adaptive rejection sampling (BARS) subdivides a sample space in order to approximate a target density in a piecewise manner by proposals that are easier to sample from. In the multivariate case the proposals are independent in each coordinate. The BARS package implements this algorithm; in this poster we describe how it is organized and give examples of its usage, showing one case where BARS is extremely effective, and another where its application is more difficult.

L'échantillonnage par rejet adaptatif binaire (BARS) divise un espace échantillonnal de façon à approximer une densité cible par morceaux au moyen de distributions instrumentales à partir desquelles il est facile d'échantillonner. Dans le cas multivarié, les lois instrumentales sont indépendantes dans chaque coordonnée.

La librairie BARS met en oeuvre cet algorithme ; dans cette affiche nous décrivons comment elle est organisée et donnons des exemples d'utilisation, notamment un cas où elle est très efficace et un autre où son application est plus difficile.

[MS-151]

The Information in One Prior Relative to Another

L'information dans une loi a priori relative à une autre

Gun Ho JANG & Michael EVANS, *University of Toronto*

A question of some interest is how to characterize the amount of information that a prior puts into a statistical analysis. Rather than a general characterization, we provide an approach to characterizing the amount of information a prior puts into an analysis, when compared to another base prior reflecting the current available information. Our purpose is to characterize priors that can be used as conservative inputs to an analysis, relative to the base prior, in the sense that they put less information into the analysis. An a priori measure of prior-data conflict is used as the basis for this development.

La quantification de la quantité d'information apportée par une loi a priori dans une analyse statistique est d'un certain intérêt. Plutôt qu'une caractérisation générale, nous donnons une approche pour la caractérisation de la quantité d'information apportée par une loi a priori dans une analyse statistique comparativement à une loi a priori de base reflétant l'information alors disponible. Notre but est de caractériser les lois a priori pouvant être utilisés de façon conservative relativement à la loi de base, dans le sens qu'elles apportent moins d'information dans l'analyse. Une mesure a priori des conflits données-a priori est utilisée comme base pour ce développement.

[MS-152]**Does Spatial Sampling Remove Spatial Autocorrelation while Preserving Inferences from Environmental Data? The Case of Roads and Forested Areas in China****L'échantillonnage spatial enlève-t-il la corrélation spatiale tout en conservant les inférences obtenues à partir de données environnementales ? Le cas des routes et forêts en Chine**

John GIBSON, *University of Waikato*, Xiangzheng DENG, *Chinese Academy of Sciences*, Scott ROZELLE, *Stanford University*, Jikun HUANG, *Chinese Academy of Sciences*

Statistical models of land cover change use samples of pixels from remote sensed satellite data. Little is known about how sampling affects spatial autocorrelation and inferences. Using forest cover data from Jiangxi province, China (166,000 1 km² pixels) we sample (i) every 4th, 8th or 16th pixel in a grid, (ii) random samples with as many pixels as the grid samples, or (iii) pixel aggregations into 4km × 4km, 8km × 8km and 16km × 16km areas. Residual spatial autocorrelation and inferences in a regression model of the effect of roads on forests are compared for each sampling strategy with results from the full dataset.

Les modèles statistiques du changement de couverture végétale utilisent des échantillons de pixels de télédétection satellite. Nous savons peu sur l'effet de l'échantillonnage sur l'autocorrélation spatiale et sur les inférences. À partir de données de couverture forestière pour la province de Jiangxi en Chine (166,000 pixels d'un km²) nous échantillonnons (i) chaque 4e, 8e ou 16e pixel dans une grille, (ii) des pixels aléatoires avec autant de pixels qu'en (i), (iii) des agrégats de pixels de 4km × 4km, 8km × 8km et 16km × 16km. L'autocorrélation spatiale résiduelle et les inférences sous un modèle de régression de l'effet des routes sur les forêts sont comparées, pour chaque stratégie d'échantillonnage, aux résultats utilisant le jeu de données en entier.

[MS-153]**Filtering Vs. Support Vector Machine Forecasters: A Simulation Study****Filtrage contre prédicteurs par machines à vecteurs de support : une étude de simulation**

Luke BORNN, *University of British Columbia*, Ingo STEINWART & Marian ANGHEL, *Los Alamos National Labs*

We examine two tools for forecasting in dynamic systems – filtering and support vector machine (SVM) forecasters. While the former relies on knowledge of the dynamics of the system as well as training data, the latter relies solely on training data. Despite the additional knowledge used by the filtering method, we demonstrate that the SVM forecaster provides better prediction accuracy under a broad range of conditions.

Nous étudions deux outils pour la prédiction dans les systèmes dynamiques – le filtrage et les prédicteurs par machine à vecteurs de support (MVS). Tandis que le premier dépend à la fois de la connaissance de la dynamique du système et des données d'apprentissage, la seconde ne dépend que des données d'apprentissages. Malgré la connaissance supplémentaire utilisée par les méthodes de filtrage, nous montrons que les prédicteurs par MVS permettent une plus grande précision de prédiction sous un vaste ensemble de conditions.

[MS-154]**A Class of Generalized Shrunken Least Squares Estimators in Linear Models****Une classe d'estimateurs par rétrécissement des moindres carrés généralisés pour modèles linéaires**

Xiaoming LIU & Liquan WANG, *University of Manitoba*

Modern data analyses often involve large number of variables, which gives rise to the problem of multicollinearity in regression models. It is well-known that in a linear model, when the design matrix X is nearly singular, then the ordinary least squares (OLS) estimator may perform poorly. To overcome this problem, many linear or nonlin-

Les analyses de données moderne impliquent souvent un grand nombre de variables, causant un problème de multicollinéarité dans les modèles de régression. Il est bien connu que dans un modèle linéaire, lorsque la matrice de plan d'expérience X est presque singulière, les estimateurs des moindres carrés ordinaires (MCO) peuvent mal performer. Pour surmonter ce problème, plusieurs estimateurs linéaires

ear biased estimators are studied. In this paper we consider a class of generalized shrunken least squares estimators that includes many other linear biased estimators recently proposed in the literature. Moreover, we compare these estimators under the matrix mean square error criterion.

ou non linéaires biaisés sont étudiés. Dans cet article nous considérons une classe d'estimateurs des moindres carrés par rétrécissement généralisés qui inclut plusieurs autres estimateurs linéaires biaisés récemment proposés dans la littérature. De plus, nous comparons ces estimateurs sous le critère de l'erreur quadratique moyenne matricielle.

[MS-155]

Designing to Estimate Certain Parameters or Parametric Functions Independently of Others

Planification afin d'estimer certains paramètres ou certaines fonctions paramétriques indépendamment des autres

Mohammad CHOWDHURY & Saumen MANDAL, *University of Manitoba*, Ben TORSNEY, *University of Glasgow*

In many regression designs it is desired to estimate certain parameters or parametric functions independently of others, by setting the covariances or correlations between relevant parameter estimates to zero. This is an example of an optimal regression design problem in which we need to calculate an optimizing probability distribution. For finding such designs we minimize the squared covariances between the estimates of the parameters of interest. To construct the designs, we use a class of multiplicative algorithms, indexed by a function which satisfies certain conditions. In conclusion, we will discuss some properties of the algorithm and report the algorithmic results.

Dans plusieurs plans de régression, il est souhaitable d'estimer certains paramètres ou fonctions paramétriques indépendamment des autres, en fixant les covariances ou corrélations entre les estimateurs des paramètres concernés à zéro. Ceci est un exemple de problème de planification de régression optimale où nous voulons calculer une distribution d'optimisation. Pour trouver de tels plans, nous minimisons les covariances carrées entre les estimateurs des paramètres d'intérêt. Pour construire les plans, nous utilisons une classe d'algorithmes multiplicatifs indexés par une fonction satisfaisant certaines conditions. En conclusion, nous allons aborder certaines propriétés de l'algorithme et présenter les résultats algorithmiques.

[MS-156]

flowClust: A Clustering Software for Automated Gating of Flow Cytometry Data

flowClust : une librairie de classification pour automatiser le portillonnage de données de cytométrie en flux

Kenneth Lo, *University of British Columbia*, Ryan BRINKMAN, *BC Cancer Research Centre*, Raphael GOTTARDO, *Institut de recherches cliniques de Montreal*

There is an increasing demand for an automated analysis platform to parallel the technological advancement of the high-throughput data-generation platform in flow cytometry (FCM). Currently, the analysis relies on sequentially selecting regions in 2-D graphical projections to extract cell populations of interest, which is manually time-consuming and ignores the high-dimensionality of FCM data. In view of this, we have developed an R package flowClust to implement a robust model-based clustering methodology to automate the gating analysis. Graphical functionalities are available to users for visualizing the results of such an analysis. The software is available for download from Bioconductor (<http://bioconductor.org>).

Il y a une demande croissante pour une plateforme d'analyse automatisée pour accompagner le développement technologique de la plateforme de génération de données à haut débit de cytométrie en flux (CMF). Présentement, l'analyse est basée sur la sélection séquentielle de régions dans des projections graphiques bidimensionnelles afin d'extraire les cellules d'intérêt, ce qui demande beaucoup de temps et ignore la dimensionnalité élevée des données de CMF. À cette fin, nous avons développé une librairie R flowClust pour mettre en oeuvre une méthodologie robuste de classification à base de modèles pour automatiser l'analyse de portillonnage. Des fonctionnalités graphiques sont disponibles pour que les usagers puissent visualiser les résultats d'une telle analyse. La librairie est disponible pour téléchargement sur le site de Bioconductor (<http://bioconductor.org>).

[MS-157]**Estimating and Pretesting in Multi-factors Ornstein-Uhlenbeck Processes****Estimation et pré-test dans les processus multivariés d'Ornstein-Uhlenbeck**Sévérien NKURUNZIZA & Ejaz AHMED, *University of Windsor*

We consider the estimation problem of the drift parameter matrix in a multi-factors Vasicek model (Ornstein-Uhlenbeck process). We also develop estimation theory for the drift parameters under restrictions. The asymptotic properties of both unrestricted and restricted maximum likelihood estimators (MLE) are established. Further, we consider shrinkage and pretest estimators when the restrictions are suspected. We examine the relative performance of the estimators under an asymptotic setup. We appraise the properties of the listed estimators based on a simulation study. Finally, we illustrate the use of suggested strategies by analyzing the interbank interest rates data of three European countries.

Nous considérons un problème d'estimation paramètre matriciel de dérive dans un modèle multi-facteurs de Vasicek (ou Ornstein-Uhlenbeck). Plus précisément, nous nous intéressons au cas où le paramètre de dérive serait susceptible d'être restreint. Les propriétés asymptotiques de l'estimateur du maximum de vraisemblance sont établies. Aussi, nous élaborons les estimateurs robustes à rétrécissement lorsque les restrictions sont suspectées et évaluons les performances relatives de ces estimateurs. Les résultats théoriques obtenus concordent bien avec ceux des simulations. Finalement, nous illustrons la mise application de la méthode suggérée en analysant les données réelles des taux d'intérêt interbancaires des trois pays européens.

Session 07**Mardi 2 juin • Tuesday, June 2, 13:30 – 14:30****IRC 6****Canadian Journal of Statistics Award Address****Allocution du récipiendaire du prix de la Revue canadienne de statistique****[MS-158]****Proportional Hazards Models based on Biased Samples and Estimated Selection Probabilities****Modèles à risques proportionnels basés sur des échantillons biaisés et les probabilités de sélection estimées**Qing PAN, *George Washington University*, Douglas SCHAUBEL, *University of Michigan*

Observational studies often consist of unrepresentative (biased) samples. Although bias can be avoided by weighting included subjects by the inverse of their respective selection probabilities, the weights are often unknown. We propose a two-stage method for estimating Cox model parameters in the presence of biased samples. At the first stage, selection probabilities are estimated using logistic regression. At the second stage, a weighted proportional hazards model is fitted to the biased sample. Asymptotic properties are derived, accounting for the randomness of the weights. Finite sample properties are assessed through simulation. The proposed methods are applied to national kidney transplant data.

Les études observationnelles consistent souvent en un ensemble d'échantillons non représentatifs (biaisés). Bien que le biais puisse être évité en pondérant les sujets inclus par l'inverse de leurs probabilités de sélection respectives, les poids sont souvent inconnus. Nous proposons une méthode à deux étapes pour l'estimation des paramètres du modèle de Cox en présence d'échantillons biaisés. Dans un premier temps, les probabilités de sélection sont estimées par régression logistique. Ensuite, un modèle de risques proportionnels pondéré est ajusté à l'échantillon biaisé. Des propriétés asymptotiques sont obtenues en tenant compte du caractère aléatoire des poids. Les propriétés sous échantillon fini sont évaluées par simulations. Les méthodes proposées sont appliquées à des données nationales de transplantation de reins.

Session 08A**Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****IRC 1****Quantitative and Qualitative Challenges in Climate Change and Ecosystem Sustainability Research****Problèmes quantitatifs et qualitatifs en changement climatique et recherches sur la durabilité des écosystèmes****[MS-159]****Eco-epidemiological Enquiry under Global Ecological Change****Enquête éco-épidémiologique sous changements écologiques globaux**Colin SOSKOLNE, *University of Alberta*, Maud HUYNEN, *Maastricht University*, Brian LADD, *Alberta Health Services*, Pim MARTENS, *Maastricht University*

Eco-epidemiology addresses the relationships between human health and the dynamics of global ecological change. Under escalating global change, new conceptual and analytical tools are needed if we are to contribute to the development of prevention and adaptation strategies for reducing negative health effects of changing ecosystems. This presentation provides an overview of methodological approaches and tools for advancing eco-epidemiological research. Several Integrated Assessment methods are discussed (integrated modeling, scenario analysis, and participatory methods), each combining knowledge elements from various disciplines into an analytical framework for identifying the cross-linkages and interactions among different contextual circumstances and processes underlying complex (eco-epidemiological) phenomena.

L'éco-épidémiologie traite des relations entre la santé humaine et la dynamique des changements écologiques globaux. Sous des changements globaux croissants, de nouveaux outils conceptuels et analytiques sont requis pour le développement de stratégies de prévention et d'adaptation visant à réduire les effets négatifs sur la santé des changements dans les écosystèmes. Cette présentation fait la revue des approches et outils méthodologiques pour faire avancer la recherche éco-épidémiologique. Plusieurs méthodes d'évaluation intégrée sont abordées (modélisation intégrée, analyse de scénarios, méthodes participatives), combinant chacune des éléments de connaissance de diverses disciplines dans un cadre analytique pour l'identification de couplages croisés et d'interactions entre différentes circonstances contextuelles et processus sous-jacents aux phénomènes (éco-épidémiologiques) complexes.

[MS-160]**Testing for Possible Climate Change-Caused Shifts in Forest Fire Ignitions****Tester l'hypothèse d'une évolution due aux changements climatiques dans le déclenchement de feux de forêts**Charmaine DEAN, Douglas WOOLFORD & Jiguo CAO, *Simon Fraser University*, David MARTELL, *University of Toronto*

We discuss methods for assessing the potential impact of climate change on forest fire risk. We use generalized additive mixture models and propose mixture models with normal and extreme components, and mixtures of binomials. In addition, we discuss a test for a shift in the proportion of observations arising from the component monitoring extreme events. We illustrate these methods using data for lightning-caused forest fire ignitions over a period of 42 years in a boreal forest of Ontario.

Nous discutons de méthodes d'évaluation de l'impact potentiel des changements climatiques sur le risque de feux de forêts. Nous utilisons des modèles de mélanges additifs généralisés et proposons des modèles de mélanges à composantes normales et extrêmes, et des mélanges de binomiales. De plus, nous discutons d'un test pour un changement dans la proportion d'observations provenant de la composante surveillant les événements extrêmes. Nous illustrons ces méthodes au moyen de données de feux de forêts déclenchés par la foudre sur une période de 42 ans dans une forêt boréale ontarienne.

[MS-161]**March Winds and April Showers: Impacts on Coastal Sockeye Salmon****Vents de mars et pluies d'avril : impacts sur le saumon rouge côtier**Rick ROUTLEDGE, *Simon Fraser University*

The talk will describe the role of statistics in ecosystem research in the context of an interdisciplinary research project on a British Columbia fjordal ecosystem. The research is aimed at substantiating a statistical association between March and April river flows and marine survival of the sockeye salmon that subsequently migrate down the inlet. We used functional regression and principal component analyses to establish the statistical correlation. The presentation will highlight the use of (i) such analyses in searching for ecosystem connections between physical and biological observations, and (ii) emerging statistical techniques for combining stochastic observations with deterministic differential equation models.

Cet exposé décrira le rôle de la statistique en recherche sur les écosystèmes dans le contexte d'un projet de recherche pluridisciplinaire sur un écosystème fjordien en Colombie-Britannique. La recherche vise à corroborer une association statistique entre l'écoulement des rivières en mars et avril et la survie marine du saumon rouge côtier qui migre sub-séquent dans le fjord. Nous avons utilisé la régression fonctionnelle et des analyses en composantes principales pour établir la corrélation statistique. Cette présentation va souligner l'utilisation de (i) telles analyses dans la recherche de liens entre observations physiques et biologiques dans un écosystème et (ii) les techniques statistiques émergentes pour combiner des observations stochastiques et des modèles à équations différentielles déterministes.

[MS-162]**Discussant****Modérateur**Peter GUTTORP, *University of Washington***Session 08B****Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****IRC 4****Quality Improvement in Healthcare****Amélioration de la qualité en soins de santé****[MS-163]****Overcoming Barriers to the Use of SPC by Front Line Quality Improvement Teams****Surmonter les barrières à l'utilisation de la MSP par des équipes sur le terrain d'amélioration de la qualité**Paul CASCAGNETTE, *Health Quality Council*

Shewhart charts provide superior sensitivity to process change. Implementation of control charts requires an aptitude for and facility with the use of computer software. Often, persons on the front lines have neither the time nor training resources to learn to use control charts. The run chart provides an easy to implement solution for quality improvement initiatives. Statistical evidence around required subgroup size is well documented for control charts, but not for run charts. Simulations showing minimum subgroup size and type I error

Les cartes de Shewart fournissent une sensibilité accrue aux changements de processus. La mise en oeuvre de cartes de contrôle nécessite une aptitude pour l'utilisation de programmes informatiques. Souvent, les personnes aux premières lignes n'ont ni le temps ni les ressources pour apprendre à utiliser des cartes de contrôle. Le graphe des séquences est alors une solution facile à mettre en oeuvre pour des projets d'amélioration de la qualité. L'évidence statistique concernant la taille requise pour les sous-groupes est bien documentée en ce qui concerne les cartes de contrôle, mais pas pour les graphes de séquences. Nous présentons

rates for run charts for common distributions and desired levels of improvement will be presented.

des simulations montrant la taille minimale des sous-groupes et la probabilité d'erreur de type I pour des graphes de séquences pour des distributions usuelles.

[MS-164]

Experiences With Quality Improvement in a New Not-for-profit Nursing home

Expériences en amélioration de la qualité dans une nouvelle maison de soins infirmiers sans but lucratif

Charlie GOLDSMITH, *McMaster University*

Shortly after opening, a nursing home with 210 clients was approached to have sessions for the staff to conduct quality improvement projects. A team of 11 staff was assembled with a list of 70 projects, whittled down to 5 projects; we started on one: reducing staff injuries from lifting clients. A second project: client restraints and a third: client walking distances were started and are in various stages of completion. This talk will outline each project, show the statistical and management tools used for each and discuss how the just-in-time teaching works to improve nursing home care quality.

Peu après son ouverture, une maison de soins infirmiers à 210 patients fut abordée dans le but d'organiser des projets d'amélioration de la qualité. Une équipe de 11 membres du personnel fut assemblée avec une liste de 70 projets, réduite par la suite à 5 projets. Nous avons commencé par un premier projet visant à réduire les blessures chez le personnel causées par le soulèvement des patients. Un deuxième projet portant sur les mesures de contention des patients, et un troisième, portant sur la distance de marche des patients, ont débuté et sont rendus à divers stades. Cet exposé va décrire les grandes lignes de chaque projet, présenter les outils statistiques et de gestion utilisés pour chacun et discuter de l'enseignement de type juste-à-temps dans le but d'améliorer la qualité des soins.

[MS-165]

Case-crossover Designs for Studies of Medical Errors and 'Near Misses'

Plans cas-croisés pour études d'erreurs médicales et de quasi-erreurs

Malcolm MACLURE, *University of BC and BC Ministry of Health Services*

'Patient Safety Culture' shifts from blaming people to investigating circumstances. Likewise, 'Why them?' (the question in case-control studies) shifts to 'Why now?' in crossover studies. Case-crossover studies resemble retrospective crossover experiments. Between-person confounding and control-selection bias are eliminated because each subject provides self-matched control-time(s). Non-cases and non-crossers are automatically excluded. Studies of sharps-related injuries in healthcare workers and insulin-dosing errors by intensive-care nurses illustrate estimation of relative risks associated with fatigue. A similar 'case-series method' - cohort analysis conditioned on likelihoods of becoming cases - is advantageous for studying well-measured single-point exposures, e.g. vaccine safety.

La 'culture de la sécurité des patients' nous incite à étudier les circonstances plutôt que de blâmer certaines personnes. Pareillement, 'pourquoi eux?' (la question en études cas-contrôles) devient 'pourquoi maintenant?' en études croisées. Les études cas-croisés ressemblent aux expériences rétrospectives croisées. La confusion inter-individus et le biais de sélection des contrôles sont éliminés, car chaque sujet apporte son propre temps contrôle. Les non-cas et les cas non croisés sont automatiquement exclus. L'étude de blessures causées par objets tranchants chez les travailleurs en soins de santé et d'erreurs de dosage d'insuline par infirmières en soins intensifs illustre l'estimation des risques relatifs associés à la fatigue. Une méthode 'cas-séries' similaire - l'analyse cohorte conditionnée sur la vraisemblance de devenir des cas - est avantageuse pour l'étude d'expositions ponctuelles bien mesurées, comme par exemple la vaccino-vigilance.

Session 08C**Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****MSL 102****Stochastic Modeling of Disordered Spatial Patterns****Modélisation stochastique des structures spatiales désordonnées****[MS-166]****The Statistics of Self-avoiding Walk and Polygon Models of Polymers****La statistique des modèles de marches et polygones autoévitant pour polymères**Christine SOTEROS, *University of Saskatchewan*

Self-avoiding walk (SAW) and polygon (SAP) models are used to study polymers. For this, distinct SAWs (or SAPs) on a lattice represent conformations of a linear (or ring) polymer. Without interactions, equal-length conformations are equally likely and average conformational properties are of interest. For example, motivated by modelling DNA, there is interest in average entanglement complexity. With interactions, phase transitions are of interest. For random copolymers (having more than one type of monomer) interactions depend on monomer type and hence phase transitions and average conformations differ from the homopolymer case. I will review applications of stochastic modelling to such models.

Les modèles de marches (MAÉ) et de polygones (PAÉ) autoévitant sont utilisés en étude des polymères. Dans ce contexte, des MAÉ (ou PAÉ) distincts sur un treillis représentent les conformations d'un polymère linéaire (ou anneau). Sans interactions, les conformations de mêmes longueurs sont équiprobables et les propriétés moyennes des conformations sont d'intérêt. Par exemple, en modélisation de l'ADN, on s'intéresse à la complexité moyenne des accrochages entre les chaînes. En présence d'interactions, on s'intéresse aux transitions entre les phases. Pour des copolymères aléatoires (ayant plus d'un type de monomère) les interactions dépendent du type de monomère et les transitions de phases et les conformations moyennes diffèrent donc du cas homopolymère. Je fais une revue des applications de la modélisation stochastique à de tels modèles.

[MS-167]**Determinantal Processes – Models Point Repulsion****Processus déterminantaux – modèles ponctuels avec répulsion**Balint VIRAG, *University of Toronto*

The simplest probabilistic models for point processes with repulsion are determinantal processes. They arise naturally in many forms, such as eigenvalues of random matrices, zeros of random functions, certain statistics of random permutations, Coulomb gas particles and even random spanning trees. There is a surprisingly elegant and simple mathematical theory behind these processes that makes them easy to simulate and understand. I will give a brief overview.

Les modèles probabilistes les plus simples pour processus ponctuels avec répulsion sont les processus déterminantaux. Ils se retrouvent naturellement sous plusieurs formes, comme les valeurs propres de matrices aléatoires, les zéros de fonctions aléatoires, certaines statistiques de permutations aléatoires, les particules de gaz de Coulomb et même les arbres de longueur aléatoire. Il y a une théorie mathématique surprenamment simple et élégante derrière ces processus les rendant faciles à simuler et à comprendre. Je vais en présenter un bref résumé.

Session 08D**Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****IRC 3****Analysis of Correlated Data with Excessive Zeros****Analyse de données corrélées à zéros excessifs****[MS-168]****Compound Poisson Distributions for Multivariate Frailty
Distributions de Poisson composées pour fragilité multivariée**Odd AALEN, *University of Oslo, Norway*, Tron A MOGER, *University of Oslo*

Frailty models are often used to model heterogeneity in survival analysis. The compound Poisson distribution is a useful formulation of frailty variation in such models. Often one would expect frailty to be zero for some individuals, corresponding to lack of susceptibility to a certain disease or event. The compound Poisson distribution naturally incorporates such a zero group. At the same time this distribution fits well with the standard frailty framework due to explicit representation of the Laplace transform. The compound Poisson distribution is extended to the multivariate setting by randomizing the Poisson parameter in the compound Poisson distribution.

Les modèles de fragilité sont souvent utilisés pour modéliser l'hétérogénéité en analyse de survie. La distribution de Poisson composée est une formulation utile de la variation de la fragilité dans de tels modèles. On s'attendrait à ce que la fragilité soit nulle pour certains individus, correspondant à un manque de susceptibilité à une certaine maladie ou à un certain événement. La distribution de Poisson composée incorpore naturellement un tel groupe. Cette distribution a aussi l'avantage de bien s'intégrer au cadre général des modèles de fragilité grâce à la représentation explicite de sa transformée de Laplace. La distribution de Poisson composée est étendue au cas multivarié en randomisant le paramètre de Poisson dans la distribution de Poisson composée.

[MS-169]**Modeling Heterogeneity in Clustered Count Data With Extra Zeros Using Compound Poisson Random Effect****Modélisation de l'hétérogénéité dans des données de dénombrement en grappes avec surplus de zéros au moyen d'un effet aléatoire de Poisson composé**M. Tariqul HASAN & Renjun MA, *University of New Brunswick*, Gary SNEDDON, *Memorial University of Newfoundland*

In medical studies, heterogeneities in clustered count data have been traditionally modeled by positive random effects in Poisson mixed models; however, excessive zeros often occur in clustered count data. In this work, we consider a three level random effects Zero-inflated Poisson model for health care utilization data where data are clustered by both subjects and families. To accommodate zero and positive components in the count responses, we model subject-level random effects by compound Poisson distribution. Our model displays variance components decomposition which clearly reflects the hierarchical structure of clustered data. A quasi-likelihood approach has been developed to estimate regression parameters.

En études médicales, l'hétérogénéité parmi des données de dénombrement en grappes a traditionnellement été modélisée par des effets aléatoires positifs dans des modèles mixtes de Poisson. Cependant, des zéros excessifs sont souvent rencontrés en données de dénombrement en grappes. Dans ce travail, nous considérons un modèle de Poisson à effets mixtes à trois niveaux avec surplus de zéros pour des données d'utilisation des soins de santé où les données sont regroupées par sujets et par familles. Afin de tenir compte des composantes nulles et positives dans les réponses de dénombrement, nous modélisons les effets aléatoires au niveau des sujets par la distribution de Poisson composée. Notre modèle montre une décomposition des composantes de la variance reflétant clairement la structure hiérarchique des données en grappes. Une approche par quasi-vraisemblance a été développée afin d'estimer les paramètres de régression.

[MS-170]

Sample Size Determination for Mixed-Effects Longitudinal Logistic Regression Models**Détermination de la taille échantillonnale pour modèles de régression logistique à effets mixtes longitudinaux**

Dulal BHAUMIK & Kush KAPUR, *University of Illinois at Chicago*, Anindya ROY, *University of Maryland Baltimore County*, Robert GIBBONS, *University of Illinois at Chicago*

We consider the problem of sample size determination for testing the group by time interaction parameter for two-level longitudinal mixed-effects logistic regression models. We propose a new method to estimate model parameters and compare the performance of this method in the context of hypothesis testing via simulation with marginal maximum likelihood (MML) and the generalized estimating equation (GEE) methods. We also derive a closed form solution for determining the necessary sample size in order to achieve a power of 80% while maintaining the Type I error rate (e.g. 5%). Finally, we consider the effect of attrition on the power function.

Nous considérons le problème de la détermination de la taille échantillonnale pour tester le paramètre d'interaction groupe par temps dans des modèles de régression logistique longitudinaux à effets mixtes à deux niveaux. Nous proposons une nouvelle méthode pour estimer les paramètres du modèle et comparons, au moyen de simulations, sa performance dans le contexte de tests d'hypothèses aux méthodes du maximum de vraisemblance marginale et des équations d'estimation généralisées. Nous obtenons également une solution analytique au problème de la détermination de la taille échantillonnale nécessaire pour atteindre une puissance de 80% en fixant le taux d'erreur de type I (par exemple à 5%). Finalement, nous considérons l'effet de l'attrition sur la fonction de puissance.

Session 08E**Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****IRC G65/66****Analysis and Modelling of Marine Ecological Data****Analyse et modélisation de données écologiques marines**

[MS-171]

Application of Sequential Data Assimilation to a Biogeochemical Ocean Model**Application de l'assimilation de données séquentielle à un modèle biogéochimique de l'océan**

Jann Paul MATTERN, Katja FENNEL & Mike DOWD, *Dalhousie University*

Large ocean datasets make data assimilation for marine models an attractive prospect to oceanographers and statisticians. Data assimilation from sources such as satellites or scientific cruises into high dimensional ocean models is a challenging task. We present an application of sequential data assimilation techniques, including the Sequential Importance Resampling and the ensemble Kalman filter to a biogeochemical ocean model. Following parameter optimization, we explore the effects of stochastic model parameters on model output for different ensemble runs. We present results of a cross validation for assimilation with different ensemble sizes in comparison with results of the model without data assimilation.

Les grands jeux de données océaniques rendent l'assimilation de données pour modèle maritimes attrayante aux yeux des océanographes et statisticiens. L'assimilation de données provenant de sources tels les satellites et les expéditions scientifiques dans des modèles océaniques de dimension élevée est une tâche difficile. Nous présentons une application de techniques d'assimilation de données séquentielles, incluant le rééchantillonnage d'importance séquentiel et le filtre d'ensemble de Kalman, à un modèle océanique biogéochimique. Après l'optimisation des paramètres, nous explorons les effets de paramètres stochastiques sur les sorties du modèle. Nous présentons les résultats d'une validation croisée pour l'assimilation avec différentes tailles d'ensembles en comparaison avec les résultats du modèle sans assimilation de données.

[MS-172]**Modelling Spring Phytoplankton Bloom Dynamics in a British Columbia Fjord****Modélisation de l'efflorescence dynamique du phytoplancton dans un fjord de la Colombie-Britannique**

Désirée TOMMASI, *University of British Columbia*, Rick ROUTLEDGE, *Simon Fraser University*, Brian HUNT & Evgeny PAKHOMOV, *University of British Columbia*

The phytoplankton dynamics in Rivers Inlet, a British Columbia fjord through which sockeye salmon migrate, were monitored daily in spring and early summer of 2007 and 2008. The goal was to assess interannual variability in the timing and biomass of the spring phytoplankton bloom. Maximum likelihood techniques were used to fit a curve describing the pattern of bloom initiation in each year and to assess spatial variability within the inlet. In both years, the bloom was synchronous across inlet stations, but in the stormy early spring of 2007, it was over two weeks later than in 2008.

La dynamique du phytoplancton de Rivers Inlet, un fjord de Colombie-Britannique à travers lequel migre le saumon rouge, a été surveillée quotidiennement au printemps et au début de l'été en 2007 et 2008. Le but était d'évaluer la variabilité interannuelle du temps et de la biomasse de l'efflorescence printanière du phytoplancton. La technique du maximum de vraisemblance a été utilisée pour ajuster une courbe décrivant le processus d'efflorescence pour chaque année et pour évaluer la variabilité spatiale à l'intérieur du fjord. Les deux années, l'efflorescence était synchrone aux différentes stations dans le fjord, mais elle a été retardée de deux semaines au printemps orageux de 2007 comparativement à 2008.

[MS-173]**Parameter Augmented Particle Filter/Smother Methods for Animal Tracking Data****Méthodes de filtres à particules/lisseuses augmentées de paramètres pour données de pistage d'animaux**

Greg BREED, Joanna MILLS FLEMMING, Mike DOWD & Christopher FIELD, *Dalhousie University*, Dan COSTA, *University of California Santa Cruz*

Electronic tags are popular means to measure movement of animals. Data they generate, however, are problematic because observations are often highly erroneous, confounding complex and stochastic behavioural dynamics. Recently, a state-space approach using MCMC in WinBUGS allowed fitting of behaviourally informative correlated random walk models to tracking data (Jonsen et al. 2005). Here we fit similar models to tracking data using a particle filter and smoother, resampling routine of Arulampalam et al. (2002), and parameter augmentation (Chen et al. 2005). The method performs as well as MCMC, but is faster and allows models to incorporate dynamic environmental fields as covariates.

Les marqueurs électroniques sont un moyen populaire de mesurer le déplacement d'animaux. Les données qu'ils génèrent, cependant, sont problématiques, car les observations sont souvent hautement erronées, confondant des dynamiques comportementales complexes et stochastiques. Récemment, une approche espaces-états par MCMC utilisant WinBUGS a permis d'ajuster de modèles de marches aléatoires corrélés à des données de pistage (Jonsen et al. 2005). Ici, nous ajustons des modèles similaires à des données de pistage au moyen d'un filtre à particules et d'une fonction lisseuse, du programme de rééchantillonnage d'Arulampalam et al. (2002), et l'augmentation paramétrique (Chen et al. 2005). La méthode performe aussi bien que les MCMC, mais est plus rapide et permet d'incorporer au modèle des champs environnementaux dynamiques comme covariables.

[MS-174]

Functional Data Analysis in Ecosystem Research: the Decline of Oweekeno Lake Sockeye**Aanalyse de données fonctionnelles en recherche en écosystèmes : le déclin du saumon rouge du lac Oweekeno**Laurie AINSWORTH, Rick ROUTLEDGE & Jiguo CAO, *SFU*

Functional regression is a natural tool for exploring the potential impact of the physical environment (continuously monitored) on biological performance (often assessed yearly). Functional regression analysis is considered in the context of a depressed sockeye salmon population in Rivers Inlet, BC. We discuss the use of generalized cross validation (GCV) for guiding decisions regarding the smoothing parameter. A formula is derived for simultaneous confidence bands for the functional regression coefficients and coverage probabilities are compared with those of the pointwise confidence bands. The analysis suggests a substantial, negative correlation between early spring river flow and sockeye salmon marine survival.

La régression fonctionnelle est un outil naturel pour explorer l'impact potentiel de l'environnement physique (continuellement surveillé) sur la performance biologique (souvent évaluée annuellement). L'analyse de régression fonctionnelle est considérée dans le contexte d'une population affaiblie de saumon rouge à Rivers Inlet, Colombie-Britannique. Nous discutons de l'utilisation de la validation croisée généralisée pour guider la prise de décision relative au paramètre de lissage. Une formule est obtenue pour des bandes de confiance simultanées pour les coefficients de régression fonctionnelle et leurs probabilités de couverture sont comparées à celles des bandes de confiance point-par-point. L'analyse suggère une corrélation négative substantielle entre le débit de début du printemps de la rivière et la survie du saumon rouge.

Session 08F**Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****MSL 101****Estimation Methods for Survey Data****Méthodes d'estimation pour données d'enquête**

[MS-175]

Changing the Methodology from Generic to Detail Allocation for Incorporated Business Tax Data**Changer la méthodologie de la désagrégation des données génériques aux détails pour les données fiscales des entreprises incorporées**Jessica ANDREWS, *Statistics Canada*

Tax data is steadily becoming a more important tool for Statistics Canada in the fight to reduce response burden and increase response rates. Before tax data can be used, however, many fields must be imputed as only 8 of 700 on the financial statements are required by the Canada Revenue Agency. The methodology which imputes missing fields has recently been changed resulting in many improvements to the data but also several implementation issues. This talk will discuss the impact of the new methodology and how new problems have arisen and been solved.

Les données fiscales deviennent un outil de plus en plus important pour Statistique Canada dans sa lutte pour réduire le fardeau de réponse et améliorer les taux de réponse. Avant que ces données puissent être utilisées, plusieurs champs doivent être imputés parce que l'Agence du Revenu du Canada n'exige que seulement 8 des 700 champs des états financiers soient remplis. La méthodologie pour imputer ces champs manquants a changé récemment et a entraîné plusieurs améliorations aux données mais aussi quelques problèmes lors de sa mise en place. L'impact de la nouvelle méthodologie et les solutions aux problèmes rencontrés seront présentés.

[MS-176]**Ratio Estimators Improved by Rao-Blackwellization under Adaptive Cluster Sampling
Estimateurs par le quotient améliorés par Rao-Blackwellisation dans le cadre de l'échantillonnage en grappes adaptatif**Chang-Tai CHAO, *National Cheng Kung University, Taiwan*

It is often that certain auxiliary variables would be observed in addition to the variable of primary interest in a sampling survey, and ratio estimator is the commonly used inference method which takes advantage of the correlation between the primary variable and the auxiliary variable for better estimation result. Various ratio estimators are possible and also helpful under adaptive cluster sampling. Still, the ratio estimators for adaptive cluster sampling proposed in the past can be further improved by Rao-Blackwellization method. Several improved ratio estimators will be proposed in this paper, and their performance will be discussed as well.

Il est fréquent que certaines variables auxiliaires soient observées en plus de la principale variable d'intérêt d'une enquête par sondage, et l'estimateur par le quotient est la méthode d'inférence courante utilisée pour tirer profit de la corrélation entre la variable principale et la variable auxiliaire pour produire de meilleures estimations. Plusieurs estimateurs par le quotient sont possibles et utiles dans le cadre de l'échantillonnage en grappes adaptatif. Malgré tout, les estimateurs par le quotient pour échantillonnage en grappes adaptatif proposés par le passé peuvent être améliorés par la Rao-Blackwellisation. Plusieurs estimateurs par le quotient améliorés seront présentés dans cet article, et leur performance sera aussi abordée.

[MS-177]**Linearization Variance Estimation for Two-phase Sampling under Mass Imputation
Estimation de la variance de linéarisation pour échantillonnage à deux phases avec imputation massive**Abdellatif DEMNATI, *Statistique Canada, J.N.K. RAO, Carleton University*

We consider two-phase sampling in which values of a variable of interest are observed only in the second-phase sub-sample. Values for the units not sampled in the second phase are mass imputed, using values from an administrative file when available and regression imputation otherwise. We study both naïve and design-consistent estimation of a total and obtain associated linearization variance estimators using Demnati-Rao (Survey Methodology, 2004) unified approach. The case of missing sub-sample values is also studied. Sample allocation issues are also studied.

Nous considérons l'échantillonnage à deux phases où les valeurs d'une variable d'intérêt ne sont observées que dans le sous-échantillon de seconde phase. Les valeurs pour les unités non échantillonnées dans la deuxième phase sont massivement imputées en utilisant des valeurs d'un fichier administratif lorsque disponible et l'imputation par régression sinon. Nous étudions à la fois l'estimation naïve et celle cohérente avec le plan d'échantillonnage de totaux et obtenons les estimateurs de variance par linéarisation associés en utilisant l'approche unifiée Demnati-Rao (Techniques d'enquête, 2004). Le cas de valeurs manquantes parmi le sous-échantillon est aussi étudié. Des enjeux touchant la répartition de l'échantillon sont aussi abordés.

[MS-178]**Applications of Quasi-Monte Carlo Methods in Inference for Complex Survey Data
Applications de méthodes quasi Monte-Carlo en inférence pour données d'enquête complexes**Stanislav KOLENIKOV, *University of Missouri, Columbia, USA*

We propose a new method for generating (approximately) balanced designs to estimate variances of complex survey estimators. The new method is based on quasi-Monte Carlo methods, such as mul-

Nous proposons une nouvelle méthode pour générer des plans de sondage (approximativement) équilibrés pour l'estimation des variances d'estimateurs de données d'enquêtes complexes. La nouvelle méthode est basée sur des méthodes

tidimensional Halton sequences. The method can be effectively used to create replication schemes in complex surveys with designs more complex than 2 PSU/strata. Several possible implementations of variance estimators for complex surveys are offered. A simulation shows that the QMC methods with additional balancing can achieve performance comparable to that of the balanced bootstrap.

quasi Monte-Carlo, telles les suites d'Halton multidimensionnelles. La méthode peut être utilisée efficacement pour créer des schémas de réplication en enquêtes complexes à plan plus complexe que deux UPÉ/strates. Plusieurs mises en oeuvre possibles des estimateurs de variance pour enquêtes complexes sont offertes. Une simulation montre que les méthodes QMC avec équilibrage additionnel peuvent atteindre une performance comparable à celle du bootstrap équilibré.

[MS-179]

Alternative Criteria for Data Quality Filtering in the ACS **Critères alternatifs en filtrage des données au sein de l'ACS**

Alfredo NAVARRO & Michael STARSINIC, *US Census Bureau*

The American Community Survey (ACS) filters out standard data tables based on a reliability criterion designed to minimize the release of unreliable estimates. As a result, about 25 percent of the tables are deemed to be unacceptable to be released. The quality filtering criteria relies on a series of rules for special cases such as medians and zero estimates. The main focus of this paper is to show the effects of the current filtering rules. It also explores the effects of alternative treatments for these estimates.

L'American Community Survey filtre les tableaux de données standards à partir d'un critère de fiabilité visant à minimiser la diffusion d'estimations peu fiables. Par conséquent, environ 25 % des tableaux sont jugés inacceptables pour publication. Le critère de filtrage est basé sur une série de règles pour cas particuliers telles les médianes et estimations de zéros. Le but principal de cet article est de montrer les effets des règles de filtrage actuelles. Il explore également les effets de traitements alternatifs pour ces estimations.

[MS-180]

Design Variations in Adaptive Web Sampling **Variations de plans d'expérience en échantillonnage en toile adaptatif**

Kyle VINCENT, & Steve THOMPSON, *Simon Fraser University*

Adaptive Web Sampling was introduced by Thompson (2006) in order to sample from hard-to-reach populations such as injection drug users. The strategy takes advantage of social relationships/links to recruit new members from the target population. The new methods produce estimates for population parameters which are comparable to those produced by standard methods such as simple random sampling. However, the adaptive techniques are less costly in time and effort. This talk explores new variations in the adaptive approach to sampling such populations. Simulation results show that these produce more efficient estimates than given by the original adaptive design.

L'échantillonnage adaptatif en toile a été présenté par Thompson (2006) afin d'échantillonner des populations difficiles à rejoindre, tels les utilisateurs de drogues injectables. Cette stratégie tire parti des relations et liens sociaux pour recruter de nouveaux sujets parmi la population cible. Les nouvelles méthodes produisent des estimés pour les paramètres de population comparables à ceux obtenus avec les méthodes habituelles comme l'échantillonnage aléatoire simple. Cependant, les techniques adaptatives sont moins coûteuses en termes de temps et d'efforts. Cet exposé explore de nouvelles variations dans l'approche adaptative de l'échantillonnage de telles populations. Des résultats de simulation montrent que ces méthodes produisent des estimés plus efficaces que ceux donnés par le plan adaptatif original.

Session 08G**Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****CHE 150****Statistical Genetics and Genomics II****Génétique et génomique statistiques II****[MS-181]****Generalized Association Study for Subjects Sampled from a Complex Pedigree****Étude d'association généralisée pour sujets à généalogie complexe**

Zeny FENG, *University of Guelph*, William WONG, *University of Waterloo*, Gao XIN, *York University*, Flavio SCHENKEL, *University of Guelph*

In genetic study, family-based association study is used to avoid the confounding effects due to the population structure. Recently, founder/isolated population based case-control association study is proposed as it effectively controls this confounding effect. We propose a generalized quasi-likelihood scoring approach for testing the association between markers and both the quantitative and binary trait when subjects are sampled from a complex pedigree or from an isolated/founder population. Our method will be applied to analysis the Holstein Cattle data set in North America in which complicated relationships among cattles and inbreeding factor of each cattle will be taking into account.

Les études d'association à base de familles sont utilisées pour éviter les effets confondants dus à la structure de la population. Récemment, les études d'association cas-contrôle basées sur des populations fondatrices ou isolées ont été proposées, car elles contrôlent cet effet confondant. Nous proposons une approche de scores basés sur la quasi-vraisemblance généralisée pour tester l'association entre les marqueurs et des traits quantitatifs ou binaires lorsque les sujets sont tirés de familles complexes ou de populations fondatrices ou isolées. Notre méthode sera appliquée à l'analyse du jeu de données de bovins Holstein en Amérique du Nord dans lequel de relations complexes entre les bêtes et le facteur de consanguinité de chaque bête seront pris en considération.

[MS-182]**Impact of Informative Prior in Discrete Bayesian Graphical Models: Application to Genome-Wide Association Studies****Impact de lois a priori informatives dans les modèles graphiques discrets bayésiens : application aux études sur l'ensemble du génome**

Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute & University of Toronto*, Jinnan LIU, *Samuel Lunenfeld Research Institute*, Adrian DOBRA, *University of Washington*, Helene MASSAM, *York University*

High-throughput sequencing technologies together with clinical and physiological data produce large amounts of biological information that is used for molecular phenotyping of complex diseases. Combined with efficient stochastic search algorithms, Bayesian graphical models allow an efficient evaluation of a large number of multi-marker models, where each model embodies complex interactions among the markers. Our goal is to use this framework to perform genome-wide association studies and identify important combinations of biomarkers that can predict complex disease-related phenotypes. Here, we propose a new approach to integrate expert biological knowledge into this framework and evaluate its impact on new genes discovery.

Les technologies de séquençage à haut débit combinées aux données cliniques et physiologiques produisent de grandes quantités d'informations utilisées pour le phénotypage moléculaire de nombreuses maladies. Associés à des algorithmes efficaces pouvant explorer l'espace des modèles très rapidement, les modèles graphiques bayésiens permettent une évaluation efficace d'un grand nombre de modèles à marqueurs multiples, où chaque modèle contient des interactions complexes entre les marqueurs. Notre but est d'utiliser cette approche pour les études d'association sur l'ensemble du génome et identifier les ensembles importants de biomarqueurs qui peuvent prédire des phénotypes liés aux maladies complexes. Ici, nous proposons une nouvelle approche pour intégrer des informations biologiques spécialisées dans ce cadre et évaluons leur impact sur la découverte de gènes nouveaux.

[MS-183]**Analysis of Family-based Genetic Association Study Data, a Bayesian Approach****Une approche bayésienne à l'analyse de données d'une étude d'association génétique basée sur les familles**

Xuekui ZHANG & Kevin MURPHY, *University of British Columbia*, Peter PARÉ & Andrew SANDFORD, *The James Hogg iCAPTURE Center, University of British Columbia*

We propose a Bayesian network to model Family-based genetic association study data. Our model can (1) incorporate available prior knowledge; (2) handle complex family structure (multiple children, missing parents, and multiple generations); (3) account for measurement (genotype and phenotype) errors; (4) give reasonable results when the number of informative families (or minor allele frequency) is low; (5) adjust for important covariates. By simply changing the component of our Bayesian network, our model can handle binary/numerical/survival/multivariate outcomes using a uniform model. Finally, we assess our methods using simulated datasets and a real dataset.

Nous proposons un modèle de réseau bayésien pour modéliser des données d'association génétique basée sur les familles. Notre modèle peut (1) incorporer des connaissances a priori disponibles ; (2) traiter des structures familiales complexes (enfants multiples, parents manquants, générations multiples) ; (3) tenir compte des erreurs de mesure de génotype et de phénotype ; (4) donner des résultats raisonnables lorsque le nombre de familles informatives (ou la fréquence d'allèles mineurs) est petit ; (5) ajuster pour des covariables importantes. En changeant simplement la composante de notre réseau bayésien, notre modèle peut traiter des réponses binaires/numériques/de survie/multivariée en utilisant un modèle uniforme. Finalement, nous évaluons nos méthodes au moyen de jeux de données simulés et d'un jeu de données réel.

[MS-184]**Effect Estimates in 2-stage Design: from Genome-wide Association to Sequencing****Estimateurs des effets en plans à deux étapes : de l'association sur tout le génome au séquençage**

Laura FAYE, *University of Toronto*, Shelley B BULL, *University of Toronto & Samuel Lunenfeld Research Institute, Mount Sinai Hospital*, Lei SUN, *University of Toronto*

Advancements in sequencing technology have made targeted sequencing following a genome wide association (GWA) study a feasible way to localize a potential causal variant. However, sequencing using all or part of the GWA samples lead to upward bias in effect estimate, while tagging SNPs leads to attenuation. We derive analytical expressions for such biases and provide numerical illustrations for both case-control and quantitative trait studies. Understanding this issue is informative for proper study design and analysis of sequencing data.

Grâce aux progrès en technologies de séquençage, le séquençage ciblé suivant une étude d'association sur tout le génome (ATG) est une façon possible de localiser une variante causale potentielle. Cependant, le séquençage utilisant les échantillons d'ATG mène à une surestimation dans les estimateurs des effets, alors que marquer les SNP mène à une atténuation. Nous obtenons des expressions analytiques pour de tels biais et les illustrons numériquement dans le cas d'études de traits quantitatifs et cas-contrôles. Comprendre ce problème est d'intérêt pour une bonne planification et analyse d'expériences avec données de séquençage.

[MS-185]**Can Gene-by-environment Interaction be Inferred from Parent-case Transmission Rates?****Peut-on inférer l'interaction gène-environnement à partir de taux de transmission parent-cas ?**

Ji-Hyung SHIN, Brad McNENEY & Jinko GRAHAM, *Simon Fraser University*

Most complex diseases result from an interplay between genes (G) and environmental attributes (E). Statistical interaction between G and E occurs

La plupart des maladies complexes résultent d'une interaction entre les gènes (G) et des caractéristiques environnementales (E). Il y a interaction statistique entre G et E

when genotype relative risks of disease vary with E . Since GxE interaction leads to variation with E in transmission rates of the risk allele, transmission rates have been used to make inference about GxE interaction. We investigate the validity of this practice by deriving theoretical transmission rates under standard population genetic assumptions and show that variation with E does not, in general, reflect the truth about GxE interaction.

lorsque les risques relatifs de contracter la maladie pour un génotype donné varient avec E . Puisque l'interaction GxE mène à une variation avec E des taux de transmission de l'allèle à risque, les taux de transmission ont été utilisés pour faire de l'inférence au sujet de l'interaction GxE. Nous étudions la validité de cette pratique en obtenant les taux de transmission théoriques sous présupposés habituels au sujet de la génétique de la population et montrons qu'une variation avec E ne reflète pas, de façon générale, la vérité au sujet de l'interaction GxE.

Session 08H

Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00

CHE 250

Bayesian Statistics and Applications Statistique et applications bayésiennes

[MS-186]

Posterior Consistency of Species Sampling Priors

Cohérence de la loi a posteriori pour lois a priori d'échantillonnage des espèces

Gun Ho JANG, *University of Toronto*, Jaeyong LEE & Sangyeol LEE, *Seoul National University*

Recently there has been increasing interest in species sampling priors, the nonparametric priors defined as the directing random probability measures of the species sampling sequences. We show that not all of the species sampling priors produce consistent posteriors. In particular, in the class of Pitman-Yor process priors, the only priors rendering posterior consistency are essentially the Dirichlet process priors. Under certain conditions, we also give a set of necessary and sufficient conditions for the posterior consistency for the general species sampling prior. Considered examples include the normalized inverse-Gaussian process, Poisson-Kingman partition and Gibbs partition.

Il y a un intérêt croissant pour les lois a priori d'échantillonnage des espèces, les lois a priori non paramétriques définies comme les mesures de probabilité aléatoires des suites d'échantillonnage des espèces. Nous montrons que toutes les lois a priori d'échantillonnage des espèces ne produisent pas des lois a posteriori cohérentes. En particulier, dans la classe des processus a priori de Pitman-Yor, les seules lois a priori résultant en une cohérence a posteriori sont essentiellement les processus a priori de Dirichlet. Sous certaines conditions, nous donnons aussi un ensemble de conditions nécessaires et suffisantes pour la cohérence a posteriori des a priori d'échantillonnage des espèces générales. Les exemples considérés comprennent les processus gaussiens inverses normalisés, la partition de Poisson-Kingman et la partition de Gibbs.

[MS-187]

Analysis of Tropical Cyclone Trends in the Atlantic

Analyse de la tendance des cyclones tropicaux dans l'Atlantique

Anne-Sophie CHAREST, *Carnegie-Mellon University*

One of the alleged signs of global warming is a believed increase in Tropical Cyclones (TC's), of which hurricanes are special cases. The difficulty in substantiating this claim is that the technical ability to detect hurricanes has increased markedly over the years. Thus, the increased number of recorded TC's is explained, at least in part, by bet-

Le réchauffement climatique serait à l'origine d'une augmentation du nombre de cyclones tropicaux, mieux connus sous le nom d'ouragans dans leur forme la plus intense. Il est toutefois difficile de valider rigoureusement cette hypothèse car les méthodes de détection plus efficaces peuvent expliquer, en partie, l'augmentation du nombre de cyclones recensés dans les dernières années. Nous proposons d'utiliser des

ter detection, and possibly also by an increased frequency of TC's. This study uses Bayesian methods to show the effect of different beliefs about trends in TC detection and trends in TC frequency.

méthodes bayésiennes pour étudier l'impact de différentes croyances concernant la fréquence des cyclones tropicaux et nos capacités de détection.

[MS-188]

Calibrating Predictions Based on a Selected Subset of Features from Bayesian Gaussian Classification Models

Calibration de prévisions basée sur un sous-ensemble de caractéristiques de modèles de classification bayésiens gaussiens

Longhai LI, *University of Saskatchewan*

There is an increasing demand for efficient, accurate and precise methods for predicting a discrete response with a great number of features. For computational reasons, it is necessary to select a subset of features before fitting a statistical model. However, such selection procedure will result in overconfident predictions. The challenge in applying the general solution presented by Li, Zhang, and Neal (2008) is the computation of the probability of a feature failing to pass the selection filter. In this paper, I show that efficient computation of the adjustment factor is possible for Bayesian Gaussian classification models for continuous features.

Il y a une demande croissante pour des méthodes efficaces et précises de prévision d'une variable réponse discrète ayant de nombreuses caractéristiques. Pour des raisons calculatoires, il est nécessaire de sélectionner un sous-ensemble de caractéristiques avant d'ajuster un modèle statistique. Cependant, une telle méthode de sélection entraînera des prévisions trop confiantes. Le défi dans l'application de la solution générale présentée par Li, Zhang, et Neal (2008) est de calculer la probabilité qu'une caractéristique ne passe pas le filtre de sélection. Dans cet article, je montre que le calcul efficace du facteur d'ajustement est possible pour des modèles de classification bayésiens gaussiens pour caractéristiques continues.

[MS-189]

Bayesian Analysis of Ordinal Response Survey Data using Dirichlet Process Priors

Analyse bayésienne de données d'enquête ordinales avec processus de Dirichlet comme loi a priori

Saman MUTHUKUMARANA & Tim SWARTZ, *Simon Fraser University*

This paper presents a Bayesian latent variable model to analyze ordinal response survey data. We treat the data as a sample that arises from a class of continuous latent variables with unknown cut points that convert latent variables to ordinal response data. We are interested in average response for each of the questions in the survey and their correlation structure. Furthermore, each respondent is characterized by two parameters that have Dirichlet process as their prior distributions. This mechanism allows us to adjust for classes of personal traits. The model is applied to student survey data in teaching and course evaluations.

Cet article présente un modèle bayésien à variable latente pour l'analyse de données d'enquête à réponse ordinale. Nous traitons les données comme un échantillon provenant d'une classe de variables latentes continues avec points de coupure inconnus convertissant les variables latentes en réponses ordinales. Nous nous intéressons à la réponse moyenne pour chacune des questions de l'enquête et à leur structure de corrélation. De plus, chaque répondant est caractérisé par deux paramètres ayant comme loi a priori des processus de Dirichlet. Ce mécanisme nous permet de faire des ajustements pour des classes de traits personnels. Le modèle est appliqué à des données d'enquête auprès des étudiants portant sur l'évaluation de cours.

[MS-190]**Adjustable Bayesian Information Criterion for Model Selection
Critère d'information bayésien ajustable pour sélection de modèles**Changjiang XU & A. Ian McLEOD, *University of Western Ontario*

We investigate an adjustable Bayesian information criterion, denoted as BICq. The criterion has an adjustable parameter, and is consistent for each fixed parameter. It is shown that the BICq includes not only Schwartz' BIC as a special case but also the extended BIC proposed recently for model selection with large model spaces. The BICq can also be adjusted to be equivalent to Akaike's AIC for each fixed sample size. A procedure is given to determine the adjustable parameter. We illustrate the performance of the BICq with simulation experiments and with an application to hospital manpower data.

Nous étudions un critère d'information bayésien ajustable, dénoté BICq. Ce critère a un paramètre ajustable, et est convergent pour chaque valeur du paramètre. Nous montrons que le BICq comprend le BIC de Schwartz et le BIC étendu récemment proposé en sélection de modèles pour grands espaces modèles comme cas particuliers. Le BICq peut aussi être adapté pour être équivalent à l'AIC d'Akaike pour chaque taille échantillonnale fixe. Nous illustrons la performance du BICq par des expériences de simulation et une application à des données de main-d'oeuvre d'hôpital.

[MS-191]**Effects of Different Prior Distributions on the Bayesian Predictive Inference
Effets de différentes distributions a priori sur l'inférence bayésienne**Azizur RAHMAN, *University of Canberra*

Predictive inference is one of the oldest methods of statistical inference and it is based on the observable data. Prior information plays an important role in the Bayesian predictive inference. Researchers in this area are often subjective to exercise non-informative prior distribution. This paper examines the effects of a range of prior distributions on predictive inference for different modelling situations such as linear regression models under normal and Student-*t* errors. Findings reveal that different choices of priors not only provide different prediction distributions of the future response(s) but also change the location and/or scale or shape parameters of the prediction distributions.

L'inférence prédictive est l'une des méthodes d'inférence statistique les plus anciennes et est basée sur des données observables. L'information a priori joue un rôle important dans l'inférence bayésienne. Les chercheurs dans ce domaine font souvent preuve de subjectivité avec des distributions a priori non informatives. Cette présentation s'intéresse aux effets d'un ensemble de distributions a priori différentes sur l'inférence bayésienne dans le cadre de modélisations de type régression linéaire à erreur normale ou de Student. On montre que des choix différents dans les distributions a priori donnent non seulement des distributions prédictives différentes, mais changent aussi les paramètres de position, d'échelle ou de forme des distributions prédites.

Session 08I**Mardi 2 juin • Tuesday, June 2, 13:30 – 15:00****IRC Lobby****Poster Session III****Séance d'affichage III****[MS-192]****Spatio-Temporal Modelling for Lupus Incidence in Toronto, Canada since 196****Modélisation spatio-temporelle de l'incidence de Lupus à Toronto, Canada depuis 196**Ye LI, *University of Toronto*

We have data containing locations of Toronto lupus cases for several census periods. Fine grids were created over entire the GTA and a Gaussian Markov Random Field is fit on the grids assuming risk surface does not change over time. A posterior sample of risk surfaces is used to construct an estimated risk surface and detect areas likely to be 'clusters' of excess risk. Reporting bias is examined by including distance from the clinic as a covariate, allowing for the clinic changing locations in the 1980's. Both MCMC and Integrated Nested Laplace Approximation [Rue] are used for each model and the results are compared.

Nous avons des données contenant la position de cas de lupus à Toronto pour plusieurs périodes de recensement. Une grille fine couvrant toute la région métropolitaine de Toronto a été créée, et un champ aléatoire gaussien a été ajusté à la grille en supposant que la surface de risque ne varie pas avec le temps. Un échantillon a posteriori de surfaces de risque est ensuite utilisé pour construire une surface de risque estimée et détecter les zones qui sont vraisemblablement des agrégats de risque excessif. Le biais de déclaration est examiné en incluant la distance à la clinique comme covariable et en traitant le changement d'emplacement de cliniques dans les années 1980. Des méthodes MCMC et par approximation de Laplace emboîtée intégrée [Rue] sont utilisées pour chaque modèle et les résultats obtenus sont comparés.

[MS-193]**Model-Based Classification in Food Authenticity Studies****Classification basée sur un modèle dans les études d'authenticité des aliments**Michelle STEANE, Paul McNICHOLAS & Rickey YADA, *University of Guelph*

When considering food authenticity problems, we can consider that an authentic food is one that is exactly what it is said to be. A model-based classification technique is developed, based on the mixtures of multivariate t-factors model, and a variation of the EM algorithm is used to estimate model parameters. This new technique is applied to data on chemical measurements on maple syrup samples from seven different geographical regions in North America. The results of this analysis are discussed and compared to other classification techniques.

Dans des problèmes d'authenticité des aliments, nous considérons qu'un aliment est authentique s'il est exactement ce qu'il affirme être. Nous développons une technique de classification à base de modèles, basée sur les mélanges de modèles multivariés de t- facteurs, et nous utilisons une variation de l'algorithme EM pour estimer les paramètres du modèle. Cette nouvelle technique est appliquée à des données de mesures chimiques d'échantillons de sirop d'érable de sept emplacements géographiques différents en Amérique du Nord. Nous discutons des résultats de cette analyse et comparons notre méthode à d'autres techniques de classification.

[MS-194]**Direct Parametric Regression Models on Cumulative Incidence Function for Competing Risk Data
Modèles paramétriques de régression directe sur la fonction d'incidence cumulative pour données de risques concurrents**Ying QU, *AQL Management Consulting Inc.*, Keumhee CARRIERE, *University of Alberta*

We consider a direct parametric regression model on cumulative incidence functions for a competing-risk data analysis with exclusive failure types. Weibull model has been considered previously (Jeong, 2007). We modified it to substitute the improper Gompertz baseline sub-distribution with another that can be more flexibly applied. Our extended model is compared with that of Jeong (2007) by a simulation study on cumulative incidence functions with known baseline hazards. It demonstrates that the extended model is more flexible and has wider application than the existing approach. Numerical example demonstrates the utility and benefit of our method.

Nous considérons un modèle de régression paramétrique directe sur les fonctions d'incidence cumulative pour une analyse de données à risques concurrents à types de défaillance exclusifs. Le modèle de Weibull a déjà été considéré (Jeong, 2007). Nous l'avons modifié afin de substituer la sous-distribution de base impropre de Gompertz par une autre plus facile à appliquer. Notre modèle étendu est comparé à celui de Jeong (2007) par une étude de simulation des fonctions d'incidence cumulative avec défaillance de base connue. Elle montre que notre modèle étendu est plus flexible et est plus applicable que l'approche existante. Des exemples numériques démontrent l'utilité et les avantages de notre méthode.

[MS-195]**Optimization Heuristics for Unimodal Kernel Density Estimation by Data Sharpening
Heuristiques d'optimisation pour estimation de densités unimodales par noyaux avec affinage de données**Mark WOLTERS & W. John BRAUN, *University of Western Ontario*

Data sharpening refers to methods where data are perturbed so that the application of a given non-parametric method gives estimates which satisfy constraints such as monotonicity or unimodality. Despite its potential, the usefulness and generality of data sharpening is presently limited by optimization difficulties. We present new strategies for performing sharpening, using unimodal kernel density estimation as a test case. A greedy algorithm is first proposed; we then explore how this algorithm can be incorporated into a stochastic, population-based search heuristic. The new methods are illustrated through simulations and examples.

L'affinage de données fait référence à des méthodes où les données sont perturbées de façon à ce que l'application d'une méthode non paramétrique donnée donne des estimateurs satisfaisant des contraintes telles la monotonie ou l'unimodalité. Malgré son potentiel, l'utilité et la généralité de l'affinage de données sont présentement limitées par des difficultés en optimisation. Nous présentons de nouvelles stratégies pour l'affinage, utilisant l'estimation d'une densité par noyaux comme cas test. Un algorithme avare est proposé; nous explorons ensuite comment cet algorithme peut être incorporé à une heuristique de recherche stochastique basée sur la population. Les nouvelles méthodes sont illustrées au moyen d'exemples et de simulation.

[MS-196]**A Statistical Method for High-throughput Screening of Predicted Orthologs
Une méthode statistique pour le tri à haut débit d'orthologues prédits**Jeong Eun MIN, Brad McNENEY & Jinko GRAHAM, *Simon Fraser University*

Orthologs are genes in different species that evolved from a common ancestral gene after speciation. Their identification is critical for reliable prediction of gene function in newly sequenced genomes.

Les orthologues sont des gènes de différentes espèces ayant évolué d'un gène ancestral commun après spéciation. Leur identification est critique pour la prévision fiable de la fonction de gènes dans des génomes récemment séquencés. Les

Orthologous genes are usually identified by a high-throughput method called Reciprocal-Best BLAST (RBB). As RBB is subject to errors from incomplete sequencing or gene loss in a species, a bioinformatics tool called Ortholuge was developed that identifies RBB-predicted orthologs with atypical genetic divergence. However, declaring the cutoff for atypical divergence is very computationally-intensive, and so we propose an alternate statistical procedure and examine its properties by simulation.

gènes orthologues sont habituellement identifiés par une méthode à haut débit appelée BLAST réciproquement meilleur (BRM). Comme BRM est sujet à des erreurs provenant de séquençages incomplets ou de la perte de gènes dans une espèce, un outil bio-informatique appelé Ortholuge a été développé afin d'identifier les orthologues identifiés par BRM à divergence génétique atypique. Cependant, déterminer le point de coupure pour divergence atypique demande beaucoup de puissance de calcul, et nous proposons donc une méthode statistique alternative et examinons ses propriétés par simulation.

[MS-197]

Interval-censored Regression

Régression censurée par intervalles

Xiuli KANG, & W. John BRAUN, *University of Western Ontario*

Local likelihood regression is a popular method for fitting regression models when the responses are completely observed. In this talk, we aim to describe several methods for handling nonparametric regression data with interval-censored responses and show that a local likelihood method can be better than interval midpoint imputation and conditional expectation imputation. The performance of the local likelihood method is also illustrated using real data sets.

La régression par vraisemblance locale est une méthode populaire pour ajuster des modèles de régression lorsque les réponses sont complètement observées. Dans cet exposé, nous visons à décrire plusieurs méthodes pour traiter des données de régression non paramétrique avec réponse censurée par intervalles et à montrer qu'une méthode par vraisemblance locale peut être meilleure que l'imputation par point milieu d'intervalle et l'imputation par espérance conditionnelle. La performance de la méthode par vraisemblance locale est aussi illustrée au moyen de jeux de données réels.

[MS-198]

Statistical Analysis of Trends in Red River Water Quality at the International Boundary Over a 45-year Period

Analyse statistique des tendances de la qualité d'eau de la rivière Rouge à la frontière internationale pendant une période de 45 ans

Carrie PAQUETTE, *University of Manitoba*

Trend analysis of water quality indicators is important in assessing the impact of changes in the landscape and land usage as well as the impact of precipitation within watersheds. We perform a statistical evaluation of trends in ion concentrations for the Red River within Manitoba over a 45-year period. The analysis is done using two different methods, a parametric flow-weighted method (developed by Aldo Vecchia, USGS) and nonparametric methods consisting of the Kruskal-Wallis test for seasonality and the Mann-Kendall test for trend. Comparisons of the methods are made.

L'analyse de la tendance des indicateurs de la qualité de l'eau est importante pour évaluer l'impact des changements du paysage et de l'utilisation du terrain ainsi que l'impact de la précipitation dans les bassins-versants. Nous effectuons une évaluation statistique des tendances de concentrations d'ions dans la rivière Rouge au Manitoba sur une période de 45 ans. L'analyse est exécutée avec deux méthodes différentes. La méthode paramétrique d'écoulement alourdi (développée par Aldo Vecchia, USGS) et deux méthodes non paramétriques : le test de saisonnalité de Kruskal-Wallis et le test de Mann-Kendall pour les tendances. Nous comparons les différentes méthodes.

[MS-199]**A Bayesian Survival Model of Phenology of High-value Perennial Agricultural Crops in the Okanagan, British Columbia****Un modèle de survie bayésien pour la phénologie de cultures vivaces de grande valeur dans l'Okanagan, Colombie-Britannique**

Song CAI, *University of British Columbia*, Denise NEILSEN & Nathaniel NEWLANDS, *Agriculture and Agri-Food Canada*, James ZIDEK, *University of British Columbia*

Climate change is expected to alter the suitability of land areas to agricultural crops. We explore how the rate of change in temperature is related to phenology, as the timing of seasonal events, such as bloom date, for apple, sweet cherry and wine grapes in the Okanagan region. We apply Bayesian survival analysis methods to model the phenological event sequence of these perennial crops. The value of the approaches as well as substantive findings on the effect of temperature are reported.

Les prévisions sur les changements climatiques montrent une altération de l'adaptation de zones terrestres aux cultures agricoles. Nous explorons comment le taux de variation de la température est lié à la phénologie, c'est-à-dire au temps de déclenchement d'événements saisonniers, telle la date d'éclosion, pour les pommes, les cerises douces et les raisins à vin dans la région d'Okanagan. Nous appliquons des méthodes d'analyse de survie bayésienne pour modéliser les suites d'événements phénologiques de ces cultures vivaces. L'utilité de ces approches et les découvertes au sujet de l'effet de la température sont présentées.

[MS-200]**Evaluating Strategies to Estimate the Association of Obesity with Mortality via a Markov Model****Évaluer des stratégies d'estimation de l'association entre obésité et mortalité au moyen d'un modèle markovien**

Bingshu CHEN, *Queen's University*, Barry GRAUBARD, *NIH/NCI*, Katherine FLEGAL, *Centers for Disease Control*, Mitchell GAIL, *NIH/NCI*

We used a first order discrete Markov model to investigate strategies to obtain unbiased estimates of the relative mortality hazard for comparing obese with normal weight subjects. This hazard ratio is confounded because subjects can migrate over time from their initial classification on obesity and health status. Recommended strategies, such as limiting the analysis to subjects classified as well initially failed to control bias adequately. We conclude that understanding the health status of subjects is important for obtaining good descriptive estimates of the effects of obesity on mortality, and updating the information on obesity and health status reduces bias.

Nous avons utilisé un modèle de Markov discret de premier ordre pour étudier les stratégies d'obtention d'estimateurs sans biais des risques de mortalité relative pour comparer des sujets obèses et de poids normal. Ce rapport de risques est confondu car les sujets peuvent migrer avec le temps et changer de catégorie de poids ou d'état de santé. Les stratégies recommandées (par exemple, limiter l'analyse aux sujets classés comme en santé initialement) n'ont pas réussi à contrôler adéquatement le biais. Nous concluons que la compréhension de l'état de santé des sujets est importante pour l'obtention de bons estimateurs descriptifs des effets de l'obésité sur la mortalité, et que mettre à jour l'information sur l'état de santé et d'obésité réduit le biais.

Session 09A**Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00****IRC 6**

Special Invited Session of the Biostatistics Section

Allocution de l'invité d'honneur du groupe de biostatistique

[MS-201]**Misspecifying a Random Effects Distribution: Why Getting it Wrong Probably Doesn't Matter****Pourquoi une mauvaise spécification d'une distribution à effets aléatoires n'est probablement pas très problématique**Charles McCulloch & John Neuhaus, *UCSF*

Statistical models that include random effects are commonly used to analyze longitudinal and correlated data, often with strong and parametric assumptions about the random effects distribution. In this talk we examine the impact of misspecification on a variety of inferences, including prediction of expected values, inference about covariate effects, and prediction of random effects. We first present a variety of motivating examples. Then, using both theoretical calculations and simulation, we identify situations where the specification is and is not important.

Les modèles statistiques comprenant des effets aléatoires sont souvent utilisés pour l'analyse de données longitudinales corrélées, souvent en faisant des hypothèses paramétriques fortes sur la distribution des effets aléatoires. Dans cet exposé, nous examinons l'impact d'une mauvaise spécification sur une variété d'inférences, incluant la prévision de valeurs espérées, l'inférence sur l'effet des covariables et la prévision d'effets aléatoires. Nous présentons d'abord une gamme d'exemples motivant ce problème. Ensuite, au moyen de calculs théoriques et de simulation, nous identifions les situations où la spécification est ou n'est pas importante.

Session 09B**Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00****IRC 1**

Innovative Uses of Statistical Software in the (Remote) Classroom

Utilisations novatrices des logiciels statistiques dans la salle de classe (à distance)

[MS-202]**Interactive Demonstrations of Statistics with JMP****Démonstrations statistiques interactives avec JMP**Mark Bailey, *SAS Institute Inc.*

Instructors strive to discover imaginative ways to explain statistical concepts to their students. Their proven methods become less effective when they teach in a remote classroom with limited resources. Instructors teaching JMP courses use graphical illustrations and simulations developed using the JMP scripting language (JSL). Students see demonstrations in the same environment they will use to solve their own problems. This presentation will show how these interactive demonstrations can be used to effectively teach hard-to-explain statistical concepts in both face-to-face and remote classrooms with four examples: Central Limit Theorem, Box-Cox transformation, collinearity, and simulated drug stability data.

Les enseignants cherchent des façons innovatrices d'expliquer les concepts statistiques à leurs étudiants. Les méthodes usuelles deviennent moins efficaces dans le cadre de l'enseignement à distance, les ressources étant limitées. Les enseignants donnant des cours utilisant JMP se servent d'illustrations graphiques et de simulations développées au moyen du langage de JMP (JSL). Les étudiants voient les illustrations dans le même environnement qu'ils utiliseront pour résoudre leurs problèmes. Cette présentation montrera comment ces démonstrations interactives peuvent être utilisées pour enseigner efficacement des concepts difficiles à expliquer en classe ou à distance avec quatre exemples : le théorème central limite, la transformation de Box-Cox, la colinéarité et des données simulées à propos de la stabilité des médicaments.

[MS-203]**Using SPSS in Teaching Predictive Analytics****Utiliser SPSS dans l'enseignement de l'analytique prédictive**Tim DACIUK, *SPSS Inc.*

Student learning research has proven that success in remote classrooms depends upon instructors' ability to clearly communicate topics. SPSS makes it is easier for academics to educate in statistics. SPSS includes menu interfaces, and a Custom Dialog Builder, enabling the creation of dialogs that walk through analyses, illustrating complex concepts, or automating tasks so students spend time interpreting results and applying theory. SPSS includes access to the R programming library, offering an array of statistical tests. SPSS Viz Designer, produces visual displays of statistical results. This presentation will demonstrate these features and other enhancements of interest to the academic community.

Des recherches en pédagogie ont prouvé que l'efficacité de l'enseignement à distance dépend de la capacité des instructeurs à clairement communiquer la matière. SPSS facilite l'enseignement statistique en milieu académique. SPSS comprend une interface à base de menus et un constructeur de dialogues sur mesure permettant la création de dialogues pour accompagner les étudiants à travers les analyses, l'illustration de concepts complexes ou l'automatisation de tâches afin de permettre aux étudiants de prendre le temps d'interpréter les résultats et d'appliquer la théorie. SPSS comprend un accès à la librairie de programmation R, offrant toute une gamme de tests statistiques. SPSS Viz Designer présente, sous forme graphique, les résultats statistiques. Cette présentation démontrera ces caractéristiques et d'autres améliorations d'intérêt pour la communauté académique.

[MS-204]**Teaching Stata in the Remote Classroom****Enseigner l'utilisation de Stata à distance**Roberto GUTIERREZ, *StataCorp LP*

Stata remains committed to the advancement of technology applied to both statistical analysis and to the teaching of statistics in the classroom. Stata is capable of searching the web for user-written modules, installing such remote software, and loading datasets remotely. Since the keys to success in teaching statistics in a distance-learning environment are having readily available data and having the ability to expand your software as needed, this makes Stata an ideal choice for such courses. We will demonstrate these capabilities, as well as other key recent developments from StataCorp.

Stata demeure dédié aux progrès technologiques appliqués à l'analyse statistique et à l'enseignement de la statistique en classe. Stata est capable d'effectuer des recherches sur le web pour des modules créés par les usagers, d'installer de tels modules et de charger des jeux de données à distance. Comme la clef du succès dans l'enseignement statistique à distance est d'avoir des données à portée de la main et de pouvoir étendre les capacités du programme utilisé au besoin, Stata est un choix idéal pour de tels cours. Nous allons démontrer ces capacités, ainsi que d'autres développements récents de StataCorp.

Session 09C**Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00****IRC 4**

Statistics/Mathematics Finance 1

Statistique/Mathématique Finance-1

[MS-205]**A Multiname First Passage Model for Credit Risk****Un modèle de premier passage à plusieurs noms pour le risque de crédit**Adam METZLER, *University of Western Ontario*, Don L. McLEISH, *University of Waterloo*

In multiname extensions of the seminal Black-Cox model, dependence between corporate defaults is typically introduced by correlating the Brownian motions driving firm values. Despite its significant intuitive appeal, such a framework is simply not capable of describing market data. In this talk we investigate an alternative framework, in which dependence is introduced via stochastic trend and volatility in obligors' credit qualities. We find that the model is capable of describing market data for synthetic CDO tranches, and compare calibrated parameters from both 2006 and 2008.

Dans les extensions à plusieurs noms du modèle de Black-Cox, la dépendance entre les défaillances d'entreprises est habituellement introduite en corrélant les mouvements browniens autour desquels sont centrées les valeurs des firmes. Malgré son attrait intuitif significatif, un tel cadre n'est tout simplement pas capable de décrire les données des marchés. Dans cet exposé, nous étudions un cadre alternatif, où la dépendance est introduite au moyen d'une tendance stochastique et d'une volatilité dans les qualités de crédit. Ce modèle est capable de décrire les données des marchés pour tranches CDO synthétiques, et nous comparons les paramètres calibrés à partir de données de 2006 et 2008.

[MS-206]**The Use of Statistical Models in Dynamic Pricing Problems****L'utilisation de modèles statistiques pour la fixation dynamique des prix**Xikui WANG, *University of Manitoba*

We explore the use of sequential statistical decision methods to understand dynamic pricing as well as investment and consumption problems. The demand function is unknown for the dynamic pricing problem and the payoff distribution is unknown for the investment and consumption problem. Several statistical models such as the bandit processes, hidden Markov models and regime switching models are discussed.

Nous explorons l'utilisation de méthodes séquentielles de décision statistique pour comprendre les problèmes de fixation dynamique des prix d'investissement et de consommation. La fonction de demande est inconnue pour le problème de la fixation dynamique des prix et la distribution des gains est inconnue pour le problème d'investissement et de consommation. Plusieurs modèles statistiques tels les processus du bandit, les modèles de Markov cachés et les modèles à changement de régime sont discutés.

[MS-207]**Equity Risk Premium and Volatility: A Correlation Structure****Prime de risque sur capitaux propres et volatilité : une structure de corrélation**Yonggan ZHAO, *Dalhousie University*

This paper investigates the relation between stock market returns and volatility using a bivariate factor model governing the evolution of volatility and the market price of risk. Daily risk premiums on a market portfolio are derived from an intertemporal constant correlation structure between a volatility

Cet article porte sur la relation entre rendements de marchés boursiers et volatilité au moyen d'un modèle factoriel bivariable de l'évolution de la volatilité et du prix courant du risque. Des primes de risque journalières pour un portefeuille sont obtenues à partir d'une structure de corrélation constante inter temporelle entre un indicateur de volatilité et le prix cou-

indicator and the market price of risk. The model-implied volatility measured by the conditional standard deviation of equity returns is compared with the predictable volatility measured by the expected value of a volatility indicator.

rant du risque. La volatilité impliquée par le modèle telle que mesurée par l'écart type conditionnel des rendements est comparée à la volatilité prédictive mesurée par la valeur espérée d'un indicateur de volatilité.

Session 09D

Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00

IRC 5

Small Area Estimation at Statistics Canada

Estimation relatives aux secteurs restreints chez Statistique Canada

[MS-208]

A SPREE Small Area Procedure for Estimating Population Counts

Une procédure SPREE pour petits domaines afin d'estimer les comptes spécifiques de la population

Emily BERG & Wayne A. FULLER, *Iowa State University*

The Canadian Labour Force Survey uses a hierarchical system to classify employment into occupations. A small area procedure is developed under the assumption that the margins of the two-way table (province and occupations) have acceptable reliabilities. The procedure obtains more stable estimates of the counts in the two-way table and preserves the estimates of the marginal totals. Auxiliary information from the previous Census is incorporated through the SPREE estimates of the cell totals. Estimators of the random model are used to combine the estimators of the proportions arising from the SPREE estimators of the totals with the direct estimates.

L'Enquête sur la population active du Canada se sert d'un système hiérarchique afin de classer l'emploi en occupations. Nous développons une méthode SPREE pour les petites régions en supposant que les marges du tableau bidimensionnel, provinces et occupations, aient des fiabilités acceptables. Cette méthode nous permet d'avoir des estimations plus stables pour les comptes à l'intérieur de ce tableau bidimensionnel tout en respectant les totaux marginaux. De l'information auxiliaire provenant des données d'un recensement précédent est incorporée aux estimateurs des totaux SPREE. On se sert d'estimateurs basés sur un modèle aléatoire afin de combiner les estimations des proportions provenant des totaux découlant des estimateurs SPREE avec les totaux dérivés à partir d'estimations directes.

[MS-209]

Estimation for Occupations Counts Using the Labour Force Survey

Estimations des comptes d'occupation en se servant de l'Enquête sur la population active du Canada

Mike HIDIROGLOU & Zdenek PATAK, *Statistics Canada*

The Canadian Labour Force Survey uses a probability sample based on a stratified multi-stage design. Employment estimates include detailed breakdowns by demographic characteristics, industry and occupation, job tenure, and usual and actual hours worked. We illustrate the use of the SPREE procedure developed by Berg and Fuller (2008) to improve the reliability and smoothness of the two digit occupation series.

L'Enquête sur la population active du Canada se sert d'un échantillon probabiliste basé sur un plan de sondage stratifié à deux degrés. Les estimations pour l'emploi sont produites par secteur d'activité et par profession, ainsi que selon la durée de chômage, le genre de travail recherché et le nombre d'heures travaillées. Nous illustrons la méthode SPREE élaborée par Berg et Fuller (2008) afin de rendre la série des occupations codées à deux chiffres plus lisse et d'en améliorer la fiabilité.

[MS-210]**Area Level Modeling Approaches to Small Area Estimation with Applications
Estimation pour petits domaines au moyen de modèles régionaux avec exemples**

Yong You, *Statistics Canada*

In this paper we discuss the Fay-Herriot model and extended area level models including unmatched models, time series models and spatial correlation models. We also discuss sampling variance smoothing and modeling methods as well as benchmarking small area estimates using EBLUP and HB approaches. Then we will apply the hierarchical Bayes approach with Gibbs sampling method to various area level models and discuss the use of small area models in different surveys for small area estimation at Statistics Canada. Comparison of model-based and direct estimates will be presented. Finally we will give conclusion and some suggestions.

Cet article étudie le modèle Fay-Herriot et des modèles régionaux plus étendus tels que les modèles sans correspondance, de séries chronologiques et de corrélation spatiale. Nous discutons du nivellement de la variance de l'échantillon et des méthodes de modélisation, ainsi de l'étalonnage des estimations sur petits domaines par le biais de la MPLSBE et du HB. Nous utilisons l'approche hiérarchique bayésienne et la méthode d'échantillonnage de Gibbs avec divers modèles régionaux dans le cadre de diverses enquêtes de Statistique Canada. Enfin, nous comparons l'estimation par modèle et l'estimation directe et présentons des conclusions et des recommandations.

[MS-211]**Small Area Estimation: A Business Survey Application
L'estimation de petits domaines appliquée à une enquête auprès des entreprises**

Wesley YUNG & Susana RUBIN-BLEUER, *Statistics Canada*

Statistics Canada's Survey of Employment, Payrolls and Hours (SEPH) combines data from a monthly payroll survey with administrative data from the Canada Revenue Agency (CRA). Using this administrative data allows SEPH to produce estimates for domains with small sample sizes. SEPH is currently investigating regression composite estimation as a means to further improve the efficiency of its estimates. Despite these efforts, there are still domains of interest for which reliable estimates can not be produced. In this presentation we propose an estimator that combines small area estimation methods with the regression composite estimator and illustrate it using data from SEPH.

L'Enquête sur l'emploi, la rémunération et les heures de travail (EERH) de Statistiques Canada combine des données issues d'une enquête mensuelle sur la rémunération et des données administratives de l'Agence du revenu du Canada (ARC). En utilisant ces données, l'EERH peut produire des estimations 'relativement aux domaines' en prenant des échantillons de petite taille. L'EERH évalue présentement l'estimateur composite de régression afin de déterminer s'il permettrait d'augmenter la précision de ses estimations. En dépit de ces efforts, il reste des domaines d'intérêt pour lesquels on ne dispose d'aucun estimateur fiable. Cette présentation se penche sur un estimateur qui allie les méthodes d'estimation de petits domaines à celles de l'estimateur composite de régression. Les données de l'EERH serviront d'étalon pour l'évaluation de cet estimateur.

Session 09E**Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00****MSL 101**

Special Session in Honour of Keith Worsley (1951-2009)

Séance en la mémoire de Keith Worsley (1951-2009)

[MS-212]**Keith Worsley: his Early Years****Keith Worsley : ses jeunes années**Alastair SCOTT, *University of Auckland*

Keith Worsley displayed the intense intellectual curiosity and almost childlike enthusiasm that characterized all his work from a very early age. In this talk, I describe his formative years and how they contributed to his development into a great scientist.

Keith a démontré dès un très jeune âge l'intense curiosité intellectuelle et l'enthousiasme candide qui caractérisa l'ensemble de son travail. Au cours de cette conférence, je décrirai ses années de formation et la façon dont elles contribuèrent à le faire évoluer en un scientifique de premier plan.

[MS-213]**Keith Worsley's Contributions to Brain Mapping****Les contributions de Keith Worsley à l'imagerie cérébrale**Jonathan TAYLOR, *Stanford University*

Keith laid much of the statistical groundwork for signal detection in smooth random fields, inspired by problems in neuroimaging, multiple comparisons and astrophysics. In this memorial talk, I will try to describe his seminal contributions in these fields.

Keith fut l'originateur d'une grande part des résultats fondamentaux sur la détection de signaux dans les champs aléatoires lisses. Il trouvait son inspiration dans les problèmes d'imagerie cérébrale et l'astrophysique. Dans cet exposé commémoratif, je tenterai de décrire ses contributions charnières dans ces domaines.

[MS-214]**A Photo Album for Keith Worsley****Un album de photos pour Keith Worsley**George STYAN, *McGill University*

Our Photo Album for Keith Worsley starts at a dinner party he hosted in December 1982 and continues through several meetings of the SSC that Keith attended from 1990 onwards. In 2004 Keith was awarded the SSC Gold Medal for "landmark contributions in the topology and geometry of smooth random fields, the application of random field theory to the statistics of human brain mapping and astrophysics, the field of multiple comparisons, and the field of change-point analysis".

Notre album de photos pour Keith Worsley débute lors d'un dîner donné par lui en 1982, et se poursuit au fil de plusieurs congrès de la SSC auxquels il participa à partir de 1990. En 2004, Keith reçut la Médaille d'or de la SSC «pour ses contributions marquantes en topologie et géométrie des champs aléatoires lisses, dans l'application de la théorie des champs aléatoires aux statistiques de la cartographie du cerveau humain et à l'astrophysique, dans le domaine des comparaisons multiples et le domaine de l'analyse des points de rupture».

Session 09F **Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00** **IRC G65/66**

Design and Analysis of Computer Experiments for Complex Systems

Planification et analyse d'expériences informatiques pour systèmes complexes

[MS-215]

Fast Calibration of Complex Computer Models

Calibration rapide de modèles informatiques complexes

Matthew PRATOLA, *Simon Fraser University*, Stephan SAIN, *National Center for Atmospheric Research*, Derek BINGHAM, *Simon Fraser University*

Computer models enable scientists to investigate real-world phenomena in a virtual laboratory using computer experiments. Recently, statistical calibration enabled scientists to incorporate field data. However, the practical application is hardly straightforward. For instance, large and non-stationary computer model output is not well addressed.

We present a computationally efficient approach using a criterion that measures discrepancy between the computer model and field data. One can then construct empirical distributions for the parameters and perform sequential design. The strength of this approach is its simple computation using existing algorithms. Our method also provides good parameter estimates for large and non-stationary data.

Les modèles informatiques permettent aux scientifiques d'étudier des phénomènes réels en laboratoire virtuel au moyen d'expériences informatiques. Récemment, la calibration statistique a permis aux scientifiques d'y incorporer des données réelles. Cependant, son application pratique est rarement directe. Par exemple, les sorties de modèles complexes et non stationnaires ne sont pas bien prises en considération.

Nous présentons une approche efficace (en terme de temps de calcul) utilisant un critère mesurant l'incohérence entre le modèle informatique et les données réelles. Il est alors possible de construire des distributions empiriques pour les paramètres et d'utiliser un plan séquentiel. La force de cette approche est la simplicité des calculs utilisant des algorithmes existants. Notre méthode fournit aussi de bons estimateurs des paramètres pour données de grande taille et non stationnaires.

[MS-216]

A New Algorithm for the Construction of Computer Experiments with Mixed Levels

Un nouvel algorithme pour la construction d'expériences informatiques à niveaux mixtes

Ryan LEKIVETZ, Randy SITTER & Derek BINGHAM, *Simon Fraser University*, M. HAMADA, L. MOORE & J. WENDELBERGER, *Los Alamos National Laboratory*

Orthogonal arrays are frequently used in computer experiments. Depending on level combinations, an orthogonal array may not exist, in which case a nearly-orthogonal array can be used. Orthogonal and nearly-orthogonal arrays can be difficult to find. This poster will introduce a new algorithm for the construction of orthogonal and nearly-orthogonal arrays with desirable statistical properties, and compare the new algorithm to a pre-existing algorithm for small run sizes.

Les tableaux orthogonaux sont souvent utilisés en expériences informatiques. D'après les combinaisons des niveaux, il peut ne pas exister de tableau orthogonal. Il sera alors possible d'utiliser un tableau presque orthogonal. Les tableaux orthogonaux et presque orthogonaux peuvent être difficiles à trouver. Cette affiche présentera un nouvel algorithme pour la construction de tableaux orthogonaux et presque orthogonaux ayant des propriétés statistiques souhaitables, et compare le nouvel algorithme à un algorithme pré-existant pour petites expériences.

[MS-217]**Sequential Design for Solving Inverse Problem for Expensive Deterministic Computer Simulators****Plan séquentiel pour résoudre le problème inverse pour simulateurs informatiques déterministes complexes**Pritam RANJAN, *Acadia University*, Derek BINGHAM, *Simon Fraser University*, George MICHAILIDIS, *University of Michigan*

In many branches of physical science, when complex physical phenomena are very expensive or time-consuming to observe, observations from their corresponding computer models are used instead. In such experiments, scientists are often interested in identifying different pre-specified process features (e.g., process optima, threshold values and quantiles). The poster presents a sequential design approach for estimating a contour (or level-set) in fewer number of simulator evaluations.

Dans plusieurs branches des sciences physiques, lorsque des phénomènes physiques complexes sont très coûteux ou demandent beaucoup de temps pour être observés, les observations provenant de leur modèle physique correspondant sont plutôt utilisées. Dans de telles expériences, les scientifiques sont souvent intéressés à identifier différentes caractéristiques préspecifiées du processus (par exemple, le maximum du processus, ses valeurs seuils et ses quantiles). Cette affiche présente une approche de planification séquentielle pour l'estimation de contours avec un plus petit nombre d'évaluations du simulateur.

Session 09G**Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00****IRC 3****Analysis of Spatial and Temporal Data****Analyse de données spatiales et temporelles****[MS-218]****Longitudinal Spatial Point Processes and Residential Histories****Processus ponctuel spatial longitudinal et histoires résidentielles**Patrick BROWN, *Cancer Care Ontario*, Peter A HENRYS, *JBA Consulting*

A typical question motivating an analysis of a spatial point process is whether event locations (ie disease incidence) are independent given their underlying intensity (or population density). Chronic diseases with long latency periods should lead researchers to consider each subject's entire residential history, with the goal of assessing the independence or clustering of the residential paths. This talk presents a statistical model for residential histories based on an inhomogeneous Poisson process, and defines a test of independence based on the marked point process K function. Results of a simulation study and an application to data from a study in Winnipeg are presented.

Une question nécessitant souvent l'analyse d'un processus spatial ponctuel concerne l'indépendance des événements rendant compte de l'intensité hétérogène du processus. En étude de maladies chroniques, les chercheurs devraient considérer l'histoire résidentielle des sujets, avec le but de découvrir l'indépendance ou le groupement des histoires individuelles. Cette présentation introduira un modèle pour les histoires résidentielles, fondé sur un processus ponctuel de Poisson, et obtient un test d'indépendance basé par le moment réduit d'ordre 2 d'un processus ponctuel marqué. Des résultats de simulations et une application avec des données d'une étude à Winnipeg sont présentés.

[MS-219]**Spatial CAR Models And Compositional Data****Modèles ARC spatiaux et données compositionnelles**Carolyn HUSTON & Carl SCHWARZ, *Simon Fraser University*

The current research estimates proportion curves using a spatial CAR model. CAR models are often used to smooth a response when information between samples is related. Compositional data analysis focuses on estimating parameters based on compositions, having here a sum-to-one constraint on estimated proportion curves. The methodology represents a novel approach combining elements of compositional data analysis with the improved estimation that occurs when using a spatial CAR model. The method is more effective than currently existing alternatives for the given conditions. This research is motivated and demonstrated by a salmon fisheries management problem in the Fraser River.

Les courbes de proportion sont actuellement estimées par un modèle autoregressif conditionnel (ARC) spatial. Les modèles ARC sont souvent utilisés pour lisser la réponse lorsque l'information entre échantillons est liée. L'analyse de données compositionnelles est centrée sur l'estimation de paramètres basée sur les compositions, ayant ici une contrainte de somme unitaire pour les courbes de proportion estimées. Cette méthode représente une nouvelle approche combinant des éléments de l'analyse de données compositionnelles à l'estimation améliorée associée à l'utilisation de modèles spatiaux ARC. Cette méthode est plus efficace que les alternatives présentement utilisées sous les conditions nous intéressant. Cette recherche est motivée par un problème de gestion des pêcheries de saumon dans le fleuve Fraser.

[MS-220]**Hot Spot Detection in Disease Mapping****Détection de points chauds en cartographie des maladies**Xin FENG, *Simon Fraser University*, Charmaine DEAN, *Simon Fraser University*

In disease mapping studies, often there is interest in the identification of extreme risk areas or relative-risk "hot spots" in order to investigate causes of mortality or for surveillance purposes, or perhaps for efficient allocation of health funding. The focus of the paper is to rank the risks of isolated regions and identify hot spots or emerging hot spots. In combination with spatial models, we used the technique of the weighted rank squared error loss function (WRSEL) introduced by Wright et al. (2003) for such ranking as well as ranking functions based on exceedence probabilities.

En étude de cartographie des maladies, il est souvent d'intérêt d'identifier les zones de risque extrême ou «points chauds» afin d'étudier les causes de la mortalité ou à des fins de surveillance ou d'allocation efficace du financement aux soins de santé. Cet article porte sur le classement des risques de régions isolées et l'identification de points chauds ou de points chauds émergents. En plus de modèles spatiaux, nous utilisons la technique de perte quadratique pondérée basée sur les rangs présentée par Wright et al. (2003) pour de tels classements ainsi que pour des classements basés sur les probabilités de dépassement.

[MS-221]**Influence of Imperfect Diagnostic Tests on Spatial Patterns in Regional Epidemiological Data****L'influence de tests diagnostics imparfaits sur les motifs spatiaux dans des données épidémiologiques régionales**Olaf BERKE, *University of Guelph*

Spatial epidemiology is concerned with the investigation of spatial patterns in the health status of populations. Generally it is assumed that the given data about the health status is measured without

L'épidémiologie spatiale traite de l'étude de motifs spatiaux dans l'état de santé de populations. Les données d'état de santé sont habituellement supposées sans erreurs de mesure, c'est-à-dire que les tests diagnostics sont considérés par-

error, i.e diagnostic tests are deemed perfect. In this work the effect of an imperfect diagnostic test on spatial pattern of disease occurrence in regional count data is investigated. Specifically the effect on the semivariogram, Moran's I statistic and the spatial scan test are evaluated for the situation of West Nile virus infections among dead birds sampled from the 30 health units of southern Ontario in 2005.

[MS-222]

Spatio-temporal Modeling Using Spline for Disease Mapping: Analysis of Childhood Cancer Trends **Modélisation spatio-temporelle au moyen de splines en cartographie des maladies : une analyse des tendances du cancer pédiatrique**

Mahmoud TORABI & Rhonda J. ROSYCHUK, *University of Alberta*

We construct a generalized linear mixed model (GLMM) for the analysis of geographic and temporal variability of cancer rates. Spatio-temporal models that use conditional autoregressive smoothing across the spatial dimension and B-spline and autoregressive smoothing over the temporal dimension are considered. The patterns of incidence rates over time are studied to identify areas with consistently high rate estimates as areas for further investigation. We compare three methods, penalized quasi-likelihood, marginal quasi-likelihood and Bayesian approach (MCMC), to estimate the model parameters. We illustrate this approach using a yearly data set of pediatric cancer diagnoses in the province of Alberta, during 1983-2004.

faits. Dans ce travail, l'effet d'un test diagnostique imparfait sur les motifs spatiaux de maladies dans données de dénombrements régionaux est étudié. Plus spécifiquement, nous évaluons l'effet sur le semi-variogramme, la statistique I de Moran et le test de balayage spatial dans le contexte des infections au virus du Nil occidental chez les oiseaux morts parmi 30 unités de santé en Ontario du sud en 2005.

Nous construisons un modèle linéaire généralisé mixte (MLGM) pour l'analyse de la variabilité géographique et temporelle des taux de cancer. Les modèles spatio-temporaux utilisant le lissage autorégressif conditionnel sur la dimension spatiale et le lissage autorégressif et par B-splines sur la dimension temporelle sont considérés. Les motifs de taux d'incidence à travers le temps sont étudiés afin d'identifier les zones à taux estimés presque toujours élevés comme zones nécessitant une attention particulière. Nous comparons trois méthodes d'estimation des paramètres du modèle : la quasi-vraisemblance pénalisée, la quasi-vraisemblance marginale et une approche bayésienne (MCMC). Nous illustrons cette approche avec un jeu de données annuelles de diagnostics de cancer pédiatrique en Alberta de 1983 à 2004.

[MS-223]

Unaliasing of Line Components in Bivariate Time Series **Dépliage de composantes lignes en séries chronologiques bivariées**

Glen TAKAHARA, Azadeh MOGHADERI & David THOMSON, *Queen's University*

We present a method for determining the frequencies of line components in stationary time series, past the sampling frequency determined Nyquist limit, beyond which detected frequencies would normally be aliased with (indistinguishable from) frequencies within the Nyquist band, when the data arise from a bivariate time series in which one of the components is a time delayed version of the other. We apply our technique to both simulated data and to solar flux data collected at 4 observatories located across the globe. We compare some of the (unaliased) frequencies found in this data to frequencies previously reported in the literature.

Nous présentons une méthode de détermination des fréquences de composantes lignes de séries chronologiques stationnaires au-delà de la limite de Nyquist déterminée par la fréquence d'échantillonnage. Les fréquences détectées au-delà de cette limite sont normalement indistingables des fréquences dans la bande de Nyquist. Nous nous intéressons au cas de données de série chronologique bivariée où une des composantes est une version décalée dans le temps de l'autre. Nous appliquons notre technique à des données simulées et à des données de flux solaire recueillies à 4 observatoires autour du globe. Nous comparons certaines des fréquences décrénelées trouvées dans ces données aux fréquences précédemment mentionnées dans la littérature.

Session 09H**Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00****MSL 102****Bayesian Methods in Biostatistics****Méthodes bayésiennes en biostatistique****[MS-224]****Hormone Therapy and Mortality in Younger Postmenopausal Women: Synthesizing the Evidence using Bayesian Meta-analysis****Thérapie hormonale et mortalité chez les plus jeunes femmes postménopause : synthétiser l'information par méta-analyse bayésienne**

Ji CHENG, *McMaster University and St. Joseph's Healthcare Hamilton*, Shelley SALPETER, *Santa Clara Valley Medical Center and Stanford University*, Lehana THABANE, *McMaster University and St. Joseph's Healthcare Hamilton*, Nicholas BUCKLEY, *California Institute of Technology*, Edwin SALPETER, *Cornell University*

Bayesian methods combine different sources of information through prior specification. We performed a Bayesian meta-analysis to synthesize the effect of hormone therapy on mortality in younger postmenopausal women from 19 RCTs using non-informative, informative and sceptical priors and misleading priors from the WHI trial. Results show that posterior estimates of relative risk and 95 percent credible interval changed respectively: 0.73(0.52,0.96), 0.72(0.62,0.82), 0.83(0.68,1.00), 1.09 (0.97,1.23). We concluded that although the evidence from RCTs shows a reduction of mortality, peoples' views on this issue may change by combining different information.

Les méthodes bayésiennes combinent différentes sources d'information à travers la spécification de la loi a priori. Nous avons mené une méta-analyse bayésienne pour synthétiser l'effet de la thérapie hormonale sur la mortalité chez les plus jeunes femmes postménopause de 19 essais randomisés utilisant des lois a priori non informatives, informatives et sceptiques et des lois a priori trompeuses pour l'étude WHI. Les résultats montrent que les estimateurs a posteriori du risque relatif et intervalle de crédibilité à 95% changent respectivement : 0.73(0.52,0.96), 0.72(0.62,0.82), 0.83(0.68,1.00), 1.09(0.97,1.23). Nous avons conclu que, bien que l'évidence des essais randomisés montre une réduction de mortalité, l'opinion des gens à ce sujet peut changer en combinant de l'information de différentes sources.

[MS-225]**Hierarchical Priors for Bias Parameters in Bayesian Adjustment for Unmeasured Confounding****Lois a priori hiérarchiques pour paramètres de biais en ajustement bayésien pour confusion non mesurée**

Lawrence McCANDLESS, *Simon Fraser University*

Recent years have witnessed new innovation in Bayesian techniques to adjust for unmeasured confounding. A challenge with existing methods is that the user is often required to elicit prior distributions for high dimensional parameters that model competing bias scenarios. This can render the methods unwieldy. We propose a novel methodology that assigns default priors for bias parameters. The confounding effects of measured and unmeasured confounders are modelled as exchangeable within a hierarchical Bayesian framework. We apply the method in a data example from pharmacoepidemiology and explore the impact of different priors for bias parameters on the analysis results.

Les dernières années ont vu de nouvelles innovations en techniques bayésiennes d'ajustement pour effets confondants non mesurés. Un défi avec les méthodes existantes est que l'utilisateur doit souvent obtenir des distributions a priori pour des paramètres de dimension élevée modélisation des scénarios de biais concurrents. Ceci peut rendre ces méthodes peu pratiques. Nous proposons une méthodologie innovatrice assignant des lois a priori par défaut aux paramètres de biais. Les effets confondants de facteurs confondants mesurés et non mesurés sont modélisés de façon échangeable dans un cadre bayésien hiérarchique. Nous appliquons la méthode à un exemple de pharmacoépidémiologie et explorons l'impact de différentes lois a priori pour les paramètres de biais sur les résultats de l'analyse.

[MS-226]**Informed Priors for Between-study Variability in Bayesian Meta-analyses****Lois a priori informées pour la variabilité inter-études en méta-analyse bayésienne**

Eleanor PULLENAYEGUM, *McMaster University*, Kevin WONG, *University of Toronto*, Savio YU & Anne HOLBROOK, *McMaster University*

Bayesian meta-analyses require specification of prior beliefs as to the distribution of the between-study variability. When there are many trials included in the analysis, a non-informative prior can be used. If, however, there are few trials included in the analysis, standard choices of non-informative priors can lead to excessive variability. Alternative informed priors based on the I2 measure of between-study variability are suggested. The impact of the various choices of prior will be examined through simulation, and through application to a recent meta-analysis.

Les méta-analyses bayésiennes nécessitent une spécification des croyances a priori quant à la distribution de la variabilité inter-études. Lorsque plusieurs études sont incluses dans l'analyse, une loi a priori non-informative peut être utilisée. Si, cependant, il y a peu d'études dans l'analyse, les choix habituels de lois a priori non-informatives peuvent mener à une variabilité excessive. Des lois a priori informées alternatives basées sur la mesure I2 de variabilité inter-études sont suggérées. L'impact de divers choix de lois a priori sera examiné par simulations et application à une méta-analyse récente.

[MS-227]**Bayesian Hierarchical Mixture Modeling for Clustered Count Data with Extra Zeros****Modélisation bayésienne hiérarchique de mélanges pour données de dénombrement en grappes avec excès de zéros**

Guohua YAN & Renjun MA, *University of New Brunswick*

Clustered count data with extra zeros occur in many disciplines. In medical and health studies, this excessive zero phenomenon is likely associated with a considerable portion of individuals who have zero-susceptibility to certain diseases during a given period of time. To accommodate both positive and zero susceptibilities, we model the count data with a mixture of a Poisson distribution and a point mass at zero. Multilevel heterogeneities and covariate effects are incorporated into the model hierarchially.

Les données de dénombrement en grappes avec excès de zéros se retrouvent dans plusieurs disciplines. En études médicales et en santé, la surreprésentation de zéros est probablement associée à une proportion considérable d'individus ayant une susceptibilité nulle à certaines maladies durant une période de temps donnée. Pour traiter les susceptibilités nulles et non nulles, nous modélisons les données de dénombrement avec un mélange d'une loi de Poisson et d'une masse de probabilité au point zéro. Des hétérogénéités à plusieurs niveaux et l'effet de covariables sont incorporés hiérarchiquement au modèle.

[MS-228]**Bayesian Adjustment for Covariate Measurement Errors: A Flexible Parametric Approach****Ajustement bayésien pour erreurs de mesures dans les covariables : une approche paramétrique flexible**

Shahadut HOSSAIN, *BC Cancer Research Centre*, Paul GUSTAFSON, *University of British Columbia*

Measurement errors in covariates/exposures are unavoidable in most epidemiological studies. If ignored, such covariate measurement errors can seriously distort the estimated disease-exposure association. Thus, adjustments for such measurement errors are necessary to have accurate association estimates. In this research, we suggest a Bayesian

La présence d'erreur de mesure dans les covariables/exposition est inévitable dans la plupart des études épidémiologiques. Si elles sont ignorées, de telles erreurs peuvent sérieusement affecter les associations maladie-exposition estimées. Il est donc nécessaire d'ajuster pour de telles erreurs de mesure pour obtenir des estimateurs d'association précis. Dans cette recherche, nous suggérons une approche bayé-

flexible parametric approach for avoiding the bias arising in the association estimates due to measurement errors in continuous covariates in logistic regression model. Though emphasis is put on logistic regression model, the proposed method is unified and is applicable to the other generalized linear models, and to the other types of non-linear regression models as well.

sienne paramétrique flexible pour éviter le biais en estimation d'associations du aux erreurs de mesures de covariables continues dans le modèle de régression logistique. Bien que nous mettions l'emphase sur le modèle de régression logistique, la méthode proposée est unifiée et applicable aux autres modèles linéaires généralisés, et à d'autres types de modèles de régression non linéaire.

Session 09I

Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00

CHE 150

Statistical Theory Théorie statistique

[MS-229]

On the Test of Ghouidi, Khoudraji and Rivest

À propos du test de Ghouidi, Khoudraji et Rivest

Noomen El Beji BEN GHORBAL & Christian GENEST, *Université Laval*, Johanna NEÖLEHOV, *ETH Zürich*

Ghouidi, Khoudraji & Rivest (1998) show how to test whether the dependence structure of a pair of continuous random variables is characterized by an extreme-value copula. The test is based on a U-statistic whose large- and finite-sample variances will be presented. Variance estimates will be proposed and compared to the jackknife estimate of Ghouidi, Khoudraji & Rivest (1998) through simulations. The finite-sample and asymptotic power of the test will also be examined under various alternatives. The results will be illustrated using financial and geological data.

Ghouidi, Khoudraji & Rivest (1998) montrent comment tester si la structure de dépendance d'un couple d'aléas continus est caractérisée par une copule de valeurs extrêmes. Le test s'appuie sur une U-statistique dont on présentera la variance asymptotique et à taille finie. Des estimations de cette variance seront proposées et comparées à l'estimateur jackknife de Ghouidi, Khoudraji & Rivest (1998) à l'aide de simulations. Les puissances à taille finie et asymptotique du test seront également étudiées sous diverses contre-hypothèses. Le propos sera illustré au moyen de données financières et géologiques.

[MS-230]

Pitman Closeness of Order Statistics to Population Quantiles

Proximité de Pitman des statistiques d'ordre par rapport aux quantiles de population

Katherine DAVIES, *University of Manitoba*, N. BALAKRISHNAN, *McMaster University*, Jerome P. KEATING, *University of Texas at San Antonio*

In this talk, Pitman closeness of sample order statistics to population quantiles of a location-scale family of distributions is discussed. Explicit expressions are derived for some specific families such as uniform, exponential and power function. Numerical results are then presented for these families for sample sizes $n = 10, 15$, and for the choices of $p = 0.10, 0.25, 0.75, 0.90$. The Pitman- closest order statistic is also determined in these cases and presented.

Dans cet exposé, nous discutons de la proximité de Pitman de statistiques d'ordre échantillonnelles aux quantiles de population d'une famille position-échelle. Des expressions explicites sont obtenues pour certaines familles spécifiques telles les lois uniforme, exponentielle et fonction de puissance. Des résultats numériques sont ensuite présentés pour ces familles pour tailles échantillonnelles $n = 10, 15$ et pour choix de $p = 0, 10; 0, 25; 0, 75; 0, 90$. La statistique d'ordre de proximité de Pitman maximale est aussi déterminée dans ces cas.

[MS-231]**On Fitting a Mixture of two von Mises Distributions****Ajustement d'un mélange de deux distributions von Mises**Qifeng JIANG, Michael STEPHENS & Richard LOCKHART, *Simon Fraser University*

The von Mises distribution is used to model directions, expressed as points around a unit circle. The circle may also be used as a clock, and the points are times of events. Data sets sometimes suggest two clusters of directions or times; in this talk we discuss how to fit a mixture of two von Mises distributions, and how to test fit for the mixture.

La distribution de von Mises est utilisée pour modéliser des directions exprimées comme points autour d'un cercle de rayon unitaire. Ce cercle peut aussi être utilisé comme horloge et les points comme temps d'événements. Les données suggèrent parfois deux grappes de directions ou de temps. Dans cet exposé, nous discutons de l'ajustement d'un mélange de deux distributions de von Mises et de la façon de tester l'adéquation du mélange.

[MS-232]**Nonparametric Tests for Conditional Independence Using Conditional Distribution and Quantile Regression****Tests non paramétriques d'indépendance conditionnelle basés sur la distribution conditionnelle et la régression de quantile**Taoufik BOUEZMARNI & Roch ROY, *Université de Montréal*, Abderrahim TAAMOUTI, *Universidad Carlos III de Madrid Calle Madrid, Spain*

In this paper, we propose two different statistics to test for conditional independence and Granger non-causality. The first one compares the conditional distribution functions and the second one is based on the quantile regression functions. These two test statistics are defined using the L2 metric. We use the weighted Nadaraya-Watson (WNW) estimator to estimate the conditional distribution functions and the quantile regressions. We establish the asymptotic size and power properties of the conditional independence tests and we motivate the validity of the local bootstrap. We ran a simulation study to investigate the finite sample properties of tests.

Dans ce travail, nous proposons deux différentes statistiques pour tester l'indépendance conditionnelle et la non-causalité de Granger. La première compare les fonctions de répartition conditionnelles et la seconde est basée sur les fonctions de régression de quantile. Ces deux statistiques de test sont définies en utilisant la distance L2. Nous employons la méthode de Nadaraya-Watson pondérée pour estimer les fonctions de répartition conditionnelles et régressions de quantile. Nous établissons le niveau asymptotique et les propriétés de puissance des tests de l'indépendance conditionnelle. Une étude de simulation qui montre la performance du test pour une taille d'échantillon finie est effectuée.

[MS-233]**Estimation of Testing fit for the von Mises Distribution with the Grouped Data****Estimation et test d'adéquation pour distribution von Mises avec données groupées**Zheng SUN, *University of Victoria*, Michael STEPHENS, *Simon Fraser University*

In many applications of the von Mises distribution, the data are grouped into cells around the circle. The talk will focus on how to estimate the parameters on a fitted von Mises distribution or a fitted mixture of the von Mises and Uniform distributions or the fitted mixture of two von Mises distributions. In addition, the problem of testing fit to the model using Watson's statistic will be considered.

Dans plusieurs applications de la distribution von Mises, les données sont groupées en cellules autour du cercle unitaire. Cet exposé discutera de l'estimation des paramètres d'une distribution von Mises ou d'un mélange de la loi von Mises et de la loi uniforme ou d'un mélange de deux distributions von Mises. De plus, les tests d'adéquation de modèle utilisant la statistique de Watson seront abordés.

[MS-234]**Information Ratio Test for Model Misspecification****Test du rapport des informations pour mauvaise spécification de modèle**

Qian ZHOU, *University of Waterloo*, Peter X.-K.SONG, *University of Michigan*, Mary THOMPSON, *University of Waterloo*

In quasi-likelihood inference, correct specification of the first two moments leads to the information unbiasedness, which states the analogue of Fisher information is equivalent to the Godambe information. By comparing model-based and robust covariance estimators, obtained from these two forms of information matrices respectively, we proposed an information ratio test for model misspecifications of variance/covariance structure under the correctly specified mean structure, in the context of linear regression models, generalized linear regression models and generalized estimating equations. Through simulation studies, the proposed test statistics appear more powerful than the classical information matrix test proposed by White (1982).

En inférence par quasi-vraisemblance, il n'y a pas de biais d'information lorsque les deux premiers moments sont correctement spécifiés, ce qui veut dire que l'analogue de l'information de Fisher est équivalent à l'information de Godambe. En comparant les estimateurs de covariance à base de modèle et robuste, obtenus respectivement à partir de ces deux formes de matrices d'informations, nous proposons un test du rapport des informations pour mauvaise spécification de structure de variance/covariance sous structure de moyenne correctement spécifiée, dans le contexte de modèles de régression linéaires et linéaires généralisés, et d'équations d'estimation généralisées. Dans les études de simulation, les statistiques de test proposées semblent plus puissantes que le test de matrices d'information classique proposé par White (1982).

Session 09J**Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00****CHE 250****Computational Methods and Algorithms****Méthodes et algorithmes informatiques****[MS-235]****GLAM: An Introduction to Array Methods in Statistics****MLGT : une introduction aux méthodes tabulaires en statistique**

Iain CURRIE, *Heriot-Watt University, UK*

A generalized linear array model (GLAM) is a low storage, high speed method for fitting a generalized linear model (GLM) when the data are in the form of an array and the model can be written as a Kronecker product. Smoothing arrays is one important application since such smoothing is particularly susceptible to runaway problems with storage and computational time. The GLAM algorithm is very fast and in large problems can be orders of magnitude faster than the usual GLM approach. We describe the GLAM algorithms and give some examples of smoothing with GLAM.

Un modèle linéaire généralisé tabulaire (MLGT) est une méthode à haute vitesse et peu volumineuse d'ajustement de modèles linéaires généralisés (MLG) lorsque les données sont organisées en un tableau et que le modèle peut être écrit comme un produit de Kronecker. Les tableaux de lissage sont une application importante puisqu'un tel lissage est particulièrement sensible à des problèmes de fuite relatifs au stockage et au temps de calcul. L'algorithme MLGT est très rapide et sa vitesse peut être de plusieurs ordres de grandeur au-dessus de celle de l'approche MLG habituelle pour des problèmes de grande taille. Nous décrivons les algorithmes MLGT et donnons quelques exemples de lissage avec MLGT.

[MS-236]**Confid2: Statistics Canada's New Tabular Data Confidentiality Software****Confid2 : nouveau logiciel de confidentialité des données tabulaires de Statistique Canada**Olena FROLOVA, Jean-Marc FILLION & Jean-Louis TAMBAY, *Statistics Canada*

Statistics Canada is obligated to protect the confidentiality of respondents' data under the Statistics Act. Cell suppression is the technique that is often used to protect tabulated magnitude data. The automated confidentiality software Confid was developed at Statistics Canada in the 1980s. A new SAS-based confidentiality software, Confid2, was developed recently to satisfy the increased users' needs. It creates a suppression pattern for large tables regardless of the number of dimensions. This paper describes the methodology, functionality and features of Confid2. It also presents the performance results and preliminary feedback from users.

En vertu de la Loi sur la statistique, Statistique Canada est tenue de protéger la confidentialité des données de ses répondants. Pour ce faire, on recourt souvent à la technique de suppression de cases, qui permet de protéger les données tabulaires quantitatives. À l'heure actuelle, on se sert du logiciel de confidentialité automatisé Confid, développé à Statistique Canada dans les années 1980, pour faire cette opération. Toutefois, les besoins en la matière n'ont cessé de croître. Afin d'y répondre, Statistique Canada a récemment développé un nouveau logiciel de confidentialité fondé sur SAS : Confid 2. Ce dernier applique un modèle de suppression aux grands tableaux indépendamment de leur nombre de dimensions. Cette présentation couvre la méthodologie, la fonctionnalité et les caractéristiques de Confid2. De plus, elle fait état de la performance du logiciel ainsi que de la rétroaction préliminaire de ses utilisateurs.

[MS-237]**Efficient, Adaptive Cross Validation for Tuning and Comparing Models****Validation croisée adaptative efficace pour mise au point et comparaison de modèles**Hui SHEN & William J. WELCH, *University of British Columbia*, Jacqueline M. HUGHES-OLIVER, *North Carolina State University*

Cross validation (CV) is widely used for tuning a model with respect to user-selected parameters and for selecting a "best" model. For multiple large sets of data, reliable CV comparisons are computationally demanding, or even infeasible.

In this paper we develop an efficient sequential methodology for model comparison based on CV. The number of models is reduced via an adaptive, multiplicity-adjusted sequential algorithm, where poor performers are quickly eliminated. By exploiting matching of individual observations, it is sometimes even possible to establish the statistically significant inferiority of some models with just one execution of CV.

La validation croisée (VC) est fréquemment utilisée pour ajuster les paramètres choisis par l'utilisateur d'un modèle et pour sélectionner un «meilleur» modèle. Lorsque nous avons plusieurs jeux de données de grande taille, les comparaisons basées sur la VC demandent beaucoup de puissance de calcul ou deviennent simplement impossibles.

Dans cet article nous développons une méthode séquentielle efficace pour la comparaison de modèles basée sur la VC. Le nombre de modèles est réduit au moyen d'un algorithme séquentiel adaptatif ajusté pour la multiplicité, éliminant rapidement les modèles moins performants. En exploitant le jumelage d'observations individuelles, il est parfois même possible d'établir l'infériorité de certains modèles de façon statistiquement significative avec une seule exécution de la VC.

[MS-238]**Analyzing the 2001 FMD Outbreak Using Dynamic Graph Models****Analyse de la poussée de fièvre aphteuse de 2001 au moyen de modèles de graphes dynamiques**Yasaman HOSSEINKASHI, Christopher SMALL & Shoja'eddin CHENOURI, *University of Waterloo*

Foot-and-mouth disease (FMD) is an infectious disease which can spread quickly among livestock. The 2001 outbreak of FMD in the UK involved slaughtering over four million animals and caused severe economic consequences. A dynamic network for infected farms is built on the data provided by the department of environment, food, and rural affairs (DEFRA). The dynamic features of this network are analyzed through a stochastic graph model. The effect of various factors in influencing the FMD outbreak will be discussed using this model.

La fièvre aphteuse est une maladie infectieuse pouvant se propager rapidement parmi le bétail. L'éclosion de 2001 de fièvre aphteuse au Royaume-Uni a nécessité d'abattre plus de quatre millions d'animaux et a eu de graves conséquences économiques. Un réseau dynamique pour fermes infectées est construit à partir des données fournies par le ministère de l'Environnement, de l'alimentation et des affaires rurales. Les caractéristiques dynamiques du réseau sont analysées avec un modèle de graphe stochastique. Les effets de divers facteurs sur l'éclosion de fièvre aphteuse seront abordés au moyen de ce modèle.

[MS-239]**The Crime Severity Index - Finally, the One, True Measure of Crime...?****L'indice de sévérité des crimes - finalement la vraie mesure du crime... ?**Colin BABYAK, Asma ALAVI, Krista COLLINS, Amanda HALLADAY & Dawn TAPPER, *Statistics Canada*

Statistics Canada has created a new measurement of police-reported crime, the Crime Severity Index (CSI). Extensive development has endeavoured to ensure it is methodologically sound, easily understandable, and robust with respect to changes in crime. We examine the methodology of the index including the overall index structure, the linkage procedure for updating the weights and how the weights are calculated. The performance of the index is examined using real and simulated changes in crime volume and sentencing patterns. In conclusion we attempt to answer the question, "Is the Crime Severity Index the one, true measure of crime?"

Statistique Canada a créé une nouvelle mesure des crimes signalés à la police, l'Indice de sévérité des crimes (ISC). Un développement approfondi nous a permis de nous assurer qu'elle est méthodologiquement fiable, facile à comprendre et robuste par rapport à des changements dans la criminalité. Nous examinons la méthodologie de l'indice incluant sa structure générale, la procédure de couplage pour mettre à jour les poids et le détail du calcul des poids. La performance de l'indice est examinée en utilisant des changements réels et simulés dans le volume des crimes et de motifs dans les condamnations. En conclusion, nous tentons de répondre à la question «l'Indice de sévérité des crimes est-il la seule vraie mesure du crime?»

[MS-240]**A Modified Power Iteration for Nonnegative Matrix Factorization****Une méthode des puissances modifiée pour la factorisation de matrices non négatives**Ali GHODSI, Stephen VAVASIS & Michael BIGGS, *University of Waterloo*

Nonnegative matrix factorization (NMF) was popularized as a tool for data mining in 1999. NMF attempts to approximate a matrix with nonnegative entries by a product of two low-rank matrices, also with nonnegative entries. We propose an algorithm called rank-one downdate (R1D) for computing an

La factorisation en matrices non négatives (FMN) fut popularisée comme outil de fouille de données en 1999. La FMN tente d'approximer une matrice à valeurs non négatives par un produit de deux matrices de petit rang, également à valeurs non négatives. Nous proposons un algorithme nommé rank-one downdate (R1D) pour calculer la FMN. Cet algo-

NMF. This algorithm computes the dominant singular values and vectors of adaptively determined submatrices of a matrix. On each iteration, R1D extracts a rank-one submatrix from the original matrix according to an objective function. We establish a theoretical result that maximizing this objective function corresponds to correctly classifying articles in a nearly separable corpus.

rithme calcule les valeurs et vecteurs singuliers dominants de sous-matrices d'une matrice déterminées de manière adaptative. À chaque itération, R1D extrait une sous-matrice de rang un de la matrice originale en se basant sur une fonction objectif. Nous établissons un résultat théorique qui stipule que la maximisation de cette fonction objectif équivaut à classer des articles en corpus presque séparables.

Session 09K

Mardi 2 juin • Tuesday, June 2, 15:30 – 17:00

IRC Lobby

Poster Session IV

Séance d'affichage IV

[MS-241]

Reliability Confidence Intervals for Oil Spills in the Gulf of Mexico

Intervalle de confiance de la fiabilité pour déversements d'hydrocarbures dans le golfe du Mexique

William HARPER, *Otterbein College*, Ted ESCHENBACH, *TEG Consulting*, Thomas JAMES, *Otterbein College*

An extensive study [Eschenbach and Harper (2006)] of offshore oil spills in the Gulf of Mexico with extensions to the northern seas of Alaska involved the estimation of the likelihood of oil spill volumes in the Gulf of Mexico for both pipeline and platform spills. This paper develops spill volume reliability confidence intervals based on the 3-parameter Weibull distribution. The statistical aspects are discussed along with applications of developed Excel VBA functions. The Excel routines are available free on the web at <http://faculty.otterbein.edu/WHarper/>. The functions are illustrated with actual Gulf of Mexico oil spill data.

Une étude approfondie [Eschenbach et Harper (2006)] des déversements d'hydrocarbures dans le golfe du Mexique avec extensions aux mers de l'Alaska impliquait l'estimation de la vraisemblance des différents volumes de déversements d'hydrocarbures provenant de plates-formes et de pipelines dans le golfe du Mexique. Cet article développe des intervalles de confiance pour la fiabilité des déversements basés sur la distribution de Weibull à 3 paramètres. Nous discutons des aspects statistiques et des applications de fonctions Excel VBA développées. Les fonctions Excel sont gratuites et disponibles en ligne à l'adresse <http://faculty.otterbein.edu/WHarper/>. Les fonctions sont illustrées au moyen de données réelles de déversements d'hydrocarbures dans le golfe du Mexique.

[MS-242]

Bayesian Ecological Modeling of Small Area Suicide Rates

Modélisation écologique bayésienne de taux de suicide dans de petits domaines

Yi LIN & Ying MACNAB, *University of British Columbia*

We present a Bayesian ecological analysis of small area suicide rates using data from British Columbia, Canada. An Empirical Bayes - Principal Component Analysis - Fully Bayesian combined approach was taken to explore regional risk factors and spatial risk patterns. Penalized quasi-likelihood method and Principal component method were used for variable selection and for the development of ecological indicators. Fully Bayesian MCMC method was used to quantify

Nous présentons une analyse écologique bayésienne de taux de suicide dans de petits domaines utilisant des données de Colombie-Britannique, Canada. Une approche combinée bayésienne empirique - analyse en composantes principales - complètement bayésienne a été adoptée pour explorer les facteurs de risque régionaux et les motifs de risque spatiaux. La méthode de quasi-vraisemblance pénalisée et l'analyse en composantes principales furent utilisées pour la sélection de variables et pour l'obtention d'indicateurs écologiques. La méthode complètement bayésienne MCMC fut utilisée pour

uncertainties associated with the estimation of all model parameters, particularly the regional relative risks. Differential regional characteristics were observed to affect suicide hospitalization and mortality rates among different age groups and genders.

quantifier les incertitudes associées à l'estimation de tous les paramètres du modèle, particulièrement des risques régionaux relatifs. Les caractéristiques régionales affectent les taux d'hospitalisation et de mortalité liés au suicide parmi les différents groupes d'âge et de genre.

[MS-243]

Multiple Hypothesis Testing Procedures with applications to Epidemiologic Studies of Genetic Susceptibility Procédures pour tester de multiples hypothèses avec applications à des études épidémiologiques de susceptibilité génétique

Conghui QU & Jinko GRAHAM, *Simon Fraser University*, John SPINELLI, *B.C. Cancer Agency*

Epidemiologic studies of genetic susceptibility often involve the testing of a large number of hypotheses with test statistics that are potentially dependent. This research investigates multiple testing procedures to control the family-wise error rate and false discovery rate. We applied and compared several multiple hypothesis testing procedures by using data from a case-control study of non-Hodgkin Lymphoma. The methods compared included Bonferroni, FDR by Benjamini and Hochberg (1995), and re-sampling based procedures that account for dependence among test statistics (Dudoit & van der Laan 2008).

Les études épidémiologiques de susceptibilité génétique impliquent souvent de tester un grand nombre d'hypothèses avec des statistiques de test potentiellement dépendantes. Cette recherche s'intéresse aux procédures de tests multiples contrôlant la proportion d'erreurs de type I global et la proportion de fausses découvertes. Nous appliquons et comparons plusieurs procédures de tests multiples avec des données d'une étude cas-contrôle de lymphome non-Hodgkin. Les méthodes comparées comprennent Bonferroni, la PFD de Benjamini et Hochberg (1995), et des procédures basées sur le rééchantillonnage tenant compte de la dépendance entre statistiques de test (Dudoit & van der Laan 2008).

[MS-244]

Confidence Intervals for Binomial Probability Predictions Intervalles de confiance pour prévisions de probabilités binomiales

Paul NGUYEN & Duncan MURDOCH, *The University of Western Ontario*

We studied methods for assessing the accuracy of detection probability estimates as a function of range to the target for side scan sonar systems used by the Canadian Navy. By simulation, we evaluated the coverage probabilities and lengths of several confidence intervals for binomial proportions used in the traditional binning approach, linear and piecewise linear logistic models, and cubic regression splines. We determined that the best approach for modeling data collected from the sonar system was B-splines with careful knot selection.

Nous avons étudié les méthodes d'évaluation de la précision d'estimateurs de probabilité de détection en fonction de la distance à la cible pour les systèmes de sonar latéraux utilisés par la Marine canadienne. Par simulation, nous avons évalué les probabilités de couverture et la longueur de plusieurs intervalles de confiance pour proportions binomiales construits par l'approche traditionnelle de regroupement (binning), par régression logistique linéaire et linéaire par morceaux, et par splines de régression cubiques. Nous avons déterminé que la meilleure approche pour la modélisation des données collectées par le système de sonar était les splines B avec noeuds minutieusement choisis.

[MS-245]

Measuring Household Consumption: A Survey Methods Experiment in Tanzania

Mesurer la consommation des ménages : une expérience en méthodes d'enquête en Tanzanie

John GIBSON, *University of Waikato*, Kathleen BEEGLE, *World Bank*, Joachim DEWEERDT, *EDI*, Jed FRIEDMAN, *World Bank*

This poster reports on a survey methods experiment where eight consumption modules were randomized within 168 enumeration areas (in 5 rural and 2 urban districts in Tanzania) with 3 households per method per EA. Modules vary by method of data capture (3 diaries and 5 recall questionnaires), by level of respondent (1 individual diary and 2 household diaries), by length of reference period (7 days, 14 days and typical month), and by the number of items in the recall list (58 detailed foods and beverages nested within 11 broad categories, and also the 17 most important individual foods and beverages).

Cette affiche présente une expérience en méthodes d'enquête où huit modules de consommation ont été randomisés parmi 168 zones de recensement (dans 5 districts ruraux et 2 districts urbains en Tanzanie) avec 3 ménages par méthodes par zone de recensement. Les modules varient par méthode de cueillette des données (3 journaux et 5 questionnaires de rappel), par niveau du répondant (1 journal individuel et 2 journaux de ménage), par longueur de la période de référence (7 jours, 14 jours et la longueur d'un mois typique), et par le nombre d'items dans la liste de rappel (58 aliments et boissons précis emboîtés dans 11 grandes catégories, et les 17 aliments et boissons les plus importants).

[MS-246]

Hypothesis Tests for Functional Data Based on Distances: a Study Using Splines

Tests d'hypothèses pour données fonctionnelles basés sur les distances : une étude utilisant des splines

Camila PEDROSO ESTEVAM DE SOUZA, *University of British Columbia*, Ronaldo DIAS, *University of Campinas*

Hypothesis Testing is a problem of great interest in Functional Data Analysis. For observations from a sample of curves generated according to some stochastic process we want to test if the mean of the process is equal to a given function. For observations from samples of curves from two stochastic processes, we wish to test the equality of the means of the two processes. Test statistics based on the Kullback-Leibler, Hellinger and L1 distances and a test statistic based on integrated squared differences are introduced and their properties are studied through simulation.

Les tests d'hypothèses sont un problème d'un grand intérêt en analyse de données fonctionnelles. Pour des observations provenant d'un échantillon de courbes générées par un processus stochastique, nous voulons tester si la moyenne du processus est égale à une certaine fonction. Pour des observations provenant d'échantillons de courbes générées par deux processus stochastiques, nous voulons tester l'égalité des moyennes des deux processus. Des statistiques de test basées sur les distances de Kullback-Leibler, d'Hellinger et L1 et une statistique de test basée sur les différences carrées intégrées sont présentées et leurs propriétés étudiées par simulation.

Session 10

Mercredi 3 juin • Wednesday, June 3, 08:30 – 09:30

IRC 6

Pierre Robillard Award Address

Allocution du récipiendaire du prix Pierre-Robillard

[MS-247]

Statistical Methods for Multi-State Analysis of Incomplete Longitudinal Data

Méthodes statistiques pour l'analyse multi-états de données longitudinales incomplètes

Baojiang CHEN, *University of Washington*, Grace YI & Richard COOK, *University of Waterloo*

Longitudinal studies often feature incomplete data. In the first part of the talk I introduce a Markov model for the analysis of longitudinal categorical

Les études longitudinales sont souvent à données incomplètes. Dans la première partie de cet exposé, je présente un modèle de Markov pour l'analyse de données longitu-

data which facilitates modeling marginal means and association structures. A likelihood formulation is employed for inference, so the resulting estimators enjoy properties such as optimal efficiency and consistency. The second part focuses on semiparametric analyses of longitudinal data. Here I consider use of augmented inverse probability weighted generalized estimating equations to deal with incomplete data due to missing responses and missing covariates. An application to a smoking prevention study is given for illustration.

dinales catégorielles facilitant la modélisation de moyennes marginales et de structures d'association. L'inférence est basée sur une formulation de la vraisemblance, permettant aux estimateurs résultants de bénéficier de propriétés telles l'efficacité optimale et la cohérence. La deuxième partie porte sur l'analyse semi-paramétrique de données longitudinales. Je considère ici l'utilisation d'équations d'estimation généralisées augmentées et pondérées par probabilités inverses afin de traiter des données incomplètes à réponses et covariables manquantes. Une application à une étude de prévention du tabagisme est donnée à des fins d'illustration.

Session 11A Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00 IRC 1
Statistics/Mathematics Finance 2
Statistique/Mathématique Finance-2

[MS-248]

Bond Valuation Under Regime-switching models

Évaluation d'obligations sous modèles à changement de régimes

Alex BADESCU & Robert ELLIOT, *University of Calgary*, Tak Ken SIU, *Curtin University of Technology*

This paper studies the bond valuation problem when the short term rate follows a discrete-time Markovian regime-switching model. The states of the Markov chain can be interpreted as various states of the economy. Since the market is incomplete, we propose a change of measure that incorporates both the market price of factor risk and the market price of regime shift risk. Closed-form solutions for zero-coupon bonds are derived based on the semi-martingale decomposition of the Markov chain. Numerical examples and comparisons for different model specifications are also considered.

Cet article étudie le problème de l'évaluation d'obligations lorsque le taux à court terme suit un modèle markovien à changement de régime à temps discret. Les états de la chaîne de Markov peuvent être interprétés comme différents états de l'économie. Puisque le marché est incomplet, nous proposons un changement de mesure qui incorpore le prix du marché du risque associé aux facteurs et le prix du marché associé au risque associé aux changements de régime. Des solutions analytiques pour obligations coupon zéro basées sur la décomposition en semi-martingale de la chaîne de Markov sont obtenues. Des exemples numériques et des comparaisons pour différentes spécifications du modèle sont aussi considérés.

[MS-249]

Pricing Basket Options With Barriers

Fixation des prix d'options sur panier en présence de barrières

Adam KOLKIEWICZ, *University of Waterloo*, Joonghee HUH, *Prudential Financial*

In this paper, we consider computational methods of pricing basket contracts whose payoffs depend on breaching a barrier. While the problem of pricing barrier options on a single asset has received a lot of attention in the literature, very few results deal with portfolios of assets. We adopt recently developed sharp large deviations techniques in order to approximate exit time probabilities of a mul-

Dans cet article, nous considérons des méthodes numériques de fixation des prix de contrats sur panier dont le gain dépend du franchissement d'une barrière. Bien que la fixation de prix d'options à barrière sur un seul actif ait fait l'objet de beaucoup d'attention dans la littérature, très peu de résultats font référence à des portefeuilles d'actifs. Nous adoptons des techniques récemment développées de grandes déviations afin d'approximer les probabilités de temps de sortie d'un

tivariate Brownian bridge. We use this approach in conjunction with simulation techniques to formulate an efficient algorithm of pricing such contracts. We illustrate the method with several numerical examples.

pont brownien multivarié. Nous utilisons cette approche en conjonction avec des techniques de simulation pour formuler un algorithme efficace de fixation des prix de tels contrats. Nous illustrons la méthode avec plusieurs exemples numériques.

[MS-250]

Swinging in Liquidity: The Use of Swing Options to Manage Liquidity Risk

L'utilisation d'options swing pour gérer le risque d'illiquidité

Mark REESOR, *University of Western Ontario*, Shudan LIU, *TD*

The evaporation of liquidity in certain markets has caused major financial stresses. One example is the asset-backed commercial paper market in Canada. Issuers were unable to rollover maturing securities due to a lack of buyers for the new issue. Here we propose the use of "swing" or "take-or-pay" options to manage funding liquidity risk. Swing options have long been used in commodity markets as a means for consumers to control the delivery amount and timing of the underlying asset. An example in which the underlying asset is a bond is analysed from both the issuer's and consumer's perspectives.

L'évaporation de la liquidité dans certains marchés a causé des stress financiers majeurs. Un exemple est le marché des billets de trésorerie adossés à des créances au Canada. Les émetteurs ne pouvaient refinancer les billets à maturité à cause d'un manque d'acheteurs pour de nouveaux billets. Nous proposons ici l'utilisation d'options «swing» ou «à prendre ou à payer» pour gérer le risque d'illiquidité. Les options swing ont longtemps été utilisées en bourses de marchandises comme moyen pour les consommateurs de contrôler la quantité et le moment de livraison de l'actif sous-jacent. Un exemple où l'actif sous-jacent est un bon est analysé de la perspective de l'émetteur et du consommateur.

Session 11B

Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00

IRC 4

Modelling and Measuring Pollution and Natural Systems

Modélisation et mesure de la pollution et des systèmes naturels

[MS-251]

Probabilistic Weather Prediction for Local Weather Users based on Statistical Ensembles of Forecasts

Prévision probabiliste météorologique pour usagers locaux basée sur des ensembles de prévisions statistiques

Yulia GEL & Bei CHEN, *University of Waterloo*

Probabilistic weather forecasting aims to produce a predictive interval for a future atmospheric quantity with a corresponding confidence level rather than a point forecast. This is of particular importance to weather users who need not only a reliable forecast but an assessment of a related uncertainty for various risk management tasks, e.g. estimating a chance of snowfall during the Olympic games in Vancouver 2010 or a drop in temperature for planning ice-wine production in Ontario etc. In this talk we discuss two procedures for generating statistical ensembles of surface temperature forecasts using a hierarchical geostatistical model and sieve bootstrap.

La prévision météorologique probabiliste vise à produire un intervalle de prévision pour une quantité atmosphérique future avec un niveau de confiance correspondant plutôt qu'une prévision ponctuelle. Ceci est d'une importance particulière pour les utilisateurs de la météo ayant besoin non seulement d'une prévision fiable, mais aussi d'une estimation de l'incertitude pour diverses tâches de gestion des risques, comme l'estimation de la probabilité de chute de neige durant les Jeux olympiques de Vancouver ou d'une diminution de la température pour planifier la production de cidre de glace en Ontario. Dans cet exposé, nous discutons de deux procédures pour générer des ensembles statistiques de prévisions de température de surface en utilisant un modèle géostatistique hiérarchique et le bootstrap tamis.

[MS-252]**Industrial Sources Influence Air Concentrations of Hydrogen Sulfide and Sulfur Dioxide in Rural Areas of Western Canada****Des sources industrielles influencent la concentration atmosphérique de sulfure d'hydrogène et de dioxyde de soufre des zones rurales de l'ouest canadien**

Hyang Mi KIM, *University of Calgary*, Igor BURSTYN, Ambikaipakan SENTHILSELVAN & Nicola CHERRY, *University of Alberta*, Elise PIETRONIRO & Cheryl WALDNER, *University of Saskatchewan*

A survey of monthly average concentrations of sulfur dioxide and hydrogen sulfide at rural locations in western Canada (Alberta, British Columbia, and Saskatchewan) was conducted in 2001-2002, as part of an epidemiological study of the effects of oil and gas industry emissions on the health of cattle. The regulatory authorities supplied the information on location of the different oil and gas facilities during the study period and, for Alberta, provided data on hydrogen sulfide content of wells and flaring volumes. The results suggest that oil and gas extraction activities contribute to air pollution in rural areas of western Canada.

Une étude des concentrations mensuelles moyennes de dioxyde de soufre et de sulfure d'hydrogène à diverses localités rurales de l'ouest canadien (Alberta, Colombie-Britannique, Saskatchewan) a été menée en 2001-2002, au sein d'une étude épidémiologique des effets des émissions de l'industrie du pétrole et du gaz naturel sur la santé du bétail. Les organismes de régulation ont fourni la position des différentes installations pétrolières et gazières durant la période de l'étude, ainsi que de données sur le contenu en sulfure d'hydrogène des puits et sur le volume brûlé à la torche en Alberta. Les résultats suggèrent que les activités d'extraction pétrolière et gazière contribuent à la pollution atmosphérique en zone rurale de l'ouest canadien.

[MS-253]**A Bayesian Capture-recapture Model With Premature Radio-tag Failure****Un modèle bayésien de capture-recapture avec défaillance prématurée des radio-étiquettes**

Laura COWEN, *University of Victoria*

Capture-recapture studies are used to estimate population parameters such as abundance and survival. Animals are captured, given unique tags, and released. If tags are lost during the study, parameter and variance estimates may be biased. An example of tag loss occurred during a radio-telemetry study of chinook smolts on the Columbia River. Here radio-tags failed before the end of the study. Cowen and Schwarz (2005) developed a model to incorporate secondary radio-failure data to adjust survival estimates. We look at how to move this model into a Bayesian framework.

Les études de capture-recapture sont utilisées pour l'estimation de paramètres de populations comme l'abondance et le taux de survie. Les animaux sont capturés et une étiquette unique leur est donnée avant qu'ils ne soient relâchés. Si les étiquettes sont perdues pendant l'étude, les estimations des paramètres et de leurs variances seront biaisées. Un exemple où les étiquettes furent perdues est survenu pendant une étude de radio-télémétrie sur les saumons chinook juvéniles dans la rivière Columbia. Dans ce cas, les radios-étiquettes avaient cessé de fonctionner avant la fin de l'étude. Cowen et Schwarz (2005) ont développé un modèle incluant des données secondaires de radio-défaillance afin d'ajuster les estimations de taux de survie. Nous regardons comment formuler ce modèle dans un cadre bayésien.

Session 11C Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00 IRC 5

Multivariate Survival Analysis
Analyse de survie à variables multiples

[MS-254]

Cure Models for Clustered Survival Data

Modèles avec fraction non à risque pour données de survie en grappes

Yingwei PENG, *Queen's University*, Jeremy TAYLOR, *University of Michigan*

The cure model is a class of models for survival data with a surviving or cure fraction. There are extensive discussions in the literature on this model for independent survival data. However, cure models for clustered survival data have not received much attention until recently. In this talk, I will review some existing models proposed recently by other researchers, and provide a detailed discussion on the marginal and random effects model approaches to deal with this type of data, including our recent work as well. The application of some of the models is demonstrated using a tonsil cancer data.

Les modèles avec fraction non à risque forment une classe de modèles pour données de survie avec une fraction survivante ou non à risque. L'utilisation de ce modèle pour données de survie indépendantes est abondamment abordée dans la littérature. Cependant, les modèles avec fraction non à risque pour données de survie en grappes n'ont pas fait l'objet de beaucoup d'attention jusqu'à récemment. Dans cet exposé, je fais la revue de certains modèles existants proposés par d'autres chercheurs, et fournis une discussion détaillée des approches du modèle marginal et du modèle à effets aléatoires pour traiter ce type de données, incluant nos travaux récents. Nous démontrons l'application de certains de ces modèles avec des données sur le cancer des amygdales.

[MS-255]

A Copula-based Mixed Poisson Model for Bivariate Recurrent Events under Event-Dependent Censoring

Un modèle mixte de Poisson basé sur les copules pour événements récurrents bivariés avec censure dépendante des événements

Richard COOK & Jerald LAWLESS, *University of Waterloo*

We describe a bivariate mixed Poisson model in which a copula function is used to model the association between two gamma distributed random effects. The model is motivated by trials involving event-dependent censoring based on one type of event, which arises when one event is sufficiently serious that its occurrence may influence the decision of whether to withdraw a patient from a study. The asymptotic biases of marginal estimators of rate and mean functions are discussed, as well as associated treatment effects and the proposed methods are shown to perform well. An application illustrates the proposed methods.

Nous décrivons un modèle de Poisson mixte bivarié dans lequel une copule est utilisée pour modéliser l'association entre deux effets aléatoires de distribution gamma. Le modèle est motivé par des essais impliquant une censure dépendante des événements basée sur un type d'événement, se produisant lorsqu'un événement est assez sérieux pour influencer la décision de retirer un patient de l'étude lorsqu'il se produit. Nous discutons des biais asymptotiques des estimateurs marginaux des fonctions de taux et de moyenne, ainsi que des effets de traitement associés. Les méthodes proposées performant bien, ce qui est illustré au moyen d'une application.

[MS-256]

The Multivariate Kaplan-Meier Estimator via an Eigenfunction Problem**L'estimateur de Kaplan-Meier multivarié au moyen d'un problème de fonctions propres**Arusharka SEN, *Concordia University*, Winfried STUTE, *Justus-Liebig University, Giessen, Germany*

The problem of constructing a multivariate version of the well-known Kaplan-Meier estimator, under multivariate random censoring, has a long history. We show that the survivor function of a random vector is an eigenfunction of its cumulative hazard measure. This leads to the Kaplan-Meier estimator in arbitrary dimensions, which has the product-limit structure in one dimension and asymptotic efficiency in general. The estimator was also found in an incomplete form by Prentice, Moody and Wu (2004). We also provide a Greenwood's formula for the variance of the estimator, some simulation results and an illustration with data on kidneys.

Le problème de la construction d'une version multivariée du célèbre estimateur Kaplan-Meier sous censure aléatoire multivariée a une longue histoire. Nous montrons que la fonction de survie d'un vecteur aléatoire est une fonction propre de sa mesure de risque cumulative. Ceci mène à l'estimateur Kaplan-Meier de dimension arbitraire, ayant une structure de type produit limite en dimension un et asymptotiquement efficace de façon générale. L'estimateur a également été obtenu de façon incomplète dans Prentice, Moody et Wu (2004). Nous présentons également une formule de Greenwood pour la variance de l'estimateur, quelques résultats de simulation et une illustration avec données rénales.

Session 11D**Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00****IRC 3****Theory and Application of the Finite Mixture Models****Théorie et application des modèles de mélange fini**

[MS-257]

Financial Market Regimes and Mixture Models**Régimes de marchés financiers et modèles de mélanges**Leonard MACLEAN & Yonggan ZHAO, *Dalhousie University*

There is growing evidence that equity markets exhibit regimes. Within a regime the standard Geometric Brownian motion model for equity returns is appropriate, but the parameters in that model depend on the regime. The transition between regimes can be defined by a Markov process with transition probabilities depending on state variables. A mixture model is considered for the returns on equities in the US financial market. Using daily data from 1990 - 2008, the conditional parameters and the mixing coefficients are estimated using the EM algorithm. The number of regimes is determined from the Bayesian Information Criterion.

La présence de régimes dans les marchés des actions est du plus en plus vraisemblable. Dans un régime le modèle de mouvement brownien géométrique standard pour rendements est approprié, mais les paramètres de ce modèle dépendent du régime. Les transitions entre régimes peuvent être définies par un processus de Markov avec probabilités de transition dépendant de variables d'état. Un modèle de mélanges est considéré pour les rendements des capitaux propres dans le marché financier américain. Les paramètres conditionnels et les coefficients de mélange sont estimés à partir de données journalières de 1990 à 2008 au moyen de l'algorithme EM. Le nombre de régimes est déterminé à l'aide du critère d'information bayésien.

[MS-258]**Efficient Adaptive Mixture Estimation
Estimation adaptative efficace de mélanges**

Jiayang SUN, *Case Western Reserve University*, Peng LIU, *Eli Lilley, USA*, Chen JIAHUA, *University of British Columbia*

Data mining is important in scientific research and knowledge discovery. In this article, we study new adaptive estimation procedures, the partial EM and its Bayesian variants for analyzing large or streaming data from heterogeneous populations. Our procedures are fast, and provide good alternatives when a full EM procedure can or can not be run. The partial EM is also consistent, and has a super-efficiency when the distribution of second batch of the data has an extra component to that of the first batch. The procedures are applied to a small data set and a network intrusion data.

La forage de données est importante en recherche scientifique et en découverte de connaissances. Dans cet article, nous étudions de nouvelles procédures d'estimation adaptative, l'algorithme EM partiel et ses variantes bayésiennes pour l'analyse de jeux de données de grande taille provenant de populations hétérogènes. Nos procédures sont rapides et sont de bonnes alternatives lorsque l'algorithme EM complet peut ou ne peut pas être utilisé. L'algorithme EM partiel est aussi convergent et peut être super-efficace lorsque la distribution d'un deuxième groupe de données a une composante additionnelle par rapport à la distribution d'un premier groupe. Les procédures sont appliquées à un petit jeu de données et à des données d'intrusion dans un réseau.

[MS-259]**Testing the Order of a Finite Mixture
Tester l'ordre d'un mélange fini**

Pengfei LI, *University of Alberta*, Jiahua CHEN, *University of British Columbia*

Hypothesis test on the order of the finite mixture model has met substantial challenge. There are few general results on the test of order and some elegant theoretical results have operational difficulties in applications. We propose an EM-test for the general problem of testing the null model that the order is m_0 versus the alternative model that the order is larger than m_0 . For any positive integer m_0 , the null limiting distribution of the EM-test is found to be a mixture of χ^2 distributions. Simulation shows that the limiting distributions closely match the finite sample distributions of the EM-test.

Tester une hypothèse sur l'ordre d'un modèle de mélange fini comporte des défis substantiels. Il y a peu de résultats généraux sur le test de l'ordre d'un mélange et certains résultats théoriques élégants, lorsque mis en application, présentent des difficultés opérationnelles. Nous proposons un test EM pour le problème général de confronter l'hypothèse nulle que l'ordre est de m_0 à l'hypothèse alternative que l'ordre est supérieur à m_0 . Nous trouvons que, pour n'importe quel entier positif m_0 , la distribution nulle limite du test EM est un mélange de distributions χ^2 . Une simulation montre que les distributions limites sont très près des distributions du test EM sous échantillon fini.

Session 11E Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00 CHE 150

Models for Count Data and Excessive Zeros

Modèles pour données de dénombrement et zéros excessifs

[MS-260]

Sample Size Calculation For Hierarchical Poisson Regression Models

Détermination de la taille échantillonnale dans un modèle hiérarchique de régression de Poisson

Nabil CHANNOUF, *Group for Research in Decision Analysis (GERAD)*, Marc FREDETTE, *HEC Montreal*, Brenda MACGIBBON, *UQAM*

Sample size calculations are necessary for classical tests about the parameters in linear and generalized linear regression models, but few methods are given for models with correlated data. In some biomedical applications the responses are counts and the Poisson distribution is often used to model the data in this case. Because the outcomes are frequently correlated, some stochastic regressors are added to the model as latent variables.

We investigate an approach to sample size calculations proposed by Shieh (2001) and we propose an extension to the hierarchical Poisson generalized model (using gamma random effects). We also study the effect of additional covariates on the model.

La détermination de la taille d'un échantillon est largement utilisée avec les tests classiques dans les modèles de régression linéaire simple ou généralisée. Cependant très peu d'études sont réalisées dans le cas de modèles avec observations corrélées. Dans certaines applications biomédicales, la régression Poisson est souvent utilisée quand la variable étudiée représente le nombre d'événements. Souvent, ces observations sont corrélées et donc des effets aléatoires non observés sont ajoutés au modèle.

Nous proposons une extension de l'approche de Shieh (2001) pour le modèle hiérarchique de régression de Poisson avec la distribution Poisson-gamma et nous étudions l'effet de variables explicatives additionnelles.

[MS-261]

Statistical Inferences of Childhood Injury Data using Mixed-Effects Zero-Inflated Poisson Models

Inférences statistiques sur des données de blessures chez les enfants au moyen de modèles à effets mixtes de Poisson avec surreprésentation de zéros

Jeffrey DOWDEN, *NL Centre for Health Information*, Khokan SIKDAR & Reza ALAGHEHBANDAN, *NL Centre for Health Information, Memorial University*

Count data exhibiting excess zeros are common in epidemiological and public health research. Such data are routinely modeled using zero-inflated Poisson (ZIP) regression. In many cases, these data exhibit correlation due to clustering (e.g. patients observed within hospitals) whereby zero-inflation and lack of independence occur simultaneously. In this presentation, we use a mixed-effects zero-inflated Poisson (MEZIP) model to account for both the high proportion of zeros and correlation of observations. A simulation study is performed to examine the performance of parameter estimates. An application of MEZIP is demonstrated using hospital admission data on childhood injury in Newfoundland and Labrador.

Les données de dénombrement à excès de zéros sont fréquentes en recherche épidémiologique et en santé publique. De telles données sont souvent modélisées par régression de Poisson à surreprésentation de zéros (ZIP). Dans plusieurs cas, ces données présentent une corrélation due à des grappes (par exemple, des patients provenant des hôpitaux); le manque d'indépendance et la surreprésentation de zéros surviennent alors simultanément. Dans cette présentation, nous avons utilisé un modèle de Poisson à effets mixtes et à surreprésentation de zéros (MEZIP) pour tenir compte à la fois de la haute proportion de zéros et de la corrélation dans les observations. Une étude de simulation est effectuée afin d'examiner la performance des estimateurs paramétriques. L'application de MEZIP est illustrée à l'aide de données d'hospitalisation d'enfants à Terre-Neuve et au Labrador.

[MS-262]

Non-linear Poisson Regression Using Neural Networks: Application to Cognitive Score Changes in Elderly Régression de Poisson non linéaire utilisant des réseaux de neurones : application aux changements de scores cognitifs chez les personnes âgées

Nader FALLAH, Kenneth ROCKWOOD & Arnold MITNITSKI, *Dalhousie University*

We described an extension of the Poisson regression model by an artificial neural network (a multi-layer perceptron). This relaxed the assumptions of traditional Poisson regression model while including it as a special case. From the simulation study we found that the Poisson regression models worked well when the linearity assumption held but the neural network models could largely improve the prediction in nonlinear situations. We investigated the changes in cognition by comparing neural network extension of Poisson regression with linear Poisson regression and found that the accuracy rate of prediction by the proposed model was superior to a Poisson regression.

Nous décrivons une extension au modèle de régression de Poisson par un réseau de neurones artificiel (un perceptron à plusieurs couches). Ceci assouplit les hypothèses du modèle de régression de Poisson traditionnel tout en l'incluant comme cas particulier. Une étude de simulation nous montre que le modèle de régression de Poisson linéaire fonctionne bien lorsque la condition de linéarité est satisfaite, mais le modèle de réseaux neuronaux peut grandement améliorer la prévision en présence de non-linéarité. Nous avons étudié les changements en cognition en comparant l'extension par réseaux de neurones de la régression de Poisson à la régression de Poisson linéaire et avons trouvé que la précision de la prévision du modèle proposé était supérieure à celle de la régression de Poisson.

[MS-263]

Modeling Count Data with an Application to Adverse Drug Reaction in Hospitalized Patients Modélisation de données de dénombrement avec une application à la réaction adverse aux médicaments chez les patients hospitalisés

Khokan SIKDAR, Jeffrey DOWDEN & Reza ALAGHEHBANDAN, *Newfoundland and Labrador Centre for Health Information, Memorial University of Newfoundland*, Don MACDONALD, *Newfoundland and Labrador Centre for Health Information*, Veeresh GADAG, *Memorial University of Newfoundland*

Research undertaken in the healthcare system often employs event counts, such as the number of hospitalization or days spent in hospital. We considered the problem of modeling count data having a large number of zeros through the use of various models such as Poisson, negative binomial, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB). These models are discussed with a focus on specification, estimation and testing, and which model is most appropriate to use in studying predictors of adverse drug reaction among hospitalized patients. Analytical and graphical methods are used to aid in the model selection and interpretation.

La recherche menée dans le système de santé utilise souvent des dénombrements d'évènements, tels le nombre d'hospitalisation ou de jours passés à l'hôpital. Nous avons considéré le problème de la modélisation de données de dénombrement ayant un grand nombre de zéros au moyen de divers modèles tels le modèle de Poisson, négatif binomial, de Poisson à surreprésentation de zéros et négatif binomial à surreprésentation de zéros. Ces modèles sont abordés sous l'optique de la spécification, de l'estimation et des tests, et l'étude de prédicteurs de réactions adverses aux médicaments parmi des patients hospitalisés. Des méthodes analytiques et graphiques sont utilisées pour aider à la sélection et à l'interprétation du modèle.

[MS-264]**Managing the Essential Zeros in Quantitative Fatty Acid Signature Analysis****Gérer les zéros essentiels en analyse quantitative de signature d'acides gras**

Connie STEWART, *University of New Brunswick Saint John*, Christopher FIELD, *Dalhousie University*

Quantitative fatty acid signature analysis (QFASA) is a recently developed diet estimation technique that depends on statistical methods. QFASA has been used successfully to obtain point estimates of the diet of predators such as seals and seabirds. In this talk, inference for a population of predators is considered. As the data is compositional and often with zeros, special techniques are required. We propose a mixture model approach to manage the zeros. Our methods are assessed through simulations using "pseudo-seals". Finally, a real-life data set involving seals will be presented to illustrate the usefulness of our methods in practice.

L'analyse quantitative de signature d'acides gras (AQSAG) est une technique d'estimation diététique récemment développée dépendant de méthodes statistiques. L'AQSAG a été utilisée avec succès pour obtenir des estimateurs ponctuels de la diète de prédateurs tels les phoques et les oiseaux marins. Dans cet exposé, nous considérons l'inférence pour une population de prédateurs. Comme les données sont compositionnelles avec des zéros, des techniques particulières sont nécessaires. Nous proposons une approche par modèles de mélanges pour traiter les zéros. Nos méthodes sont évaluées par simulations utilisant des «pseudo-phoques». Finalement, un jeu de données réelles impliquant des phoques sera présenté afin d'illustrer l'utilité pratique de notre méthode.

[MS-265]**Probit- Log- Skew-Normal Mixture Model for longitudinal outcomes with excess zeros****Modèle de mélanges probit/log-normaux-assymétrique pour réponses longitudinales avec excès de zéros**

Sadia MAHMUD, *Aga Khan University, Karachi, Pakistan*, Neil JOHNSTON, *McMaster University*

In this paper we present the extension of the probit/log- skew normal mixture model (Chai and Bailey 2008) to a longitudinal study design. The correlation among measurements on the same subject was modeled by including two random intercepts in the linear predictors of probit and log-skew normal components, assuming the random effects follow a bivariate normal distribution. The proposed model was fitted using a real dataset from a study titled 'Diary of Asthma and Viral Infection Study' a 12 - month cohort study conducted to monitor infection and respiratory symptoms in children aged 5-11 years, with and without asthma.

Dans cet article, nous présentons une étendue du modèle de mélange probit/log-normale- asymétrique (Chai et Bailey 2008) à un plan d'expérience longitudinal. La corrélation entre les mesures du même sujet a été modélisée en incluant deux ordonnées à l'origine aléatoires dans les prédicteurs linéaires des composantes probit et log-normale-asymétrique, en supposant que les effets aléatoires suivent une distribution normale bivariée. Le modèle proposé a été ajusté à des données réelles d'une étude appelée 'Diary of Asthma and Viral Infection Study', une étude cohorte de 12 mois visant à surveiller les infections et symptômes respiratoires chez les enfants de 5 à 11 ans avec et sans asthme.

Session 11F Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00 CHE 250**Biostatistical Applications, Issues, and Interpretations
Applications, problèmes et interprétations biostatistiques****[MS-266]****Modifying a Modifier, Confounding a Modifier, Confounding a Confounder
Modifier un modificateur, confondre un modificateur et confondre un confondant**Gordon FICK, *University of Calgary*

Take a scenario from epidemiology with interest in a disease-exposure relationship with age and gender as possible confounders / modifiers. The investigator wishes to model the (log) odds of disease. The comparison of the available models that address exposure, age and gender will include the consideration of various forms of modification and confounding. With some illustrations, it will be shown that there are interesting possible extensions of the notions of confounding and modification that can be considered via modeling that are not directly seen with stratified analyses.

Prenons un scénario en épidémiologie où on s'intéresse à la relation maladie-exposition avec l'âge et le genre comme possibles confondeurs/modificateurs. L'investigateur souhaite modéliser le log de la cote de la maladie. Nous comparons les modèles disponibles traitant l'exposition, l'âge et le genre et prenons en considération différentes formes de modification et de confusion. Avec quelques illustrations, nous allons montrer qu'il y a des extensions possibles aux notions de confusion et de modification pouvant être considérées par modélisation qui ne sont pas directement apparentes avec des analyses stratifiées.

[MS-267]**Statistical Issues in a Meta-analysis of Studies of Integrated Treatment Programs for Women with Substance Use Problems and Their Children****Difficultés statistiques en méta-analyses d'études de programmes de traitements intégrés pour femmes ayant des problèmes d'abus de substances et leurs enfants**Jennifer LIU & Lehana THABANE, *McMaster University*, Karen MILLIGAN, *Credit Valley Hospital*, Alison NICCOLS & Wendy SWORD, *McMaster University*, Joanna HENDERSON, *Centre for Addiction and Mental Health*, Ainsley SMITH, *McMaster University*

We conducted a meta-analysis to evaluate the effectiveness of integrated treatment programs for women with substance use issues and their children. This presentation will address some of the statistical issues we encountered, resulting primarily from varying study quality and heterogeneity in outcome measures, study design and timing of assessments. We discuss methods for combining studies using different measures of treatment effect, determining the completeness of the literature search and assessing the impact of study quality on effects of integrated treatment programs. This will be followed by a discussion of key findings, limitations and implications for clinical practice and further research.

Nous avons effectué une méta-analyse pour évaluer l'efficacité des programmes de traitement intégrés pour femmes avec problèmes d'abus de substances et leurs enfants. Cette présentation va répondre à certaines questions statistiques que nous avons rencontrées, résultant principalement de la qualité variable des études et de l'hétérogénéité dans les mesures des réponses, les plans d'expérience et les temps d'évaluation. Nous discutons de méthodes pour combiner des études utilisant différentes mesures de l'effet du traitement, déterminer la complétude de la revue de littérature et évaluer l'impact de la qualité des études sur les effets de programmes de traitement intégrés. Ceci sera suivi par une discussion des résultats-clés, des limites et des implications pour la pratique clinique et les recherches futures.

[MS-268]**Analysis and Interpretation Issues When There are Multiple Control Groups in a Case-control Study
Problèmes d'analyse et d'interprétation lorsqu'il y a plusieurs groupes contrôle dans une étude cas-témoins**Binod NEUPANE, Mark LOEB, Stephen WALTER & Paul KRUEGER, *McMaster University*

When there are two or more control groups for a case-control study, problems about analyzing and interpreting the study results arise. With reference to a case-control study with hospitalized patients and community residents as controls to identify the risk factors of pneumonia hospitalization, we will discuss the uses of multiple logistic regression fitting two models simultaneously and binary logistic regressions fitting two models separately on cases vs. hospital cases and vs. community controls. We will also investigate the discrepancies in inferences drawn using two types of controls and discuss the possible mechanisms, e.g., selection bias explaining such differences.

Lorsqu'il y a deux groupes de contrôle ou plus en études cas- contrôle, des problèmes surviennent en analyse et interprétation des résultats. En faisant référence à une étude cas-contrôle avec patients hospitalisés et résident d'une communauté comme contrôles pour identifier les facteurs de risques d'hospitalisation pour pneumonie, nous allons discuter de l'utilisation de régression logistique multiple ajustant deux modèles simultanément et de régressions logistiques binaires ajustant deux modèles séparément. Nous allons aussi étudier les différences entre les inférences à partir de deux types de contrôles et discuter des possibles mécanismes (par exemple, le biais de sélection) expliquant ces différences.

[MS-269]**Effectiveness of Control Measures for Diseases That Spread in Populations With Contact Network Structure
Efficacité des mesures de contrôle de maladies transmises par structure de réseaux de contacts**Lilia RAMIREZ RAMIREZ & Mary THOMPSON, *University of Waterloo*

Based on the pgf for the total number of infected individuals during an outbreak it is possible to obtain the explicit expression for the mean final outbreak size in populations under control (isolation, mass vaccination, ring vaccination and acquaintance vaccination).

The effectiveness of the combination of these control measures and other outbreak characteristics, such as the outbreak duration and epidemic peak position, can be studied using stochastic simulations of the outbreak.

In this talk we present some of the most relevant effects of control measures considering networks with different levels of heterogeneity and different values for the agent's parameters.

En se basant sur la fgp pour le nombre total d'individus infectés lors d'une épidémie, il est possible d'obtenir une expression explicite pour la taille finale moyenne de l'épidémie chez des populations sous contrôle (isolation, vaccination massive, vaccination en anneau, vaccination des connaissances).

L'efficacité de la combinaison de ces mesures de contrôle, ainsi que d'autres caractéristiques des éclosions telles la durée de l'éclosion et la position du pic de l'épidémie, peuvent être étudiées au moyen de simulations stochastiques de l'épidémie.

Dans cet exposé, nous présentons quelques-uns des effets les plus importants des mesures de contrôle considérant des réseaux ayant différents niveaux d'hétérogénéité et différentes valeurs de paramètres des agents.

[MS-270]**Confidence Intervals for Cost Data in Cluster Randomization Trials: Does it Make a Difference?
Intervalles de confiance pour données de coûts en essais randomisés par grappe : y a-t-il une différence ?**Julia TALEBAN & GuangYong ZOU, *University of Western Ontario*

In health economics, trials often follow clustered designs, where individuals within a cluster are more similar than individuals in other clusters.

En économie de la santé, les essais sont souvent à plans en grappes, où les individus à l'intérieur d'une grappe sont plus similaires que les individus de d'autres grappes. Les inter-

Confidence intervals on the difference between mean costs are commonly used to obtain practical interpretations in health-care policy decisions. However, traditional confidence intervals are subject to forced symmetry, despite potential asymmetric distributions of parameter estimates. Asymmetric confidence intervals based on the method of variance estimates recovery for the difference between lognormal mean costs in cluster randomization trials are presented. Numerical evaluations show that they perform better than the Satterthwaite method and confidence intervals utilizing a robust cluster-adjusted standard error.

valles de confiance pour les différences entre coûts moyens sont souvent utilisés pour obtenir des interprétations pratiques de décisions en soins de santé. Cependant, les intervalles de confiance traditionnels sont sujets à une symétrie forcée, malgré une asymétrie potentielle dans la distribution des estimateurs paramétriques. Nous présentons des intervalles de confiance asymétriques pour la différence entre coûts moyens lognormaux en essais randomisés par grappes basés sur la méthode de la récupération des estimateurs de variance. Des évaluations numériques montrent qu'ils performent mieux que la méthode Satterthwaite et que les intervalles de confiance utilisant une erreur standard robuste ajustée par grappe.

[MS-271]

Development and Evaluation of an Efficient and Valid Missing Values Imputation Method to Investigate the Health Effects of Air Pollution Levels

Développement et évaluation d'une méthode efficace et valide pour l'imputation de valeurs manquantes dans l'étude des effets sur la santé de niveaux de pollution atmosphérique

Swarna WEERASINGHE, *Dalhousie University*.

Different patterns of missing values are identified in environmental data. Reliable imputation techniques are required to obtain valid estimates of the associations with sparse health outcomes. In this presentation, I will demonstrate existing imputation methods and advantageous of a new method applicable to time series of data. Validity and efficacy of the methods in terms of accurately diagnosing the association between adverse health outcomes will also be demonstrated. Existing missing values imputation methods failed to incorporate the prediction errors. The method that incorporates prediction errors provides valid and efficient estimates of the association with adverse health outcomes.

Différents motifs de valeurs manquantes sont identifiés en données environnementales. Des techniques d'imputation fiables sont nécessaires pour obtenir des estimateurs valides des associations avec résultats de santé éparés. Dans cet exposé, je vais présenter les méthodes d'imputation existantes et les avantages d'une nouvelle méthode applicable à des données de séries chronologiques. La validité et l'efficacité des méthodes, en terme d'un diagnostic correct de l'association entre résultats de santé négatifs, seront aussi démontrées. Les méthodes existantes d'imputation des valeurs manquantes n'arrivent pas à tenir compte des erreurs de prévision. La méthode incorporant ces erreurs fournit des estimateurs valides et efficaces de l'association avec résultats de santé négatifs.

Session 11G Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00 SCR 100

Survey Design and Analysis

Planification et analyse d'enquêtes

[MS-272]

The Sample Design of the Quarterly Survey of Financial Statistics: Review And Improvements

Plan de sondage de l'Enquête trimestrielle sur les statistiques financières : revue et améliorations

Nora BOHOSSIAN, Pierre DAoust & Sylvie GAUTHIER, *Statistics Canada*

The Quarterly Survey of Financial Statistics (QFS) for Enterprises collects financial data such as revenue and assets on incorporated businesses in

L'Enquête trimestrielle sur les statistiques financières des entreprises (EFT) recueille des données financières comme le revenu et l'actif de sociétés du Canada. Son principal ob-

Canada. Its main objective is to measure financial performance and activity by various industry aggregations in the Canadian economy. The survey consists of a take-all stratum of the largest enterprises and a sample of medium-sized enterprises. Recently, various revisions were made to the sample design with the aim of achieving greater sample stability. In this paper, we present an overview of the design and discuss future improvements.

KEY WORDS: Financial Survey Methodology, Stratified Sample Design

jectif est d'évaluer le rendement et l'activité financiers de l'économie canadienne en se fondant sur diverses agrégations d'industries du pays. Cette enquête est composée d'une strate à tirage complet des grandes entreprises et d'un échantillon d'entreprises moyennes. Récemment, divers aspects de son plan de sondage ont été revus afin d'augmenter la stabilité de son échantillon. Cet article présente un aperçu de ce plan et se penche sur les améliorations qu'il serait possible d'y apporter.

MOTS CLÉ : méthodologie des enquêtes financières, plan de sondage stratifié

[MS-273]

Rent Component of the Consumer Price Index

Méthodologie de la composante loyer de l'Indice des prix à la consommation

Jeannine CLAVEAU, Jack LOTHIAN & Sylvie GAUTHIER, *Statistique Canada*

The Consumer Price Index (CPI) is one of Statistics Canada's key indicators of Canada's economic wellbeing. It represents the rate of price change for goods and services bought by Canadian consumers. The Rent component of the CPI is a component representing 5% of the CPI basket of commodities. The data underlying the rental index is derived from the Labour Force Survey (LFS). In this presentation, we discuss the current methodology used to estimate the Rental Index of the CPI and potential future survey improvements.

Indicateur essentiel pour Statistique Canada, l'Indice des prix à la consommation (IPC) représente le taux de variation des prix des biens et services achetés par les consommateurs canadiens. L'indice des prix des loyers représente environ 5% de l'IPC. Les données ayant trait aux prix des loyers payés par les locataires, intrant du calcul de l'indice, proviennent de l'échantillon des locataires emprunté à l'Enquête sur la population active (EPA). Cette présentation s'intéresse à la méthodologie appliquée à la composante loyer de l'IPC et aux améliorations qu'il serait possible d'y apporter.

[MS-274]

Recent Improvements to Statistics Canada's Business Register Industry Autocoding Process

Améliorations récentes au processus de codage automatisé de l'activité industrielle pour le Registre des entreprises de Statistique Canada

José GAUDET, *Statistique Canada*

The Business Register (BR) is used as the survey frame for most of Statistics Canada's business surveys. One of the most important pieces of information available on the BR is the industry code. Recently, efforts were made to improve the quality of the industry coding. In this presentation, we will give a brief overview of the industry coding process on the BR and of the improvements that were recently implemented. We will also discuss the testing environment that we implemented to assess the impact of those improvements.

Le Registre des entreprises (RE) est utilisé comme base de sondage par la vaste majorité des enquêtes auprès des entreprises de Statistique Canada. Une des principales informations que fournit le RE est le code d'activité industrielle. Dernièrement, des efforts ont été déployés pour améliorer le codage de cette information. Dans cette présentation, nous donnerons un bref aperçu du processus de codage de l'activité industrielle sur le RE et des améliorations qui ont récemment été apportées à ce processus. Nous discuterons également de l'environnement de test qui a été mis en place pour évaluer l'impact de ces améliorations.

[MS-275]

New Methodology for the T1 tax data processing

Nouvelle méthodologie pour le traitement des données fiscales T1

Javier OYARZUN, *Statistique Canada*

Administrative data are an important source of information on Canadian economic activity and Statistic Canada tries to make an optimal use of such data. For the last few years, T1 fiscal data from non-incorporated businesses are used to produce estimates for different sectors of the economy. For 2007, more than 75% of the financial statements for non-incorporated businesses were sent electronically, thus enabling the partial imputation of paper respondents. This talk will describe the new methodology used in the production of population estimates for non-incorporated businesses.

Les données administratives sont une source importante d'information sur l'activité économique canadienne et Statistique Canada s'efforce d'en faire une utilisation optimale. Depuis quelques années, les données fiscales T1 provenant des entreprises non incorporées sont utilisées afin de produire des estimations pour différents secteurs économiques. Pour l'année 2007, plus de 75% des déclarations financières reçues pour les entreprises non incorporées furent transmises électroniquement, permettant ainsi l'imputation partielle des répondants papiers. Cette présentation décrira la nouvelle méthodologie employée lors de la production des estimations pour la population des entreprises non incorporées.

Session 11H Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00 MSL 101

Statistical Education in Action

L'éducation statistique en action

[MS-276]

Teaching Regression using Interactive Visualization

L'enseignement de la régression en utilisant la visualisation interactive.

Mikelis BICKIS, *University of Saskatchewan*

The theory of regression analysis is closely linked to concepts in multidimensional geometry. While geometric language is commonly used in textbooks, students' intuition about such ideas may not be well developed. Computer graphics allows the creation of interactive images to illustrate concepts such as subspaces, projections, residuals, and partial correlation. A set of MATLAB programs provides an interface through which students can explore the geometric ideas and test their own intuition.

La théorie de l'analyse de régression est étroitement liée aux concepts de la géométrie multidimensionnelle. Bien que les manuels utilisent souvent le langage géométrique, l'intuition concernant de telles idées peut être mal formée chez les étudiants. L'infographie rend possible la création d'images interactives pour illustrer des concepts comme les sous-espaces, les projections, les résidus et la corrélation partielle. Un ensemble de logiciels MATLAB fournit une interface qui permet aux étudiants d'explorer les notions géométriques et de vérifier leur propre intuition.

[MS-277]

Choosing a Text for a Service Course in Introductory Statistics

Choisir un livre pour un cours de service d'introduction à la statistique

Anne Michele MILLAR, *Mount Saint Vincent University*

Most universities offer a service course in statistics. These often have large classes and several sections with different instructors. Common elements, such as a joint exam or weekly tutorials, may necessitate

La plupart des universités offrent un cours de service en statistique. Ceux-ci ont souvent un grand nombre d'étudiants et plusieurs sections avec différents chargés de cours. Des éléments communs aux sections, tels des examens communs ou

the sections be consistent, for example in the use of terminology and notation, and in the general emphasis. Hence the choice of text can be crucial to the overall success of the course. We discuss possible criteria for choosing a text, and some of the challenges in changing to a new text.

des démonstrations hebdomadaires, peuvent nécessiter une certaine cohérence entre les sections, par exemple en ce qui a trait à la terminologie, à la notation et à l'emphase générale. Le choix d'un manuel peut donc être crucial au succès global du cours. Nous discutons de critères possibles pour le choix d'un manuel, et certains des défis liés au passage à un nouvel ouvrage.

Session 11I Mercredi 3 juin • Wednesday, June 3, 08:30 – 10:00 MSL 102

Stochastic Processes and Applications Processus stochastiques et applications

[MS-278]

Modelling Trajectories via Stochastic Functional Differential Equations

Modélisation de trajectoires au moyen d'équations différentielles stochastiques fonctionnelles

David BRILLINGER, *UC Berkeley*

The stochastic differential equations employed to model the motion of objects seem generally to lead to Markov processes. They involve only the most recent time point. However a case can be made that more of the past history is required in some applications. For example, an elk drinks at a stream, then it doesn't drink again for some period. Stochastic functional differential equations (SFDEs) can involve the complete past. This talk will present a procedure for fitting SFDEs and present examples of its use.

Les équations différentielles stochastiques employées pour modéliser le mouvement d'objets semblent généralement mener à des processus markoviens, puisqu'elles impliquent uniquement le passé immédiat du processus. Cependant, il est possible d'argumenter qu'une plus grande part du passé du processus est nécessaire dans certaines applications. Par exemple, un wapiti s'abreuvant à un ruisseau n'y retournera pas boire avant une certaine période de temps.

Les équations différentielles fonctionnelles stochastiques (ÉDFS) peuvent impliquer tout le passé du processus. Cet exposé présentera une procédure pour ajuster des ÉDFS et des exemples de leurs utilisations.

[MS-279]

Fractional Integrated Process for Mortality Index k_t in Lee-Carter Model

Processus intégré fractionnel pour l'indice de mortalité k_t dans le modèle Lee-Carter

Sirlei CAVASSIN, Xiaoming LIU & Hao YU, *University of Western Ontario*

Under the Lee-Carter model framework, the time varying mortality index k_t plays an important role in mortality projection. In practice, the dynamics of k_t is often modelled by a "universal" ARIMA(0,1,0). It has been shown that the parameters of this ARIMA(0,1,0) are very sensitive to the choice of fitting period, and the resulting projection could not handle the irregular patterns properly. In this paper, we propose to use fractional integrated process to model k_t in order to deal with long-term memory issues, providing more robust yet accurate prediction. Numerical results are provided to evaluate the forecast performance.

Dans le cadre du modèle de Lee-Carter, l'indice de mortalité variant avec le temps k_t joue un rôle important dans la projection de la mortalité. En pratique, la dynamique de k_t est souvent modélisée par un modèle «universel» ARIMA(0,1,0). Il a été montré que les paramètres de ce modèle ARIMA(0,1,0) sont très sensibles au choix de la période d'ajustement, et que la projection résultante ne peut pas traiter des motifs irrégulier correctement. Dans cet article, nous proposons l'utilisation de processus intégrés fractionnels pour modéliser k_t de façon à traiter les questions de mémoire à long terme, résultant en une prévision plus robuste, mais tout de même précise. Des résultats numériques sont fournis pour évaluer la performance de prévision.

[MS-280]**Stochastic Life Annuity
Rente viagère stochastique**JISOO JANG & Xiaoming LIU, *University of Western Ontario*

In the Lee-Carter model framework, the future forces of mortality are random, driven by a time-varying mortality index k_t . As a result, the risk borne by annuity providers cannot be diversified as they increase the size of their portfolio. We first investigate the effect of uncertainty in life annuity due to the unknown future path of k_t . More specifically, approximate prediction intervals for life annuity based on Lee-Carter projections are studied. Secondly, we propose a block bootstrap that takes all sources of variability into account in mortality projection, providing more accurate prediction intervals for life annuity.

Dans le cadre du modèle de Lee-Carter, les forces de mortalité futures sont aléatoires et basées sur un indice de mortalité k_t variant avec le temps. Ceci a pour conséquence que le risque porté par les payeurs de rentes ne peut être diversifié en augmentant la taille de leur portefeuille. Nous étudions d'abord l'effet de l'incertitude sur la rente viagère due aux valeurs futures inconnues de k_t . Plus spécifiquement, des intervalles de prévision approximatifs pour la rente viagère basés sur les projections de Lee-Carter sont étudiés. Ensuite, nous proposons un bootstrap par blocs tenant compte de toutes les sources de variabilité dans la projection de mortalité, apportant des intervalles de prévision plus précis pour les rentes viagères.

[MS-281]**Series Representations for Multivariate Generalized Gamma Processes via a Scale Invariance Principle
Représentations en série pour processus gamma généralisés multivariés au moyen d'un principe d'invariance aux changements d'échelle**Mahmoud ZAREPOUR, *University of Ottawa*, Hemant ISHWARAN, *Cleveland clinic foundation*

We introduce a scale invariance property for Poisson point processes and use this property to define a series representation for a correlated bivariate gamma process. This approach is quite general and can be used to define other types of multidimensional Lévy processes with given marginals. Some important special cases are bivariate G -processes, bivariate variance gamma processes and multivariate Dirichlet processes. Using the scale invariance principle we show how to construct simple approximations to these multivariate processes.

Nous présentons une propriété d'invariance aux changements d'échelle pour processus de Poisson ponctuels et utilisons cette propriété pour définir une représentation en séries d'un processus gamma bivarié corrélé. Cette approche est très générale et peut être utilisée pour définir d'autres types de processus de Lévy avec marginales données. Quelques cas particuliers importants sont les G -processus bivariés, les processus de variance gamma bivariés, et les processus Dirichlet bivariés. Nous montrons comment construire des approximations simples pour ces trois processus multivariés en utilisant le principe d'invariance aux changements d'échelle.

[MS-282]**Testing for Carryover Effects in Recurrent Event Processes
Tester pour les effets résiduels dans des processus d'événements récurrents**Candemir CIGSAR, *University of Waterloo*

In many recurrent event settings, the event intensity increases for a limited time period after each event occurrence. We refer this feature as a carryover effect. In this study, we discuss testing for carryover effects when the null model is a Poisson process. Asymptotic properties of tests are considered when a large number of independent processes

En présence d'événements récurrents, il arrive souvent que l'intensité des événements augmente pour une période de temps limitée après chaque réalisation d'événement. Nous appelons ce phénomène effet résiduel. Dans cette étude, nous discutons de test pour effets résiduels lorsque le modèle sous l'hypothèse nulle est un processus de Poisson. Les propriétés asymptotiques des tests sont étudiées lorsqu'un

are observed and also when only a single process is observed. Simulations are performed to supplement the results and to understand the performance of test statistics under the null and alternative hypotheses.

grand nombre de processus indépendants sont observés et lorsqu'un seul processus est observé. Des simulations sont effectuées afin de compléter les résultats et de comprendre la performance des statistiques de test sous les hypothèses nulles et alternatives.

[MS-283]

Estimation with Incompletely Observed Semi-Markov Processes

Estimation avec processus semi-markoviens incomplètement observés

Lihui ZHAO & X. Joan HU, *Simon Fraser University*

We consider the estimation of the transition probabilities of the embedded Markov chain based on incompletely observed semi-Markov processes. We show that the maximum likelihood estimators of the transition probabilities are unique up to an equivalence class. Both the estimators suggested by Lagakos et al. (*Biometrika*, 1978) belong to this equivalence class. Phelan (*Statistics & Probability Letters*, 1990) proposes estimators for a class of semi-Markov processes whose semi-Markov kernel has a special multiplicative form. Based on the equivalence class, we propose bounds and further construct confidence intervals for the transition probabilities.

Nous considérons l'estimation des probabilités de transition de la chaîne de Markov emboîtée basée sur des processus semi-markoviens incomplètement observés. Nous montrons que les estimateurs du maximum de vraisemblance des probabilités de transition sont uniques à une classe d'équivalence près. Les deux estimateurs suggérés par Lagakos et al. (*Biometrika*, 1978) font partie de cette classe d'équivalence. Phelan (*Statistics & Probability Letters*, 1990) propose des estimateurs pour une classe de processus semi-markoviens dont le noyau semi-markovien a une forme multiplicative particulière. Nous proposons des bornes et construisons des intervalles de confiance pour les probabilités de transition en nous basant sur la classe d'équivalence.

Session 12A Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00

IRC 1

Ecological Processes, Mechanisms, and Statistical Evidence

Processus, mécanismes et preuves statistiques écologiques

[MS-284]

The Economic Uncertainty of Emerging Infectious Diseases and Biological Invasions

L'incertitude économique des maladies infectieuses et invasions biologiques émergentes

Christopher JERDE, *University of Notre Dame*, David FINNOFF, *University of Wyoming*, Katherine SMITH, *Brown University*, Peter DASZAK, *Consortium for Conservation Medicine*, David LODGE, *University of Notre Dame*

The global transport of species is one pathway leading to the spread of infectious diseases and biological invasions. Yet, despite the resources invested in forecasting these events, there is inherently low predictive accuracy as to when and where outbreaks or invasions will occur. Here we show how a hierarchical stochastic process can be linked to an economic evaluation of an economy's net present value. The uncertainty due to invasions or outbreaks can lead to poor decisions based on the risks posed by low probability events. Avian influenza outbreak and Eurasian watermilfoil invasion are used as examples.

Le transport global d'espèces est un des chemins menant à la diffusion des maladies infectieuses et aux invasions biologiques. Malgré les ressources investies dans la prévision de tels événements, la précision des prévisions quant à l'endroit et au moment où des épidémies ou des invasions apparaîtront reste faible. Nous montrons ici un processus stochastique hiérarchique pouvant être lié à une évaluation économique de la valeur nette actuelle d'une économie. L'incertitude due aux épidémies et aux invasions peut mener à de mauvaises décisions basées sur les risques posés par des événements improbables. L'épidémie de grippe aviaire et l'invasion des mille-feuilles aquatiques eurasiennes sont utilisées comme exemples.

[MS-285]**Revealing Spatial Population Structure With Time Series Data****Révéler des structures spatiales des populations au moyen de données de séries chronologiques**Eric WARD & Eli HOLMES, *NOAA*, Haridas CHIRAKKAL & Leah GERBER, *ASU*

Identifying spatial structure in biological data is important in assessing population viability, yet data that are commonly used to reveal structure are prohibitively expensive for biologists to collect. We present a new technique for identifying and examining synchrony between subpopulations, using multivariate state space models. The approach is flexible, and can be analyzed in a maximum likelihood or Bayesian framework. To illustrate an application of these methods, we demonstrate how to test several biological hypotheses about population structure of California sea lions in the Gulf of Mexico.

L'identification de structures spatiales dans des données biologiques est importante en évaluation de viabilité des populations, mais le coût de la collecte par des biologistes du type de données habituellement utilisé pour la révélation de structures est inabordable. Nous présentons une nouvelle technique pour identifier et examiner la synchronie entre sous-populations au moyen de modèles d'espaces d'états multivariés. L'approche est flexible et peut être analysée dans un cadre bayésien ou de vraisemblance maximale. Pour illustrer une application de ces méthodes, nous montrons comment tester plusieurs hypothèses biologiques concernant la structure de la population d'otaries de Californie dans le golfe du Mexique.

[MS-286]**How a Lack of Statistical Analysis Led to Claims that Many Animals Forage Using Similar Random Walks Described by Power-law Distributions****Comment un manque d'analyse statistique mène à l'affirmation que plusieurs animaux paissent utilisant des marches aléatoires similaires décrites par des distributions de loi de puissance**Andrew EDWARDS, *Pacific Biological Station, DFO*

What do these have in common: fishermen, grey seals, microzooplankton, wandering albatrosses, bumblebees, deer, reindeer, jackals and human hunter gatherers? Through a series of papers, it has been claimed they exhibit similar movement patterns, namely Lévy flights (scale-free random walks with step lengths from probability distributions with heavy power-law tails). However, the power laws were concluded by fitting straight lines to log-log histograms of data. I show that such methods are misleading. Using likelihood methods and Akaike weights, re-analysis of the data finds that Lévy flights are far from a universal foraging strategy (Nature, 449:1044; J. Anim. Ecol., 77:1212).

Qu'ont en commun les pêcheurs, les phoques gris, le microzooplancton, les albatros hurleurs, les bourdons, les cerfs, les rennes, les chacals et les humains chasseurs-cueilleurs? À travers une série d'articles, on a affirmé que leurs patrons de mouvement étaient similaires, soit le vol de Lévy (une classe de marches aléatoires libre d'échelles où la longueur des pas est tirée de distributions à ailes relevées de type loi de puissance). Cependant, le choix des lois de puissance a été fait en ajustant des lignes droites à des histogrammes log-log des données. Je montre que ces méthodes sont trompeuses. En utilisant des méthodes basées sur la vraisemblance et de poids d'Akaike, et en analysant à nouveau les données, je trouve que le vol de Lévy est loin d'être une stratégie de pâturage universelle (Nature, 449 :1044 ; J. Anim. Ecol., 77 :1212).

Session 12B**Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00****IRC 4**

Empirical Processes and Applications to Finance

Processus empiriques et applications à la finance

[MS-287]**Residuals in the Functional Linear Model With Application to Goodness-of-fit testing****Résidus dans le modèle fonctionnel linéaire avec application aux tests d'ajustement**Piotr KOKOSZKA, *Utah State University*, Lajos HORVATH, *University of Utah*, Robertas GABRYS, *Utah State University*

The errors in the fully functional linear model are functions. For most testing procedures, a dimension reduction which defines the residuals as finite dimensional vectors is most useful. As a result, such residuals contain not only errors due to parameter estimation, but also two additional types of errors: those introduced by projecting onto a finite dimensional subspace, and those due to the estimation of the optimal subspace. We explain the effect of these errors on the distribution of the residuals, and show how these results can be used in goodness-of fit testing.

Les erreurs dans le modèle linéaire complètement fonctionnel sont des fonctions. Pour la plupart des procédures de tests, une réduction de dimension définissant les résidus comme des vecteurs de dimension finie est très utile. Les résidus contiennent alors les erreurs dues à l'estimation paramétrique et deux types d'erreurs supplémentaires : celles introduites par la projection sur un sous- espace de dimension finie, et celle due à l'estimation du sous-espace optimal. Nous expliquons les effets de ces erreurs sur la distribution des résidus, et montrons comment ces résultats peuvent être utilisés en tests d'adéquation.

[MS-288]**Monte Carlo Methods for Testing Goodness-of-fit****Méthodes Monte Carlo pour tests d'adéquation**Bruno RÉMILLARD, *HEC Montreal*

In this talk I will present Monte Carlo methods for testing goodness-of-fit in stochastic models used in financial applications.

Dans cet exposé, je présenterai des méthodes Monte-Carlo pour des tests d'adéquation de modèles stochastiques utilisés principalement en finance.

[MS-289]**Empirical Processes of Standardized Residuals in GARCH Models****Processus empiriques de résidus standardisés en modèles GARCH**Hao YU, *University of Western Ontario*

It is well known that empirical processes constructed from unmodified residuals of GARCH models converge to Gaussian processes that are related to model parameters. This makes traditional statistical inferences such as Kolmogorov-Smirnov goodness-of-fit test invalid. In this talk, empirical processes constructed from modified residuals are proposed. We show that the suggested empirical processes will converge to Gaussian processes that are free of model parameters. Applications of the proposed residual processes will be presented.

Il est bien connu que les processus empiriques construits à partir des résidus non modifiés de modèles GARCH convergent vers des processus gaussiens reliés aux paramètres du modèle. Ceci rend invalides des inférences statistiques traditionnelles telles le test d'adéquation de Kolmogorov-Smirnov. Dans cet exposé, les processus empiriques construits à partir de résidus modifiés sont proposés. Nous montrons que les processus empiriques suggérés convergeront vers des processus gaussiens libres des paramètres du modèle. Des applications des processus résiduels proposés seront présentées.

Session 12C

Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00

IRC 5

What do Mathematical Statisticians do at Statistics Canada

Travail des statisticiens mathématiques chez Statistique Canada

[MS-290]

Synthetic Data Files and PUMFs, Two Applications Taken from Post-Censal Surveys

Les fichiers de données synthétiques et les FMGD, deux applications tirées d'enquêtes post-censitaires

Cindy UBARTAS, *Statistique Canada*

In order to respond to the needs of survey data users, Statistics Canada is offering different files such as public use microdata files (PUMF)– analytical file subsets – and synthetic files – artificial data that facilitate analytical program development.

Methodologists must pay close attention to these files in order to render data accessible while protecting the confidentiality of respondents. With the help of examples stemming from post-censal surveys on Aborigines, this presentation will detail the particularities and key steps associated with creating these types of files.

Afin de répondre aux besoins des utilisateurs de données d'enquête, Statistique Canada offre différents fichiers tels que les fichiers de microdonnées à grande diffusion (FMGD), sous-ensembles des fichiers analytiques, et les fichiers synthétiques, données artificielles permettant le développement de programmes d'analyse.

Ces fichiers nécessitent une attention particulière de la part des méthodologistes pour rendre les données accessibles tout en préservant la confidentialité des répondants. Dans le cadre de cet exposé, les particularités et les étapes essentielles à la création de tels fichiers seront abordées à l'aide d'exemples tirés des enquêtes postcensitaires autochtones.

[MS-291]

Challenges and Gains in Using Administrative Data in the Unified Enterprise Survey

Défis et avantages reliés à l'utilisation de données administratives dans le cadre de l'Enquête unifiée auprès des entreprises

Chi Wai YEUNG, *Statistics Canada*

A large portion of the annual business statistics program in Statistics Canada is covered by the Unified Enterprise Survey (UES). Results from this survey are key inputs to the calculation of Canada's GDP. Statistics Canada has long recognized the potential of using administrative data in its statistics program. However, their incorporations into the UES are known to be less than optimal due to resource and operational constraints. The goal of this presentation is to illustrate some challenges faced in incorporating more administrative data and the possible gains when the constraints are removed and sound methodology is employed.

Le programme annuel de la statistique des entreprises de Statistique Canada repose en grande partie sur l'Enquête unifiée auprès des entreprises (EUE), qui fournit des données essentielles au calcul du PIB du Canada. Par ailleurs, Statistique Canada reconnaît depuis longtemps le potentiel d'intégrer des données administratives dans ses programmes de la statistique. Or, dans le cas de l'EUE, cette démarche est limitée par des contraintes opérationnelles et des restrictions sur le plan des ressources. Cette présentation porte sur les défis inhérents à l'intégration d'une plus grande quantité de données administratives dans l'EUE. Elle évoque aussi les avantages qui en ressortiront, dès lors qu'une approche méthodologique solide aura permis de surmonter ces défis.

[MS-292]**Spatial Modelling of Geocoded Crime Data****Modélisation spatiale de données géocodées sur la criminalité**Krista COLLINS, *Statistics Canada*

How are different types of crime distributed across cities? What factors are associated with high neighbourhood crime rates? These are but a few of the questions Statistics Canada is addressing in the analysis of crime rates from city neighbourhoods, as part of the project on geocoding crime data. To ensure the analysis is accurate the effects of spatial autocorrelation must be accounted for in the model. This presentation will give an overview of the spatial autoregressive models used to analyse the crime rate in several Canadian cities and the potential influence these models have for crime reduction strategies in Canada.

Quels sont les crimes que l'on commet et à quelle fréquence, selon qu'on habite une ville plutôt qu'une autre ? Quels sont les facteurs qui expliquent que les taux de criminalité sont plus élevés dans certains quartiers ? Voilà le genre de questions qu'aborde Statistique Canada dans le cadre de l'analyse des taux de criminalité dans les quartiers des villes, l'une des ramifications du projet sur le géocodage des données de la criminalité. Afin de s'assurer que l'estimation soit fidèle, le modèle doit tenir compte des effets de l'autocorrélation spatiale. Cette présentation donne un aperçu des modèles autorégressifs spatiaux utilisés pour l'analyse du taux de criminalité dans plusieurs villes canadiennes et de leur influence possible sur les stratégies de réduction de la criminalité mises de l'avant au Canada.

Session 12D**Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00****IRC 3****Statistical Methods for Causal Inference****Méthodes statistiques pour l'inférence causale****[MS-293]****Comparing Approaches to Causal Inference for Longitudinal Data: GPS vs IPTW****Comparaison de méthodes d'inférence causale de données longitudinales : SGP vs PPIT**Ashkan ERTEFAIE,, *McGill University*

Robins (1993) has developed a set of counterfactual models, marginal structural models (MSMs), for observational data, of the causal effect of a time-dependent exposure in the presence of time-dependent covariates that are themselves affected by previous treatment. The parameters of the MSM can be consistently estimated using the Inverse Probability of Treatment Weighting (IPTW). We describe an alternative approach based on the Generalized Propensity Score (GPS). We extend the GPS approach to the longitudinal setting. We then compare the strengths and weaknesses of IPTW versus GPS for causal inference from two-interval simulated longitudinal data.

Robins (1993) a développé une série de modèles "contre-factuels", les modèles structuraux marginaux (MSMs). À partir des données observées, ils modélisent les effets causés par une exposition dépendante du temps en présence de variables dépendantes du temps qui sont elles-mêmes affectées par le traitement précédent. Les paramètres de MSM peuvent être estimés en utilisant la pondération par probabilité inverse de traitement (PPIT). Nous décrivons une approche alternative basée sur les scores généralisés de propension (SGP). Nous étendons l'approche de SGP à l'arrangement longitudinal. Nous comparons alors les forces et les faiblesses de PPIT et de SGP pour l'inférence causale à partir de données longitudinales simulées pour deux intervalles.

[MS-294]**Outcome Regression, Propensity Scores, and Instrumental Variables****Régression des résultats thérapeutiques, scores de propension et variables instrumentales**Zhiqiang TAN, *Rutgers University*, Zhiqiang TAN,

A common objective in economic and epidemiologic studies is to draw inferences about the effects of treatments and actions. We give a comparative overview and present recent developments on the three approaches referred to in the title, with the aim to understand the differences and connections. Specifically, we examine different characteristics of the outcome regression and propensity score approaches, and the nature of bias and variance reduction offered by doubly robust estimation under the so-called assumption of no (unmeasured) confounding. We also discuss regression and weighting methods under instrumental variable assumptions different from the no-confounding assumption.

Un objectif commun aux études économiques et épidémiologiques est de tirer des inférences sur les effets de traitements et d'actions. Nous faisons une revue comparative et présentons les développements récents liés aux trois approches mentionnées dans le titre, dans le but d'en comprendre les différences et les similarités. Plus spécifiquement, nous examinons différentes caractéristiques de la régression sur les résultats thérapeutiques et des scores de propension, et la nature de la réduction du biais et de la variance offerte par l'estimation doublement robuste sous l'hypothèse de l'absence d'effet confondant non mesuré. Nous discutons aussi de méthodes de régression et de pondération sous des suppositions sur les variables instrumentales différentes de la supposition d'absence d'effet confondant.

[MS-295]**Flexible Marginal Structural Models****Modèles structuraux marginaux flexibles**Robert PLATT, *McGill University*

Marginal structural models (MSMs) are regression models for the average causal effect of an exposure on outcome. I review the assumptions underlying MSMs, focusing on the need to correctly specify the functional form of the exposure-treatment relationship. I discuss consequences of violating this assumption and methods for reducing dependence on this assumption in estimating causal effects. I describe a general alternative to existing approaches that allows for flexible exposure-outcome relationships while minimizing dependence on correctly specified parametric assumptions. Using simulations, I compare flexible models to incorrectly- and correctly-specified parametric models and discuss the bias-variance tradeoffs.

Les modèles structuraux marginaux (MSM) sont des modèles de régression pour l'effet causal moyen d'une exposition sur l'issue. Je fais une revue des hypothèses sous-jacentes aux MSM, en me concentrant sur la nécessité de spécifier correctement la forme fonctionnelle de la relation exposition/traitement. Je discute des conséquences d'un non-respect de cette hypothèse et des méthodes pour réduire la dépendance à cette hypothèse dans l'estimation des effets causals. Je décris une alternative générale aux approches existantes permettant plus de flexibilité dans les relations exposition-issue tout en minimisant la dépendance à des hypothèses paramétriques correctement spécifiés. Au moyen de simulations, je compare les modèles flexibles aux modèles paramétriques correctement et incorrectement spécifiés et discute du compromis biais-variance.

Session 12E Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00 IRC 6

Session in Honor of Jack Kalbfleisch's 65th Birthday: Analysis of Life History Data Arising from Non-standard Selection and Observation Schemes

Séance en honneur du 65^e anniversaire de Jack Kalbfleisch : analyse de données de cycles de vie provenant de schémas de sélection et d'observation non standards

[MS-296]

Effect of Event Dependent Sampling in the Analysis of Recurrent Events

Effet de l'échantillonnage dépendant des événements en analyse d'événements récurrents

Kajsa KVIST, Per KRAGH ANDERSEN, Jules ANGST & Lars Vedel KESSING, *University of Copenhagen*

The effect of event-dependent sampling is investigated when studying whether the risk of recurrence increases with event count. A process is selected for study if an event occurs in a given interval as seen in studies of psychiatric patients. Subsequently, either only data from selection and onwards are used in the analysis, or, both retrospective and prospective data from selected processes are used. Two methods are proposed to correct for the selection depending on selected data. The power of the methods to correct for the bias is investigated with simulations. Furthermore, the methods are applied to data on affective disorder.

Nous étudions l'effet d'échantillonnage en fonction des événements lorsque le risque de récurrence augmente avec le nombre de fois où se produit l'évènement. Un processus est choisi pour étude si un évènement se produit dans un intervalle donné tel que vu dans l'étude de patients psychiatriques. Subséquemment, soit seules les données à partir de la sélection sont utilisées dans l'analyse, soit les données prospectives et rétrospectives sont utilisées. Deux méthodes sont proposées pour corriger pour la sélection selon les données choisies. La puissance des méthodes pour corriger le biais est étudiée par simulations. De plus, les méthodes sont appliquées à des données de troubles affectifs.

[MS-297]

Analysis of Duration Data From Longitudinal Surveys

Analyse de données de durée pour études longitudinales

Jerald LAWLESS & Dagmar MARIACA-HAJDUCEK, *University of Waterloo*

Data on education, employment, health and other life processes are collected in longitudinal surveys. The durations of spells that persons spend in specific life states are often of interest, for example, the durations of jobless spells, episodes of illness, or periods of social assistance. Features of such surveys include complex sampling designs, intermittent ascertainment of data through interviews, and losses to followup through panel attrition. In addition, successive durations for an individual are typically correlated, as are durations of individuals within clusters. This talk will discuss the analysis of duration times under such conditions.

Les données d'éducation, d'emploi, de santé et relatives à d'autres processus de la vie humaine sont recueillies dans les études longitudinales. La durée de périodes passées par les individus dans des états de vie spécifiques (par exemple, les périodes de chômage, de maladies et d'assistance sociale) sont souvent d'intérêt. Les caractéristiques de telles enquêtes comprennent des plans d'échantillonnage complexes, une cueillette de données intermittentes par entrevues, et la perte de suivi par attrition. De plus, les durées successives chez un individu sont souvent corrélées, de même que les durées chez individus d'une même grappe. Cet exposé discutera de l'analyse de temps de durées sous de telles conditions.

[MS-298]

Estimating the Effect of a Time-Dependent Therapy on Restricted Mean Lifetime using Observational Data
Estimation de l'effet d'une thérapie dépendant du temps sur l'espérance de vie restreinte au moyen de données d'observations

Douglas SCHAUBEL & Jack KALBFLEISCH, *University of Michigan*

We develop semiparametric methods to estimate the effect on restricted mean lifetime of a time-dependent treatment. In the data structure of interest, both an experimental and established form of treatment are available; pre- and post-treatment hazards are non-proportional; subjects may experience periods of treatment ineligibility; treatment assignment is not randomized. The proposed methods involve weighting results from stratified proportional hazards models fitted using a generalization of case-cohort sampling. Asymptotic and finite-sample properties of the proposed estimators are evaluated. The proposed methods are applied to data from a national organ transplant registry.

Nous développons des méthodes semi-paramétriques pour estimer l'effet sur l'espérance de vie restreinte d'un traitement dépendant du temps. Dans la structure de données d'intérêt, un traitement établi et un traitement expérimental sont disponibles, les risques pré- et post-traitement ne sont pas proportionnels, les sujets peuvent subir des périodes d'inéligibilité au traitement, et l'assignation des traitements n'est pas randomisée. Les méthodes proposées impliquent la pondération des résultats de modèles stratifiés à risques proportionnels au moyen d'une généralisation de l'échantillonnage cas-cohorte. Les propriétés asymptotiques et sous échantillon fini des estimateurs proposés sont évaluées. Les méthodes proposées sont appliquées à des données d'un registre national de transplantation d'organes.

Session 12F Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00 MSL 101

Design and Analysis for Business and Industrial Statistics
Planification et analyse pour statistique industrielle et de gestion

[MS-299]

Determining Optimal Number of Samples for Constructing Multivariate Control Chart
Détermination du nombre optimal d'échantillons pour la construction d'une carte de contrôle multivariée

Sheau-Chiann CHEN & Jeh-Nan PAN, *National Cheng Kung University, Taiwan*

Practically, samples size determination plays a crucial role in estimating the quality control costs. Thus, this paper focuses on determining optimal number of samples in Phase I (trial stage) during the construction of multivariate quality control chart using the sample mean vector and the successive difference matrix. We used the detecting performance of the T^2 control chart in Phase II (monitoring stage) to evaluate the effect of number of samples. Furthermore, a trade-off analysis between the quality control costs and the accuracy of detection is performed and optimal number of samples in Phase I is determined by our cost-effective approach.

En pratique, la détermination de tailles échantillonnelles joue un rôle crucial en estimation de coûts de contrôle de la qualité. Cet article se concentre sur la détermination du nombre optimal d'échantillons en phase I (phase d'essai) durant la construction de cartes de contrôle multivariées utilisant le vecteur de moyenne échantillonnale et la matrice des différences successives. Nous utilisons la performance de la carte de contrôle T^2 en phase II (phase de surveillance) pour évaluer l'effet du nombre d'échantillons. En outre, nous effectuons une analyse du compromis entre les coûts du contrôle de la qualité et la précision de la détection et déterminons le nombre optimal d'échantillons en phase I avec notre approche rentable.

[MS-300]**Assessment of Binary Measurement Systems
Évaluation de systèmes de mesure binaires**Oana DANILA & R. Jock MACKEY & Stefan H. STEINER, *Univeristy of Waterloo*

Binary measurement systems that classify parts as pass or fail are widely used in industry. To support production and quality improvement, it is critical that the misclassification rates of such binary measurement systems are assessed. In this talk, we focus on the situation where a routine assessment investigation of the binary system is conducted and the pass rate is known prior to the study. We assume there is no available gold standard and parts are repeatedly classified using the binary measurement system. We propose new better sampling plans for estimating the misclassification rates, and also give planning recommendations.

Les systèmes de mesure binaires classant les pièces comme étant adéquates ou inadéquates sont largement utilisés en industrie. Pour soutenir la production et l'amélioration de la qualité, il est crucial que les taux de mauvaises classifications de tels systèmes de mesure binaires soient évalués. Dans cet exposé, nous nous concentrons sur la situation où une évaluation de routine du système binaire est menée et où le taux de pièces adéquates est connu avant l'étude. Nous supposons qu'il n'y a pas d'étalon-or et que les pièces sont classées de façon répétée par le système de mesure binaire. Nous proposons de nouveaux plans d'échantillonnage améliorés pour l'estimation des taux de mauvaises classifications et de recommandations pour la planification.

[MS-301]**On informational Content of Model-Based Volatility Forecasts and Implied Volatility
Contenu en information des prévisions de volatilité basées sur des modèles et de la volatilité impliquée**Amaan MEHRABIAN, Tigran ATOYAN & David STEPHENS, *McGill University*

This paper discusses the informational content of implied volatility. We examine whether the S&P 500 implied volatility index (VIX) contains any information relevant to future volatility beyond that captured in model based forecasts (MBF), e.g., GARCH and Stochastic Volatility (SV) models. For this purpose, the standard SV/GARCH models have been extended to incorporate the realized volatility as an exogenous variable in the volatility equation. Also VIX is decomposed into two components; VIX_{MBF} which is information in VIX captured by MBF and VIX^* which is constructed to be orthogonal to VIX_{MBF} and then informational content of VIX^* is analyzed.

Cet article discute du contenu informationnel de la volatilité implicite. Nous examinons si l'indice de volatilité implicite S&P 500 (VIX) contient de l'information relative à la volatilité future en plus de celle présente dans les prévisions à base de modèles (MBF), comme les modèles GARCH et de volatilité stochastique (VS). À cette fin, les modèles usuels VS/GARCH ont été étendus afin d'incorporer la volatilité réalisée comme variable exogène dans l'équation de volatilité. Le VIX est également décomposé en deux composantes : VIX_{MBF} , l'information dans le VIX contenue dans les MBF, et VIX^* , construite de façon à être orthogonale à VIX_{MBF} . Le contenu en information de VIX^* est analysé.

[MS-302]**Bayesian Network Model of Regional Biofuel Supply Chains
Modélisation de chaînes logistiques de biocarburants au moyen de réseaux bayésiens**Nathaniel NEWLANDS, *Environmental Health, Agriculture and Agri-Food Canada*, Mohua PODDER, *Joint with University of British Columbia and Agriculture and Agri-Food Canada*, James ZIDEK, *University of British Columbia*, Lawrence TOWNLEY-SMITH & Brian MCCONKEY, *Agriculture and Agri-Food Canada*

We present a Bayesian network model of bioenergy supply chain dynamics. The model links conditional dependencies between climate variables

Nous présentons un réseau bayésien pour modéliser la dynamique des chaînes logistiques de bioénergie. Le modèle lie des dépendances conditionnelles entre variables climatiques

with cropping area selection, seeding, fertilizer and water additions, harvesting, collection, pre-treatment, storage, transport, conversion and co-products. This approach incorporates uncertainty and can be easily interrogated, adapted and expanded as knowledge of a wide variety of resource stakeholders improves. The model is tested with spatial data on yield distribution and climate for energy crops in Western Canada. We present preliminary findings comparing crop mixtures and supply logistics. This work will compare network designs against future demand scenarios and known ecological constraints.

à la sélection de terres à cultiver, l'ensemencement, l'ajout d'engrais et d'eau, la récolte, le prétraitement, l'entreposage, le transport, la conversion et les coproduits. Cette approche incorpore l'incertitude et peut aisément être questionnée, adaptée et étendue alors que nos connaissances s'améliorent. Le modèle est testé avec des données spatiales de volume de récoltes et de climat pour plantes énergétiques dans l'Ouest canadien. Nous présentons des résultats préliminaires comparant des mélanges de plantes et des logistiques d'approvisionnement. Ce travail va comparer différents plans de réseaux face à différents scénarios de demande sous contraintes écologiques connues.

[MS-303]

Semifolding Two-Level Fractional Factorial Split-Plot Designs

Plans semi-repliés à deux niveaux fractionnels factoriels avec subdivision des parcelles

Jenna TICHON, *University of Manitoba*, Robert McLEOD, *University of Winnipeg*

Foldover designs have recently been advocated as a useful follow-up strategy for de-aliasing low-order effects in fractional factorial split-plot experiments. One argument against the use of foldover designs is that they are "degree-of-freedom inefficient". Semifoldover designs, which use only one-half of the runs of a foldover, have been explored in the fractional factorial setting but are considered sparingly in the fractional factorial split-plot design context. We consider the construction of split-plot semifoldovers and investigate the impact on the semifoldover selection process due to the presence of whole-plot and subplot factors.

Les plans d'expérience avec repliement ont récemment été recommandés comme stratégie de suivi utile pour le décrénelage d'effet d'ordre inférieur en expériences fractionnelles factorielles à parcelles sous-divisées. Un argument contre l'utilisation de plans à repliement est qu'ils sont inefficaces en terme de degrés de liberté. Les plans semi-repliés, n'utilisant que la moitié des séries d'un repliement, ont été considérés dans le cadre fractionnel factoriel, mais sont peu considérés dans le contexte fractionnel factoriel à parcelles sous-divisées. Nous nous intéressons à la construction de plans semi-repliés à parcelles sous-divisées et étudions l'impact sur le processus de sélection du semi-repliement de la présence de facteurs pour parcelles entières et pour sous-division de parcelles.

Session 12G Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00 CHE 150

Mixed Models, Multistate Modeling, and Measurement Error

Modèles mixtes, modélisation à plusieurs états et erreurs de mesure

[MS-304]

Two-Part Mixed and Mixture Models for Longitudinal Data with both Random and Non-Random Missing Data

Modèle aléatoire et modèle de croissance en deux parties pour données longitudinales avec données manquantes aléatoires et non-aléatoires

Depeng JIANG, *St Michael's Hospital*, Dorcas BEATON & Kenneth TANG, *St. Michael's Hospital*, Sheilah HOGG-JOHNSON, *Institute for Work and Health*

This study proposes a two-part mixed model and two-part growth mixture model for handling random and non-random missing data in a longitudinal

Cette étude propose un modèle aléatoire en deux parties et un modèle de croissance en deux parties afin de résoudre le problème de données manquantes aléatoires et non aléa-

study. We jointly model two processes with correlated random effects, one determining the missing mechanism and the other determining the actual trajectories of the outcomes if it is non-missing. These methods were applied to a longitudinal study of workers with work-relevant musculoskeletal disorders, to show how the new approaches can help to identify the distinct trajectories of worker productivity loss and the associated prognostic factors.

toires dans une étude longitudinale. Nous modélisons deux processus à effets aléatoires corrélés, un déterminant le mécanisme manquant, et l'autre les trajectoires actuelles des valeurs résultantes lorsque les données sont présentes. Ces méthodes sont appliquées aux données longitudinales chez les travailleurs ayant des problèmes musculo-squelettiques reliés spécifiquement au travail, afin de démontrer comment les nouvelles approches peuvent aider à identifier les trajectoires distinctes de perte de productivité et les facteurs pronostiques associés.

[MS-305]

Two-Step and Likelihood Methods for Semiparametric Nonlinear Mixed-effects Models with Measurement Errors and Missing Covariates

Méthodes par vraisemblance et à deux étapes pour modèles à effets mixtes non linéaires semi-paramétriques avec erreurs dans les mesures et covariables manquantes

Wei LIU, *York University*, Lang WU, *University of British Columbia*

Semiparametric nonlinear mixed-effects (NLME) models are very flexible in modeling complex longitudinal data. In the presence of measurement errors and missing data in time-dependent covariates, we show that the commonly used two-step method may give approximately unbiased estimates but may under-estimate standard errors, based on some theoretical arguments with both finite-sample and large-sample asymptotic results. We propose a likelihood method, implemented by a Monte-Carlo EM (MCEM) algorithm along with Gibbs sampler, and we compare the methods via simulation. A real dataset is analyzed in detail to illustrate the methods.

Les modèles non linéaires à effets mixtes (NLEM) semi-paramétriques sont très flexibles pour modéliser des données longitudinales complexes. En présence d'erreurs de mesure et de données manquantes parmi les covariables dépendant du temps, nous montrons que la méthode à deux étapes habituellement utilisée peut donner des estimateurs approximativement sans biais, mais sous-estime les erreurs standard à partir d'arguments théoriques avec résultats asymptotiques et sous échantillon fini. Nous proposons une méthode basée sur la vraisemblance, mise en oeuvre par un algorithme EM Monte-Carlo (EMMC) et l'échantillonneur de Gibbs, et nous comparons les méthodes par simulation. Un jeu de données réelles est analysé en détail pour illustrer les méthodes.

[MS-306]

A Multi-state Modeling Approach to Changes in Cognitive Scores Over a Fixed Time Interval

Une approche multi-états à la modélisation de changements dans les scores cognitifs sur un intervalle de temps fixe

Arnold MITNITSKI & Nader FALLAH, *Dalhousie University*, Charmaine DEAN, *Simon Fraser University*, Kenneth ROCKWOOD, *Dalhousie University*

We present a multistate model to describe longitudinal changes in cognitive scores over a fixed period of time. The scores are modeled according to a Poisson distribution, conditional on survival, with the Poisson mean dependent upon the baseline score and covariates. The model provides a unified treatment of the distribution of cognitive scores, taking into account baseline scores and survival, and the simultaneous estimation of covariate

Nous présentons un modèle multi-états décrivant les changements longitudinaux dans les scores cognitifs sur une période de temps fixe. Les scores sont modélisés suivant une loi de Poisson, conditionnellement à la survie, avec la moyenne de Poisson dépendant du score de base et de covariables. Le modèle apporte un traitement unifié de la distribution des scores cognitifs et prend en considération les scores de base et la survie et estime simultanément les effets des covariables. Le modèle permet d'estimer les probabilités de tran-

effects. The model allows estimates of the probabilities of transitions in different directions: improvements, declines, and death. The model application is illustrated using data from the Canadian Study of Health and Aging.

sitions vers différentes directions : amélioration, détérioration et mort. L'application du modèle est illustrée avec des données de l'Étude canadienne sur la santé et le vieillissement.

[MS-307]

Multistate Models for Hepatitis C Virus Testing, Seroconversion and Death

Modèles à plusieurs états pour le dépistage, la séroconversion et la mortalité associés au virus de l'hépatite C

Jesse RAFFA & Joel DUBIN, *University of Waterloo*

The hepatitis C virus (HCV) has become increasingly prevalent in high risk populations such as injection drug users (IDUs). Individuals at risk for HCV seroconversion are repeatedly tested over time. It has been hypothesized that in IDUs subjects with a high intensity of HCV testing and those with large gaps between tests are at increased risk for HCV seroconversion and death. Characterizing the process of HCV testing, seroconversion and mortality could improve prevention and surveillance strategies. Using a Vancouver-based cohort of IDUs, we examine these hypotheses through the development of a multi-state modeling framework that accounts for repeated testing.

Le virus de l'hépatite C (VHC) est de plus en plus prévalent parmi les populations à risque tels les utilisateurs de drogues injectées (UDI). Les individus à risque pour la séroconversion VHC sont testés de façon répétée. Une hypothèse est que les sujets UDI à intensité de tests du VHC élevée et ceux avec de grands écarts entre les tests sont à risque plus élevé de séroconversion VHC et de mort. Caractériser le processus de test du VHC, de séroconversion et de mortalité pourrait améliorer les stratégies de prévention et de surveillance. Nous examinons ces hypothèses en développant un cadre de modélisation multi-états tenant compte des tests répétés à partir d'une cohorte d'UDI de Vancouver.

[MS-308]

Generalized Imperfect Variables for Generalized Linear Models

Variables imparfaites généralisées pour modèles linéaires généralisés

Michael REGIER, *University of British Columbia*

Social epidemiology is interested in the link between the social context of people and their health. Embodiment suggests that social, physical, and psychological experiences are biologically incorporated. To reconstruct these experiences, disparate data sources are linked which presents a confluence of two problems: missing data and measurement error. These data deficiencies have received much attention in isolation, but there exists only a small body of research concerning both problems in tandem. Using a social epidemiological cancer data set as a point of inspiration, a general framework for unifying missing data and measurement error for generalized linear models will be proposed.

L'épidémiologie sociale s'intéresse au lien entre le contexte social des individus et leur santé. L'incarnation suggère que les expériences sociales, physiques et psychologiques sont biologiquement incorporées. Afin de reconstruire ces expériences, des sources de données disparates sont liées, ce qui présente une confluence de deux problèmes : des données manquantes et des erreurs de mesure. Ces deux problèmes, de façon isolée, ont été beaucoup étudiés, mais très peu de recherche a été menée au sujet des deux problèmes simultanément. À partir d'un jeu de données d'épidémiologie sociale du cancer, nous proposons un cadre général pour traiter de façon unie les données manquantes et les erreurs de mesure pour des modèles linéaires généralisés.

[MS-309]

Joint Mean-covariance Modeling With Nonlinear Random Mean Models for Longitudinal Data Analysis
Modélisation conjointe de la moyenne et de la covariance avec modèles à moyennes aléatoires non linéaires en analyse de données longitudinales

Peng ZHANG, *University of Alberta*, Peter X.-K. SONG, *University of Michigan*

We propose a generalization of mixed effects models by directly assuming the mean vector itself to be random in longitudinal data analysis. It is termed the random mean model, which includes the mixed effects model as a special case. The generalization allows us to relax the restriction that random effects have to be a subset of fixed effects and to model the mean and covariance matrix jointly. In this paper, we focus on a nonlinear mixed effects random mean model where two data examples were used to motivate the usefulness of the new model.

Nous proposons une généralisation des modèles à effets mixtes en supposant que le vecteur moyenne est lui-même aléatoire en analyse de données longitudinales. Nous l'appelons modèle à moyenne aléatoire, comprenant le modèle à effets mixtes comme cas particulier. Cette généralisation nous permet d'assouplir la restriction que les effets aléatoires doivent être un sous-ensemble des effets fixes et de modéliser la moyenne et la matrice de covariance conjointement. Dans cet article, nous nous concentrons sur un modèle non linéaire à effets mixtes à moyenne aléatoire où deux jeux de données ont été utilisés pour justifier l'utilité du nouveau modèle.

Session 12H Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00 MSL 102

Modeling, Classification, and Algorithms
Modélisation, classification et algorithmes

[MS-310]

Estimating Equations and Two-phase Sampling
Équations d'estimation et échantillonnage à deux phases

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

Most existing efficient methods for fitting regression models to data with structured missingness such as two-phase sampling designs can in theory handle second-phase data on an arbitrary number of explanatory variables but, at best, they can only handle all-individuals data on one or perhaps two continuous explanatory variables. For many potential applications this is the wrong way around. We discuss some estimating equation methods which are showing real promise for problems with all-individuals data on several continuous explanatory variables. The talk focuses on fitting linear models for a continuous outcome variable but will also be discussion of logistic regression.

La plupart des méthodes efficaces existantes pour ajuster des modèles de régression en présence de données structurellement manquantes, tels les plans d'échantillonnage à deux phases, peuvent en théorie traiter les données de seconde phase pour un nombre arbitraire de variables explicatives. En pratique, elles ne peuvent traiter des données présentes chez tous les individus que pour une ou deux variables explicatives continues. Pour plusieurs applications potentielles, ceci est problématique. Nous discutons de méthodes d'équations d'estimation très prometteuses pour des problèmes avec plusieurs variables explicatives continues présentes chez tous les individus. Cet exposé est centré sur l'ajustement de modèles linéaires pour une variable réponse continue, mais abordera aussi la régression logistique.

[MS-311]**Adjustment for Selection Bias in Constructing and Evaluating Prediction Models****Ajustement pour le biais de sélection en construction et évaluation de modèles de prévision**

Rod WALKER, *Group Health Center for Health Studies, Seattle, WA*, Sebastien HANEUSE, *Group Health Center for Health Studies; University of Washington*

We consider the general problem of constructing and evaluating risk prediction models when the data are subject to selection bias. Specific attention is paid to small-sample settings that preclude splitting the data into training and validation datasets. Towards this, we propose an algorithm based on the bootstrap that simultaneously ensures honest assessments of predictive performance and valid estimates of overall uncertainty. Inverse probability weighting, imbedded with the bootstrap, is used to account for selection bias. We address both threshold-specific performance measures and overall performance. The approach is illustrated with autopsy data from the Adult Changes in Thought study.

Nous considérons le problème général de la construction et de l'évaluation de modèles de prévision du risque lorsque les données sont sujettes à un biais de sélection. Une attention particulière est apportée au contexte de petits échantillons ne permettant pas de séparer les données en un jeu d'apprentissage et un jeu de validation. Nous proposons à cette fin un algorithme basé sur le bootstrap qui assure simultanément l'évaluation honnête de la performance prédictive et l'estimation valide de l'incertitude. La pondération par probabilités inverses, emboîtée dans le bootstrap, et utilisée pour tenir compte du biais de sélection. Nous abordons les mesures de performance spécifiques à certains seuils et la performance globale. L'approche est illustrée avec des données d'autopsie de l'Étude Adult Changes in Thought.

[MS-312]**Issues with Rank-Based Inverse Normal Transformations****Problèmes avec les transformations normales inverses basées sur les rangs**

Mark BEASLEY, *University of Alabama at Birmingham*

Applying parametric tests to rank-based inverse normal transformations (INTs), as a means to correct for violations of linear model assumptions, has gained popularity among applied researchers. We demonstrate that the use of INTs can drastically inflate the Type 1 error rate in both simple and more complex designs. In other circumstances, the use of INTs can show less power than nonparametric competitors. We contend that blindly applying parametric procedures to INTs has the same issues inherent with ranks, and therefore, there is a lack of justification for the growing misuse of INTs.

L'application de tests paramétriques à des transformations inverses normales (TIN) basées sur les rangs, afin de corriger un non-respect d'hypothèses de linéarité du modèle, a gagné en popularité en recherche appliquée. Nous démontrons que l'utilisation de TNI peut mener à une augmentation dramatique du taux d'erreur de type I sous des plans simples comme complexes. Dans d'autres circonstances, l'utilisation de TNI peut mener à une baisse de puissance par rapport à d'autres compétiteurs non paramétriques. Nous concluons que l'application aveugle de procédures paramétriques aux TNI pose les mêmes problèmes inhérents aux rangs, et qu'il n'y a donc pas de justification pour la mauvaise utilisation croissante des TNI.

[MS-313]**Prognostic and Predictive Biomarker Model Building for Time to Event Outcomes****Construction de pronostics et de modèles prédictifs basés sur biomarqueurs pour données de temps d'évènements**

Keyue DING, *Queen's University*, Keramat NOURIJELYANI & Philip TWUMASI-ANKRAH, *NCIC CTG, Queen's University*

Increasing knowledge in cancer biology leads to more accurate diagnosis, prognosis assessment and development of targeted therapy. To maximize patients benefit in apply those achievements, an important issue is to identify prognostic and predictive markers for clinical trial design and for patients' treatment selection. In this paper, we discuss issues in prognostic and predictive markers model building, and propose a practical approach for the purpose. And using our proposed approach, we show that P53 IHC protein expression level predicts survival benefit from adjuvant chemotherapy in NSCL Stage IB and Stage II patients.

Des connaissances plus approfondies sur les aspects biologiques du cancer mènent à des diagnostics, pronostics et développement de thérapies visées plus précises. Afin d'appliquer ces développements de façon à maximiser le bénéfice des patients, une question importante est l'identification de marqueurs prédictifs et pronostics pour la planification d'essais cliniques et pour la sélection des traitements des patients. Dans cet article, nous discutons de problèmes en construction de modèles pour pronostics et marqueurs prédictifs, et proposons une approche pratique à cette fin. En utilisant notre approche, nous montrons que les niveaux d'expression de protéines P53 IHC permettent de prédire l'amélioration de la survie apportée par la chimiothérapie adjuvante pour patient atteints de cancer du poumon à grandes cellules de phase IB et II.

[MS-314]**Kernel-Induced Random Forest Classification for functional Gene Analysis****Classification aléatoire de forêts induite par noyaux en analyse de gènes fonctionnelle**

Jiguo CAO, *Simon Fraser University*, Guangzhe FAN, *University of Waterloo*

We introduced a new type of classification algorithm for functional gene analysis. It is kernel-induced random forest model where functional data are used to construct kernels. Then the random forest procedure is applied to the kernels to classify genes. The approach is flexible and shows advantage over traditional discriminant methods in functional gene analysis with evidence in real and simulated data.

Nous présentons un nouveau type d'algorithme de classification pour analyse fonctionnelle génétique. Il s'agit d'un modèle de forêts aléatoire à base de noyaux où les données fonctionnelles sont utilisées pour construire les noyaux. La procédure des forêts aléatoires est alors appliquée aux noyaux pour classer les gènes. Cette approche est flexible et présente certains avantages sur les méthodes discriminantes traditionnelles en analyse de gène fonctionnelle, tel que démontré par l'analyse de données réelles et simulées.

[MS-315]**Customizing the rpart library for multivariate gaussian outcomes: the longRPart library****Customiser la librairie rpart pour réponses gaussiennes multivariées : la librairie longRPart**

Sam STEWART & Mohamed ABDOLELL, *Dalhousie University*, Michael LEBLANC, *Fred Hutchinson Cancer Research Center*

The work of Abdoell et al 2002 (SIM, 21(22) 3395-3409) on binary partitioning for continuous repeated measures outcomes is implemented and extended to the context of recursive partitioning utilizing the nlme and rpart libraries in R, pre-

Les travaux d'Abdoell et al. 2002 (SIM, 21(22) 3395-3409) sur le partitionnement binaire de mesures répétées continues est mis en oeuvre et étendu au contexte du partitionnement récursif au moyen des librairies R nlme et rpart, sous la forme de la librairie R longRPart. La librairie longRPart

sented as an R library `longRPart`. The `longRPart` library accommodates the inclusion of ordered and unordered categorical covariates as well as continuous covariates. This presentation focuses on the challenges of implementing new splitting criteria using the `Rpart` library as the core partitioning algorithm. A plotting function is developed that graphically annotates terminal nodes.

traite l'inclusion de covariables catégorielles ordonnées et non ordonnées et de covariables continues. Cet exposé est centré sur les défis de la mise en oeuvre de nouveaux critères de séparation en utilisant la librairie `Rpart` comme mécanisme principal de partitionnement. Une fonction annotant graphiquement les noeuds terminaux est développée.

Session 12I Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00 CHE 250

Statistical Inference Inférence statistique

[MS-316]

Random Set Modeling of Forest Fires

Modélisation par ensembles aléatoires de feux de forêts

Jeffrey PICKA, *University of New Brunswick*

The modeling of the evolution of a single forest fire requires a random set model which evolves over time. Suitable random set processes are likely to be cellular automata or to be based on dynamical systems. Once these models are fit to data, the fit must be assessed in order to ensure that it is not superficial. Comparison of these models with linear models suggests how the development of fire models may evolve, and what type of data will be required to properly fit them.

La modélisation de l'évolution d'un unique feu de forêt nécessite un modèle d'ensemble aléatoire évoluant avec le temps. Des processus d'ensembles aléatoires adéquats seront habituellement des automates cellulaires ou basés sur des systèmes dynamiques. Une fois ces modèles ajustés aux données, l'ajustement doit être évalué pour s'assurer qu'il n'est pas superficiel. La comparaison de ces modèles aux modes linéaires suggère comment le développement de modèles pour feux de forêts pourrait évoluer, et quel type de données sera nécessaire à leur ajustement.

[MS-317]

A Surprising Comparison of Tests of Fit

Une comparaison surprenante de tests d'adéquation

Michael STEPHENS & Richard LOCKHART, *Simon Fraser University*, Federico O'REILLY, *UNAM*

The Gibbs sampler is used to generate co-sufficient samples (samples from the conditional distribution of a sample, given a minimal sufficient statistic). Following the suggestion of Lehmann, as the correct way to test fit, these samples are used to estimate the distributions of Cramer-von Mises statistics, and the resulting tests are exact. However, comparison of the p-values of many samples with the approximate values given by the parametric bootstrap reveals a surprisingly high correlation. Is the age of elegance therefore dead, at least for testing fit?

L'échantillonneur de Gibbs est utilisé pour générer des échantillons co-exhaustifs (des échantillons de la distribution conditionnelle d'un échantillon, étant donnée une statistique exhaustive minimale). D'après la suggestion de Lehmann au sujet de la façon correcte de tester l'adéquation, ces échantillons sont utilisés pour estimer les distributions de statistiques de Cramer-von Mises, et les tests résultants sont exacts. Cependant, la comparaison des seuils observés de plusieurs échantillons aux valeurs approximatives données par bootstrap paramétrique révèle une corrélation étonnamment élevée. L'âge de l'élégance tire-t-il à sa fin, du moins en ce qui concerne les tests d'adéquation ?

[MS-318]**Testing Marginal Homogeneity Against Uniform Stochastic Ordering in a Contingency Table****Tester l'homogénéité marginale contre un classement stochastique uniforme dans une table de contingence**Chul PARK, *Carleton University*, Pranab SEN, *University of North Carolina*

Observations or measurements are often jointly taken from experimental units and displayed in a contingency table. When the two variables have the same ordinal categorical levels, it is interesting to test whether their marginal distributions are ordered in some way. In this talk, we discuss a testing problem for the equality of the marginal distributions against the hypothesis that they are uniformly stochastically ordered. Based on constrained estimators, a likelihood ratio statistic is suggested and its asymptotic null distribution is derived. This testing method is illustrated with a data set used in other research.

Les observations ou mesures sont souvent conjointement tirées d'unités expérimentales et présentées dans un tableau de contingence. Lorsque deux variables ont les mêmes niveaux ordinaux catégoriels, il est intéressant de tester si leurs distributions marginales sont ordonnées d'une certaine façon. Dans cet exposé, nous discutons du problème de tester l'égalité des distributions marginales avec contre-hypothèse qu'elles sont uniformément ordonnées stochastiquement. Nous proposons une statistique du rapport des vraisemblances basée sur des estimateurs sous contraintes, et obtenons sa distribution nulle asymptotique. Cette méthode de test est illustrée avec un jeu de données utilisé dans une autre recherche.

[MS-319]**On Average Predictive Comparisons and Interactions****Comparaison et interactions prédictives moyennes**Juxin LIU, *University of Saskatchewan*, Paul GUSTAFSON, *University of British Columbia*

In a regression context, consider the difference in expected outcome associated with a particular difference in one of the input variables. If the true regression relationship involves interactions, then this predictive comparison can depend on the values of the other input variables. Therefore, one may wish to consider an average predictive comparison as a target of inference, where the averaging is with respect to the population distribution of the input variables. We consider inferences about such targets, with emphasis on inferential performance when the regression model is misspecified.

Dans le contexte de la régression, considérons la différence en réponse espérée associée à une différence particulière dans une des variables prédictives. Si la relation réelle implique des interactions, alors cette comparaison prédictive peut dépendre des valeurs des autres variables prédictives. Nous pourrions alors considérer les comparaisons prédictives moyennes comme cible d'inférence, où l'espérance est prise par rapport à la distribution des variables prédictives parmi la population. Nous considérons de telles inférences, en mettant l'accent sur la performance inférentielle lorsque le modèle de régression est mal spécifié.

[MS-320]**Statistical Inference on Recall, Precision and Average Precision****Inférence statistique sur le rappel, la précision et la précision moyenne**Wanhua SU & Peng ZHANG, *University of Alberta*

The objective of a rare target detection problem is to identify the rare targets as early as possible. Recall, precision and average precision are three popular performance measures for evaluating different detection methods. However, there is little literature on the statistical properties of these three measures. We develop a framework for conducting

L'objectif du problème de détection de cible rare est d'identifier les cibles rares le plus tôt possible. Le rappel, la précision et la précision moyenne sont trois mesures de performance populaires pour évaluer différentes méthodes de détection. Cependant, il y a peu de littérature au sujet des propriétés statistiques de ces trois mesures. Nous développons un cadre pour faire de l'inférence statistique au sujet du rappel, de la

statistical inference on recall, precision and average precision through establishing their asymptotic properties. Simulations and real data analysis are used to illustrate the ideas. The proposed methods can also be applied in other areas where ranking systems need to be evaluated, such as information retrieval.

précision et de la précision moyenne en établissant leurs propriétés asymptotiques. Des analyses de données réelles et de simulation sont utilisées pour illustrer ces idées. Les méthodes proposées peuvent également être utilisées dans d'autres domaines où des systèmes de classement en rangs doivent être évalués, telle la récupération d'information.

[MS-321]

On the Moments and the Moment-generating Function of the Maximum of a Negatively Drifting Random Walk

Moments et fonction génératrice des moments du maximum d'une marche aléatoire à dérive négative

Kai YU, *National Institutes of Health*

Let S_n be a negatively drifting random walk starting from the origin, that is, $S_n = X_1 + \dots + X_n$ and $S_0 = 0$, where X_n are independent and identically distributed random variables with negative expectation. Let $M = \sup_{n \geq 0} S_n$. The moments and the moment-generating function of M are studied in relation to the distribution of X_n^+ . In particular a necessary and sufficient condition on X_n^+ is obtained for the finiteness of the moment-generating function of M for some positive argument. An application of this result will be given.

Soit S_n , une marche aléatoire à dérive négative débutant à l'origine, c'est-à-dire que $S_n = X_1 + \dots + X_n$ et $S_0 = 0$, où les X_i sont des variables aléatoires indépendantes identiquement distribuées d'espérance négative. Soit $M = \sup_{n \geq 0} S_n$. Les moments et la fonction génératrice des moments de M sont étudiés en relation avec la distribution de X_n^+ . En particulier, nous obtenons une condition nécessaire et suffisante sur X_n^+ pour que la fonction génératrice des moments de M soit finie pour un certain argument positif. Une application de ce résultat sera donnée.

Session 12J Mercredi 3 juin • Wednesday, June 3, 10:30 – 12:00 SCR 100

Mixture Models

Modèles de mélanges

[MS-322]

Robust Mixture Regression Models

Mélanges de modèles de régression robustes

Shaheena BASHIR, *The Aga Khan University Karachi, Pakistan*, Shagufta SULTAN, *University of Ottawa*, E CARTER, *University of Guelph*

Mixture of regression models are used for modeling when the data are a mixture of subgroups to allow for heterogeneity of the population. In the presence of outliers, standard mixture of linear regression models fail. To study mixtures of linear regression models, we introduce a class of robust estimators, called S-estimators. We investigate their breakdown point in mixture of linear regression models. It is expected that the robust S-estimators can achieve the high breakdown point in the contaminated data in this setting.

Les mélanges de modèles de régression sont utilisés pour modéliser des données qui sont elles-mêmes des mélanges de sous-groupes afin de traiter l'hétérogénéité de la population. En présence d'observations aberrantes, les mélanges de modèles de régression habituels ne fonctionnent plus. Pour étudier les mélanges de modèles de régression, nous présentons une classe d'estimateurs robustes, les S-estimateurs. Nous étudions leur point de rupture sous des mélanges de modèles de régression linéaire. Nous nous attendons à ce que les S-estimateurs robustes aient un point de rupture élevé dans ce contexte.

[MS-323]**Mixtures of Normal, Gamma, and Gumbel Distributions****Mélanges de lois normales, gamma et Gumbel**

Guillaume ÉVIN, *Université du Québec - Institut national de la recherche scientifique - Eau, Terre et Environnement (INRS-ETE)*, James MERLEAU & Luc PERREAU, *Institut de recherche d'Hydro-Québec (IREQ)*

We study mixtures of two distributions with normal, gamma and Gumbel components. Moving away from the standard normal setting, gamma mixtures are developed in order to model strictly positive environmental data and Gumbel mixtures for extreme variates. Since the data analyzed can exhibit dependency through time, we treat both the independent and dependent cases, where the latter is modelled through a Markov process. A unified approach is adopted for the different distributions and the problem is treated from the Bayesian perspective, which enables us to use Bayes factors to compare the adequacy of different models for a given data set.

Nous étudions les mélanges de deux distributions avec des composantes normales, gamma et Gumbel. Sortant du cadre gaussien, les mélanges gamma sont développés pour modéliser des données strictement positives et les mélanges Gumbel pour les variables extrêmes. Les données environnementales analysées pouvant afficher une dépendance temporelle, nous traitons les cas indépendant et dépendant, où ce dernier est modélisé via un processus markovien. Une méthode unifiée est adoptée pour les différentes distributions et le problème est traité sous le paradigme bayésien, permettant d'utiliser les facteurs de Bayes afin de comparer l'adéquation des différents modèles pour un jeu de données particulier.

[MS-324]**Extending Mixtures of t-Factors for Model-Based Clustering****Étendue des mélanges de t-facteurs pour la classification à base de modèle**

Jeffrey ANDREWS & Paul McNICHOLAS, *University of Guelph*

A family of mixture models based on the mixtures of multivariate t-factor analyzers will be introduced, two of which are novel. These models present a unified, robust, clustering framework and parameter estimation is achieved via an alternating expectation-conditional maximization (AECM) algorithm. Details of the AECM algorithms are presented along with brief discussion of model selection issues. This new family of mixture models is then applied to real data where good clustering performance is observed.

Une famille de modèles de mélanges basée sur les mélanges d'analyseurs t-facteurs multivariés sera présentée. Ces modèles représentent un cadre robuste et unifié de classification, et l'estimation paramétrique est effectuée au moyen d'un algorithme alternant espérance-maximisation conditionnelle (AEMC). Les algorithmes AEMC ainsi que quelques questions en sélection de modèles seront abordés. Cette nouvelle famille de modèles de mélanges sera ensuite appliquée à des données réelles, où nous observerons une bonne performance de classification.

[MS-325]**Feature Selection in Mixture-of-Experts Models****Sélection de caractéristiques en modèles de mélanges d'experts**

Abbas KHALILI, *University of British Columbia*

Feature selection is unavoidable in many regression problems. We study the feature selection problem in mixture-of-experts models. These models are extension of finite mixture models where both the mixture components (experts) and the mixing probabilities (gating network) are functions of poten-

La sélection de caractéristiques est inévitable dans plusieurs problèmes de régression. Nous l'étudions dans le contexte de modèles de mélanges d'experts. Ces modèles sont une extension des modèles de mélanges finis où les composantes (les experts) et les probabilités (le réseau de synchronisation) du mélange sont des fonctions de caractéristiques po-

tial features. In applications, often a large number of features are considered, and their contributions toward the response variable vary between the experts and gating network. This results in a complex feature selection problem. We introduce a new penalized likelihood approach to this problem. Statistical properties of the new method are studied theoretically and via simulations.

tentielles. Dans les applications, un grand nombre de caractéristiques sont considérées, et leur contribution à la variable réponse varie entre les experts et le réseau de synchronisation. Ceci résulte en un problème de sélection de caractéristiques complexe. Nous présentons une nouvelle approche par vraisemblance pénalisée pour ce problème. Les propriétés statistiques de la nouvelle méthode sont étudiées théoriquement et par simulation.

[MS-326]

Analysis of Bivariate FDI Panel Count Data: a Markov Bivariate Zero-inflated Negative Binomial Regression Approach

Analyse de données de dénombrement bivariées d'IÉD : une approche bivariée markovienne par régression négative binomiale à surreprésentation de zéros

Peiming WANG, *Auckland University of Technology*, Joseph ALBA, *Nanyang Technological University*

This paper proposes a Markov bivariate zero-inflated negative binomial regression approach for the analysis of bivariate foreign direct investment (FDI) panel count data with the presence of a high proportion of zero counts and a positive correlation between two response variables of counts, and provides an estimation method based on the EM and quasi-Newton algorithms. An application to the joint analysis of mergers and acquisitions (M&A) and non-M&A FDI projects with panel firm-level data is given, which provides evidence supporting the hypothesis of relative access to bank credit for both M&A and non-M&A FDI.

Cet article propose un modèle bivarié de régression négative binomiale de Markov à surreprésentation de zéros pour l'analyse de données de dénombrement bivariées d'investissement direct étranger (IDÉ) en présence d'une proportion élevée de zéros et d'une corrélation positive entre deux variables réponses de dénombrements, et fournit une méthode d'estimation basée sur les algorithmes EM et quasi-Newton. Une application à l'analyse conjointe des projets d'IDÉ de type fusions et acquisitions (F&A) et non-F&A avec données de dénombrement de firmes est présentée, soutenant l'hypothèse d'un accès relatif au crédit bancaire pour IDÉ F&A et non-F&A.

[MS-327]

Modelling Longitudinal Count Data With Excess Zeros Using a Compound Poisson Random Effect

Modélisation de données de dénombrement longitudinales avec surreprésentation de zéros au moyen d'un effet aléatoire de Poisson composé

Gary SNEDDON, *Memorial University*, M. Tariqul HASAN & Renjun MA, *University of New Brunswick*

In clustered medical count data, excess zeros often occur, so a traditional Poisson mixed model may not be suitable for the data. We consider a two level random effects zero-inflated Poisson model, where the subject level random effects are modelled by a compound Poisson distribution. A quasi-likelihood approach to estimation is discussed, and we illustrate the method with an analysis of skin cancer data.

Dans des données médicales de dénombrement en grappes, on retrouve souvent des zéros excessifs, rendant inappropriée l'utilisation d'un modèle de Poisson à effets mixtes traditionnel. Nous considérons un modèle de Poisson à effets aléatoires à deux niveaux avec surplus de zéros, où les effets aléatoires au niveau du sujet sont modélisés par une distribution de Poisson composée. Une approche d'estimation par quasi-vraisemblance est abordée, et nous illustrons la méthode avec une analyse de données du cancer de la peau.

Session 13 Mercredi 3 juin • Wednesday, June 3, 13:30 – 14:30 IRC 6

Address of the winner of the Centre de Recherches Mathématiques/Statistical Society of Canada Prize

Allocution du récipiendaire du prix Centre de Recherches Mathématiques/Société Statistique du Canada

[MS-328]

Statistical Learning With Trees

Apprentissage statistique avec arbres

Hugh CHIPMAN, *Acadia University*

Tree models, also known as recursive partitioning models, provide an intuitive way to fit predictive models. By combining together many trees, via random forests or boosting, predictive accuracy can be improved. As originally introduced, these models are more algorithms than full statistical models. This talk will explore efforts to answer statistical questions about trees and ensembles of trees, and demonstrate some cases in which formulation as a statistical model gives some real benefits.

Les modèles d'arbres, aussi connus comme modèles de partitionnement récursif, nous apportent une façon intuitive d'ajuster des modèles prédictifs. En combinant plusieurs arbres au moyen de forêts aléatoires ou de boosting, la précision prédictive peut être améliorée. Tels qu'originellement présentés, ces modèles sont plutôt des algorithmes que des modèles statistiques complets. Cet exposé va explorer les efforts visant à répondre aux questions statistiques au sujet des arbres et ensembles d'arbres, et présenter certains cas où la formulation en tant que modèle statistique apportera des avantages réels.

Session 14A Mercredi 3 juin • Wednesday, June 3, 13:30 – 15:00 IRC 1

Statistics in Sports

La statistique dans le monde des sports

[MS-329]

Small Sample Statistics in Baseball - the Batter-pitcher Matchup

Statistique en présence de petits échantillons au baseball - la rencontre batteur-lanceur

Hal STERN, *University of California, Irvine*

One aspect of the sport of baseball that makes it a popular topic for statistical analysis is that the game is comprised of a series of individual batter versus pitcher matchups. These matchups are at the heart of the large data sets that are analyzed by baseball enthusiasts. Data available for each matchup include measures of both batter and pitcher ability. This talk considers ways of combining this information to provide accurate assessments of the likely outcome of each matchup. The data analyzed provide some insight on the reliability of the small sample inferences that are common in baseball.

Un aspect du baseball le rendant un sujet populaire d'analyse statistique est que chaque partie est composée d'une série de rencontres individuelles entre frappeur et lanceur. Ces rencontres sont au coeur des grands jeux de données qui sont analysés par les fans de baseball. Les données disponibles pour chaque match comprennent des mesures de l'habileté des frappeurs et des lanceurs. Cet exposé considère des façons de combiner ces informations pour fournir une estimation précise du résultat vraisemblable de chaque rencontre. Les données analysées apportent des précisions sur la fiabilité des inférences en présence de petits échantillons, qui sont fréquentes au baseball.

[MS-330]

A Graduate Course in Statistic in Sport

Un cours de cycle supérieur de statistique dans le sport

Tim SWARTZ, *Simon Fraser University*

This talk describes aspects of a graduate course in Statistics in Sport that was first taught at Simon Fraser University in 2004 and taught again in 2008. It is argued that the course allows graduate students to gain valuable expertise in (a) statistical modelling, (b) the reading of scientific papers and (c) statistical methodology. Some student projects have resulted in published papers and we report on some of this work.

Cet exposé décrit certains aspects d'un cours de cycle supérieur de Statistique dans le sport qui a été donné pour la première fois en 2004 à l'Université Simon Fraser et qui fut donné à nouveau en 2008. Le cours permet aux étudiants de cycles supérieurs d'acquérir de l'expérience en (a) modélisation statistique, (b) lecture d'articles scientifiques et (c) méthodologie statistique. Certains projets des étudiants ont mené à des articles publiés.

[MS-331]

NHL Scoring and Fighting: Before and After 2005

Compter et cogner dans la LNH : avant et après 2005

Paramjit GILL, *UBC Okanagan*

The NHL in 2005 approved a set of rule changes to reduce hooking, holding, tripping, slashing, cross checking and interference. This has resulted in more penalties being awarded and thus leading to more power play goals. We present a Bayesian model for analyzing the even-strength and power play goals in a game. Did the rule changes result in less fighting on the ice? Yes, but only for a while. Fighting in hockey has been and will always be part of the game. What can a statistician do about it? Fit a statistical model to answer some interesting questions about fights.

La LNH a approuvé en 2005 un ensemble de changements aux règles du jeu réduisant les accrochages, les retenues, le trébuchage, le cinglage, le double-échec et l'obstruction. Ceci a eu pour conséquence une augmentation du nombre de pénalités et donc de buts marqués en avantage numérique. Nous présentons un modèle bayésien pour analyser les buts marqués en et hors avantage numérique. Les changements aux règlements ont-ils réduit les combats sur la glace ? Oui, mais seulement pendant un certain temps. Les combats au hockey ont toujours fait partie et feront toujours partie du jeu. Que peut y faire un statisticien ? Ajuster un modèle statistique pour répondre à certaines questions d'intérêt à propos des combats.

[MS-332]

Discussant

Modérateur

Shane REESE, *Brigham Young University*

Session 14B Mercredi 3 juin • Wednesday, June 3, 13:30 – 15:00 IRC 5

Statistical Inference in Studies with Large p and Small n Inférence statistique dans les études à grand p et petit n

[MS-333]

Optimal Shrinkage Estimation of Variances With Applications to Microarray Data Analysis

Estimation optimale par rétrécissement de variances avec applications à l'analyse de données de micropuces

Yuedong WANG, *University of California - Santa Barbara*, Tiejun TONG, *University of Colorado*

Microarray technology allows a scientist to study genome-wide patterns of gene expression. Thousands of individual genes are measured with relatively little replication. The gene-specific estimates of variances are not reliable and gene-by-gene tests have low power. We propose a family of shrinkage estimators for variances raised to a fixed power. We derive optimal shrinkage parameters under both Stein and the squared loss functions. Our results show that the standard sample variance is inadmissible under both loss functions. We construct F statistics using these shrinkage variance estimators and apply them to detect differentially expressed genes in a microarray experiment.

Les micropuces permettent aux scientifiques d'étudier les motifs d'expression génétique sur l'ensemble du génome. Des milliers de gènes individuels sont mesurés avec relativement peu de réplication. Les estimateurs de la variance spécifique aux différents gènes ne sont pas fiables et les tests gène-par-gène sont de faible puissance. Nous proposons une famille d'estimateurs à rétrécissement pour les variances élevées à une puissance fixe. Nous obtenons les paramètres de rétrécissement optimaux sous les fonctions de perte quadratique et de Stein. Nos résultats montrent que la variance échantillonnale standard est inadmissible sous ces deux fonctions de perte. Nous construisons des statistiques F au moyen de ces estimateurs à rétrécissement et les appliquons à la détection de gènes d'expression différentielle dans une expérience à micropuce.

[MS-334]

Improving Differential Expression Analysis With the Consideration of Genome-wide Co-expression Information

Améliorer l'analyse de l'expression différentielle en considérant l'information de co-expression sur tout le génome

Yinglei LAI, *The George Washington University*

The control of false positives in differential expression analysis remains a major challenge although many statistical methods have been proposed for its improvement. Since genes interact with each other during cellular and molecular processes, an efficient incorporation of genome-wide co-expression information may significantly improve the detection of differential expression. We will address our recent research progress in this direction.

Le contrôle de faux positifs en analyse d'expression différentielle reste un défi majeur bien que plusieurs méthodes statistiques aient été proposées pour son amélioration. Puisque les gènes interagissent ensemble durant les processus cellulaires et moléculaires, une incorporation efficace de l'information de coexpression pour tout le génome peut significativement améliorer la détection de l'expression différentielle. Nous allons discuter de nos récents progrès dans cette direction.

[MS-335]**Shrinkage Estimation of Means for High-dimensional Data****Estimation par rétrécissement de moyennes pour données de haute dimension**

Tiejun TONG, *University of Colorado at Boulder*, Liang CHEN, *University of Southern California*, Hongyu ZHAO, *Yale University*

High-dimensional data have created new challenges to traditional statistical methods. In particular, the feature-specific estimates of means are unreliable due to the small sample size. In this paper, we propose a family of shrinkage estimators for means under the assumption of unequal and unknown variances. We show that the proposed estimators are minimax under the squared loss function. Our method is general and widely applicable, whereas we illustrate it in the framework of discriminant analysis. Specifically, we propose a shrinkage-based diagonal discriminant rule and demonstrate its improvement over the original rule through both simulation and real data analysis.

Les données de dimension élevée apportent de nouveaux défis aux méthodes statistiques traditionnelles. En particulier, les estimateurs de moyenne spécifiques à certaines caractéristiques sont peu fiables due à la faible taille échantillonnale. Dans cet article, nous proposons une famille d'estimateurs à rétrécissement pour les moyennes sous l'hypothèse de variances inégales et inconnues. Nous montrons que les estimateurs proposés sont minimax sous la perte quadratique. Notre méthode est générale et largement applicable, ce que nous illustrons dans le cadre de l'analyse discriminante. Spécifiquement, nous proposons une règle discriminante diagonale basée sur le rétrécissement et démontrons qu'elle est une amélioration de la règle originale au moyen d'analyses de données réelles et simulées.

[MS-336]**Clustering Strategies for Large p , Small n Problems****Stratégies de classification pour problèmes à grand p et petit n**

Paul McNICHOLAS, *University of Guelph*

The clustering of data exhibiting the large p , small n structure is discussed. Special attention is given to a family of mixture models that is developed using a latent Gaussian model which is closely related to the factor analysis model. Model selection, convergence criteria and other implementation issues are discussed, including details of a coarse-grain parallel implementation. Possible extensions of these models are mentioned before they are applied to real data where they give good clustering performance.

Nous discutons de la classification de données pour grand p et petit n . Une attention particulière est apportée à une famille de modèles de mélanges développée à partir d'un modèle gaussien latent et liée au modèle d'analyse factorielle. La sélection de modèles, les critères de convergences et d'autres questions de mise en oeuvre sont abordés, incluant les détails d'une mise en oeuvre parallèle à gros-grain. Des étendues possibles pour ces modèles sont mentionnées avant leur application à des données réelles démontrant leur bonne performance de classification.

Session 14C Mercredi 3 juin • Wednesday, June 3, 13:30 – 15:00 MSL 102

Recent Developments in Order-Restricted Inference

Évolution récente de l'inférence d'ordre restreint

[MS-337]

Recent Advances in Constrained Statistical Inference in Parametric Models

Développements récents en inférence statistique sous contraintes pour modèles paramétriques

Mervyn SILVAPULLE, *Monash University, Australia*

I will discuss a selection of advances in constrained statistical inference in parametric models, with particular emphasis on the recent developments. Further, I will also discuss an example where there are inequality constraints in the null hypothesis and it appears impossible to identify the least favourable null configuration. Although this example originated in the study of income inequality, the statistical issues are of general interest.

Je vais discuter de certaines avancées dans le domaine de l'inférence sous contraintes en statistique paramétrique, avec une emphase particulière sur les développements récents. De plus, je vais discuter d'un exemple où il y a des contraintes d'inégalité sous l'hypothèse nulle et où il semble impossible d'identifier la configuration la moins favorable sous l'hypothèse nulle. Bien que cet exemple provienne de l'étude de l'inégalité des revenus, les questions statistiques soulevées sont d'intérêt général.

[MS-338]

Approximately Unbiased Tests for Cone Shaped Regions via Multiscale Bootstrap

Tests approximativement sans biais pour régions en forme de cône au moyen de bootstrap multi-échelles

Hidetoshi SHIMODAIRA, *Tokyo Institute of Technology*

The null hypotheses appearing in multiple comparisons (MC) problems are typically represented as cone-shaped regions in the parameter space. The type-I error of standard MC methods matches the nominal significance level at the vertex of cone, but it is often very conservative as it moves away from the vertex. We discuss a resampling-based method for computing adjusted p-values so that the type-I error is controlled approximately at any points on the boundary of the cone. The new p-value is calculated by extrapolating the bootstrap probability to $n' = -n$; n is the sample size and n' is that for bootstrap replicates.

Les hypothèses nulles apparaissant dans des problèmes de comparaisons multiples (CM) sont typiquement représentées par des régions en forme de cônes dans l'espace paramétrique. L'erreur de type I des méthodes de CM usuelles atteint le niveau de signification nominal au sommet du cône, mais devient souvent très conservatrice en s'éloignant du sommet. Nous discutons d'une méthode basée sur le ré-échantillonnage pour calculer des seuils observés ajustés afin que l'erreur de type I soit contrôlée approximativement à n'importe quel point de la frontière du cône. Le nouveau seuil observé est calculé en extrapolant la probabilité bootstrap que $n' = -n$, où n est la taille échantillonnale et n' est la taille des échantillons bootstrap.

[MS-339]

Stepwise Confidence Intervals for Partial Monotone Dose-Response Studies

Intervallens de confiance pas-à-pas pour études dose-réponse partiellement monotones

Jianan PENG, *Acadia University*, Lang WU, *University of British Columbia*, Weizhen WANG, *Wright State University*

Usually the dose-response curves are monotonic. However, in some situations there may be a negative response at low doses, then it is appropriate to make a partial monotone dose-response assumption. Based on the partitioning principle, we combine the new proposed one-sided multiple contrast test and t test to develop a stepwise confidence interval procedure to identify the minimum effective dose. The new stepwise confidence interval procedure is more powerful than Hsu and Berger's DR method, which is the special case of the new procedure.

Les courbes dose-réponse sont habituellement monotones. Cependant, dans certaines situations, la réponse à de faibles doses peut être négative. Il est alors approprié de faire l'hypothèse d'une dose-réponse partiellement monotone. En nous basant sur le principe du partitionnement, nous combinons le nouveau test unilatéral de contrastes multiples et le test t afin de développer une procédure d'intervalles de confiance pas-à-pas pour l'identification de la dose efficace minimale. La nouvelle procédure d'intervalles de confiance pas-à-pas est plus puissante que la méthode DR et Hsu et Berger, qui est un cas particulier de la nouvelle procédure.

Session 14D

Mercredi 3 juin • Wednesday, June 3, 13:30 – 15:00

IRC 3

Matrices and Statistics

Matrices et statistique

[MS-340]

Equality of the BLUPs for New Observations Under two Linear Models

Égalité des MPLSB pour de nouvelles observations sous deux modèles linéaires

SIMO PUNTANEN, *University of Tampere, Finland*, Stephen J. HASLETT, *Massey University, New Zealand*

We consider two linear models, A and B , say, with new unobserved future observations. We give necessary and sufficient conditions in the general situation, without any rank assumptions, that the best linear unbiased predictor (BLUP) of the new observation under the model A continues to be BLUP also under the model B . Results concerning the invariance of the BLUPs of the random effect under two mixed models follow as a special case.

Nous considérons deux modèles linéaires A et B et des observations futures non observées. Nous donnons des conditions nécessaires et suffisantes dans le cas général, sans hypothèses sur les rangs pour que le meilleur prédicteur linéaire non biaisé (BLUP) d'une nouvelle observation sous le modèle A soit aussi BLUP sous le modèle B . Des résultats concernant l'invariance des BLUP d'effets aléatoires sous deux modèles mixtes sont ensuite obtenus comme cas particuliers.

[MS-341]

Classical Rank Formulas From Projectors' Point of View

Formules classiques pour le rang du point de vue des projecteurs

Goetz TRENKLER, *Dortmund University of Technology, Germany*, Oskar M. BAKSALARY, *Adam Mickiewicz University, Poland*

For conformable matrices F and G and the corresponding orthogonal projectors $P_F = FF^\dagger$ and $P_G = GG^\dagger$, we use the identity $\text{rank}(F^*G)\text{rank}(P_F P_G)$ to derive a number of new rank formulas involving $\text{rank}(F : G)$, $\text{rank}(P_F + P_G)$ and $\text{rank}(P_F - P_G)$.

Pour des matrices conformables F et G et les projecteurs correspondants $P_F = FF^\dagger$ et $P_G = GG^\dagger$, nous utilisons l'identité $\text{rang}(F^*G)\text{rang}(P_F P_G)$ pour obtenir de nouvelles formules pour les rangs impliquant $\text{rang}(F : G)$, $\text{rang}(P_F + P_G)$ et $\text{rang}(P_F - P_G)$. Une attention particulière est apportée à la

Further attention is paid to the relationship between ranks and ranges. Some new bounds for ranks of matrices are presented which depend on their traces.

relation entre rangs et étendues. De nouvelles bornes pour les rangs de matrices dépendant de leur trace sont présentées.

[MS-342]

Core Inverse of Matrices

Inverse nucléaire de matrices

Oskar M. BAKSALARY, *Adam Mickiewicz University, Poland*, Goetz TRENKLER, *Dortmund University of Technology, Germany*

The notion of the Core inverse, as an alternative to the group inverse, is introduced and investigated. Several of its properties are derived with a perspective towards possible applications. The particular attention is paid to a matrix partial ordering based on the Core inverse.

La notion d'inverse nucléaire comme alternative à l'inverse groupe est présentée et étudiée. Plusieurs de ses propriétés sont obtenues et ses applications possibles sont mises en perspective. Une attention particulière est apportée au classement partiel de matrices basé sur leur inverse nucléaire.

The translation team did its best. Please forgive us for our misinterpretations, we are neither professional translators nor specialists in all areas of statistics.

L'équipe de traduction a fait de son mieux. Excusez-nous pour les erreurs de traduction, nous ne sommes ni des traducteurs professionnels ni des spécialistes dans tous les domaines de la statistique.

Author Index (per session)

Index des auteurs (par session)

AALEN, Odd	08D	BEAUMONT, Jean-François	02A
ABANIDA, Emmanuel	06D	BEEGLE, Kathleen	09K
ABARIN, Taraneh	02F	BELLAVANCE, François	03H
ABDOLELL, Mohamed	12H	BELLHOUSE, David	02C
ABRAHAMOWICZ, Michal	02F	BEN GHORBAL, Noomen El Beji	09I
ADEM, Alejandro	0	BERG, Emily	09D
AHMED, Ejaz	06K	BERGERON, Pierre-Jérôme	02I
AINSWORTH, Laurie	08E	BERKE, Olaf	09G
ALAGHEHBANDAN, Reza	11E	BHAUMIK, Dulal	08D
ALAVI, Asma	09J	BICKEL, Peter	04H
ALBA, Joseph	12J	BICKIS, Mikelis	11H
ALIOUM, Ahmadou	02F	BIGGIN, Mark	04H
ANDERSON-COOK, Christine	06B	BIGGS, Michael	09J
ANDREWS, Jeffrey	12J	BINGHAM, Derek	04C, 09F
ANDREWS, Jessica	08F	BLEI, David	06F
ANDRIEU, Christophe	04G	BOHOSSIAN, Nora	11G
ANGHEL, Marian	06K	BOLEY, Nathan	04H
ANGST, Jules	12E	BONNER, Simon	06J
ARMSTRONG, Paul	02H	BORNN, Luke	06K
ASOKAN, Mulayath Variyath	06H	BORUVKA, Audrey	04F
ASTON, John	04D	BOUEZMARNI, Taoufik	09I
ATHERTON, Juli	04H	BOVAS, Abraham	06H
ATOYAN, Tigran	12F	BRANSCUM, Adam	04B
AYLWARD, Bruce	06D	BRANT, Rollin	03A
BABYAK, Colin	09J	BRAUN, W. John	06G, 08I
BADESCU, Alex	11A	BREED, Greg	08E
BAI, Yan	04G	BREHENY, Patrick	02E
BAILEY, Mark	09B	BRILLINGER, David	11I
BAKAL, Jeffrey	02H	BRINKMAN, Ryan	06K
BAKSALARY, Oskar M.	14D	BRIOLLAIS, Laurent	04H, 08G
BALAKRISHNAN, N.	04D, 09I	BROWN, Ben	04H
BALASOORIYA, Uditha	06H	BROWN, Patrick	02F, 02J, 09G
BALKEMA, Guus	02G	BUCKLEY, Nicholas	09H
BALSHAW, Robert	02H	BULL, Shelley B	08G
BASHIR, Shaheena	12J	BURKETT, Kelly	02J
BATAINEH, Osama	02J	BURSTYN, Igor	11B
BEASLEY, Mark	12H	CAI, Song	08I
BEATON, Dorcas	12G	CALAMAI, Peter	03D

CAMPBELL, David 02D
 CANNeCTIN 04F
 CAO, Jiguo 02D, 06G, 08A, 08E, 12H
 CAREY, John 03D
 CARRIERE, Keumhee 08I
 CARRILLO-GARCIA, Ivan 06J
 CARTER, E 12J
 CASCAGNETTE, Paul 08B
 CASKENETTE, Amanda 06H
 CAVASSIN, Sirlei 11I
 CHALLENGER, Wendell 06I
 CHAMBERS, Ray 06C
 CHAMPAGNE, Maryse 03H
 CHANG, Billy 03J
 CHANG, Hsing-Ming 04G
 CHANNOUF, Nabil 11E
 CHAO, Chang-Tai 08F
 CHAREST, Anne-Sophie 08H
 CHAUBEY, Yogendra 06I
 CHELIOTIS, Dimitris 03B
 CHEN, Baojiang 10A
 CHEN, Bei 11B
 CHEN, Bingshu 08I
 CHEN, Haimo 03J
 CHEN, Jiahua 11D
 CHEN, Liang 14B
 CHEN, Sheau-Chiann 12F
 CHEN, Yalin 03J
 CHEN, Yung-Pin 02G
 CHEN, Zhengfei 03J
 CHEN, Zhijian 03F
 CHENG, Cheng 06I
 CHENG, Ji 09H
 CHENG, Yi 04E
 CHENOURI, Shoja'eddin 02H, 03G, 09J
 CHERRY, Nicola 11B
 CHI, Yunchan 02H
 CHIPMAN, Hugh 13A
 CHIRAKKAL, Haridas 12A
 CHIU, Grace 06I
 CHOI, Yun-Hee 04H
 CHOW, Shein-Chung 02H
 CHOWDHURY, Mohammad 06K
 CHU, Rong 04F
 CHUBEY, Yogendra 06I
 CIGSAR, Candemir 11I
 CLAVEAU, Jeannine 11G
 CLEMENT, David 04B
 COLLINS, Krista 09J, 12C

COMPETE GROUP 04F
 COOK, Alex 06D
 COOK, Richard 02F, 04F, 10A, 11C
 CORMIER, Eric 03I
 COSTA, Dan 08E
 COWEN, Laura 11B
 COX, Ted 03B
 CRAIB, Kevin 03A
 CRAWFORD, Steve 06H
 CUERDEN, Meaghan 04F
 CURRIE, Iain 09J

 DACIUK, Tim 09B
 DANILA, Oana 12F
 DAOUST, Pierre 11G
 DARLINGTON, Gerarda 02I, 02J
 DASZAK, Peter 12A
 DAVIES, Katherine 06G, 09I
 DEAN, Charmaine .. 02H, 04B, 04C, 06G, 08A, 09G,
 12G
 DEARDON, Rob 06D
 DEGHAN, Mohannad Hossein 02I
 DEMETRASHVILI, Nino 03J
 DEMNATI, Abdellatif 08F
 DENG, Xiangzheng 06K
 DEN HOLLANDER, Frank 03B
 DESMOND, Anthony 02J
 DESMOND, Tony 02J
 DEVROYE, Luc 05A
 DEWEERDT, Joachim 09K
 DIAS, Ronaldo 09K
 DING, Keyue 12H
 DIXON, Stephanie 02I
 DOBRA, Adrian 06F, 08G
 DONG, Bin 06I
 DONNELLY, Christl 06D
 DOUCET, Arnaud 03H, 04G
 DOWD, Mike 08E
 DOWDEN, Jeffrey 11E
 DUBIN, Joel 03A, 06I, 12G
 DUCHESNE, Thierry 02I

 EDWARDS, Andrew 12A
 ELLIOT, Robert 11A
 EMBRECHTS, Paul 02G
 ERTEFAIE,, Ashkan 12D
 ESCHENBACH, Ted 09K
 EVANS, Michael 06K
 ÉVIN, Guillaume 12J

- FALGUEROLLES, Antoine de02C
 FALLAH, Nader 11E, 12G
 FAN, Chun-Po Steve 02F
 FAN, Guangzhe 12H
 FAVRE, Anne-Catherine 03H
 FAYE, Laura 08G
 FENG, Cindy 03I
 FENG, Xin 09G
 FENG, Zeny 08G
 FENNEL, Katja 08E
 FERLAND, Rene 02G
 FICK, Gordon 11F
 FIELD, Christopher 08E, 11E
 FILLION, Jean-Marc 09J
 FINNOFF, David 12A
 FLEGAL, Katherine 08I
 FLOURNOY, Nancy 04E
 FOX, Colin 03E
 FRANCO, Eduardo 02B
 FREDETTE, Marc 11E
 FRIEDMAN, Jed 09K
 FROLOVA, Olena 09J
 FULLER, Wayne A. 09D

 GABRYS, Robertas 12B
 GADAG, Veeresh 11E
 GAIL, Mitchell 08I
 GARG, Amit 04F
 GASASIRA, Alex 06D
 GAUDET, José 11G
 GAUTHIER, Sylvie 11G
 GEL, Yulia 11B
 GENEST, Christian 09I
 GERBER, Leah 12A
 GHAHRAMANI, Melody 03G
 GHODSI, Ali 09J
 GIBBONS, Robert 08D
 GIBBS, Alison 03I, 03J
 GIBSON, John 06K, 09K
 GILL, Paramjit 14A
 GILLIS, Daniel 02J, 06H
 GOLDSMITH, Charlie 08B
 GOTTARDO, Raphael 04H, 06K
 GOU, Tingting 06K
 GOULD, Jim 06G
 GRAHAM, Jinko 02J, 08G, 08I, 09K
 GRASSLY, Nicholas 06D
 GRAUBARD, Barry 08I
 GREEN, Jane 04H

 GUO, Qing 03J
 GUSTAFSON, Paul 03H, 09H, 12I
 GUTIERREZ, Roberto 09B
 GUTTORP, Peter 06A, 08A

 HAJJEM, Ahlem 03H
 HAKULINEN, Timo 02B
 HALLADAY, Amanda 09J
 HAMADA, M. 09F
 HAN, Lengyi 06G
 HANEUSE, Sebastien 12H
 HANLEY, James 3, 02C
 HANSON, Timothy 04B
 HARDWICK, Janis 04E
 HARFORD, Bill 06H
 HARPER, William 09K
 HASAN, M. Tariqul 08D, 12J
 HASLETT, Stephen J. 14D
 HE, Wenqing 02J, 04H
 HEARNE, Leonard 02G
 HECKMAN, Nancy 0, 02D, 03H
 HENDERSON, Joanna 11F
 HENRYS, Peter A 09G
 HIDIROGLOU, Mike 09D
 HIGDON, David 03E
 HOGG-JOHNSON, Sheilah 12G
 HOLBROOK, Anne 04F, 09H
 HOLENSTEIN, Roman 04G
 HOLMES, Eli 12A
 HOLT, John 02J
 HOOKER, Giles 02D
 HORVATH, Lajos 12B
 HOSSAIN, Shahadut 09H
 HOSSEINI, Reza 04G
 HOSSEINKASHI, Yasaman 09J
 HU, Simin 02E
 HU, X. Joan 11I
 HUANG, Jian 02E
 HUANG, Jianhua 02D
 HUANG, Jikun 06K
 HUANG, Mei Ling 06J
 HUGHES-OLIVER, Jacqueline M. 09J
 HUH, Joonghee 11A
 HUNT, Brian 08E
 HUSTON, Carolyn 03I, 09G
 HUSZTI, Ella 02F
 HUYNEN, Maud 08A
 HYUN, Seung Won 04E

 ISHWARAN, Hemant 11I

JAMES, Thomas 09K
 JANG, Gun Ho 06K, 08H
 JANG, Jisoo 11I
 JASON, Nielsen 02D
 JENKINS, Helen 06D
 JERDE, Christopher 12A
 JEWELL, Nicholas 06D
 JHANGRI, Gian 03A
 JIAHUA, Chen 11D
 JIANG, Depeng 12G
 JIANG, Hedy 02J
 JIANG, Min 06J
 JIANG, Qifeng 09I
 JOHNSON, Edward A. 06G
 JOHNSON, Wesley 04B
 JOHNSTON, Calvin 03J
 JOHNSTON, Neil 11E
 JUAREZ, Elizabeth 02H, 03I
 JUNG, Andrew J. 04I

KALBFLEISCH, Jack 12E
 KAN, Kin Hung 03G
 KANG, Xiuli 08I
 KAPUR, Kush 08D
 KAUFMAN, Cari 03E
 KEATING, Jerome P. 09I
 KESSING, Lars Vedel 12E
 KEVIN, Murphy 06J
 KHALILI, Abbas 12J
 KHAN, Shahedul 06I
 KIANI, Mehdi 03H
 KIM, Hyang Mi 11B
 KLOTZ, Greg 06H
 KOKOSZKA, Piotr 12B
 KOLENIKOV, Stanislav 08F
 KOLEOSHO-ADELEKAN, Titi 06D
 KOLKIEWICZ, Adam 11A
 KOPCIUK, Karen 04H
 KOUTRAS, Markos 04D
 KOVACEVIC, Milorad 03F, 06C, 06J
 KOWGIER, Matthew 03J
 KRAGH ANDERSEN, Per 12E
 KRUEGER, Paul 11F
 KVIST, Kajsa 12E

LABBÉ, Chantal 02G
 LADD, Brian 08A
 LAFONTAINE, Amanda 06H
 LAI, Yinglei 14B
 LAN, Kuang-Kuo 03C

LAROCQUE, Denis 03H
 LAWLESS, Jerald 02I, 03F, 11C, 12E
 LE, Nhu 04G
 LEBLANC, Michael 12H
 LEBOVIC, Gerald 02J
 LEE, Alan 12H
 LEE, Chel Hee 02F
 LEE, Jaeyong 08H
 LEE, Sangyeol 08H
 LEE, Terry 04B
 LEKIVETZ, Ryan 09F
 LI, Dongdong 03J
 LI, Jackie 06H
 LI, Jun 06I
 LI, Longhai 08H
 LI, Pengfei 11D
 LI, Ye 08I
 LI, Yi 02E
 LI, Yolanda 06G
 LIM, Hyun 02I
 LIN, Yi 09K
 LIU, Amy 03J
 LIU, Bitao 02D
 LIU, Jennifer 11F
 LIU, Jinnan 06F, 08G
 LIU, Juxin 12I
 LIU, Peng 11D
 LIU, Shudan 11A
 LIU, Wei 04B, 12G
 LIU, Xiaoming 06G, 06K, 11I
 LIU, Yan 03F
 LIX, Lisa 02J
 LO, Kenneth 04H, 06K
 LOCKHART, Richard 02D, 04C, 09I, 12I
 LODGE, David 12A
 LOEB, Mark 11F
 LOTHIAN, Jack 11G
 LU, Linghong 03J
 LYSENKO, Natalia 02G

MA, Jinhui 04F
 MA, Renjun 08D, 09H, 12J
 MACDONALD, Don 11E
 MACGIBBON, Brenda 11E
 MACKAY, R. Jock 12F
 MACLEAN, Leonard 11D
 MACLURE, Malcolm 08B
 MACNAB, Ying 09K
 MAHMUD, Sadia 11E

- MALIK, Rajat 03J
 MANDAL, Saumen 06K
 MANN, Rena 03J
 MANTEL, Harold 03F
 MARIACA-HAJDUCEK, Dagmar 03F, 12E
 MARLIN, Ben 06F
 MARRIOTT, Paul 02H
 MARTELL, David 06G, 08A
 MARTENS, Pim 08A
 MASAOUD, Elmabrok 02H
 MASSAM, Helene 06F, 08G
 MATTERN, Jann Paul 08E
 MATTHEWS, David E. 06I
 McCANDLESS, Lawrence 09H
 McCONKEY, Brian 12F
 McCULLOCH, Charles 09A
 McKINLEY, TJ 06D
 McLEISH, Don L. 09C
 McLEOD, A. Ian 08H
 McLEOD, Robert 12F
 McNENEY, Brad 02J, 08G, 08I
 McNICHOLAS, Paul 02J, 08I, 12J, 14B
 MEANEY, Christopher 03J
 MEEK, Christopher 06F
 MEHRABIAN, Amaan 04G, 12F
 MERLEAU, James 12J
 METZLER, Adam 09C
 MICHAEL, Friedlander 06J
 MICHAILIDIS, George 09F
 MICHALETZ, Sean 06G
 MILLAR, AnneMichele 11H
 MILLER, Arden 04C
 MILLIGAN, Karen 11F
 MILLS FLEMMING, Joanna 08E
 MIN, Jeong Eun 08I
 MITNITSKI, Arnold 11E, 12G
 MIZERA, Ivan 06E
 MOGER, Tron A 08D
 MOGHADERI, Azadeh 09G
 MOORE, L. 09F
 MOULTON, David 03E
 MUELLER, Hans-Georg 02D
 MULAYATH VARIYATH, Asokan 03G
 MURDOCH, Duncan 02J, 09K
 MUROCH, Duncan 06K
 MURPHY, Kevin 03H, 06F, 08G
 MUTHUKUMARANA, Saman 08H
 NATHOO, Farouk 06G
 NAVARRO, Alfredo 08F
 NEILSEN, Denise 08I
 NEÖLEHOV, Johanna 09I
 NEUHAUS, John 09A
 NEUPANE, Binod 11F
 NEWLANDS, Nathaniel 08I, 12F
 NG, Michelle 04I
 NGUYEN, Paul 09K
 NICCOLS, Alison 11F
 NING, Kaida 04H
 NKURUNZIZA, Sévérien 06K
 NORDHAUSEN, Klaus 06E
 NOURIJELYANI, Keramat 12H
 NUEL, Gregory 04D
 OJA, Hannu 06E
 OLDFORD, Wayne 06B
 O'REILLY, Federico 12I
 OYARZUN, Javier 11G
 PAHWA, Punam 03A
 PAINDAVEINE, Davy 06E
 PAKHOMOV, Evgeny 08E
 PAN, Jeh-Nan 12F
 PAN, Qing 07A
 PAQUETTE, Carrie 08I
 PARÉ, Peter 08G
 PARFREY, Pat 04H
 PARK, Chul 12I
 PATAK, Zdenek 09D
 PEDROSO ESTEVAM DE SOUZA, Camila 09K
 PENG, Defen 03J
 PENG, Jianan 14C
 PENG, Jyh-Ying 04D
 PENG, Yingwei 02I, 11C
 PERREAULT, Luc 12J
 PHRI STATISTICS GROUP 04F
 PICKA, Jeffrey 12I
 PIETRONIRO, Elise 11B
 PLATT, Robert 12D
 PODDER, Mohua 12F
 POGUE, Janice 04F
 PRATOLA, Matthew 09F
 PROSCHAN, Michael 03C
 PROVOST, Serge B. 06J
 PULLENAYEGUM, Eleanor 04F, 09H
 PUNTANEN, Simo 14D
 QU, Conghui 09K
 QU, Ying 08I

QUANTIN, Catherine02F
 QUESSY, Jean-François 03H

 RÉMILLARD, Bruno12B
 RAFFA, Jesse12G
 RAHMAN, Azizur 08H
 RAMEZAN, Reza 02H
 RAMIREZ RAMIREZ, Lilia 11F
 RAMSAY, Jim 02D
 RANJAN, Pritam09F
 RAO, J. Sunil02E
 RAO, J.N.K.04C, 08F
 RAYMOND, Kyle 02F
 REESE, Shane 03E, 14A
 REESOR, Mark 03G, 11A
 REGIER, Michael 12G
 RÉMILLARD, Bruno 02G
 RENAUD, Jean-François 02G
 ROBERTS, G.O.04G
 ROBERTSON, Gordon 04H
 ROCKWOOD, Kenneth11E, 12G
 ROONEY, Neil06H
 ROSENTHAL, J.S.04G
 ROSYCHUK, Rhonda J.09G
 ROUTLEDGE, Rick08A, 08E
 ROY, Anindya 08D
 ROY, Roch 09I
 ROZELLE, Scott06K
 RUBIN-BLEUER, Susana09D

 SAAB, Rabih03J
 SAIDI, Abdelnasser 03F
 SAIN, Stephan 03E, 09F
 SAJOBI, Tolulope 02J
 SALAMH, Mustafa Abdel-Aziz 03I
 SALEEM, Muhammad 03H
 SALIBIÁN-BARRERA, Matías 03H
 SALPETER, Edwin09H
 SALPETER, Shelley09H
 SANDFORD, Andrew08G
 SARRASI, F.J. 03G
 SCHAUBEL, Douglas07A, 12E
 SCHENKEL, Flavio 08G
 SCHMIDT, Mark 06J
 SCHWARZ, Carl 06I, 06J, 09G
 SCOTT, Alastair 09E, 12H
 SEN, Arusharka 11C
 SEN, Pranab06I, 12I
 SENTHILSELVAN, Ambikaipakan 11B
 SHAIKH, Mateen 02J

SHAN, Janet 03I
 SHEN, Hui09J
 SHIH, Weichung03C
 SHIMODAIRA, Hidetoshi14C
 SHIN, Ji-Hyung 08G
 SIKDAR, Khokan11E
 SILVAPULLE, Mervyn 14C
 SITTER, Randy 09F
 SIU, Tak Ken 11A
 SMALL, Christopher 09J
 SMITH, Ainsley11F
 SMITH, Katherine 12A
 SNEDDON, Gary08D, 12J
 SONG, Peter X.-K.09I, 12G
 SORIANO, Juan 03G
 SOSKOLNE, Colin 08A
 SOTEROS, Christine 08C
 SPINELLI, John 09K
 STAFFORD, James 02F
 STARSINIC, Michael 08F
 STEANE, Michelle 08I
 STEINER, Stefan H.12F
 STEINWART, Ingo 06K
 STEPHENS, David 04G, 12F
 STEPHENS, Michael09I, 12I
 STERN, Hal 14A
 STEWART, Connie11E
 STEWART, Sam 12H
 STONE, John 03D
 STONE, Ryan 03J
 STOUT, Quentin 04E
 STRAWDERMAN, Robert 04B
 STRYHN, Henrik02H
 STUTE, Winfried 11C
 STYAN, George09E
 SU, Wanhua 12I
 SULTAN, Shagufta 12J
 SUN, Chengchen03J
 SUN, Jiayang 11D
 SUN, Lei 08G
 SUN, Zheng 03I, 09I
 SWARTZ, Tim 08H, 14A
 SWORD, Wendy 11F
 SYKES, Jenna 03J

 TAAMOUTI, Abderrahim 09I
 TABET, Aline 03H
 TAKAHARA, Glen 09G
 TALEBAN, Julia 11F

- TAMBAY, Jean-Louis 09J
 TAN, Zhiqiang 12D
 TANG, Kenneth 12G
 TAPPER, Dawn 09J
 TASKINEN, Sara 06E
 TAYLOR, Jeremy 11C
 TAYLOR, Jonathan 09E
 TAYLOR, Steve W. 06G
 TEY, Jasper 06H
 THABANE, Lehana 04F, 09H, 11F
 THAVANESWARAN, Aerambamoorthy 03G
 THIESSON, Bo 06F
 THOMPSON, Mary 09I, 11F
 THOMPSON, Steve 08F
 THOMSON, David 09G
 THORPE, Kevin 03A
 TICHON, Jenna 12F
 TOMLINSON, George 02J
 TOMMASI, Désirée 08E
 TONG, Tiejun 14B
 TORABI, Mahmoud 09G
 TORSNEY, Ben 06K
 TOWNLEY-SMITH, Lawrence 12F
 TRENKLER, Goetz 14D
 TRIANTAFYLLOU, I. 04D
 TURNER, Amanda 03B
 TURNER, Elizabeth 02C
 TWUMASI-ANKRAH, Philip 12H

 UBARTAS, Cindy 12C

 VAN DEN BERG, Ewout 06J
 VAVASIS, Stephen 09J
 VINCENT, Kyle 08F
 VIRAG, Balint 08C
 VIVEROS-AGUILERA, Roman 0
 VRUGT, Jasper 03E

 WALDNER, Cheryl 11B
 WALKER, Rod 12H
 WALTER, Stephen 11F
 WANG, Chong 06F
 WANG, Liangliang 03H
 WANG, Liquan 02F, 06K
 WANG, Peiming 12J
 WANG, Steven 03J
 WANG, Weizhen 14C
 WANG, Xikui 09C
 WANG, Yuan 03J
 WANG, Yuedong 14B

 WARD, Eric 12A
 WARRIYAR, K.V, Vineetha 06H
 WATIER, Francois 02G
 WEERASINGHE, Swarna 11F
 WELCH, William J. 09J
 WELDON, Larry 06B
 WELSH, Robert 02H
 WENDELBERGER, J. 09F
 WESTERHOUT, Cynthia 02H
 WICKHAM, Charlotte 03I
 WILD, Chris 01A, 12H
 WOLTERS, Mark 08I
 WONG, Kevin 09H
 WONG, William 08G
 WOODALL, William 04A
 WOOLFORD, Douglas 06G, 08A
 WU, Beilei 03J
 WU, Changbao 03F
 WU, Jianrong 06I
 WU, Lang 04B, 12G, 14C
 WU, Yujun 03C

 XIA, Michelle C. 03J
 XIN, Gao 08G
 XIONG, Juan 04H
 XU, Changjiang 08H
 XU, Jieqing 04I
 XU, Ruimin 02J
 XU, Xiaojian 03H

 YADA, Rickey 08I
 YAN, Guohua 09H
 YANG, Lan-Yan 02H
 YAO, Ping 04E
 YEHUDA, Dayan 06H
 YEUNG, Chi Wai 12C
 YI, Grace 04B, 10A
 YI, Yanqing 04E
 YILMAZ, Yildiz 06J
 YOU, Yong 09D
 YU, Hao 11I, 12B
 YU, Kai 12I
 YU, Savio 09H
 YUAN, Yan 02I
 YUE, Li Hua 02J
 YUEN, Wai Kong 04G
 YUNG, Wesley 09D

 ZAREPOUR, Mahmoud 11I
 ZENG, Leilei 04B

ZHANG, Jessica	04I
ZHANG, Jiajia	02I
ZHANG, Jinning	03I
ZHANG, Peng	12G, 12I
ZHANG, Xu	02I
ZHANG, Xuekui	04H, 08G
ZHANG, Yanqiao	03G
ZHANG, Zheng	06J
ZHAO, Hongyu	14B
ZHAO, Lihui	11I
ZHAO, Lin	03H
ZHAO, Yonggan	09C, 11D
ZHOU, Jonathan	04I
ZHOU, Qian	09I
ZHU, Qiaohao	02H
ZHU, Yifan	03J
ZHUANG, Meiyang	04I
ZIDEK, James	04G, 08I, 12F
ZOU, GuangYong	11F
ZUMBO, Bruno	03F
ZUO, Yijun	06E

Author Index • Index des auteurs

- | | | | |
|---------------------------------|----------|-----------------------------------|------------------------|
| AALEN, Odd | 143 | BELLHOUSE, David | 73 |
| ABANIDA, Emmanuel | 122 | BEN GHORBAL, Noomen El Beji | 170 |
| ABARIN, Taraneh | 77 | BERG, Emily | 161 |
| ABDOLELL, Mohamed | 208 | BERGERON, Pierre-Jérôme | 84 |
| ABRAHAMOWICZ, Michal | 78 | BERKE, Olaf | 166 |
| ADEM, Alejandro | 70 | BHAUMIK, Dulal | 144 |
| AHMED, Ejaz | 138 | BICKEL, Peter | 116 |
| AINSWORTH, Laurie | 146 | BICKIS, Mikelis | 191 |
| ALAGHEHBANDAN, Reza | 184, 185 | BIGGIN, Mark | 116 |
| ALAVI, Asma | 174 | BIGGS, Michael | 174 |
| ALBA, Joseph | 213 | BINGHAM, Derek | 107, 164, 165 |
| ALIOUM, Ahmadou | 78 | BLEI, David | 125 |
| ANDERSON-COOK, Christine | 120 | BOHOSSIAN, Nora | 189 |
| ANDREWS, Jeffrey | 212 | BOLEY, Nathan | 116 |
| ANDREWS, Jessica | 146 | BONNER, Simon | 132 |
| ANDRIEU, Christophe | 113 | BORNN, Luke | 136 |
| ANGHEL, Marian | 136 | BORUVKA, Audrey | 111 |
| ANGST, Jules | 200 | BOUEZMARNI, Taoufik | 171 |
| ARMSTRONG, Paul | 81 | BOVAS, Abraham | 130 |
| ASOKAN, Mulayath Variyath | 130 | BRANSCUM, Adam | 104 |
| ASTON, John | 108 | BRANT, Rollin | 91 |
| ATHERTON, Juli | 116 | BRAUN, W. John | 126, 155, 156 |
| ATOYAN, Tigran | 202 | BREED, Greg | 145 |
| AYLWARD, Bruce | 122 | BREHENY, Patrick | 75 |
| BABYAK, Colin | 174 | BRILLINGER, David | 192 |
| BADESCU, Alex | 178 | BRINKMAN, Ryan | 137 |
| BAI, Yan | 113 | BRIOLLAIS, Laurent | 116, 149 |
| BAILEY, Mark | 158 | BROWN, Ben | 116 |
| BAKAL, Jeffrey | 81 | BROWN, Patrick | 78, 88, 89, 165 |
| BAKSALARY, Oskar M. | 219, 220 | BUCKLEY, Nicholas | 168 |
| BALAKRISHNAN, N. | 107, 170 | BULL, Shelley B | 150 |
| BALASOORIYA, Uditha | 128 | BURKETT, Kelly | 87 |
| BALKEMA, Guus | 80 | BURSTYN, Igor | 180 |
| BALSHAW, Robert | 82 | CAI, Song | 157 |
| BASHIR, Shaheena | 211 | CALAMAI, Peter | 94 |
| BATAINEH, Osama | 90 | CAMPBELL, David | 75 |
| BEASLEY, Mark | 207 | CANNECTIN | 112 |
| BEATON, Dorcas | 203 | CAO, Jiguo | 75, 126, 139, 146, 208 |
| BEAUMONT, Jean-François | 71 | CAREY, John | 94 |
| BEEGLE, Kathleen | 177 | CARRIERE, Keumhee | 155 |
| BELLAVANCE, François | 100 | CARRILLO-GARCIA, Ivan | 134 |

- CARTER, E 211
 CASCAGNETTE, Paul 140
 CASKENETTE, Amanda 129
 CAVASSIN, Sirlei 192
 CHALLENGER, Wendell 131
 CHAMBERS, Ray 121
 CHAMPAGNE, Maryse 101
 CHANG, Billy 104
 CHANG, Hsing-Ming 115
 CHANNOUF, Nabil 184
 CHAO, Chang-Tai 147
 CHAREST, Anne-Sophie 151
 CHAUBEY, Yogendra 131
 CHELIOTIS, Dimitris 92
 CHEN, Baojiang 177
 CHEN, Bei 179
 CHEN, Bingshu 157
 CHEN, Haimo 104
 CHEN, Jiahua 183
 CHEN, Liang 217
 CHEN, Sheau-Chiann 201
 CHEN, Yalin 104
 CHEN, Yung-Pin 79
 CHEN, Zhengfei 104
 CHEN, Zhijian 97
 CHENG, Cheng 131
 CHENG, Ji 168
 CHENG, Yi 109
 CHENOURI, Shoja'eddin 82, 99, 174
 CHERRY, Nicola 180
 CHI, Yunchan 83
 CHIPMAN, Hugh 214
 CHIRAKKAL, Haridas 195
 CHIU, Grace 130
 CHOI, Yun-Hee 116
 CHOW, Shein-Chung 83
 CHOWDHURY, Mohammad 137
 CHU, Rong 111
 CIGSAR, Candemir 193
 CLAVEAU, Jeannine 190
 CLEMENT, David 105
 COLLINS, Krista 174, 198
 COMPETE GROUP 111
 COOK, Alex 123
 COOK, Richard 77, 110, 111, 177, 181
 CORMIER, Eric 103
 COSTA, Dan 145
 COWEN, Laura 180
 COX, Ted 91
 CRAIB, Kevin 91
 CRAWFORD, Steve 129
 CUERDEN, Meaghan 110
 CURRIE, Iain 172
 DACIUK, Tim 159
 DANILA, Oana 202
 DAOUST, Pierre 189
 DARLINGTON, Gerarda 84, 88
 DASZAK, Peter 194
 DAVIES, Katherine 126, 170
 DEAN, Charmaine .. 82, 106, 107, 126, 139, 166, 204
 DEARDON, Rob 123
 DEGHAN, Mohammad Hossein 86
 DEMETRASHVILI, Nino 104
 DEMNATI, Abdellatif 147
 DENG, Xiangzheng 136
 DEN HOLLANDER, Frank 92
 DESMOND, Anthony 88
 DESMOND, Tony 87, 90
 DEVROYE, Luc 118
 DEWEERDT, Joachim 177
 DIAS, Ronaldo 177
 DING, Keyue 208
 DIXON, Stephanie 84
 DOBRA, Adrian 124, 149
 DONG, Bin 132
 DONNELLY, Christl 122
 DOUCET, Arnaud 102, 113
 DOWD, Mike 144, 145
 DOWDEN, Jeffrey 184, 185
 DUBIN, Joel 91, 130, 205
 DUCHESNE, Thierry 86
 EDWARDS, Andrew 195
 ELLIOT, Robert 178
 EMBRECHTS, Paul 80
 ERTEFAIE,, Ashkan 198
 ESCHENBACH, Ted 175
 EVANS, Michael 135
 ÉVIN, Guillaume 212
 FALGUEROLLES, Antoine de 73
 FALLAH, Nader 185, 204
 FAN, Chun-Po Steve 78
 FAN, Guangzhe 208
 FAVRE, Anne-Catherine 101
 FAYE, Laura 150
 FENG, Cindy 103
 FENG, Xin 166

- FENG, Zeny 149
 FENNEL, Katja 144
 FERLAND, Rene 81
 FICK, Gordon 187
 FIELD, Christopher 145, 186
 FILLION, Jean-Marc 173
 FINNOFF, David 194
 FLEGAL, Katherine 157
 FLOURNOY, Nancy 109
 FOX, Colin 95
 FRANCO, Eduardo 72
 FREDETTE, Marc 184
 FRIEDMAN, Jed 177
 FROLOVA, Olena 173
 FULLER, Wayne A. 161

 GABRYS, Robertas 196
 GADAG, Veeresh 185
 GAIL, Mitchell 157
 GARG, Amit 110
 GASASIRA, Alex 122
 GAUDET, José 190
 GAUTHIER, Sylvie 189, 190
 GEL, Yulia 179
 GENEST, Christian 170
 GERBER, Leah 195
 GHAHRAMANI, Melody 98
 GHODSI, Ali 174
 GIBBONS, Robert 144
 GIBBS, Alison 103
 GIBSON, John 136, 177
 GILL, Paramjit 215
 GILLIS, Daniel 90, 129
 GOLDSMITH, Charlie 141
 GOTTARDO, Raphael 115, 117, 137
 GOU, Tingting 135
 GOULD, Jim 128
 GRAHAM, Jinko 87, 150, 155, 176
 GRASSLY, Nicholas 122
 GRAUBARD, Barry 157
 GREEN, Jane 116
 GUO, Qing 104
 GUSTAFSON, Paul 102, 169, 210
 GUTIERREZ, Roberto 159
 GUTTORP, Peter 119, 140

 HAJJEM, Ahlem 100
 HAKULINEN, Timo 72
 HALLADAY, Amanda 174
 HAMADA, M. 164

 HAN, Lengyi 126
 HANEUSE, Sebastien 207
 HANLEY, James 71, 73
 HANSON, Timothy 104
 HARDWICK, Janis 110
 HARFORD, Bill 129
 HARPER, William 175
 HASAN, M. Tariqul 143, 213
 HASLETT, Stephen J. 219
 HE, Wenqing 86, 117
 HEARNE, Leonard 79
 HECKMAN, Nancy 70, 74, 102
 HENDERSON, Joanna 187
 HENRYS, Peter A 165
 HIDIROGLOU, Mike 161
 HIGDON, David 95
 HOGG-JOHNSON, Sheilah 203
 HOLBROOK, Anne 111, 169
 HOLENSTEIN, Roman 113
 HOLMES, Eli 195
 HOLT, John 90
 HOOKER, Giles 75
 HORVATH, Lajos 196
 HOSSAIN, Shahadut 169
 HOSSEINI, Reza 114
 HOSSEINKASHI, Yasaman 174
 HU, Simin 76
 HU, X. Joan 194
 HUANG, Jian 75
 HUANG, Jianhua 75
 HUANG, Jikun 136
 HUANG, Mei Ling 134
 HUGHES-OLIVER, Jacqueline M. 173
 HUH, Joonghee 178
 HUNT, Brian 145
 HUSTON, Carolyn 103, 166
 HUSZTI, Ella 78
 HUYNEN, Maud 139
 HYUN, Seung Won 109

 ISHWARAN, Hemant 193

 JAMES, Thomas 175
 JANG, Gun Ho 135, 151
 JANG, Jisoo 193
 JASON, Nielsen 74
 JENKINS, Helen 122
 JERDE, Christopher 194
 JEWELL, Nicholas 122
 JHANGRI, Gian 91

- JIAHUA, Chen 183
 JIANG, Depeng 203
 JIANG, Hedy 89
 JIANG, Min 133
 JIANG, Qifeng 171
 JOHNSON, Edward A. 127
 JOHNSON, Wesley 104
 JOHNSTON, Calvin 104
 JOHNSTON, Neil 186
 JUAREZ, Elizabeth 82, 103
 JUNG, Andrew J. 118

 KALBFLEISCH, Jack 201
 KAN, Kin Hung 99
 KANG, Xiuli 156
 KAPUR, Kush 144
 KAUFMAN, Cari 96
 KEATING, Jerome P. 170
 KESSING, Lars Vedel 200
 KEVIN, Murphy 133
 KHALILI, Abbas 212
 KHAN, Shahedul 130
 KIANI, Mehdi 102
 KIM, Hyang Mi 180
 KLOTZ, Greg 129
 KOKOSZKA, Piotr 196
 KOLENIOV, Stanislav 147
 KOLEOSHO-ADELEKAN, Titi 122
 KOLKIEWICZ, Adam 178
 KOPCIUK, Karen 116
 KOUTRAS, Markos 107
 KOVACEVIC, Milorad 96, 98, 121, 134
 KOWGIER, Matthew 104
 KRAGH ANDERSEN, Per 200
 KRUEGER, Paul 188
 KVIST, Kajsa 200

 LABBÉ, Chantal 80
 LADD, Brian 139
 LAFONTAINE, Amanda 129
 LAI, Yinglei 216
 LAN, Kuang-Kuo 92
 LAROCQUE, Denis 100
 LAWLESS, Jerald 85, 97, 181, 200
 LE, Nhu 114
 LEBLANC, Michael 208
 LEOVIC, Gerald 88
 LEE, Alan 206
 LEE, Chel Hee 78
 LEE, Jaeyong 151

 LEE, Sangyeol 151
 LEE, Terry 106
 LEKIVETZ, Ryan 164
 LI, Dongdong 104
 LI, Jackie 128
 LI, Jun 131
 LI, Longhai 152
 LI, Pengfei 183
 LI, Ye 154
 LI, Yi 76
 LI, Yolanda 127
 LIM, Hyun 85
 LIN, Yi 175
 LIU, Amy 104
 LIU, Bitao 74
 LIU, Jennifer 187
 LIU, Jinnan 124, 149
 LIU, Juxin 210
 LIU, Peng 183
 LIU, Shudan 179
 LIU, Wei 105, 204
 LIU, Xiaoming 126, 136, 192, 193
 LIU, Yan 98
 LIX, Lisa 89, 90
 LO, Kenneth 115, 137
 LOCKHART, Richard 74, 107, 171, 209
 LODGE, David 194
 LOEB, Mark 188
 LOTHIAN, Jack 190
 LU, Linghong 104
 LYSENKO, Natalia 80

 MA, Jinhui 111, 112
 MA, Renjun 143, 169, 213
 MACDONALD, Don 185
 MACGIBBON, Brenda 184
 MACKAY, R. Jock 202
 MACLEAN, Leonard 182
 MACLURE, Malcolm 141
 MACNAB, Ying 175
 MAHMUD, Sadia 186
 MALIK, Rajat 104
 MANDAL, Saumen 137
 MANN, Rena 104
 MANTEL, Harold 97
 MARIACA-HAJDUCEK, Dagmar 97, 200
 MARLIN, Ben 125
 MARRIOTT, Paul 82
 MARTELL, David 126, 139

- MARTENS, Pim 139
MASAOUD, Elmabrok 82
MASSAM, Helene 124, 149
MATTERN, Jann Paul 144
MATTHEWS, David E. 132
McCANDLESS, Lawrence 168
McCONKEY, Brian 202
McCULLOCH, Charles 158
McKINLEY, TJ 123
McLEISH, Don L. 160
McLEOD, A. Ian 153
McLEOD, Robert 203
McNENEY, Brad 87, 150, 155
McNICHOLAS, Paul 87, 88, 154, 212, 217
MEANEY, Christopher 104
MEEK, Christopher 125
MEHRABIAN, Amaan 114, 202
MERLEAU, James 212
METZLER, Adam 160
MICHAEL, Friedlander 133
MICHAILIDIS, George 165
MICHALETZ, Sean 127
MILLAR, AnneMichele 191
MILLER, Arden 106
MILLIGAN, Karen 187
MILLS FLEMMING, Joanna 145
MIN, Jeong Eun 155
MITNITSKI, Arnold 185, 204
MIZERA, Ivan 123
MOGER, Tron A 143
MOGHADERI, Azadeh 167
MOORE, L. 164
MOULTON, David 95
MUELLER, Hans-Georg 74
MULAYATH VARIYATH, Asokan 99
MURDOCH, Duncan 86, 176
MUROCH, Duncan 135
MURPHY, Kevin 102, 125, 150
MUTHUKUMARANA, Saman 152

NATHOO, Farouk 127
NAVARRO, Alfredo 148
NEILSEN, Denise 157
NEÖLEHOV, Johanna 170
NEUHAUS, John 158
NEUPANE, Binod 188
NEWLANDS, Nathaniel 157, 202
NG, Michelle 118
NGUYEN, Paul 176

NICCOLS, Alison 187
NING, Kaida 117
NKURUNZIZA, Sévérien 138
NORDHAUSEN, Klaus 123
NOURIJELYANI, Keramat 208
NUEL, Gregory 108

OJA, Hannu 123
OLDFORD, Wayne 120
O'REILLY, Federico 209
OYARZUN, Javier 191

PAHWA, Punam 91
PAINDAVEINE, Davy 123
PAKHOMOV, Evgeny 145
PAN, Jeh-Nan 201
PAN, Qing 138
PAQUETTE, Carrie 156
PARÉ, Peter 150
PARFREY, Pat 116
PARK, Chul 210
PATAK, Zdenek 161
PEDROSO ESTEVAM DE SOUZA, Camila 177
PENG, Defen 104
PENG, Jianan 219
PENG, Jyh-Ying 108
PENG, Yingwei 85, 181
PERREAULT, Luc 212
PHRI STATISTICS GROUP 112
PICKA, Jeffrey 209
PIETRONIRO, Elise 180
PLATT, Robert 199
PODDER, Mohua 202
POGUE, Janice 112
PRATOLA, Matthew 164
PROCHAN, Michael 93
PROVOST, Serge B. 133
PULLENAYEGUM, Eleanor 111, 169
PUNTANEN, Simo 219

QU, Conghui 176
QU, Ying 155
QUANTIN, Catherine 78
QUESSY, Jean-François 101

RÉMILLARD, Bruno 196
RAFFA, Jesse 205
RAHMAN, Azizur 153
RAMEZAN, Reza 82
RAMIREZ RAMIREZ, Lilia 188

- RAMSAY, Jim 75
 RANJAN, Pritam 165
 RAO, J. Sunil 76
 RAO, J.N.K. 106, 147
 RAYMOND, Kyle 77
 REESE, Shane 95, 215
 REESOR, Mark 99, 179
 REGIER, Michael 205
 RÉMILLARD, Bruno 80
 RENAUD, Jean-François 80
 ROBERTS, G.O. 113
 ROBERTSON, Gordon 117
 ROCKWOOD, Kenneth 185, 204
 ROONEY, Neil 129
 ROSENTHAL, J.S. 113
 ROSYCHUK, Rhonda J. 167
 ROUTLEDGE, Rick 140, 145, 146
 ROY, Anindya 144
 ROY, Roch 171
 ROZELLE, Scott 136
 RUBIN-BLEUER, Susana 162
- SAAB, Rabih 104
 SAIDI, Abdelnasser 98
 SAIN, Stephan 95, 164
 SAJOBI, Tolulope 89
 SALAMH, Mustafa Abdel-Aziz 103
 SALEEM, Muhammad 102
 SALIBIÁN-BARRERA, Matías 102
 SALPETER, Edwin 168
 SALPETER, Shelley 168
 SANDFORD, Andrew 150
 SARRASI, F.J. 99
 SCHAUBEL, Douglas 138, 201
 SCHENKEL, Flavio 149
 SCHMIDT, Mark 133
 SCHWARZ, Carl 131, 132, 166
 SCOTT, Alastair 163, 206
 SEN, Arusharka 182
 SEN, Pranab 131, 210
 SENTHILSELVAN, Ambikaipakan 180
 SHAIKH, Mateen 87
 SHAN, Janet 103
 SHEN, Hui 173
 SHIH, Weichung 93
 SHIMODAIRA, Hidetoshi 218
 SHIN, Ji-Hyung 150
 SIKDAR, Khokan 184, 185
 SILVAPULLE, Mervyn 218
- SITTER, Randy 164
 Siu, Tak Ken 178
 SMALL, Christopher 174
 SMITH, Ainsley 187
 SMITH, Katherine 194
 SNEDDON, Gary 143, 213
 SONG, Peter X.-K. 172, 206
 SORIANO, Juan 99
 SOSKOLNE, Colin 139
 SOTEROS, Christine 142
 SPINELLI, John 176
 STAFFORD, James 78
 STARSINIC, Michael 148
 STEANE, Michelle 154
 STEINER, Stefan H. 202
 STEINWART, Ingo 136
 STEPHENS, David 114, 202
 STEPHENS, Michael 171, 209
 STERN, Hal 214
 STEWART, Connie 186
 STEWART, Sam 208
 STONE, John 94
 STONE, Ryan 104
 STOUT, Quentin 110
 STRAWDERMAN, Robert 105
 STRYHN, Henrik 82
 STUTE, Winfried 182
 STYAN, George 163
 SU, Wanhua 210
 SULTAN, Shagufta 211
 SUN, Chengchen 104
 SUN, Jiayang 183
 SUN, Lei 150
 SUN, Zheng 103, 171
 SWARTZ, Tim 152, 215
 SWORD, Wendy 187
 SYKES, Jenna 104
- TAAMOUTI, Abderrahim 171
 TABET, Aline 102
 TAKAHARA, Glen 167
 TALEBAN, Julia 188
 TAMBAY, Jean-Louis 173
 TAN, Zhiqiang 199
 TANG, Kenneth 203
 TAPPER, Dawn 174
 TASKINEN, Sara 123
 TAYLOR, Jeremy 181
 TAYLOR, Jonathan 163

- TAYLOR, Steve W. 127
 TEY, Jasper 129
 THABANE, Lehana 111, 112, 168, 187
 THAVANESWARAN, Aerambamoorthy 98
 THIESSON, Bo 125
 THOMPSON, Mary 172, 188
 THOMPSON, Steve 148
 THOMSON, David 167
 THORPE, Kevin 91
 TICHON, Jenna 203
 TOMLINSON, George 88
 TOMMASI, Désirée 145
 TONG, Tiejun 216, 217
 TORABI, Mahmoud 167
 TORSNEY, Ben 137
 TOWNLEY-SMITH, Lawrence 202
 TRENKLER, Goetz 219, 220
 TRIANTAFYLLOU, I. 107
 TURNER, Amanda 92
 TURNER, Elizabeth 73
 TWUMASI-ANKRAH, Philip 208

 UBARTAS, Cindy 197

 VAN DEN BERG, Ewout 133
 VAVASIS, Stephen 174
 VINCENT, Kyle 148
 VIRAG, Balint 142
 VIVEROS-AGUILERA, Roman 70
 VRUGT, Jasper 95

 WALDNER, Cheryl 180
 WALKER, Rod 207
 WALTER, Stephen 188
 WANG, Chong 125
 WANG, Liangliang 102
 WANG, Liquan 77, 78, 136
 WANG, Peiming 213
 WANG, Steven 104
 WANG, Weizhen 219
 WANG, Xikui 160
 WANG, Yuan 104
 WANG, Yuedong 216
 WARD, Eric 195
 WARRIYAR, K.V, Vineetha 130
 WATIER, Francois 81
 WEERASINGHE, Swarna 189
 WELCH, William J. 173
 WELDON, Larry 119
 WELSH, Robert 81

 WENDELBERGER, J. 164
 WESTERHOUT, Cynthia 81
 WICKHAM, Charlotte 103
 WILD, Chris 70, 206
 WOLTERS, Mark 155
 WONG, Kevin 169
 WONG, William 149
 WOODALL, William 104
 WOOLFORD, Douglas 126, 139
 WU, Beilei 104
 WU, Changbao 96
 WU, Jianrong 131
 WU, Lang 105, 204, 219
 WU, Yujun 93

 XIA, Michelle C. 104
 XIN, Gao 149
 XIONG, Juan 117
 XU, Changjiang 153
 XU, Jieqing 118
 XU, Ruimin 88
 XU, Xiaojian 101

 YADA, Rickey 154
 YAN, Guohua 169
 YANG, Lan-Yan 83
 YAO, Ping 109
 YEHUDA, Dayan 129
 YEUNG, Chi Wai 197
 YI, Grace 105, 177
 YI, Yanqing 110
 YILMAZ, Yildiz 134
 YOU, Yong 162
 YU, Hao 192, 196
 YU, Kai 211
 YU, Savio 169
 YUAN, Yan 85
 YUE, Li Hua 86
 YUEN, Wai Kong 114
 YUNG, Wesley 162

 ZAREPOUR, Mahmoud 193
 ZENG, Leilei 106
 ZHANG, Jessica 118
 ZHANG, Jiajia 85
 ZHANG, Jinning 103
 ZHANG, Peng 206, 210
 ZHANG, Xu 85
 ZHANG, Xuekui 117, 150
 ZHANG, Yanqiao 100

ZHANG, Zheng	134
ZHAO, Hongyu	217
ZHAO, Lihui	194
ZHAO, Lin	101
ZHAO, Yonggan	160, 182
ZHOU, Jonathan	118
ZHOU, Qian	172
ZHU, Qiaohao	83
ZHU, Yifan	104
ZHUANG, Meiyang	118
ZIDEK, James	114, 157, 202
ZOU, GuangYong	188
ZUMBO, Bruno	98
ZUO, Yijun	124