

Table of Contents • Table des matières

Table of Contents • Table des matières	3
Welcome • Bienvenue	4
Sponsors • Commanditaires	6
Organizers • Organisateurs	7
Exhibitors • Exposants	8
General Information • Information générale	9
Committees and Meetings • Comités et réunions	14
List of Sessions • Liste des sessions	18
Program • Programme	23
Abstracts • Résumés	61
Author Index • Index des auteurs	179
Author Index (per session) • Index des auteurs (par session)	184

Welcome • Bienvenue

Welcome to the City of London and to the University of Western Ontario. Whether you are a presenter or just listening to the sessions, we hope that the conference is an enjoyable and fruitful experience for you.

Founded in 1878, Western is not only one of Canada's oldest universities, but also has one of the most beautiful campuses. In 2004 Western received the top award in the Parks and Grounds division of the Communities in Bloom competition. This competition is a nation-wide beautification program that judges cities, towns, villages and large institutions in a wide range of landscaping and architectural categories. During your leisure time, stroll around the campus and enjoy its beauty.

London is the hub of Southwestern Ontario. There are many things to do in the city outside the scheduled conference activities. In the downtown area there is an eclectic blend of restaurants, museums, pubs, parks and art galleries. Many restaurants and bars can be found along Richmond Row. Near the forks of the Thames, in the heart of downtown, is London's art gallery. Nearby is Eldon House, London oldest residence. Across the river from the art gallery is Labatt Park, home of the oldest ballpark in North America on its original site. For your convenience there is a London Tourism booklet in your conference bag.

For those who enjoyed the walk through the campus, the city has more than 35 kilometers of hiking-biking trails through park and riverbank areas. You can walk to downtown from the university residence Elgin Hall in about 45 minutes with only an initial brief step on a city street. The trail begins on the south side of University drive at Sunset Street. On Sunset there is a quick turn onto The Parkway. Soon, on the right side of the street, there is a paved trail into a parkland area. You are downtown when you reach the forks of the Thames.

Should the attractions of the city and campus not distract you too much from the conference itself and should you look at the program more closely, you will find that there is an exciting array of workshops, talks and presentations over the next few days.

Enjoy.

Bienvenue à la ville de London et à l'Université Western Ontario. Que vous soyez conférencier ou simple participant aux séances, nous espérons que ce Congrès vous sera agréable et fructueux.

Fondée en 1878, Western est l'une des plus anciennes universités du Canada et elle se vante de l'un des plus magnifiques campus. En 2004, Western a reçu le premier prix dans la division Parcs et espaces verts du concours Collectivités en fleurs. Ce concours est un programme national d'embellissement qui juge villes, villages et grandes institutions dans diverses catégories d'aménagement paysager et d'architecture. Pendant vos moments de loisir, prenez le temps de vous promener sur le campus et d'en apprécier la beauté.

London est le centre du sud-ouest de l'Ontario. Il y a bien plus à y faire que les activités organisées dans le cadre du Congrès. Le centre-ville abrite un mélange éclectique de restaurants, de musées, de pubs, de parcs et de galeries d'art. Richmond Row est connu pour ses nombreux restaurants et bars. Près de la fourche de la Thames, au cœur du centre-ville, vous trouverez le musée d'art de London et tout près, Eldon House, la plus vieille résidence privée de la ville. Sur l'autre rive de la rivière, vous apercevez Labatt Park, qui abrite le plus ancien stade de base-ball de l'Amérique du nord encore sur son site original. Pour plus de

détails sur ces attractions et d'autres, vous trouverez une brochure de London Tourism dans votre trousse du congrès.

Pour ceux d'entre vous qui avez apprécié la promenade sur le campus, la ville offre plus de 35 kilomètres de sentiers pédestres/pistes cyclables sillonnant les parcs et les bords de rivière. Vous pouvez vous rendre à pied au centre-ville depuis la résidence universitaire Elgin Hall en 45 minutes environ, en évitant presque entièrement les rues. Le sentier débute au sud de University Drive à Sunset Street. Sur Sunset, prenez immédiatement un virage pour retrouver The Parkway. Bientôt, à droite de la rue, vous verrez un sentier pavé qui donne sur un parc. Vous arriverez au centre-ville à la fourche de la Thames.

Si les attractions de la ville et du campus ne vous distraient pas trop du Congrès lui-même et si vous examinez le programme de celui-ci plus en détail, vous découvrirez qu'il y a également une variété d'ateliers, d'exposés et de présentations passionnants auxquels participer pendant ces quelques jours.

Profitez-en donc !

Sponsors • Commanditaires

SSC 2006 thanks the following sponsors for their extremely generous contributions for the success of this meeting:

Le Congrès SSC 2006 remercie les commanditaires suivants pour leurs très généreuses contributions qui ont mené au succès du congrès :

- Centre de recherches mathématiques
- Department of Statistical and Actuarial Sciences,
The University of Western Ontario
- Faculty of Science, The University of Western Ontario
- The Fields Institute for Research in Mathematical Sciences
- MITACS
- National Program on Complex Data Structures (NPCDS)
- Pacific Institute for the Mathematical Sciences
- Research Western, The University of Western Ontario
- Robarts Research Institute
- SAS Institute, Inc.

Organizers • Organisateurs

Local Arrangements Committee • Comité organisateur local

Chair • Président: David Bellhouse

Too many cooks spoil the broth, so the saying goes. In this case, had there not been so many cooks, there would have been no broth.

Many in the Department of Statistical and Actuarial Sciences were of great assistance to me. I would like to single out Duncan Murdoch who handled the web and a few other duties, Doug Woolford who ordered bags, pens, pads of paper and other odds and ends, and Jennifer Dungavell who straightened out a number of financial and other issues, and was involved in the design of the covers for the conference book. Throughout the conference you will see a number of people wearing red t-shirts, over 20 in number and too many to mention by name. They are undergraduate and graduate student volunteers who have helped with registration, room set-ups, audiovisual equipment and a host of other things.

The conference book itself would not have been possible without the work of Christian Léger and staff at CRM, in particular, André Montpetit, Daniel Ouimet, and Louise Letendre. The program with abstracts was produced using a database at CRM. It was Kinkos that printed what CRM produced.

Félix Labrecque-Synnot and Jihad Jomaa, two students from Université de Montréal, have translated most abstracts under the supervision of Hélène Bérard (Statistics Canada), Chair of the Bilingualism Committee. Members of the Bilingualism Committee have reviewed the abstracts while Christian Léger did final revisions.

Beginning at a leisurely pace well over two years ago and lately in a much more intense fashion, I have worked closely with Melanie Harvey of Conference Services at the University of Western Ontario. She and her staff have taken care of the conference registration process, the room bookings and several other logistical issues.

Chef Michael Coleby and his staff of Great Hall Catering at the University of Western Ontario were responsible for much of the food, including the banquet, the welcoming reception and various committees' breakfasts, lunches and dinners.

Finally, I am grateful to my French horn teacher, Sasha Gorbasesw of the Don Wright Faculty of Music at Western for arranging the musical entertainment throughout the conference.

Trop de mains à la pâte, ça la gâte, dit-on. Mais dans notre cas, il n'y aurait pas eu de pâte sans toutes ces mains.

De nombreux collègues du Département de sciences statistiques et actuarielles m'ont grandement aidé. Je tiens à remercier tout particulièrement Duncan Murdoch, qui s'est chargé du site Web et d'autres responsabilités, Doug Woolford, qui a commandé les sacs, stylos, bloc-notes et autres fournitures, et Jennifer Dungavell, qui a résolu divers problèmes financiers et autres et a aidé à créer la couverture du cahier du Congrès. Pendant tout le Congrès, vous croiserez une équipe de plus de vingt personnes en t-shirt rouge, trop nombreux pour être nommés individuellement ici. Il s'agit de bénévoles, étudiants de premier cycle

et des cycles supérieurs, qui ont aidé pour les inscriptions, l'organisation des salles, l'équipement audiovisuel, etc.

Le cahier du Congrès n'aurait jamais existé sans les efforts de Christian Léger et du personnel du CRM, tout particulièrement, André Montpetit, Daniel Ouimet et Louise Letendre. Le programme, qui inclut des résumés, a été produit grâce à une base de données du CRM et imprimé par Kinkos.

La traduction de la plupart des résumés a été réalisée par Félix Labrecque-Synnot et Jihad Jomaa, deux étudiants de l'Université de Montréal, sous la supervision d'Hélène Bérard (Statistique Canada), présidente du Comité du bilinguisme. Les membres du Comité du bilinguisme ont révisé les résumés. Christian Léger a fait une révision finale.

Depuis plus de deux ans déjà, mais de façon beaucoup plus intensive ces derniers temps, j'ai travaillé en étroite collaboration avec Melanie Harvey du Service de conférences de l'Université Western Ontario. Elle et son équipe se sont chargées du processus d'inscription au congrès, des réservations de salles et d'autres problèmes logistiques.

Le chef cuisinier Michael Coleby et son personnel du service de restauration Great Hall de l'Université Western Ontario sont responsables de la plupart des repas, y compris du banquet, de la réception d'accueil et des petits-déjeuners, déjeuners et dîners de divers comités.

Enfin, je suis reconnaissant à mon professeur de cor d'harmonie, Sasha Gorbasew de la Faculté de musique Don Wright de Western, pour avoir organisé les divertissements musicaux pendant tout le Congrès.

Program Committee • Comité du programme

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

Members • Membres:

Biostatistics Section • Groupe de biostatistique

Michael Escobar, *University of Toronto*

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

François Pageau, *SNC Technologies*

Survey Methods Section • Groupe de méthodologie d'enquête

Patricia Whitridge, *Elections Canada*

Probability Section • Groupe de probabilité

Andre Dabrowski, *University of Ottawa*

Exhibitors • Exposants

Displays of books for examination and purchase may be found in room 2333 of the Social Science Centre. Representatives of John Wiley, Pearson, Thomson Nelson and W.H. Freeman will be present. Cambridge University Press did not send a representative, but their material may be viewed in the book display area. Salford Systems will also have a display of software in room 2333.

Vous trouverez des étagères de livres à consulter ou acheter dans la salle 2333 du Centre des sciences sociales. Des représentants de John Wiley, Pearson, Thomson Nelson et W.H. Freeman seront présents. Cambridge University Press n'a pas envoyé de représentant, mais vous pouvez consulter leurs ouvrages dans la salle d'exposition. Salford Systems exposera également ses logiciels dans la salle 2333.

General Information • Information générale

Registration will be in Elgin Hall on Saturday, May 27 from 14:00 to 22:00 and on Sunday, May 28 from 11:00 to 19:00. From Monday to Wednesday, the registration desk will be located in the Social Science Centre near room 2050. The desk will be open Monday, May 29 from 8:00 to 15:00. On Tuesday and Wednesday the desk will be open from 8:00 to 11:00.

Les inscriptions se feront à Elgin Hall, samedi 27 mai de 14 heures à 22 heures et dimanche 28 mai de 11 heures à 19 heures. De lundi à mercredi, le bureau des inscriptions sera situé dans le Centre des sciences sociales, près de la salle 2050. Le bureau sera ouvert lundi 29 mai de 8 heures à 15 heures, ainsi que mardi et mercredi de 8 heures à 11 heures.

Directions

There are campus maps in every conference bag so that you may easily find your way around the campus. The map is included in a brochure that provides a self-guided walking tour of the campus, which is well worth the effort.

For those staying in Elgin Hall, the university residence assigned for the conference, or staying in downtown hotels, the main entrance to the University of Western Ontario is located on Richmond Street. The road into the university off Richmond Street is called University Drive. Elgin Hall is the building on your right as you come along University Drive from Richmond Street. There is free parking in the Medway Lot behind Elgin Hall. Conference registration on Saturday and Sunday, May 27 and 28, will be at Elgin Hall.

For those staying in Windermere Manor, the hotel is a short walk to the main campus.

The London Transit Commission has provided all conference registrants with free bus passes. Bus passes will be in the conference bag.

For those deciding to take a bus from the downtown area, there are two convenient routes. The Richmond bus (#6) leaves from Richmond and Dundas Streets approximately every 20 minutes on weekdays. The ride to the university from this intersection is approximately 15 to 20 minutes. The Wellington bus (#13) leaves Dundas and Wellington Streets approximately every 15 minutes on weekdays, again an approximate 15 to 20 minute ride to the university. For those staying at the Hilton and the Delta London Armouries, the closest bus is the Wellington bus. Both the Wellington and Richmond buses go by the Station Park Hotel.

Vous trouverez dans votre trousse de congrès un plan du campus pour vous aider à vous retrouver facilement. Ce plan est inclus dans une brochure qui présente une visite à pied autoguidée du campus, que nous vous recommandons vivement.

Pour les personnes séjournant à Elgin Hall, la résidence universitaire assignée aux congressistes, ou dans un hôtel du centre-ville, l'entrée principale de l'Université de Western Ontario est située rue Richmond. La route qui pénètre sur le campus à partir de la rue Richmond s'appelle rue University. Elgin Hall est le bâtiment à votre droite lorsque vous prenez la rue University à partir de la rue Richmond. Vous pouvez stationner gratuitement dans l'aire Medway derrière Elgin Hall. Les inscriptions au Congrès se feront samedi 27 et dimanche 28 mai à Elgin Hall.

Pour les personnes séjournant à Windermere Manor, l'hôtel est à quelques minutes à pied du campus principal.

La London Transit Commission a fourni à tous les congressistes une carte d'abonnement d'autobus gratuite, que vous trouverez dans votre trousse.

Si vous souhaitez prendre l'autobus depuis le centre-ville, vous avez le choix entre deux lignes. L'autobus Richmond (ligne 6) quitte le coin des rues Richmond et Dundas toutes les 20 minutes environ en semaine. Il faut de 15 à 20 minutes pour arriver à l'université depuis cette intersection. L'autobus Wellington (ligne 13) quitte le coin des rues Dundas et Wellington toutes les 15 minutes environ en semaine et prend aussi de 15 à 20 minutes pour arriver à l'université. Pour les personnes séjournant aux hôtels Hilton et Delta London Armouries, l'autobus le plus proche est la ligne Wellington. Les deux lignes (Wellington et Richmond) passent devant l'hôtel Station Park.

Rooms • Salles

Most scientific sessions will be held in the Social Science Centre at The University of Western Ontario. Room 2050 of the Social Science Centre is on the main floor of the building just beyond the registration desk. Several other sessions are on this floor or on the next floor up the main staircase. A few sessions will be held in Somerville House and University College. These buildings are all located in the central part of the campus. Directions to these buildings are posted on signs in various locations. Volunteers wearing red t-shirts are also on hand to give directions and general help.

Conference registration during the scientific sessions will be in the Social Science Centre.

The Board meeting on Sunday will be held at Windermere Manor. All other committee meetings held over either breakfast or lunch, as well as the Wednesday Board meeting, will be held in Somerville House.

La plupart des séances scientifiques se dérouleront dans le Centre des sciences sociales de l'Université Western Ontario. La salle 2050 du Centre des sciences sociales est située au rez-de-chaussée du bâtiment, juste derrière le bureau des inscriptions. Plusieurs autres séances se tiendront à cet étage ou à l'étage supérieur, en haut de l'escalier principal. Quelques séances seront organisées à Somerville House et à University College. Ces bâtiments sont tous situés dans la partie centrale du campus. Des panneaux d'indication seront affichés à divers endroits pour vous diriger vers ces bâtiments. Des bénévoles en t-shirt rouge seront également là pour vous orienter et vous informer.

Les inscriptions se feront au Centre des sciences sociales pendant les séances scientifiques.

La réunion du Conseil d'administration de dimanche se tiendra à Windermere Manor. Les autres réunions de comités, organisées à l'heure du petit-déjeuner ou du déjeuner, et la réunion du Conseil d'administration de mercredi, se tiendront à Somerville House.

Workshops • Ateliers

The workshops organized by the three sections will take place in the Somerville House. The Biostatistics Section and Survey Methods Section workshop both start at 9:00 while the Business and Industrial Statistics Section workshop starts at 13:00.

Les ateliers organisés par les trois groupes se tiendront à Somerville House. Les ateliers du groupe de biostatistique et du groupe des méthodes d'enquête commencent à 9 heures, tandis que celui du groupe de statistique industrielle et de gestion débute à 13 heures.

Poster Session • Séance d'affichage

The poster session will be held on Sunday, May 28 from 19:00 to 21:00 in the Great Hall of Somerville house in conjunction with the welcoming reception.

La séance d'affichage se tiendra dimanche 28 mai de 19 heures à 21 heures dans le Grand Hall de Somerville House, dans le cadre de la réception de bienvenue.

NSERC Session • Atelier offert par le CRSNG

Wednesday, May 31, 12:30–13:15

Lundi, le 31 mai, 12h30–13h15

Social Science Centre 2036

Isabelle Blain, Vice-President of Research Grants and Scholarships at NSERC will make a presentation and lead a discussion on the support of mathematics and statistics by NSERC. This includes a broad discussion of funding for major research resources such as the National Program for Complex Data Structures, as well as other funding envelope concepts. The session is open to all interested researchers.

Isabelle Blain, vice-présidente de la Direction des subventions de recherche et bourses au CRSNG, fera une présentation et animera une discussion sur le soutien de la mathématique et de la statistique par le CRSNG. Elle abordera notamment la question du financement des principales ressources de recherche, telles le Programme national sur les structures de données complexes, ainsi que d'autres concepts d'enveloppes de financement. Cette séance est ouverte à tous les chercheurs intéressés.

Welcome Reception • Réception de bienvenue

The welcoming reception will take place at the Great Hall in Somerville House on Sunday, May 28 from 19:00 to 21:00. The reception is sponsored by the National Program on Complex Data Structures. All conference attendees and companions are warmly invited to come and enjoy a social time with old and new friends in the statistical community. A string quartet comprised of students from the Western's Don Wright Faculty of Music will perform during the reception.

La réception de bienvenue se tiendra dans le Grand Hall de Somerville House, dimanche 28 mai de 19 heures à 21 heures. La réception est commanditée par le Programme national sur les structures de données complexes. Tous les participants au Congrès et leurs invités sont chaleureusement invités à venir socialiser avec leurs nouveaux et vieux amis de la communauté statistique. Un quatuor à cordes d'étudiants de la Faculté de musique Don Wright de Western jouera pendant la réception.

Food Service on Campus • Restauration sur le campus

There are three locations where lunch can be purchased on campus. Centre Spot, run by the university's Food Services, is a cafeteria on the main floor of the University Community Centre. On the second floor of the University Community Centre is The Wave, a licenced restaurant run by the University Student Council. The University Community Centre is the building next to the Social Science Centre. It is a short walk outside between these buildings. Those wanting an indoor route can take the tunnel that connects the

SH = Somerville House

SSC = Social Science Centre

UC = University College

two buildings. In the Social Science Centre, the tunnel can be reached by taking the stairs nearest room 2333, where the book displays are located, to the ground floor. A third lunch location is the Grad Club, which is located on the ground floor of Middlesex College, a five minute walk from the Social Science Centre. The Grad Club, operated by the Society of Graduate Students, is licenced and features an outdoor patio.

Other than the Grad Club, there are no facilities for dinner on campus.

Vous pouvez vous acheter à déjeuner à trois endroits différents sur le campus. Centre Spot, qui est exploité par les services d'alimentation de l'université, est une cafétéria au rez-de-chaussée du Centre communautaire de l'Université. Au deuxième étage de ce même Centre se trouve The Wave, un restaurant avec permis d'alcool exploité par le Conseil étudiant de l'Université. Le Centre communautaire de l'Université (University Community Centre) est le bâtiment adjacent au Centre des sciences sociales, à quelques pas. Si vous souhaitez vous y rendre sans sortir à l'extérieur, un tunnel relie les deux bâtiments. Depuis le Centre des sciences sociales, prenez les escaliers près de la salle 2333, où se trouve l'exposition de livres, jusqu'au rez-de-chaussée pour rejoindre ce tunnel. Vous pouvez aussi déjeuner au Grad Club, qui est situé au rez-de-chaussée de Middlesex College, à cinq minutes à pied du Centre des sciences sociales. Le Grad Club, qui est exploité par la Société des étudiants, a un permis d'alcool et se vante d'un patio extérieur.

Le Grad Club est le seul endroit où dîner le soir sur le campus.

Barbecue

The annual barbecue will be held in the Grad Club on Monday, May 29 from 18:30 to 20:30. All graduate students receive tickets to this event with their registration. All are invited to attend; tickets may be purchased at the registration desk. Featured entertainment for this event will be a jazz singer currently studying at Western's Don Wright Faculty of Music.

Le barbecue annuel est organisé au Grad Club lundi 29 mai, de 18 heures 30 à 20 heures 30. Tous les étudiants diplômés recevront un ticket pour ce repas avec leur inscription. Vous êtes tous invités à participer ; il vous suffit d'acheter un ticket au bureau des inscriptions. Un divertissement musical sera proposé par un chanteur de jazz qui étudie à la Faculté de musique Don Wright de Western.

Women in Statistics Reception • Réception pour les femmes en statistique

The Canadian Section of the Caucus for Women in Statistics and the SSC Committee on Women in Statistics cordially invite all interested men and women to attend an informal get-together in the Board room of the Grad Club at 20:30 after the barbecue. Students are especially welcome.

La section canadienne du Caucus for Women in Statistics et le Comité sur les femmes en statistique de la SSC invitent cordialement tous les hommes et toutes les femmes intéressé(e)s à participer à une réunion informelle dans la salle du Conseil du Grad Club à 20 heures 30 après le barbecue. Les étudiant(e)s sont tout particulièrement les bienvenu(e)s.

Alumni Receptions • Réceptions pour les anciens étudiants

The Department of Statistical and Actuarial Sciences of the University of Western Ontario will host an alumni reception in the Board Room of the Grad Club during the lunch break on Monday, May 29. Pizza will be served.

The Department of Statistics and Actuarial Science of the University of Waterloo will host an alumni reception in Michael's Garden located in Somerville House Room 3320 during the lunch break on Tuesday, May 30. A light lunch will be served.

Le Département de sciences statistiques et actuarielles de l'Université Western Ontario organise une réception avec pizza pour les anciens étudiants dans la salle du Conseil du Grad Club pendant la pause déjeuner, lundi 29 mai.

Le Département de sciences statistiques et actuarielles de l'Université de Waterloo organise une réception pour les anciens étudiants à Michael's Garden, qui est situé dans la salle 3320 de Somerville House, pendant la pause déjeuner, mardi 30 mai. Un déjeuner léger sera servi.

Banquet

All participants are cordially invited to attend the Conference Banquet on Tuesday evening, May 30. Banquet tickets are distributed with the registration package. An outdoor cash bar will be available from 18:30 in the 3M Patio near Somerville House. Dinner will be served in the Great Hall in Somerville House at 19:15. During the social time on the 3M Patio a brass ensemble will perform. After the meal, a singer will be featured. All performers are students at the Don Wright Faculty of Music. The Society awards will be presented at the end of the evening.

Tous les participants sont cordialement invités à participer au Banquet du Congrès, mardi 30 mai au soir. Les tickets pour le banquet seront distribués avec la trousse d'inscription. Un bar payant sera proposé à partir de 18 heures 30, sur le Patio 3M près de Somerville House. Le dîner sera servi dans le Grand Hall de Somerville House à 19 heures 15. Pendant la réception au Patio 3M, un ensemble de cuivres jouera. Après le repas, vous pourrez écouter un chanteur. Tous les musiciens sont des étudiants de la Faculté de musique Don Wright de Western. Les prix de la Société seront présentés en fin de soirée.

Internet Access • Accès Internet

Computers with internet access are available in room 256 of the Western Science Centre in the Department of Statistical and Actuarial Sciences. The Western Science Centre, located next to Middlesex College and the Grad Club, is a five minute walk from the Social Science Centre. Your conference bag contains a page with a user id and password.

Vous trouverez des ordinateurs avec accès Internet dans la salle 256 du Centre des sciences Western, dans le Département de sciences statistiques et actuarielles. Le Centre des sciences Western, qui est situé près de Middlesex College et du Grad Club, est à cinq minutes à pied du Centre des sciences sociales. Votre trousse contient une feuille indiquant un nom d'utilisateur et le mot de passe à utiliser.

Committees and Meetings

Comités et réunions

Calendar Listing • Calendrier journalier

Saturday May 27 • Samedi 27 mai

18:00–22:00 Executive Committee • *Comité exécutif* Windermere Manor,
North Meeting Room

Sunday, May 28 • Dimanche 28 mai

9:00–11:00 Finance Committee • *Comité des finances* Windermere Manor, Room A
11:00–12:00 Publications Committee • *Comité des publications* Windermere Manor, Room A
12:00–16:00 Board of Directors • *Conseil d'administration* Windermere Manor, Room A

Monday, May 29 • Lundi 29 mai

7:00–8:00 Biostatistics Section (Section Officers) • *Groupe de bio-* Michael's Garden, SH3320
statistique (directeurs du groupe)
Business and Industrial Statistics Section (Section Offi- Michael's Garden, SH3320
cers) • *Groupe de statistique industrielle et de gestion*
(directeurs du groupe)
Survey Methods Section (Section Officers) • *Groupe des* Michael's Garden, SH3320
méthodes d'enquête (directeurs du groupe)
Probability Section (Section Officers) • *Groupe de pro-* Michael's Garden, SH3320
babilité (directeurs du groupe)
12:30–13:30 Accreditation Appeals Committee • *Comité des appels* Michael's Dining Room,
d'accréditation SH3340
CJS Editorial Board • *Comité de rédaction de la RCS* Michael's Garden, SH3320
Public Relations Committee • *Comité des relations pu-* Michael's Dining Room,
bliques SH3340
Southern Ontario Regional Association (General Meet- Michael's Dining Room,
ing) • *Association régionale du sud de l'Ontario (as-* SH3340
semblée générale)
13:30–15:00 Accredited Members General Meeting (open) • *Assem-* SSC3018
blée générale des membres accrédités (public)
17:00–18:00 Business and Industrial Statistics Section (General SSC3026
Meeting) • *Groupe de statistique industrielle et de ges-*
tion (assemblée générale)

SH = Somerville House

SSC = Social Science Centre

UC = University College

Survey Methods Section (General Meeting) • <i>Groupe des méthodes d'enquête (assemblée générale)</i>	SSC2024
Biostatistics Section (General Meeting) • <i>Groupe de bio-statistique (assemblée générale)</i>	SSC2036
Probability Section (General Meeting) • <i>Groupe de probabilité (assemblée générale)</i>	SSC3018

Tuesday, Mai 30 • Mardi 30 mai

7:00–8:00	Research Committee • <i>Comité de la recherche</i>	Michael's Dining Room, SH3340
	Accreditation Committee • <i>Comité sur l'accréditation</i>	Michael's Dining Room, SH3340
	Professional Development Committee • <i>Comité sur le perfectionnement professionnel</i>	Michael's Dining Room, SH3340
12:30–13:30	NPCDS • <i>PNSDC</i>	Michael's Dining Room, SH3340
	Bilingualism Committee • <i>Comité du bilinguisme</i>	Michael's Dining Room, SH3340
	Committee on Women in Statistics • <i>Comité sur les femmes en statistique</i>	Michael's Dining Room, SH3340
	Statistical Education Committee • <i>Comité d'éducation en statistique</i>	Michael's Dining Room, SH3340
16:45–18:15	SSC General Meeting • <i>Assemblée générale de la SSC</i>	SSC2050

Wednesday, May 31 • Mercredi 31 mai

7:00–8:00	Biostatistics Section (Section Officers) • <i>Groupe de bio-statistique (directeurs du groupe)</i>	Michael's Garden, SH3320
	Business and Industrial Statistics Section (Section Officers) • <i>Groupe de statistique industrielle et de gestion (directeurs du groupe)</i>	Michael's Garden, SH3320
	Survey Methods Section (Section Officers) • <i>Groupe des méthodes d'enquête (directeurs du groupe)</i>	Michael's Garden, SH3320
	Probability Section (Section Officers) • <i>Groupe de probabilité (directeurs du groupe)</i>	Michael's Garden, SH3320
12:30–13:30	Liaison	Michael's Garden, SH3320
	Program Committee • <i>Comité du programme</i>	Michael's Garden, SH3320
17:30–19:30	Board of Directors • <i>Conseil d'administration</i>	Michael's Garden, SH3320
19:30	Executive Committee • <i>Comité exécutif</i>	Michael's Garden, SH3320

Alphabetical Listing • Liste alphabétique**Accreditation Appeals Committee • *Comité des appels d'accréditation***

Monday, May 28 • lundi 28 mai, 12:30–13:30, Michael's Dining Room, SH3340

Accreditation Committee • *Comité sur l'accréditation*

Tuesday, May 30 • mardi 30 mai, 7:00–8:00, Michael's Dining Room, SH3340

Accredited Members General Meeting (open) • Assemblée générale des membres accrédités (public)

Monday, May 29 • lundi 29 mai, 13:30–15:00, SSC3018

Bilingualism Committee • Comité du bilinguisme

Tuesday, May 30 • mardi 30 mai, 12:30–13:30, Michael's Dining Room, SH3340

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

Monday, May 29 • lundi 29 mai, 17:00–18:00, SSC2036

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

Monday, May 29 • lundi 29 mai, 7:00–8:00, Michael's Garden, SH3320

Wednesday, May 31 • mercredi 31 mai, 7:00–8:00, Michael's Garden, SH3320

Board of Directors • Conseil d'administration

Sunday, May 28 • dimanche 28 mai, 12:00–16:00, Windermere Manor, Room A

Wednesday May 31 • mercredi 31 mai, 17:30–19:30, Michael's Garden, SH3320

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

Monday, May 29 • lundi 29 mai, 17:00–18:00, SSC3026

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

Monday, May 29 • lundi 29 mai, 7:00–8:00, Michael's Garden, SH3320

Wednesday, May 31 • mercredi 31 mai, 7:00–8:00, Michael's Garden, SH3320

CJS Editorial Board • Comité de rédaction de la RCS

Monday, May 28 • lundi 28 mai, 12:30–13:30, Michael's Garden, SH3320

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

Tuesday, May 30 • mardi 30 mai, 12:30–13:30, Michael's Dining Room, SH3340

Executive Committee • Comité exécutif

Saturday May 27 • samedi 27 mai, 18:00–22:00, Windermere Manor, North Meeting Room

Wednesday May 31 • mercredi 31 mai, 19:30, Michael's Garden, SH3320

Finance Committee • Comité des finances

Sunday, May 28 • dimanche 28 mai, 9:00–11:00, Windermere Manor, Room A

Liaison

Wednesday, May 31 • mercredi 31 mai, 12:30–13:30, Michael's Garden, SH3320

NPCDS • PNSDC

Tuesday, May 30 • mardi 30 mai, 12:30–13:30, Michael's Dining Room, SH3340

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

Monday, May 29 • lundi 29 mai, 17:00–18:00, SSC3018

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

Monday, May 29 • lundi 29 mai, 7:00–8:00, Michael's Garden, SH3320

Wednesday, May 31 • mercredi 31 mai, 7:00–8:00, Michael's Garden, SH3320

Professional Development Committee • Comité sur le perfectionnement professionnel

Tuesday May 30 • mardi 30 mai, 7:00–8:00, Michael's Dining Room, SH3340

Program Committee • Comité du programme

Wednesday May 31 • mercredi 31 mai, 12:30–13:30, Michael's Garden, SH3320

Public Relations Committee • Comité des relations public

Monday, May 29 • lundi 29 mai, 12:30–13:30, Michael's Dining Room, SH3340

Publications Committee • Comité des publications

Sunday, May 28 • dimanche 28 mai, 11:00–12:00, Windermere Manor, Room A

Research Committee • Comité de la recherche

Tuesday, May 30 • mardi 30 mai, 7:00–8:00, Michael's Dining Room, SH3340

Southern Ontario Regional Association (General Meeting) • Association régionale du sud de l'Ontario (assemblée générale)

Monday, May 29 • lundi 29 mai, 12:30–13:30, Michael's Dining Room, SH3340

SSC General Meeting • Assemblée générale de la SSC

Tuesday, May 30 • mardi 30 mai, 16:45–18:45, SSC2050

Statistical Education Committee • Comité d'éducation en statistique

Tuesday, May 30 • mardi 30 mai, 12:30–13:30, Michael's Dining Room, SH3340

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

Monday, May 29 • lundi 29 mai, 17:00–18:00, SSC2024

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

Monday, May 29 • lundi 29 mai, 7:00–8:00, Michael's Garden, SH3320

Wednesday, May 31 • mercredi 31 mai, 7:00–8:00, Michael's Garden, SH3320

List of Sessions • Liste des sessions

Sunday, May 28 • Dimanche 28 mai

19:00–21:00

SH Session 0 Contributed Poster Session
Great Hall *Présentation d'affiches*

Monday, May 29 • Lundi 29 mai

8:30–8:45

SSC2050 Session 01A Inaugural Session
Séance inaugurale

8:45–10:00

SSC2050 Session 01B SSC Presidential Invited Address
Allocution de l'invité du président de la SSC

10:30–12:00

SH3345 Session 02A Current Issues in the Analysis of Longitudinal Data
Problèmes courants en analyse de données longitudinales

SSC2050 Session 02B The Statistics of Magnetic Resonance Imaging
Statistique pour l'imagerie par résonance magnétique

SSC2036 Session 02C Social Statistics
Statistique sociale

SH3315 Session 02D Parallel Computing in Statistics
Calcul parallèle en statistique

SSC2024 Session 02E The Challenges of Being an Environmental Statistician
Défis pour le statisticien environnemental

SSC3026 Session 02F Case Study I: Record Linkage and Case Study II: Obstructive Sleep Apnea
Étude de cas I : Couplage d'enregistrements et Étude de cas II : L'apnée obstructive du sommeil

SSC3006 Session 02G Applications of Statistics I
Applications de la statistique I

SSC3014 Session 02H Business and Industrial Statistics
Statistique industrielle et de gestion

SSC3018 Session 02I Design and Analysis of Clinical Trials
Planification et analyse d'essais cliniques

13:30–14:30

SSC2050 Session 03A Canadian Journal of Statistics Award Address
Allocution du récipiendaire du Prix de la Revue canadienne de statistique

13:30–15:00

SH3345 Session 03B Statistical Issues in Designing and Monitoring Clinical Trials
Problèmes statistiques dans la planification et la surveillance d'essais cliniques

SH3315 Session 03C Change-point Detection
Détection de point de changement

SSC2036 Session 03D Design and Analysis of Experiments for Complex Computer Models
Planification et analyse d'expériences pour des modèles d'ordinateur complexes

15:30–16:30

SSC2050 Session 04A Record Linkage and Counterterrorism
Couplage d'enregistrements et la lutte au terrorisme

15:30–17:00

SSC2036 Session 04B Statistical Applications in Forestry
Applications statistiques en foresterie

SH3345 Session 04C Statistics in Public Health
Statistique en santé publique

SH3315 Session 04D Statistical Applications in Marketing
Applications statistiques en marketing

SSC2024 Session 04E Limit Theorems with Applications in Statistics
Théorèmes limites avec applications en statistique

SSC3026 Session 04F Bayesian Statistics
Statistique bayésienne

SSC3006 Session 04G Methodology for Time-to-Event Data
Méthodologie pour des données de temps d'attente

SSC3014 Session 04H Nonparametric Methods
Méthodes non paramétriques

Tuesday, May 30 • Mardi 30 mai**8:30–9:45**

SSC2050 Session 05A Gold Medal Address
Allocution de la médaille d'or

10:15–11:45

SSC3018 Session 06A Some Current Developments in the Analysis of Complex Survey Data
Développements récents en analyse de données d'enquêtes complexes

SSC2024 Session 06B Regression Models: Recent Advances and Applications
Modèles de régression : Développements récents et applications

SSC2050 Session 06C Bioinformatics
Bioinformatique

- SH3345 Session 06D Meta-analysis of Randomized Controlled Trials
Méta-analyse d'essais contrôlés randomisés
- UC85 Session 06E Advances in Markov Processes
Développements dans les processus de Markov
- SSC3006 Session 06F Classical Statistics
Statistique classique
- SSC3010 Session 06G Biostatistics in Action
Biostatistique en action
- SSC3014 Session 06H Censored Data
Données censurées

13:15–14:15

- SH3345 Session 07A Isobel Loutit Invited Address on Business and Industrial Statistics
Allocution Isobel Loutit sur la statistique industrielle et de gestion

13:15–14:45

- SSC2024 Session 07B Spatial/Temporal Modelling for Marine Ecological Systems
Modèles spatiaux/temporels dans les systèmes écologiques marins
- SSC2050 Session 07C Statistical Methods in Computational Genetics
Méthodes statistiques en calculs génétiques
- SSC3006 Session 07D Probability in High Dimensions and its Applications
Probabilités en dimensions élevées et leurs applications
- SSC3018 Session 07E Innovative Teaching in Probability and Statistics
Enseignement innovateur en probabilités et statistique
- UC85 Session 07F Joint Modelling of Longitudinal and Survival Data and Applications in Cancer Patients
Modélisation conjointe de données longitudinales et de survie et applications aux patients atteints de cancer
- SSC3014 Session 07G Tail Behaviour and Elliptically Contoured Distributions
Comportement des ailes et distributions à contours elliptiques
- UC222 Session 07H Dependence, Scores and Applications
Dépendance, cotes et applications
- SSC3010 Session 07I Business Surveys
Enquêtes sur les entreprises

15:30–17:00

- SSC2050 Session 08A Data Mining Complex Structures
Structures complexes de forage de données
- SH3345 Session 08B Genomic Assessment Implications of Biologic Heterogeneity
Implications sur l'évaluation génomique de la variabilité biologique
- SSC2024 Session 08C Probability and Finance
Probabilités et finance
- SSC3018 Session 08D The History of Statistics and Probability
L'histoire de la statistique et des probabilités
- UC85 Session 08E Estimating Function Methodology
Méthodologie d'estimation de fonctions

- SSC3006 Session 08F Computational Methods in Statistics
Méthodes computationnelles en statistique
- SSC3010 Session 08G Methodology for Complex Data Structures
Méthodologie pour des données à structure complexe
- SSC3014 Session 08H Survey Analysis and Operations
Analyse d'enquêtes et opérations

Wednesday, May 31 • Mercredi 31 mai**8:30–9:30**

- SSC2050 Session 09A Pierre Robillard Award Address
Allocution du récipiendaire du prix Pierre-Robillard

8:30–10:00

- SSC2024 Session 09B Applied Probability
Probabilités appliquées
- SSC2036 Session 09C Analysis of Longitudinal Complex Survey Data Using Marginal Modelling and GEE: Theory and Practice
Analyse de données longitudinales d'enquêtes complexes utilisant la modélisation marginale et les GEE : théorie et pratique

10:30–11:30

- SSC2050 Session 10A Special Invited Session of the Biostatistics Section
Allocution de l'invité d'honneur du Groupe de biostatistique

10:30–12:00

- SSC2036 Session 10B Recent Advances in SPC and Data Driven Statistics
Développements récents en CSP et statistiques guidées par les données
- SH3345 Session 10C Analyzing Data from International Surveys
Analyse de données d'enquêtes internationales
- SSC3018 Session 10D Statistical Analysis of Non-precise Data
Analyse statistique des données imprécises
- SSC2024 Session 10E Case Study II: Obstructive Sleep Apnea
Étude de cas II : L'apnée obstructive du sommeil
- SSC3026 Session 10F Time Series
Séries chronologiques
- SSC3006 Session 10G Probability and Mathematical Statistics
Probabilité et statistique mathématique
- SSC3014 Session 10H Statistical Genetics and Genomics
Génétique statistique et génomique
- SSC Session 10I Posters for Case Study II
2nd floor hallway *Affiches pour l'étude de cas II*

12:30–13:15

- SSC2036 Support of Mathematics and Statistics by NSERC
Appui des mathématiques et de la statistique par le CRSNG

13:00–14:30

SSC2024 Session 11B Risk Theory I
Théorie du risque I

13:30–14:30

SSC2050 Session 11A Address of the winner of the Centre de recherches mathématiques/Statistical Society of Canada Prize
Allocution du lauréat du Prix du Centre de recherches mathématiques/Société statistique du Canada

15:00–16:30

- SSC2050 Session 12A Industrial Experimental Design
Planification d'expériences industrielles
- SSC2036 Session 12B Latent Variable and Multilevel Modelling Using Survey Data
Variables latentes et modélisation multiniveaux pour des données d'enquêtes
- SSC2024 Session 12C Risk Theory II
Théorie du risque II
- SH3345 Session 12D Geometry, Topology and Statistics
Géométrie, topologie et statistique
- SSC3018 Session 12E Applications of Statistics II
Applications de la statistique II
- SSC3026 Session 12F Multivariate Time Series
Séries chronologiques multivariées
- SSC3006 Session 12G Observational Studies: Methods and Applications
Méthodes et applications pour les études observationnelles
- SSC3014 Session 12H Modelling and Analysis in Sample Surveys
Modélisation et analyse pour l'échantillonnage

Program • Programme

Dimanche 28 mai • Sunday, May 28

19:00–21:00 Session 0

SH - Great Hall

Contributed Poster Session

Présentation d'affiches

Contributed Paper Session • Séance de communications libres

19:00 Paul SUTCLIFFE & André DABROWSKI, *University of Ottawa*, Kathryn WILLIAMS, Bassam HAIDAR & Yves MARCEL, *Ottawa Heart Institute*

Use of Contrasts in the Analysis of the Affymetrix Hypoalpha Dataset • L'emploi des contrastes dans l'analyse des données Affymetrix Hypolpha [MS-1]

19:00 Yu-Ling TSAI, *University of Western Ontario*, Debbie DUPUIS, *HEC Montréal*, Duncan MURDOCH, *University of Western Ontario*

A Robust Test for Asymptotic Independence in Multivariate Extreme Values • Un test robuste pour l'indépendance asymptotique de valeurs extrêmes multivariées [MS-2]

19:00 Richard HARSHMAN & Margaret E. LUNDY, *University of Western Ontario*

Valid p -values for Stepwise Regression and Other Post-hoc Model Selection Methods • Valeurs- p valides pour la régression pas-à-pas et autres méthodes de sélection post-hoc d'un modèle [MS-3]

19:00 Depeng JIANG & Debra PEPLER, *York University*

Screening for Risk Status and Statistical Power in Randomized Intervention Research • Criblage pour le statut de risque et puissance statistique dans la recherche d'intervention randomisée [MS-4]

19:00 Wai Lok Wallace HUI & Yulia R. GEL, *University of Waterloo*, Joseph GASTWIRTH, *George Washington University*, Weiwen MIAO, *Macalester College*

A New R Package for Biostatistics, Public Policy and Law • Une nouvelle logithèque R pour la biostatistique, les politiques d'intérêt public et le domaine judiciaire [MS-5]

19:00 Julia BRETTSCHEIDER, *Queen's University*, François COLLIN, *Affymetrix Inc.*, Ben BOLSTAD, *University of California at Berkeley*, Terence SPEED, *University of California at Berkeley & Walter and Eliza Hall Institute, Australia*

Quality Assessment for Microarray Gene Expression Data • Évaluation de la qualité de données d'expression génétique mesurées par micro puces [MS-6]

SH = Somerville House

SSC = Social Science Centre

UC = University College

19:00 Richard HARSHMAN, *University of Western Ontario*

Generalizing Factor Analysis and Canonical Correlation to Three-way Arrays Increases their Ability to Disentangle Information • La généralisation de l'analyse factorielle et de corrélation canonique aux tableaux à trois niveaux augmente leur capacité de démêler l'information [MS-7]

19:00 Lutong ZHOU & Willard BRAUN, *University of Western Ontario*

Some Ideas About Forestry Data Visualization • Quelques idées au sujet de la visualisation de données reliées à la foresterie [MS-8]

19:00 Yongtao ZHU & Alexander DE LEON, *University of Calgary*

ANOVA Extensions for Mixed Data • Extension de l'ANOVA aux données mixtes [MS-9]

19:00 Zilin WANG, *Wilfrid Laurier University*, Mary THOMPSON & Zhichang JIANG, *University of Waterloo*

A Resampling Approach to Multilevel Models • Une approche de rééchantillonnage pour les modèles à plusieurs niveaux [MS-10]

Lundi 29 mai • Monday, May 29

08:30–08:45 Session 01A

SSC2050

Inaugural Session

Séance inaugurale

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: David Binder

08:45–10:00 Session 01B

SSC2050

SSC Presidential Invited Address

Allocution de l'invité du président de la SSC

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: David BINDER

08:45 Ivars PETERSON, *Science News*

MythMath: From Google to Wikis and Blogs • MythMath : De Google à wikis et blogues [MS-11]

10:00–10:30 Coffee Break • Pause-café

10:30–12:00 Session 02A

SH3345

Current Issues in the Analysis of Longitudinal Data

Problèmes courants en analyse de données longitudinales

Invited Paper Session • Présentations sur invitation

National Program on Complex Data Structures • Programme national sur les structures de données complexes

Organizer and Chair • Responsable et président: Peter SONG *University of Waterloo*10:30 Jerry LAWLESS & Richard COOK, *University of Waterloo*

Estimation of Marginal Characteristics for Event History Processes • Évaluation des caractéristiques marginales pour des processus sur l'historique des événements [MS-12]

10:52 Jeremy TAYLOR & Wen YE, *University of Michigan*, Xihong LIN, *Harvard University*

Recent Developments in Joint Models for Longitudinal and Survival Data. • Développements récents de modèles conjoints pour des données longitudinales et de survie [MS-13]

11:14 Paul GUSTAFSON, *University of British Columbia*

On Measurement Error Models and Complex Data Structures • Sur les modèles d'erreur de mesure et les structures de données complexes [MS-14]

11:36 Grace YI, *University of Waterloo*

Some Methods on Longitudinal Data Analysis • Quelques méthodes sur l'analyse de données longitudinales [MS-15]

10:30–12:00 Session 02B**SSC2050**The Statistics of Magnetic Resonance Imaging
Statistique pour l'imagerie par résonance magnétique

Invited Paper Session • Présentations sur invitation

Bernoulli Society • Société Bernoulli

Organizer and Chair • Responsable et président: Fred HOPPE *McMaster University*10:30 Jonathan TAYLOR, *Stanford University*, Keith WORSLEY, *McGill University*, Robert ADLER, *Technion*Deformation Based Morphometry, Roy's Maximum Root and Recent Advances in Random Fields •
Morphométrie basée sur la déformation, la racine maximale de Roy, et des avancées récentes en
champs aléatoires [MS-16]11:15 Keith WORSLEY, *McGill University*, Jonathan TAYLOR, *Stanford University*Our entry in the Functional Imaging Analysis Contest • Notre participation dans le concours
d'analyse d'imagerie par résonance magnétique fonctionnelle [MS-17]**10:30–12:00 Session 02C****SSC2036**Social Statistics
Statistique sociale

Invited Paper Session • Présentations sur invitation

Canadian Initiative on Social Statistics • Initiative canadienne pour la statistique sociale

Organizer and Chair • Responsable et président: Gustave GOLDMAN & Georgia ROBERTS *Statistics
Canada*10:30 Rajulton FERNANDO, *University of Western Ontario*The Direct and Indirect Effects of Causal Systems with Longitudinal Data • Causalité directe et
indirecte avec les données longitudinales [MS-18]11:00 Lisa STROHSCHNEIN, *University of Alberta*Tracking the Mental Health Consequences of Divorce: A Growth Curve Model Approach • Obser-
vation des conséquences du divorce sur la santé mentale : une approche utilisant un modèle avec
courbe de croissance [MS-19]11:30 Bradley CORBETT, *Statistics Canada*Identifying Risk and Protective Factors Using Longitudinal Population Based Survey Data: Im-
proving our Knowledge of Smoking Behaviour Development • Identification du risque et des fac-
teurs protecteurs en utilisant des données longitudinales d'une population : amélioration de notre
connaissance du développement du comportement des fumeurs [MS-20]**10:30–12:00 Session 02D****SH3315**Parallel Computing in Statistics
Calcul parallèle en statistique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Hugh CHIPMAN *Acadia University*

10:30 Robert GRAMACY, *University of Cambridge*, Herbert K. H. LEE, *University of California at Santa Cruz*

Parallelism in Interface, Design, and Nonstationary Surrogate Modeling of Supercomputer Experiments • Parallélisme dans l'interface, conception, et modélisation avec suppléant non stationnaire des expériences sur super-ordinateurs [MS-21]

11:00 Theodoro KOULIS, *McGill University*

Parallel Simulations for Sea Ice Modelling • Simulations parallèles pour la modélisation de la glace de mer [MS-22]

11:30 Mark REESOR, TYSON WHITEHEAD & Matt DAVISON, *University of Western Ontario*

A Bias Reduction Technique for Monte Carlo Pricing of Early Exercise Options • Une technique de réduction de biais pour des estimateurs de prix de type Monte-Carlo pour les premières options calculées [MS-23]

10:30–12:00 Session 02E

SSC2024

The Challenges of Being an Environmental Statistician

Défis pour le statisticien environnemental

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Grace CHIU *University of Waterloo*

10:30 Michael DOWD *Dalhousie University*

Statistical Data Assimilation for Environmental Prediction • Assimilation de données statistiques à des fins de prédictions environnementales [MS-24]

11:00 Tanja SREBOTNJAK, *Yale University*

Some Challenges in Measuring and Communicating National Environmental Performance • Quelques défis associés à la mesure et à la communication de la performance environnementale nationale [MS-25]

11:30 Peter GUTTORP, *University of Washington*

Decisions, Uncertainty, and the Law: Statistical Considerations of Environmental Risk Management. • Les décisions, l'incertitude, et la loi : considérations statistiques en gestion de risque écologique [MS-26]

10:30–12:00 Session 02F

SSC3026

Case Study I: Record Linkage and Case Study II: Obstructive Sleep Apnea

Étude de cas I : Couplage d'enregistrements et Étude de cas II : L'apnée obstructive du sommeil

Case Study • Étude de cas

Organizer and Chair • Responsable et présidente: Peggy T. NG *York University* & Alison L. GIBBS *University of Toronto*

10:30 Peggy NG, *York University*

Case Study I: Record Linkage: Maintaining a Registry • Étude de cas I : Couplage d'enregistrements : maintenance d'un registre [MS-27]

10:50 Jin ZHANG & Yunna SONG, *McMaster University*

11:10 Tao SUN, Wenjie QIAO, Qiang PU, Tianshu MA & Pengfui GUO, *York University*

11:30 Alison GIBBS, *University of Toronto*

Case Study II: Obstructive Sleep Apnea • Étude de cas II : L'apnée obstructive du sommeil [MS-28]

11:40 Rachel CHU, Shuyu FAN, Kimberly FERNANDES, Shahadut HOSSAIN & Jesse RAFFA, *University of British Columbia*

11:50 Shahedul Ahsan KHAN, Enayetur RAHEEM, Fei YUAN & Michelle ZHOU, *University of Waterloo*

10:30–12:00 Session 02G

SSC3006

Applications of Statistics I

Applications de la statistique I

Contributed Paper Session • Séance de communications libres

Chair • Président: Bruce SMITH *Dalhousie University*

10:30 Nathaniel NEWLANDS, *Agriculture and Agri-Food Canada*, Harvey HILL, *Climate Impacts and Adaptation*, PFRA-Agriculture and Agri-Food Canada, Allan HOWARD, *National Agroclimate Information Service*, *Agriculture and Agri-Food Canada*

Developing a Daily Agro-climate Interpolation Grid for Canada: Data Verification and Model Validation • Développement d'une grille d'interpolation quotidienne pour l'agro-climat du Canada : vérification de données et validation des modèles [MS-29]

10:45 Zi JIN & Zheng ZHENG, *University of Toronto*

Statistical Inferences Problems in High Energy Physics • Problèmes statistiques d'inférences en physique de forte énergie [MS-30]

11:00 Jennifer ASIMIT & W. John BRAUN, *University of Western Ontario*, William SIMPSON, *Defence Research and Development Canada*

Point Process Modeling of Choice Reaction Time Data • Modélisation de données du temps de réaction à choix multiples par un processus ponctuel [MS-31]

11:15 Katherine DAVIES, *University of Western Ontario*

Statistical Behaviour of a Spatial Renewal Process • Comportement statistique d'un processus de renouvellement spatial [MS-32]

11:30 Douglas WOOLFORD, Reg KULPERGER & W. John BRAUN, *University of Western Ontario*

Modelling Lightning and Lightning Caused Forest Fires • Modélisation de la foudre et des feux de forêt qu'elle cause [MS-33]

11:45 B. M. Golam KIBRIA, *Florida International University*

Zero-truncated and Zero-inflated Regression Models to Predict the ROR Crash Data • Modèles de régression zéro-tronqués et zéro-majorés pour la prédiction de données sur les accidents ROR [MS-34]

10:30–12:00 Session 02H**SSC3014**
Business and Industrial Statistics
Statistique industrielle et de gestion

Contributed Paper Session • Séance de communications libres

 Chair • Président: Fernando CAMACHO *Damos Inc.*

 10:30 Abdulkadir HUSSEIN, *University of Windsor*

Sequential Monitoring of Process Capability Indices • Supervision séquentielle des indices de capacité d'un processus [MS-35]

 10:45 Razi GHORI & S. Ejaz AHMED, *University of Windsor*

A Robust Estimation of Process Capability Index • Une estimation robuste de l'indice de capacité d'un processus [MS-36]

 11:00 Xu Sunny WANG, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, William J. WELCH, *University of British Columbia*

Special Mixture Discriminant Model for Drug Discovery Data • Modèle de mélange discriminant en recherche de médicaments [MS-37]

 11:15 Shakhwat HOSSAIN & S. Ejaz AHMED, *University of Windsor*

Risk-reducing Stein-like Estimation in a Lifetime Censored Regression Model • Estimation de Stein réduisant le risque dans un modèle de régression pour le temps de survie censuré [MS-38]

 11:30 Crystal LINKLETTER, Chunfang LIN, Wilson LU, Pritam RANJAN, Derek BINGHAM & Richard LOCKHART, *Simon Fraser University*, Tom LOUGHIN, *Kansas State University*

Compliance Testing for Random Effects Models with Joint Acceptance Criterion • Tests de conformité pour modèles à effets aléatoires avec critère d'acceptation conjointe [MS-39]

 11:45 Krishna SAHA, *Central Connecticut State University*, Maxwell KING, *Monash University*

An Alternative Wald Type Test for Two Linear Restrictions with Applications to Non-linear Models • Un test alternatif de type Wald pour deux restrictions linéaires avec applications aux modèles non linéaires [MS-40]

10:30–12:00 Session 02I**SSC3018**
Design and Analysis of Clinical Trials
Planification et analyse d'essais cliniques

Contributed Paper Session • Séance de communications libres

 Chair • Présidente: Judy-Anne CHAPMAN *Queen's University*

 10:30 Yuanyuan LIANG & Keumhee Chough CARRIÈRE, *University of Alberta*

New Allocation Rule for Repeated Measurement Designs • Nouvelles règles de répartition pour plans à mesures répétées [MS-41]

 10:45 Yanqing YI & Xikui WANG, *University of Manitoba*

Testing Treatment Effects of Binary Response Adaptive Designs • Test sur les effets de traitement des plans adaptatifs de réponse binaire [MS-42]

11:00 Gregory POND & Lillian SIU, *Princess Margaret Hospital/University of Toronto*

Improving Efficiency in Clinical Trials Using a Markov Chain Transition Rate Model with Conditional Power • Augmentation de l'efficacité des essais cliniques par l'utilisation de chaînes de Markov avec puissance conditionnelle [MS-43]

11:15 Lisa LIX & Kathleen DEERING, *University of Manitoba*

A Robust Two-Group Step-Down Test for Multiple Outcomes • Un test descendant robuste à deux groupes pour des résultats multiples [MS-44]

11:30 Yang ZHAO, *University of Regina*

A Maximum Likelihood Estimator for the Analysis of Randomized Trials with Treatment Non-compliance and Subsequent Nonresponse • Un estimateur du maximum de vraisemblance pour l'analyse d'essais cliniques aléatoires en présence de non-observance et de non-réponse [MS-45]

11:45 Eleanor PULLENAYEGUM, *University of Waterloo*, Andrew WILLAN, *University of Toronto*

Semi-parametric Models and Efficiency in Mean Cost Estimation • Modèles semi-paramétriques et efficacité dans l'évaluation du coût moyen [MS-46]

12:00–13:30 Lunch • Dîner

13:30–14:30 Session 03A

SSC2050

Canadian Journal of Statistics Award Address

Allocution du récipiendaire du Prix de la Revue canadienne de statistique

Special Session • Conférence spéciale

Chair • Présidente: Mary THOMPSON *University of Waterloo*

13:30–15:00 Session 03B

SH3345

Statistical Issues in Designing and Monitoring Clinical Trials

Problèmes statistiques dans la planification et la surveillance d'essais cliniques

Invited Paper Session • Présentations sur invitation

Biostatistics Section • Groupe de biostatistique

Organizer and Chair • Responsable et présidente: Joan HU *Simon Fraser University*

13:30 James NEATON, *University of Minnesota*

Challenges in the Design, Implementation, and Monitoring of HIV Treatment Trials • Défis dans la conception, l'implantation, et la surveillance d'essais de traitement du VIH [MS-47]

14:15 Christopher JENNISON, *University of Bath, UK*

Adaptivity in Group Sequential Designs • L'adaptabilité dans les plans séquentiels par groupes [MS-48]

13:30–15:00 Session 03C**SH3315**

Change-point Detection

Détection de point de changement

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Edit GOMBAY *University of Alberta*13:30 Miklós CSÖRGŐ & Barbara SZYSZKOWICZ, *Carleton University*, Qiying WANG, *University of Sydney*

The Probable Error of a Mean and That of a Change in a Mean • L'erreur probable d'une moyenne et celle d'un changement de moyenne [MS-49]

14:00 Murray BURKE, *University of Calgary*

Change-point Detection in a General Regression Model • Détection de changement structurel dans un modèle de régression général [MS-50]

14:30 Edit GOMBAY & Daniel SERBAN, *University of Alberta*

Monitoring Parameter Change in Time Series Models • Surveillance du changement de paramètre dans une série chronologique [MS-51]

13:30–15:00 Session 03D**SSC2036**

Design and Analysis of Experiments for Complex Computer Models

Planification et analyse d'expériences pour des modèles d'ordinateur complexes

Invited Paper Session • Présentations sur invitation

National Program on Complex Data Structures • Programme national sur les structures de données complexes

Organizer and Chair • Responsable et président: Derek BINGHAM *Simon Fraser University*13:30 Bela NAGY, Jason LOEPPKY & William J. WELCH, *University of British Columbia*

Uncertainty in Kriging Predictions with and Without Random Error • Incertitude dans la prévision par krigeage avec et sans erreur aléatoire [MS-52]

14:00 Marc KENNEDY, Clive ANDERSON, Anthony O'HAGAN, Mark LOMAS & Ian WOODWARD, *University of Sheffield, UK*, Andreas HEINEMEYER, *University of York, UK*

Quantifying Uncertainty in the Biospheric Carbon Flux for England and Wales • Quantifier l'incertitude dans le flux du carbone biosphérique en Angleterre et au pays de Galles [MS-53]

14:30 Shane REESE, *Brigham Young University*, Derek BINGHAM, *Simon Fraser University*

Design of Experiments for Integrated Computer and Physical Experiments • Planification d'expériences intégrant l'expérimentation physique et informatique [MS-54]

15:00–15:30 Coffee Break • Pause-café

15:30–16:30 Session 04A**SSC2050**

Record Linkage and Counterterrorism

Couplage d'enregistrements et la lutte au terrorisme

Special Session • Conférence spéciale

Survey Methods Section • Groupe de méthodologie d'enquête

Organizer and Chair • Responsable et présidente: Patricia WHITRIDGE *Elections Canada*15:30 Michael LARSEN, *Iowa State University*

Record Linkage, Nondisclosure, Counterterrorism, and Statistics • Couplage d'enregistrements, non-divulgateion, contre-terrorisme et statistiques [MS-55]

15:30–17:00 Session 04B**SSC2036**

Statistical Applications in Forestry

Applications statistiques en foresterie

Invited Paper Session • Présentations sur invitation

National Program on Complex Data Structures • Programme national sur les structures de données complexes

Organizer and Chair • Responsable et président: John BRAUN *University of Western Ontario*15:30 Al TITHECOTT & Rob McALPINE, *Ontario Ministry of Natural Resources*

Forest Fire Management—Practical Problems of Uncertainty in Decision Support • Gestion des feux de forêts : problèmes pratiques d'incertitude dans le support décisionnel [MS-56]

16:00 Edward JOHNSON, *University of Calgary*

Statistics in Forest Ecology • La statistique en écologie forestière [MS-57]

16:30 Farouk NATHOO & Charmaine DEAN, *Simon Fraser University*

Mixture Models for Spatio-temporal Multi-state Processes • Modèles de mélange pour des processus spatio-temporels multi-état [MS-58]

15:30–17:00 Session 04C**SH3345**

Statistics in Public Health

Statistique en santé publique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Leilei ZENG *Simon Fraser University*15:30 Norman BRESLOW & Jon WELLNER, *University of Washington*

Weighted Likelihood for Semiparametric Models and Two Phase Stratified Samples, with Application to Cox Regression • Vraisemblance pondérée pour modèles semiparamétriques et échantillons stratifiés à deux niveaux, avec application à la régression de Cox [MS-59]

16:00 Thérèse STUKEL, *Institute for Clinical Evaluative Sciences*

Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive

Cardiac Management on AMI Survival • Analyse des études d'observation en présence d'un biais de sélection d'un traitement : effets de gestion cardiaque invasive sur la survie IMA [MS-60]

16:30 John NEUHAUS & Charles McCULLOCH, *University of California at San Francisco*

The Consequences of Misspecified Random Effects Distributions on Predictions of Random Effects • Les conséquences d'une mauvaise spécification de la distribution des effets aléatoires lors de la prédiction d'effets aléatoires [MS-61]

15:30–17:00 Session 04D

SH3315

Statistical Applications in Marketing Applications statistiques en marketing

Invited Paper Session • Présentations sur invitation

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

Organizer and Chair • Responsable et président: Fernando CAMACHO *Damos Inc.*

15:30 Alison BURNHAM, *Digital Cement*

Customer Value Segmentation—Informing Better Marketing Decisions • Segmentation de la valeur client – informer à propos de meilleures décisions de marketing [MS-62]

16:15 Anthony PERCACIO, *Municipal Property Assessment Corporation*

National Automated Valuation Models in Canada • Modèles nationaux d'évaluation automatisée au Canada [MS-63]

15:30–17:00 Session 04E

SSC2024

Limit Theorems with Applications in Statistics Théorèmes limites avec applications en statistique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Deli LI *Lakehead University*

15:30 Mei Ling HUANG & Ron KERMAN, *Brock University*

Hermite Wavelet Estimation of Probability Density • Estimation de la densité de probabilité à l'aide d'ondelettes de Hermite. [MS-64]

15:52 Uwe EINMAHL, *Vrije Universiteit Brussel, Belgium*

Some Recent Results on the Consistency of Kernel-Type Estimators • Résultats récents sur la convergence d'estimateurs à noyau [MS-65]

16:14 Tiefeng JIANG, *University of Minnesota*

How Many Entries of a Typical Orthogonal Matrix can be Approximated by Independent Normals? • Combien d'entrées d'une matrice orthogonale typique peuvent être approximées par des variables normales indépendantes? [MS-66]

16:36 Andrei VOLODIN, *University of Regina*, Andrew ROSALSKY, *University of Florida*

On the Golden ratio, Strong Law, and First Passage Problem • Sur le nombre d'or, la loi forte, et le problème du premier passage [MS-67]

15:30–17:00 Session 04F**SSC3026**

Bayesian Statistics

Statistique bayésienne

Contributed Paper Session • Séance de communications libres

Chair • Président: Bela NAGY *University of British Columbia*15:30 Mahmoud ZAREPOUR, *University of Ottawa*, Reza HABIBI, *Shiraz University*

A Quasi-Bayesian Change Point Detection with Random Exchangeable Weights • Détection quasi bayésienne de changement structurel avec poids aléatoires échangeables [MS-68]

15:45 David CAMPBELL, James O. RAMSAY, Giles HOOKER & Jiguo CAO, *McGill University*

Comparing Methods for Estimating Parameters in Dynamic Systems • Comparaison de méthodes pour estimer des paramètres dans des systèmes dynamiques [MS-69]

16:00 Sohee KANG & Michael ESCOBAR, *University of Toronto*

Semi-parametric Bayesian Logistic Regression • Régression logistique bayésienne semi-paramétrique [MS-70]

16:15 Hadas MOSHONOV & Mike EVANS, *University of Toronto*Checking for Prior-data Conflict with Hierarchically Specified Priors • Vérification de l'existence d'un conflit dans l'*a priori* avant les observations avec des lois *a priori* hiérarchiques [MS-71]16:30 Wanhua SU, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, Mu ZHU, *University of Waterloo*

Bayesian LAGO for Statistical Detection Problems • LAGO bayésienne pour des problèmes statistiques de détection [MS-72]

16:45 Sofia MOSESOVA, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*

Profile-monitoring Curve Data • Contrôle de profils pour des données fonctionnelles [MS-73]

15:30–17:00 Session 04G**SSC3006**

Methodology for Time-to-event Data

Méthodologie pour des données de temps d'attente

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Julie HORROCKS *University of Guelph*15:30 Xuewen LU & Jie SUN, *University of Calgary*

Efficient Semiparametric Estimation in a Class of Conditional Single-index Survival Models • Estimation semi paramétrique efficace dans une classe de modèles de survie conditionnels à indice simple [MS-74]

15:45 Zhaozhi FAN, *Memorial University of Newfoundland*

Marginal Hazards Model for Multivariate Failure Times with Auxiliary Covariates • Modèle des risques marginaux pour des temps de ruine multivariés avec des covariables auxiliaires [MS-75]

16:00 Yingwei PENG, *University of Michigan & Memorial University of Newfoundland*, Jeremy TAYLOR, *University of Michigan*, Binbing YU, *Information Management Services, Inc., USA*

A Marginal Regression Model for Multivariate Failure Time Data with a Surviving Fraction • Un modèle de régression marginal pour données multivariées de temps de défaillance avec une proportion qui survit [MS-76]

16:30 Dianliang DENG, *University of Regina*

On the Asymptotics of NPMLE of a Distribution Function Based on Bivariate Interval-censored Data • Propriétés asymptotiques de l'EVMNP d'une distribution basé sur des données bidimensionnelles censurées par intervalles [MS-77]

16:45 Abdul Sattar AL-KHALIDI, *Insurance Bureau of Canada*

Low-risk Dose Estimation for Environmental Toxic Agents—Weibull Hazard Model • Estimation de la dose à faible risque d'agents toxiques environnementaux – le modèle de Weibull [MS-78]

15:30–17:00 Session 04H

SSC3014

Nonparametric Methods

Méthodes non paramétriques

Contributed Paper Session • Séance de communications libres

Chair • Président: BRUNO RÉMILLARD *HEC Montréal*

15:30 Alberto CARABARIN & B. Gail IVANOFF, *University of Ottawa*

Two-dimensional Survival Analysis Under General Censoring • Analyse de survie bidimensionnelle sous censure générale [MS-79]

15:45 Mayer ALVO & Vladislav BRION, *University of Ottawa*

Nonparametric Tests of Hypotheses for Umbrella Alternatives • Tests d'hypothèse non paramétriques pour des alternatives de type parapluie [MS-80]

16:00 Xiaobin YUAN & Cheng CHENG, *St. Jude Children's Research Hospital*

Projection-directed Invariant Tests for the Multivariate Multi-sample Problem • Tests invariants par projection pour le problème multivarié à plusieurs échantillons [MS-81]

16:15 Jean-François PLANTE & James V. ZIDEK, *University of British Columbia*

Nonparametric Likelihood Weights • Poids non paramétriques pour la vraisemblance pondérée [MS-82]

16:30 Nathalie VANDAL, *Université Laval*

Comparison of Two Non Parametric Regression Methods with Application to a Mammary Density Study • Comparaison de deux méthodes de régression non paramétrique multivariées par une application à une étude portant sur la densité mammaire [MS-83]

16:45 G. ARTHUR MIHRAM, *Princeton, NJ*, DANIELLE MIHRAM, *University of Southern California*

Science, Mathematics, and Statistics • Science, mathématiques, et statistique [MS-84]

Mardi 30 mai • Tuesday, May 30

08:30–09:45 Session 05A

SSC2050

Gold Medal Address

Allocution de la médaille d'or

Special Session • Conférence spéciale

Organizer and Chair • Responsable et président: Jim RAMSAY *McGill University*08:30 David ANDREWS, *University of Toronto*

Statistical Analysis for Public Policy • L'analyse statistique au service de la politique publique [MS-85]

09:45–10:15 Coffee Break • Pause-café

10:15–11:45 Session 06A

SSC3018

Some Current Developments in the Analysis of Complex Survey Data

Développements récents en analyse de données d'enquêtes complexes

Invited Paper Session • Présentations sur invitation

MITACS and National Program on Complex Data Structures • MITACS et Programme national sur les structures de données complexes

Organizer and Chair • Responsable et président: Milorad KOVACEVIC *Statistics Canada* & Changbao WU *University of Waterloo*10:15 Ivan CARRILLO-GARCIA, *University of Waterloo*, Milorad KOVACEVIC, *Statistics Canada*, Changbao WU, *University of Waterloo*

Analysis of Longitudinal Survey Data with Missing Observations: An Application to the National Longitudinal Survey of Children and Youth (NLSCY) • Analyse de données longitudinales en présence d'observations manquantes : une application à l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) [MS-86]

10:45 Xiaojian XU, *University of Alberta*, Pierre LAVALLÉE, *Statistics Canada*

Treatments for the Problem of Identification of Links in Indirect Sampling • Traitements pour le problème de l'identification des liens en échantillonnage indirect [MS-87]

11:15 Chunfang (Devon) LIN & Randy SITTER, *Simon Fraser University*

Resampling Methods for Two-stage Sampling in Complex Surveys • Méthodes de ré-échantillonnage pour l'échantillonnage à deux phases dans les enquêtes à plans complexes [MS-88]

10:15–11:45 Session 06B**SSC2024**

Regression Models: Recent Advances and Applications

Modèles de régression : Développements récents et applications

Invited Paper Session • Présentations sur invitation

Canadian Caucus for Women in Statistics and Women in Statistics Committee of the SSC • Caucus canadien des femmes en statistique et Comité des femmes en statistique de la SSC

Organizer and Chair • Responsable et présidente: Rhonda ROSYCHUK *University of Alberta*10:15 Yun Hee CHOI, *Samuel Lunenfeld Research Institute at the Mount Sinai Hospital*, E. David MATTHEWS & Y. Yi YI, *University of Waterloo*

Modelling Dependent Bivariate Time-to-event Data Using Accelerated Life Regression • Modélisation des données dépendantes bidimensionnelles en utilisant une régression accélérée de la survie [MS-89]

11:00 Tulay KORU-SENGUL, *McMaster University*

Regression Models for Longitudinal Data and its Applications • Les modèles de régression en présence de données longitudinales et leurs applications [MS-90]

10:15–11:45 Session 06C**SSC2050**

Bioinformatics

Bioinformatique

Invited Paper Session • Présentations sur invitation

Canadian Applied and Industrial Mathematics Society • Société canadienne de mathématiques appliquées et industrielles

Organizer and Chair • Responsable et président: Daniel ASHLOCK *University of Guelph*10:15 Gary FOGEL, *Natural Selection, Inc.*

Applications of Computational Intelligence to Pattern Discovery in Biological Systems • Applications de l'intelligence informatique à la découverte de patrons reliés aux systèmes biologiques [MS-91]

10:45 Gabriel MORENO-HAGELSIEB, *Wilfrid Laurier University*, Sarath JANGA, *CCG-UNAM*

Operons and the Quality of Predictions of Functional Interactions Using Phylogenetic Profiles • Les opérons et la qualité des prévisions des interactions fonctionnelles en utilisant des profils phylogénétiques [MS-92]

11:15 Daniel ASHLOCK, *University of Guelph*

Multiclustering, Avoiding the Shape of Underlying Metrics • Éviter la forme des métriques sous-jacentes en classification [MS-93]

10:15–11:45 Session 06D**SH3345**

Meta-analysis of Randomized Controlled Trials

Méta-analyse d'essais contrôlés randomisés

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Lehana THABANE *McMaster University*10:15 Andrew WILLAN & Joseph BEYENE, *SickKids Research Institute/University of Toronto*

Combining Studies with Continuous and Binary Outcomes • Combinaison d'études avec résultats continus et binaires [MS-94]

10:45 Anna SAWKA, *University of Toronto*

Use of Bayesian Methodology in Meta-analysis • Utilisation de méthodologie bayésienne pour la méta-analyse [MS-95]

11:15 Keith O'ROURKE, *Ottawa Health Research Institute*

A Generalized Meta-analysis Model: The Combination of Observations and Studies • Un modèle généralisé de méta-analyse : la combinaison des observations et des études [MS-96]

10:15–11:45 Session 06E**UC85**

Advances in Markov Processes

Développements dans les processus de Markov

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Gail IVANOFF *University of Ottawa*10:15 David McDONALD, *University of Ottawa*, Charles BORDENAVE & Alexandre PROUTIERE, *École Normale Supérieure*

A Mean Field Analysis of Decentralized Random Access Protocols • Analyse par une approximation champ moyen d'un protocole d'accès décentralisé [MS-97]

10:45 Byron SCHMULAND, *University of Alberta*

Reversibility for diffusions via Quasi-invariance • La réversibilité pour un processus de diffusion par le biais de la quasi-invariance [MS-98]

11:15 Donald DAWSON, *Carleton University*

Catalytic Branching Networks • Réseaux avec branchement catalytique [MS-99]

10:15–11:45 Session 06F**SSC3006**

Classical Statistics

Statistique classique

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Yang ZHAO *University of Regina*10:15 Denis LAROCQUE, *HEC Montréal*, Jaakko NEVALAINEN & Hannu OJA, *University of Tampere*

A Weighted Multivariate Sign Test for Cluster Correlated Data • Un test pondéré des signes multivarié pour des données corrélées en grappes [MS-100]

10:30 Ye SUN & Augustine WONG, *York University*

Inference of the Normal Correlation Coefficient • Inférence à propos du coefficient de corrélation de variables normales [MS-101]

10:45 Hyung-Tae HA & Serge PROVOST, *University of Western Ontario*

A Moment-based Percentile Approximation Methodology • Une méthodologie d'approximation des centiles basée sur les moments [MS-102]

11:00 Chu-In LEE, *Memorial University of Newfoundland*, Jianan PENG, *Acadia University*, Lin LIU, *University of California at San Diego*

Statistical Inference for the Difference Between the Best Treatment Mean and a Control Mean • Inférence statistique sur la différence entre la moyenne du groupe avec le meilleur traitement et celle du groupe contrôle [MS-103]

11:15 Abdul KHAN, *Aligarh Muslim University, India*

Characterization of Distributions through Conditional Expectation of Functions of Generalized Order Statistics • Caractérisation de distributions par le biais d'espérances conditionnelles de fonctions de statistiques d'ordre généralisées [MS-104]

10:15–11:45 Session 06G

SSC3010

Biostatistics in Action

Biostatistique en action

Contributed Paper Session • Séance de communications libres

Chair • Président: Peter MACDONALD *McMaster University*

10:15 Kenneth McRAE & H.W. (Bud) PLATT, *Agriculture & Agri-Food Canada*

An Index for Cultivar Resistance Based on Disease Progress Curves • Un indice pour la résistance du cultivar basé sur des courbes de progrès de maladie [MS-105]

10:30 Liying ZHANG, Andrew LOBLAW & Laurence KLOTZ, *Sunnybrook & Women's College Health Sciences Centre, University of Toronto*

Modeling Prostate Specific Antigen (PSA) Kinetics in Prostate Cancer Patients during Surveillance • Modélisation de la cinétique de l'antigène spécifique de prostate (PSA) chez des patients sous surveillance souffrant de cancer de prostate [MS-106]

10:45 Liqin XU & Marek SMIEJA, *McMaster University*, Lehana THABANE, *McMaster University, Centre for Evaluation of Medicines, St Joseph's Healthcare*

Modeling Lung Disease and its Progression in a Chronic Obstructive Pulmonary Disease Cohort • Modélisation d'une maladie pulmonaire et sa progression au sein d'une cohorte obstructive chronique [MS-107]

11:00 Ji CHENG, Jack SUN & Richard WHITLOCK, *McMaster University*, Lehana THABANE, *Centre for Evaluation of Medicines, St. Joseph Healthcare & McMaster University*

A Meta-analysis on Bleeding and Cardiovascular Outcomes • Une méta-analyse sur le saignement et les résultats cardiovasculaires [MS-108]

11:15 Shih-Wa YING & Gurbakhshash SINGH, *Simon Fraser University*, Mary McBRIDE, *BC Cancer Agency*
 Analyzing Long Term Hospitalization Records From an Administrative Database • Analyse des dossiers d'hospitalisation d'une base de données administratives [MS-109]

10:15–11:45 Session 06H

SSC3014

Censored Data

Données censurées

Contributed Paper Session • Séance de communications libres

Chair • Président: Thierry DUCHESNE *Université Laval*

10:15 Valérie ROY, Thierry DUCHESNE & Belkacem ABDOUS, *Université Laval*

Smooth Nonparametric Quantile Regression with a Censored Response • Régression non paramétrique lisse des percentiles avec réponse censurée [MS-110]

10:30 David BEAUDOIN, Thierry DUCHESNE & Christian GENEST, *Université Laval*

Improving the Estimation of Kendall's tau when Censoring Affects Only One of the Variables • Amélioration de l'estimation du tau de Kendall lorsque la censure n'affecte qu'une seule des deux variables [MS-111]

10:45 Shojaeddin CHENOURI, *University of Waterloo*

A Nonparametric Two Sample Test for Bivariate Right Censored Data • Un test non paramétrique pour données bidimensionnelles censurées à droite [MS-112]

11:45–13:15 Lunch • Dîner

13:15–14:15 Session 07A

SH3345

Isobel Loutit Invited Address on Business and Industrial Statistics

Allocution Isobel Loutit sur la statistique industrielle et de gestion

Special Session • Conférence spéciale

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

Organizer and Chair • Responsable et président: François PAGEAU *SNC Tec*

13:15 Geoff VINING, *Virginia Tech*

The Challenges of Statistical Leadership • Les défis du leadership en statistique [MS-113]

13:15–14:45 Session 07B

SSC2024

Spatial/Temporal Modelling for Marine Ecological Systems

Modèles spatiaux/temporels dans les systèmes écologiques marins

Invited Paper Session • Présentations sur invitation

National Program on Complex Data Structures and Biostatistics Section • Programme national sur les structures de données complexes et le groupe de biostatistique

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

13:15 Stephen SMITH, *Bedford Institute of Oceanography*

Data Needs and Challenges for Providing Ecosystem-based Advice for the Marine Environment • Besoin en données et défis reliés à l'établissement de conseils concernant l'écosystème de l'environnement marin [MS-114]

13:45 Michael Dowd, *Dalhousie University*

State Space Approaches for Nonlinear Ecological Models • Approches d'espace d'état pour les modèles écologiques non linéaires [MS-115]

14:15 Ransom MYERS, Ian JONSEN, Greg BREED, Mike JAMES & Joanna FLEMMING, *Dalhousie University*

Using Robust Hierarchical State Space Switching Models to Elucidate the Behaviour of Seals and Turtles at Sea • Utilisation des modèles à changements de régime hiérarchiques d'espace d'état robustes pour élucider le comportement des phoques et des tortues marines [MS-116]

13:15–14:45 Session 07C

SSC2050

Statistical Methods in Computational Genetics

Méthodes statistiques en calculs génétiques

Invited Paper Session • Présentations sur invitation

Biostatistics Section • Groupe de biostatistique

Organizer and Chair • Responsable et présidente: Aurélie LABBE *Université Laval*

13:15 Mayetri GUPTA, Joseph IBRAHIM & Pingping QU, *University of North Carolina at Chapel Hill*

Inferring Gene Regulation From Integrated Bayesian Analysis of Genomic Sequence and Expression Data • Inférer la régulation des gènes à partir de l'analyse bayésienne intégrée de données de séquençage et d'expression génomique [MS-117]

13:45 Raphael GOTTARDO, *University of British Columbia*

Bayesian Analysis of ChIP-chip Experiments: Data, Questions and Partial Answers • Analyse bayésienne d'expériences ChIP-chip : données, questions et réponses partielles [MS-118]

14:15 Radu CRAIU, *University of Toronto*, Shelley BULL, *Samuel Lunenfeld Research Institute*, Andrew PATERSON, *Hospital for Sick Children, Toronto*, Lei SUN, *Departments of Public Health and Statistics, Hospital for Sick Children, Toronto*

Separating the Wheat from the Chaff: Statistical Methods for False Discovery Control • Méthodes statistiques pour le contrôle des fausses découvertes [MS-119]

13:15–14:45 Session 07D

SSC3006

Probability in High Dimensions and its Applications

Probabilités en dimensions élevées et leurs applications

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Andre VOLODIN *University of Regina*

13:15 Evarist GINÉ, *University of Connecticut*, David M. MASON, *University of Delaware*

Local U -statistic Processes and the Estimation of Densities of Functions of Several Sample Variables: Rates of Approximation in Probability • Processus local de la statistique- U et estimation de

densités de fonctions de plusieurs variables indépendantes : taux d'approximation en probabilité [MS-120]

13:45 David MASON, *University of Delaware*, Evarist GINÉ, *University of Connecticut*

LIL for the Local U -Statistic Process with Application to the Strong Consistency of a Class of Kernel Density Estimators • LLI pour le processus local de la statistique- U avec application à la convergence forte d'une classe d'estimateurs de densité par le noyau [MS-121]

14:15 Rafal KULIK, *Carleton University*

Bahadur-Kiefer Theory for Weakly and Strongly Dependent Sequences • Théorie de Bahadur-Kiefer pour séquences faiblement et fortement dépendantes [MS-122]

13:15–14:45 Session 07E

SSC3018

Innovative Teaching in Probability and Statistics

Enseignement innovateur en probabilités et statistique

Invited Paper Session • Présentations sur invitation

Education Committee of the SSC • Comité d'éducation de la SSC

Chair • Président: Román VIVEROS-AGUILERA *McMaster University*

Organizer • Responsable: Rachel ALTMAN *Simon Fraser University* & Román VIVEROS-AGUILERA *McMaster University*

13:15 Paul COREY, *University of Toronto*

Use of Simulation in the Teaching of Non-specialists • Utilisation de simulations dans l'enseignement pour non-spécialistes [MS-123]

13:45 Jim STALLARD, *University of Calgary*

The Education of Us • L'éducation des nôtres [MS-124]

14:15 Paul CABILLO, *Acadia University*

An Integrated Computer Aided Approach to Teaching Introductory Statistics • Une approche intégrée assistée par ordinateur pour enseigner la statistique de base [MS-125]

13:15–14:45 Session 07F

UC85

Joint Modelling of Longitudinal and Survival Data and Applications in Cancer Patients

Modélisation conjointe de données longitudinales et de survie et applications aux patients atteints de cancer

Invited Paper Session • Présentations sur invitation

Chair • Présidente: Tulay KORU-SENGUL *McMaster University*

Organizer • Responsable: Shesh RAI *St Jude's Hospital*

13:15 Deo Kumar SRIVASTAVA, Shesh N. RAI, Lijun ZHANG, James M. BOYETT & Sean PHIPPS, *St. Jude Children's Research Hospital*

Multivariate Modeling of Longitudinal Data and Survival Data with Application to Pediatric Cancer Study • Modélisation multivariée de données longitudinales et de survie, avec une application à l'étude du cancer pédiatrique. [MS-126]

13:45 Jianguo (Tony) SUN, Xingwei TONG & Xin HE, *University of Missouri*

Regression Analysis of Longitudinal Data with Informative Observation and Follow-up Times • Analyse par régression de données longitudinales en présence d'observations informatives et des temps de suivi [MS-127]

14:15 Do-Hwan PARK, *University of Nevada-Reno*, Liuquan SUN, *Chinese Academy of Sciences*

Gap Time Analysis of Longitudinal Data with the Application to Cancer Studies • Analyse des temps d'intervalle pour des données longitudinales avec une application aux études sur le cancer [MS-128]

13:15–14:45 Session 07G

SSC3014

Tail Behaviour and Elliptically Contoured Distributions

Comportement des ailes et distributions à contours elliptiques

Contributed Paper Session • Séance de communications libres

Chair • Président: José GARRIDO *Concordia University*

13:15 Taraneh ABARIN, *University of Manitoba*

A Simulation Study of the GMM and the Second-order Least Squares Estimators in Nonlinear Models • Une étude de simulation pour les estimateurs MMG et des moindres carrés du second ordre sous des modèles non linéaires [MS-129]

13:30 Debbie DUPUIS, *HEC Montréal*, Maria-Pia VICTORIA-FESER, *Université de Genève*

Robust Prediction Error Criterion for Pareto Modeling of Upper Tails • Un critère robuste de prévision pour la modélisation Pareto des ailes supérieures [MS-130]

13:45 Alexandru ASIMIT & Bruce JONES, *University of Western Ontario*

On the Extreme Behaviour of Continuous Bivariate Elliptical Distributions • Le comportement extrême des distributions elliptiques bidimensionnelles continues [MS-131]

14:15 Shahjahan KHAN, *University of Southern Queensland*, A K Md Ehsanes SALEH, *Carleton University*

Estimation of Parameters of the Simple Multivariate Linear Model with Student-*t* Errors • Estimation des paramètres du modèle linéaire multivarié simple avec des erreurs *t* de Student [MS-132]

14:30 Mahmoud TORABI & J.N.K. RAO, *Carleton University*

On Small Area Estimation Under a Sub-area Level Model • L'estimation de petites régions à l'aide d'un modèle défini pour les sous-régions [MS-133]

13:15–14:45 Session 07H

UC222

Dependence, Scores and Applications

Dépendance, cotes et applications

Contributed Paper Session • Séance de communications libres

Chair • Président: Sudhir PAUL *University of Windsor*

13:15 Sohail KHAN & S. Ejaz AHMED, *University of Windsor*

Estimating Intra-class Correlation Coefficients from the Class of Elliptically Contoured Distributions • Estimation des coefficients de corrélation intra-classe pour la famille des distributions à courbes de niveau elliptiques [MS-134]

13:30 Lajmi LAKHAL-CHAIEB, *University of Waterloo*, Louis-Paul RIVEST & David BEAUDOIN, *Université Laval*

Estimating Kendall's tau Under Bivariate Censoring • Estimation du tau de Kendall à partir de données censurées [MS-135]

13:45 Miao WANG, Dongsheng TU & Deb FELDMAN-STEWART, *Queen's University*

Analysis of Quality of Life Data (QoL) from Cancer Clinical Trials based on Item Response Theory • Analyse de mesures de qualité de vie (QdV) d'essais cliniques basées sur la théorie de réponse aux items [MS-136]

14:00 Mahbulul HAQ, *Centre for the Evaluation of Medicines (Hamilton, Ontario)*, Lisa DOLOVICH, *Centre for the Evaluation of Medicines, Leslie Dan Faculty of Pharmacy (University of Toronto), Pharmacy Department (St. Joseph's Healthcare, Hamilton, Ontario), Department of Clinical Epidemiology and Biostatistics (McMaster University, Hamilton, Ontario), Department of Family Medicine (McMaster University, Hamilton, Ontario)*, Lisa McCARTHY, *Centre for the Evaluation of Medicines (Hamilton, Ontario)*, Janusz KACZOROWSKI, *Centre for the Evaluation of Medicines (Hamilton, Ontario), Department of Clinical Epidemiology and Biostatistics (McMaster University, Hamilton, Ontario), Department of Family Medicine (McMaster University, Hamilton, Ontario)*, Lehana THABANE, *Centre for the Evaluation of Medicines (Hamilton, Ontario), Department of Clinical Epidemiology and Biostatistics (McMaster University, Hamilton, Ontario)*

Self Reported Risk for Drug Related Problems in Ambulatory Elderly Patients Referred for Family Practice Pharmacist Assessment • Le risque auto rapporté des problèmes reliés à la médication de patients âgés référés pour une évaluation d'un pharmacien de famille praticien [MS-137]

14:15 Mi YAN, *University of Waterloo*, Edward GIBSON, Lehana THABANE & A.C. Peter POWLES, *McMaster University*, Susan O'BRIEN, *Ancaster High School*, Danielle Sirriani MOLNAR, *Brock University*

Sleepiness in Adolescence: A Survey of 3200 Canadian Students • Somnolence chez les adolescents : une enquête sur 3200 étudiants canadiens [MS-138]

13:15–14:45 Session 07I

SSC3010

Business Surveys

Enquêtes sur les entreprises

Contributed Paper Session • Séance de communications libres

Chair • Présidente: Susana RUBIN-BLEUER *Statistics Canada*

13:15 Zachary PRITCHARD & Charles TARDIF, *Statistics Canada*

The Use of The Goods and Services Tax by the Monthly Restaurants, Caterers and Taverns Survey • L'utilisation de la taxe sur les produits et services pour l'Enquête mensuelle sur les restaurants, traiteurs et tavernes [MS-139]

13:30 ÉRIC PELLETIER, Wisner JOCELYN & Lenka MACH, *Statistique Canada*

Comparison between Two Data Sources: Tax Data and Respondent Data of the Unified Enterprise Survey of Statistics Canada • Comparaisons entre deux sources de données : données fiscales et données des répondants de l'enquête unifiée auprès des entreprises de Statistique Canada [MS-140]

13:45 Rose-Carline EVRA & Sylvie DEBLOIS, *Statistique Canada*

Review of the Collection Strategy of Statistics Canada's Unified Enterprise Survey • Revue des stratégies de collecte de l'enquête unifiée auprès des entreprises de Statistique Canada [MS-141]

14:00 Cynthia BOCCI & Jean-François BEAUMONT, *Statistique Canada*

Dealing with the Problem of Combined Reports at the Sampling Design Stage for the Workplace and Employee Survey • Le traitement du problème des rapports combinés à l'étape du plan de sondage à l'Enquête sur le milieu de travail et les employés [MS-142]

14:15 Stuart PURSEY, Paul HUNSBERGER, Zeeshan ZEESHAN & Stuart PURSEY, *Statistics Canada*

Recent Developments in the Redesign of Statistics Canada's Business Register • Développements récents dans le remaniement du registre des entreprises de Statistique Canada [MS-143]

14:45–15:15 Coffee Break • Pause-café

15:15–16:45 Session 08A

SSC2050

Data Mining Complex Structures

Structures complexes de forage de données

Invited Paper Session • Présentations sur invitation

National Program on Complex Data Structures • Programme national sur les structures de données complexes

Organizer and Chair • Responsable et président: Antonio CIAMPI *McGill University* & Hugh CHIPMAN *Acadia University*

15:15 Shuying SUN, Celia GREENWOOD & Radford NEAL, *University of Toronto*

Haplotype Inference Using a Hidden Markov Model with Efficient Markov Chain Sampling • Inférence sur le haplotype en utilisant un modèle de Markov caché avec un échantillonnage efficace par chaînes de Markov [MS-144]

15:45 Sam ROWEIS, *University of Toronto*, Amir GLOBERSON, *MIT*

Visualizing Binary Pairwise Similarity Data • Visualisation de données binaires de similarité deux-à-deux [MS-145]

16:15 David BANKS, *Duke University*

Robustness in Data Mining • Robustesse en forage de données [MS-146]

15:15–16:45 Session 08B

SH3345

Genomic Assessment Implications of Biologic Heterogeneity

Implications sur l'évaluation génomique de la variabilité biologique

Invited Paper Session • Présentations sur invitation

Biostatistics Section • Groupe de biostatistique

Organizer and Chair • Responsable et présidente: Judy-Anne CHAPMAN *Queen's University*

15:15 Debashis GHOSH, *University of Michigan*

Implications of Tumor Heterogeneity on Assessment of Differential Expression using Genomic Data • Conséquences de l'hétérogénéité des tumeurs pour déterminer les différences d'expression en utilisant des données génomiques [MS-147]

15:45 Guilherme ROSA, *Michigan State University*

Design and Analysis of Genetical Genomics Studies Involving Complex Traits • Conception et analyse des études des génomes génétiques impliquant des traits complexes [MS-148]

16:15 Judy-Anne CHAPMAN, *NCIC Clinical Trials Group*, David AXELROD, *Rutgers University*, Naomi MILLER, *University of Toronto*, Jin QIAN, *University of Waterloo*, William CHRISTENS-BARRY, *Equipoise Imaging LLC*, Yan YUAN, *University of Waterloo*, Yuejiao FU, *York University*

Clinical Implications of Tumor Heterogeneity Demonstrated Pathologically and with Image Analysis • Implications cliniques de l'hétérogénéité des tumeurs démontrées pathologiquement et par analyse d'image [MS-149]

15:15–16:45 Session 08C

SSC2024

Probability and Finance

Probabilités et finance

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: André DABROWSKI *University of Ottawa*

15:15 Alexander MELNIKOV & Yulia ROMANIUK, *University of Alberta*

Stochastic Methods and Statistical Technique for Equity-linked Life Insurance • Méthodes stochastiques et technique statistique pour l'assurance-vie liée à des indices boursiers [MS-150]

15:45 Wei SUN, *Concordia University*, Ze-Chun HU, *Nanjing University, China*, Zhi-Ming MA, *Institute of Applied Mathematics, Chinese Academy of Sciences*

Extensions of Levy-Khintchine Formula and Beurling-Deny Formula in Semi-Dirichlet Forms Setting • Extensions de la formule de Levy-Khintchine et de la formule Beurling-Deny dans le contexte des formes semi-Dirichlet [MS-151]

16:15 Steve DREKIC, *University of Waterloo*

Optimal Dividends Under a Ruin Probability Constraint • Dividendes optimaux sous une contrainte de probabilité de ruine [MS-152]

15:15–16:45 Session 08D

SSC3018

The History of Statistics and Probability

L'histoire de la statistique et des probabilités

Invited Paper Session • Présentations sur invitation

Institute of Mathematical Statistics • Institut pour la statistique mathématique

Organizer and Chair • Responsable et président: Duncan MURDOCH *University of Western Ontario*

15:15 Michael FRIENDLY, *York University*

André-Michel Guerry and the Rise of Moral Statistics • André-Michel Guerry et la naissance de la statistique morale [MS-153]

15:45 Christian GENEST, *Université Laval*, David R. BELLHOUSE, *University of Western Ontario*

Abraham De Moivre: Genius in Exile • Abraham de Moivre : génie en exil [MS-154]

16:15 George STYAN, *McGill University*, Simo PUNTANEN, *University of Tampere*

Some Comments About Issai Schur (1875–1941) and The Early History of Schur Complements • Quelques commentaires sur Issai Schur (1875-1941) et le développement des compléments de Schur [MS-155]

15:15–16:45 Session 08E

UC85

Estimating Function Methodology

Méthodologie d'estimation de fonctions

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et présidente: Mary THOMPSON *University of Waterloo*

15:15 Peter SONG & Qian ZHOU, *University of Waterloo*

Information Matrix: From Fisher to Godambe • Matrices d'information : de Fisher à Godambe [MS-156]

15:45 Bent JØRGENSEN, *University of Southern Denmark*

Estimating Functions and Analysis of Missing Data • Fonctions d'estimation et analyse des données manquantes [MS-157]

16:15 Jack KALBFLEISCH, Yining YE & Douglas SCHAUBEL, *University of Michigan*

Estimating Functions for Correlated Recurrent and Terminal Events • Fonctions d'estimation pour événements récurrents et terminaux corrélés [MS-158]

15:15–16:45 Session 08F

SSC3006

Computational Methods in Statistics

Méthodes computationnelles en statistique

Contributed Paper Session • Séance de communications libres

Chair • Président: Paul PENG *University of Michigan and Memorial University of Newfoundland*

15:15 Justin HARRINGTON, *University of British Columbia*

An Introduction to Linear Grouping Analysis • Introduction à l'analyse linéaire de groupes [MS-159]

15:30 Peter MACDONALD, *McMaster University*, Ruochu GAO, *University of Western Ontario*

The EM Algorithm for Finite Mixtures with Subsampling • Un algorithme EM pour des mélanges finis de distributions avec sous-échantillonnage [MS-160]

15:45 Johan VAN HOREBEEK, *CIMAT*, Hugh CHIPMAN, *Acadia University*

Context Sensitive Regression Models with Binary Predictors • Modèles de régression contextuels avec prédicteurs binaires [MS-161]

16:00 Hui SHEN & William J. WELCH, *University of British Columbia*

Estimating Prediction Error in Linear Regression by Cross Validation • Estimer l'erreur de prévision en régression linéaire par validation croisée [MS-162]

16:15 Yu ZHANG, *University of Regina*

Bivariate Exponential Distributions and its Application on Matched Pair Data • Distributions exponentielles bidimensionnelles et leur application dans un contexte de données appariées [MS-163]

16:30 Adeniyi ADEWALE & Douglas P. WIENS, *University of Alberta*

Robust Designs for Logistic Model with Overdispersion • Plans robustes pour un modèle logistique avec hyperdispersion [MS-164]

15:15–16:45 Session 08G

SSC3010

Methodology for Complex Data Structures

Méthodologie pour des données à structure complexe

Contributed Paper Session • Séance de communications libres

Chair • Président: Kenneth McRAE *Agriculture and Agri-Food Canada*

15:15 Eshetu ATENAFU, *University of Toronto*

Joint Analysis of Longitudinal and Survival Data with Application to Cystic Fibrosis • Analyse conjointe de données longitudinales et de survie, avec une application à la fibrose kystique [MS-165]

15:30 David TOLUSSO & Richard COOK, *University of Waterloo*

Flexible Parametric Models for Truncated and Interval-censored Multi-state Data • Modèles paramétriques flexibles pour des données tronquées, à intervalle censuré et multi-état [MS-166]

16:00 Jiguo CAO, James O. RAMSAY & Giles HOOKER, *McGill University*

Fitting HIV Mixed-effects Dynamic Models to Clinical Trials • Ajustement de modèles dynamiques à effets mixtes pour des essais cliniques pour le VIH [MS-167]

16:15 Mohua PODDER, William J. WELCH & Ruben H. ZAMAR, *University of British Columbia*

Dynamic Variable Selection in SNP Classification Based on an APEX Microarray Platform • Sélection de variable dynamique en classification de SNP basée sur une plate-forme de puces APEX [MS-168]

16:30 Qian ZHOU, *University of Waterloo*, S. Stanley YOUNG, *National Institute of Statistical Sciences*, Mu ZHU, *University of Waterloo*

Sequential Learning of SVMs for Target Identification in Drug Discovery • Apprentissage séquentielle de SVMs pour l'identification de cible dans la découverte de médicament [MS-169]

16:45 Julie HORROCKS, Azim BHAMANI, Hoki PONG, Marianne VAN DEN HEUVEL, Francis TEKPEY & B. Anne CROY, *University of Guelph*

Classification of Longitudinal Data with Missing Values • Classification de données longitudinales avec des valeurs manquantes [MS-170]

15:15–16:45 Session 08H

SSC3014

Survey Analysis and Operations
Analyse d'enquêtes et opérations

Contributed Paper Session • Séance de communications libres

Chair • Président: Louis-Paul RIVEST *Université Laval*

15:15 Harold MANTEL & Edward J. CHEN, *Statistics Canada*

Methodological Innovations for the Canadian Labour Force Survey • Innovations méthodologiques dans l'Enquête canadienne sur la population active [MS-171]

15:45 Hansheng XIE & Walter MUDRYK, *Statistics Canada*

A Methodological Approach to Controlling Inter-monitor Variability in Computer Assisted Telephone Interviewing Operations • Une approche méthodologique du contrôle de la variabilité entre les surveillants dans les opérations d'entrevues téléphoniques assistées par ordinateur [MS-172]

16:00 Pat NEWCOMBE, *Statistics Canada*

What to do when Survey Weights Weigh you Down • Quoi faire quand les pondérations (les poids) d'une enquête pèsent sur vous [MS-173]

16:45–18:15

SSC2050

Assemblée générale de la SSC
SSC General Meeting

Mercredi 31 mai • Wednesday, May 31

08:30–09:30 Session 09A

SSC2050

Pierre Robillard Award Address

Allocution du récipiendaire du prix Pierre-Robillard

Special Session • Conférence spéciale

 Organizer and Chair • Responsable et président: Paul GUSTAFSON *University of British Columbia*

08:30–10:00 Session 09B

SSC2024

Applied Probability

Probabilités appliquées

Invited Paper Session • Présentations sur invitation

 Chair • Président: Reg KULPERGER *University of Western Ontario*

 Organizer • Responsable: Reg KULPERGER & David STANFORD *University of Western Ontario*

 08:30 David BRILLINGER, *University of California at Berkeley*, Brent STEWART, *Hubbs-Sea World Research Institute*, Charles LITNAN, *Pacific Islands Fisheries Science Center, NOAA Fisheries*

Three Months Journeying of a Hawaiian Monk Seal • Le voyage de trois mois d'un phoque moine hawaïen [MS-174]

 09:00 Bruce SMITH, *Dalhousie University*

Return Period Models for High Sea Levels • Modèles de la période de retour pour les niveaux élevés de la mer [MS-175]

 09:30 Glen TAKAHARA, *Queen's University*, Chris LEITH, *Royal Bank of Canada*

Finite Memory Polya Urns • Urnes de Polya à mémoire finie [MS-176]

08:30–10:00 Session 09C

SSC2036

Analysis of Longitudinal Complex Survey Data Using Marginal Modelling and GEE: Theory and Practice

Analyse de données longitudinales d'enquêtes complexes utilisant la modélisation marginale et les GEE : théorie et pratique

Invited Paper Session • Présentations sur invitation

Survey Methods Section • Groupe de méthodologie d'enquête

 Organizer and Chair • Responsable et présidente: Susana RUBIN-BLEUER *Statistics Canada*

 08:30 Georgia ROBERTS, *Statistics Canada*, J.N.K. RAO, *Carleton University & Statistics Canada*, Qunshu REN, *Carleton University*

Goodness-of-fit Tests of Marginal Logistic Models for Longitudinal Complex Survey Data • Tests d'adéquation de modèles logistiques marginaux pour des données longitudinales d'enquêtes complexes [MS-177]

08:52 Abdelnasser SAIDI, *Statistics Canada*, Diane STUKEL, *UNESCO*, Susana RUBIN-BLEUER, *Statistics Canada*

Efficiency Comparisons of GEE versus IEE for Longitudinal Complex Survey Data with Ordinal Responses • Comparaisons de l'efficacité de la méthode des EEG par rapport à la méthode des EEI pour des données longitudinales ordinales d'enquêtes complexes [MS-178]

09:14 Punam PAHWA, *Dept. of Community Health and Epidemiology*, Chandima KARUNANAYAKE & Helen McDUFFIE, *Inst. of Agricultural Rural and Environmental Health*

Modelling of Longitudinal Polytomous Outcomes From Complex Survey Data: Using GEE and Missing Data Analysis • Modélisation des résultats polychotomiques longitudinaux pour des données complexes d'une enquête : utilisation du EEG et de l'analyse de données manquantes [MS-179]

09:36 Sunita GHOSH & Punam PAHWA, *University of Saskatchewan*, Geert MOLENBERGHS, *Hasselt University*

Design-based Versus Model-based Methods: A Comparative Study Using Longitudinal Survey Data. • Les méthodes basées sur le plan de sondage versus les méthodes basées sur un modèle : étude comparative en utilisant des données longitudinales d'une enquête. [MS-180]

10:00–10:30 Coffee Break • Pause-café

10:30–11:30 Session 10A

SSC2050

Special Invited Session of the Biostatistics Section

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

Special Session • Conférence spéciale

Biostatistics Section • Groupe de biostatistique

Organizer and Chair • Responsable et président: Michael ESCOBAR *University of Toronto*

10:30 Donald BERRY, *The University of Texas M.D. Anderson Cancer Center*

Weaknesses and Strengths of a Bayesian Approach in Clinical Research • Faiblesses et forces d'une approche bayésienne en recherche clinique [MS-181]

10:30–12:00 Session 10B

SSC2036

Recent Advances in SPC and Data Driven Statistics

Développements récents en CSP et statistiques guidées par les données

Invited Paper Session • Présentations sur invitation

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

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

10:30 Douglas HAWKINS, *University of Minnesota*

Statistical Process Control and the Problem of Unknown Parameters • Difficultés causées par la présence de paramètres inconnus en contrôle des procédés. [MS-182]

11:00 Willem ALBERS & Wilbert C.M. KALLENBERG, *University of Twente*

Shewhart Control Charts in New Perspective • Une nouvelle perspective pour les cartes de contrôle Shewhart [MS-183]

11:30 Lingyun ZHANG, *Massey University*, Gemai CHEN, *University of Calgary*

On t and EWMA t Charts for Monitoring Changes in the Process Mean • Sur des cartes de contrôle t et EWMA t pour étudier le changement dans la moyenne du processus [MS-184]

10:30–12:00 Session 10C

SH3345

Analyzing Data from International Surveys

Analyse de données d'enquêtes internationales

Invited Paper Session • Présentations sur invitation

Survey Methods Section • Groupe de méthodologie d'enquête

Organizer and Chair • Responsable et présidente: Diane STUKEL *UNESCO*

10:30 Scott MURRAY, *UNESCO Institute for Statistics*

Examples From Large-scale Skill Assessment • Exemples à partir d'évaluations d'habiletés à grande échelle [MS-185]

11:00 Miles CORAK, *Statistics Canada*

Child Poverty and Changes in Child Poverty • La pauvreté chez l'enfant et les changements dans la pauvreté chez l'enfant [MS-186]

11:30 Edilberto LOAIZA, *UNICEF*

Analysis and Availability of Data Collected Via International Household Surveys Programs • Analyse et disponibilité des données provenant de programmes d'enquêtes internationales sur les ménages [MS-187]

10:30–12:00 Session 10D

SSC3018

Statistical Analysis of Non-precise Data

Analyse statistique des données imprécises

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Mayer ALVO *University of Ottawa*

10:30 Reinhard VIERTL, *Vienna University of Technology*

Fuzzy Information and Statistics • L'information floue et la statistique [MS-188]

11:00 François THÉBERGE & Mayer ALVO, *University of Ottawa*

Classification of Non-precise Data • Classification de données non précises [MS-189]

11:30 Maria GIL, Gil GONZALEZ-RODRIGUEZ & Ana COLUBI, *Universidad de Oviedo, Spain*

A Classical Goodness-of-fit Test Based on Fuzzy Random Variables • Un test classique d'ajustement basé sur des variables aléatoires floues [MS-190]

10:30–12:00 Session 10E**SSC2024**

Case Study II: Obstructive Sleep Apnea

Étude de cas II : L'apnée obstructive du sommeil

Case Study • Étude de cas

Organizer and Chair • Responsable et présidente: Allison L. GIBBS *University of Toronto*10:30 Sharon CHUNG, *Toronto Western Hospital*

Presentation of the Data Set • Présentation du jeu de données

10:40 Meng DU, Zi JIN & Longhai LI, *University of Toronto*10:50 Linglong KONG, Yuanyuan LIANG, Mengzhe WANG & Qiaohao ZHU, *University of Alberta*11:00 Song LI & Yaling YIN, *University of Saskatchewan*11:10 Yin CUI, Ahmad A. FAROOQI & Tasneem ZAIHRA, *University of Windsor*11:20 Ryan BROWNE, Charlotte GRIEVE, Kristen MADDALENA & Erik YOUNGSON, *University of Waterloo*11:30 Lin FANG, David HAN & Wang LI, *McMaster University*11:40 MingXia DENG, *University of New Brunswick*11:50 Rena (Jie) SUN & Cynthia (Xin) ZHENG, *University of Calgary***10:30–12:00 Session 10F****SSC3026**

Time Series

Séries chronologiques

Contributed Paper Session • Séance de communications libres

Chair • Président: Reg KULPERGER *University of Western Ontario*10:30 Asokan MULAYATH VARIYATH, BOVAS ABRAHAM & Jiahua CHEN, *University of Waterloo*

Variable Selection in Generalized Linear Models by Empirical Likelihood • Sélection de variables par la vraisemblance empirique dans un modèle linéaire généralisé [MS-191]

10:45 Alex BADESCU & Reg KULPERGER, *University of Western Ontario*, Emese LAZAR, *University of Reading*

Normal Mixtures GARCH Models and Option Pricing • Modèles GARCH de mélanges normaux et évaluation d'option [MS-192]

11:00 Mohamedou OULD HAYE & A. K. Md. E. SALEH, *Carleton University*

Quasi-empirical Bayes Methods of Estimation in ARMA models • Méthodes quasi bayésiennes dans l'estimation des modèles ARMA [MS-193]

11:15 Khadija BOUZAACHANE, *Ecole Mohammadia d'Ingénieurs, Maroc*, Youssef BENGHABRIT, *Ecole Nationale Supérieure des Arts et Métiers, Meknès*, Mostaf HARTI, *Université Sidi Mohammed Ben Abdellah, Maroc*

Algorithm of Maximum Likelihood Parameters Estimation for First-order Superdiagonal Bilinear Time Series • Algorithme d'estimation par maximum de vraisemblance des paramètres du modèle bilinéaire superdiagonal d'ordre un [MS-194]

11:30 Zahirul HOQUE, *University of Newcastle, Australia*, Richard GERLACH, *University of Sydney, Australia*

Improved Estimation for Dynamic Linear Regression Model • Estimation améliorée pour un modèle de régression linéaire dynamique [MS-195]

11:45 Ian McLEOD, *University of Western Ontario*, Ying ZHANG, *Acadia*

Subset Autoregression: A New Approach • Une nouvelle approche pour modèles autorégressifs sur sous-ensembles [MS-196]

10:30–12:00 Session 10G

SSC3006

Probability and Mathematical Statistics

Probabilité et statistique mathématique

Contributed Paper Session • Séance de communications libres

Chair • Président: Miklós CsÖRGÖ *Carleton University*

10:30 Cody HYNDMAN, *University of Calgary*

Forward-backward Stochastic Differential Equations and Transforms for Affine Diffusions • Les équations différentielles stochastiques directes et rétrogrades et les transformations pour diffusions affines [MS-197]

10:45 François WATIER & René FERLAND, *Université du Québec à Montréal*

Maximizing Expected Utility Through Forward-backward Stochastic Differential Equations • Optimisation de l'utilité espérée par le biais d'équations différentielles stochastiques directes et rétrogrades [MS-198]

11:00 Adriana JORDAN & Gail IVANOFF, *University of Ottawa*

A Two-sample Copula Process • Un processus de copules pour deux échantillons [MS-199]

11:15 Hanna JANKOWSKI, *University of Washington*

Empirical Central Limit Theorems under Model Misspecification • Versions empiriques du théorème central limite pour un modèle mal spécifié [MS-200]

11:30 Raluca BALAN, *University of Ottawa*, Ingrid-Mona ZAMFIRESCU, *Baruch College, City University of New York*

Strong Approximation for Mixing Sequences with Infinite Variance • Approximation forte pour des suites mélangées avec variance infinie [MS-201]

11:45 Yuliya MARTSYNYUK, *Carleton University*

New Central Limit Theorems via Studentization in Linear Structural Error-in-variables Models • Nouveaux théorèmes limites centraux basés sur la studentisation dans les modèles linéaires avec erreurs de mesure [MS-202]

10:30–12:00 Session 10H

SSC3014

Statistical Genetics and Genomics
Génétique statistique et génomique

Contributed Paper Session • Séance de communications libres

Chair • Président: Angelo CANTY *McMaster University*

10:30 M. Tariqul HASAN & Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital.*

Group Sequential Approach for Penetrance Estimation of Genes Involved in Complex Diseases Based on Weighted Log-rank Statistics • Estimation de la pénétrance des gènes impliqués dans des maladies complexes : une approche séquentielle par groupes basée sur des statistiques de (log) rang pondérées [MS-203]

10:45 Yun Hee CHOI & Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital*

Estimating the Lifetime Risk Associated with Mutated Genes Involved in Complex Diseases • Estimation du risque associé aux gènes mutants impliqués dans des maladies complexes [MS-204]

11:00 Longyang WU, *Samuel Lunenfeld Research Int.*, LongYang WU, *SLRI*, Lei SUN, *University of Toronto*, Bull SHELLEY, *SLRI*

Genetic Effect Estimation via Bootstrap in Linkage Analysis of Quantitative Trait Loci • Estimation d'effet génétique par bootstrap dans l'analyse de liaison du trait quantitatif du loci (emplacement des gènes) [MS-205]

11:15 Wenqing HE & Ian McLEOD, *University of Western Ontario*

A Method for Analysis of CGH Microarray Data • Une méthode pour l'analyse de données de micropuces CGH [MS-206]

11:30 Lihua AN & S. Ejaz AHMED, *University of Windsor*

Improved Estimation for Differential Gene Expression by Shrinking Variance Components • Estimation améliorée de l'expression de gènes par le rétrécissement des composantes de variance [MS-207]

11:45 Adam ZAGDANSKI & Rafal KUSTRA, *University of Toronto*

Improving the Relevance of Microarray Data Clustering through Integration with Gene Ontology • Amélioration de la pertinence du regroupement des données de micropuces par l'intégration avec l'ontologie du gène [MS-208]

12:00–16:30 Session 10I

SSC - 2nd floor hallway

Posters for Case Study II
Affiches pour l'étude de cas II

Case Study • Étude de cas

12:00 Dan HE & Weiling WANG, *McMaster University*

12:00 Matthew DAVIS, Lei HE & Chao QIU, *University of Calgary*

12:00–13:30 Lunch • Dîner

SH = Somerville House	SSC = Social Science Centre	UC = University College
-----------------------	-----------------------------	-------------------------

12:30–13:15**SSC2036**

Appui des mathématiques et de la statistique par le CRSNG

Support of Mathematics and Statistics by NSERC

Organizer and Chair • Responsable et présidente: Isabelle BLAIN, *NSERC/CRSNG***13:30–14:30 Session 11A****SSC2050**

Address of the winner of the Centre de recherches mathématiques/Statistical Society of Canada Prize

Allocution du lauréat du Prix du Centre de recherches mathématiques/Société statistique du Canada

Special Session • Conférence spéciale

Organizer and Chair • Responsable et présidente: Charmaine DEAN *Simon Fraser University***13:00–14:30 Session 11B****SSC2024**

Risk Theory I

Théorie du risque I

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Steve DREKIC *University of Waterloo*13:00 David STANFORD, *University of Western Ontario*, Vaidyanathan RAMASWAMI, *ATT Research*, Douglas WOOLFORD, *University of Western Ontario*

Erlangized Fluid Queues and their Relation to Finite-time Ruin Probabilities • Les files d'attente de fluides erlangisées et leur relation aux probabilités de ruine en temps fini [MS-209]

13:30 Andrei BADESCU, *University of Waterloo*, Vaidyanathan RAMASWAMI, *AT&T Labs*, Soohan AHN, *University of Seoul*

Risk Models with Constant Dividend Barrier • Modèles de risque avec barrière de dividende constante [MS-210]

14:00 David LANDRIAULT, *University of Waterloo*

Constant Dividend Barrier in a Risk Model With Time-dependent Claim Sizes • Barrière de dividende constante dans un modèle de risque où le montant des réclamations dépend du temps [MS-211]

14:30–15:00 Coffee Break • Pause-café**15:00–16:30 Session 12A****SSC2050**

Industrial Experimental Design

Planification d'expériences industrielles

Invited Paper Session • Présentations sur invitation

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

Organizer and Chair • Responsable et président: John BREWSTER *University of Manitoba*

15:00 Peter PARKER, *National Aeronautics and Space Administration*

Second-order Split-plot Design Construction and Evaluation • Plan d'expérience à parcelles subdivisées du 2^e ordre : construction et évaluation [MS-212]

15:30 Rob McLEOD, *University of Winnipeg*, John BREWSTER, *University of Manitoba*

Optimal Foldover Plans for Two-level Fractional Factorial Split-plot Designs • Plans de repliement optimaux pour plans factoriels fractionnels à deux niveaux à parcelles subdivisées [MS-213]

16:00 Pritam RANJAN & Derek BINGHAM, *Simon Fraser University*, George MICHAILIDIS, *University of Michigan*

Sequential Experiment Design for Contour Estimation from Complex Computer Codes. • Plan d'expérience séquentiel pour l'estimation d'une courbe de niveau à partir d'un programme informatique complexe [MS-214]

15:00–16:30 Session 12B

SSC2036

Latent Variable and Multilevel Modelling Using Survey Data

Variabes latentes et modélisation multiniveaux pour des données d'enquêtes

Invited Paper Session • Présentations sur invitation

Survey Methods Section • Groupe de méthodologie d'enquête

Organizer and Chair • Responsable et président: Changbao Wu *University of Waterloo*

15:00 Sophia RABE-HESKETH, *University of California, Berkeley*, Anders SKRONDAL, *London School of Economics*

Pseudo Maximum Likelihood Estimation of Multilevel and Latent Variable Models • Estimation par pseudo-maximum de vraisemblance de modèles à niveaux multiples et à variables latentes [MS-215]

15:30 Eric LACOURSE & Véronique DUPÉRÉ, *Université de Montréal*

Modeling Change in Longitudinal Studies With Individual Trajectories: Parametric and Nonparametric Multilevel Models • Méthodes d'analyse du changement fondées sur la trajectoire de développement individuelle : modèles multiniveaux paramétriques et non paramétriques [MS-216]

16:00 Milorad KOVACEVIC & Rong HUANG, *Statistics Canada*

Fitting Multi-level Models to Survey Data • Modélisation de données d'enquête complexes à l'aide de modèles à multi-niveaux [MS-217]

15:00–16:30 Session 12C

SSC2024

Risk Theory II

Théorie du risque II

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: David STANFORD *University of Western Ontario*

15:00 Gordon WILLMOT, *University of Waterloo*

On the Discounted Penalty Function in the Renewal Risk Model with Arbitrary Interclaim Times •

Sur la fonction de pénalité escomptée dans le modèle de risque de renouvellement avec des temps d'inter-réclamation arbitraires [MS-218]

15:30 JOSÉ GARRIDO, *Concordia University*, MANUEL MORALES, *Université de Montréal*

The Gerber-Shiu Function for Risk Models Based on Lévy Processes • La fonction de Gerber-Shiu pour des modèles de risque basés sur des processus de Lévy [MS-219]

16:00 KRISTINA SENDOVA, *University of Western Ontario*, X. Sheldon LIN, *University of Toronto*

The Compound Poisson Risk Model with Multiple Thresholds • Le modèle de risque composé de Poisson à seuils multiples [MS-220]

15:00–16:30 Session 12D

SH3345

Geometry, Topology and Statistics

Géométrie, topologie et statistique

Invited Paper Session • Présentations sur invitation

Organizer and Chair • Responsable et président: Peter KIM *University of Guelph*

15:00 VICTOR PATRANGENARU & Samanmalee SUGATHADASA, *Texas Tech University, USA*

Reconstruction of 3D Scenes and Projective Shape Analysis • Reconstruction de scènes en trois dimensions et analyse projective des formes [MS-221]

15:30 Theodore CHANG, *University of Virginia*

Exploratory Data Analysis for Human Motion Data • Analyse exploratoire du mouvement humain [MS-222]

16:00 Peter BUBENIK, *Cleveland State University*, Peter KIM, *University of Guelph*

A Statistical Approach to Persistent Homology • Une approche statistique sur l'homologie persistante [MS-223]

15:00–16:30 Session 12E

SSC3018

Applications of Statistics II

Applications de la statistique II

Contributed Paper Session • Séance de communications libres

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

15:00 Jeffrey PICKA, *University of New Brunswick*

Exploratory Analysis of Sphere Packings By Shape-based Statistics • Analyse exploratoire d'empilements de sphères avec des statistiques basées sur les formes [MS-224]

15:15 Karim OUALKACHA & Louis-Paul RIVEST, *Université Laval*

A New Statistical Model for Random Unit Vectors • Un nouveau modèle statistique pour des vecteurs unitaires aléatoires [MS-225]

15:30 Pengfei LI, *University of Waterloo*, Yuejiao FU, *York University*, Jiahua CHEN, *University of Waterloo*

Modified Likelihood Ratio Test for Homogeneity in a Mixture of von Mises Distributions • Test du

rapport de vraisemblance modifié pour l'homogénéité en présence d'un mélange de distributions von Mises [MS-226]

15:45 Shahjahan KHAN, *University of Southern Queensland, Australia*

Optimal Tolerance Regions for Future Regression Vector and Residual Sum of Squares of Multiple Regression Model with Multivariate Spherically Contoured Errors • Régions de tolérance optimales pour un futur vecteur de régression et sommes de carrés résiduels sous un modèle de régression multiple avec erreurs multivariées aux contours sphériques [MS-227]

16:15 Anwar SAGR, *University of Al-Jabal Al-Gharbi*

Theories in Linear Regression Model • Théorie en régression linéaire [MS-228]

15:00–16:30 Session 12F

SSC3026

Multivariate Time Series

Séries chronologiques multivariées

Contributed Paper Session • Séance de communications libres

Chair • Président: Ian McLEOD *University of Western Ontario*

15:00 Mohanad AL-KHASAWNEH & S. Ejaz AHMED, *University of Windsor*

Empirical Bayes Analysis for a Hierarchical Negative Binomial Generalized Linear Model • Analyse bayésienne empirique sous un modèle binomial négatif hiérarchique linéaire généralisé [MS-229]

15:30 Shaheena BASHIR & Edward CARTER, *University of Guelph*

Fixed Point Clustering in Robust Mixture Discriminant Analysis • Regroupement point fixe dans l'analyse discriminante de mélange robuste [MS-230]

15:45 Christian GAGNÉ & Pierre DUCHESNE, *Université de Montréal*

On Robust Forecasting in Dynamic Vector Time Series Models • Sur les prévisions robustes dans les modèles de séries chronologiques vectoriels dynamiques [MS-231]

16:00 Abdessamad SAIDI & Roch ROY, *Université de Montréal*

Robust Optimal Tests for Non-causality in Multivariate Time Series • Tests optimaux et robustes de non-causalité pour des séries chronologiques multivariées [MS-232]

16:15 Jennifer POULIN & Pierre DUCHESNE, *Université de Montréal*

Empirical Study of Portmanteau Test Statistics Designed for Multivariate Time Series • Étude empirique de tests de type portemanteau pour séries chronologiques multivariées [MS-233]

15:00–16:30 Session 12G

SSC3006

Observational Studies: Methods and Applications

Méthodes et applications pour les études observationnelles

Contributed Paper Session • Séance de communications libres

Chair • Président: Noel CADIGAN *Department of Fisheries and Oceans*

- 15:00 Russell STEELE, *McGill University*, Ian SHRIER, *Jewish General Hospital*
Approximating Relative Risk Using Logistic Regression Revisted • Approximation du risque relatif en utilisant la régression logistique : un second aperçu [MS-234]
- 15:15 Lawrence McCANDLESS & Paul GUSTAFSON, *University of British Columbia*, Peter C AUSTIN, *University of Toronto*
Bayesian Propensity Score Analysis for Observational Data • Analyse bayésienne de score de propension pour des données d'observation [MS-235]
- 15:30 Hyang Mi KIM, Yutaka YASUI & Igor BURSTYN, *University of Alberta*
Estimation of Parameters in Logistic Regression Models with Multiplicative Measurement Error • Estimation des paramètres d'un modèle de régression logistique en présence d'erreur de mesure multiplicative [MS-236]
- 15:45 Maria LORENZI, *BC Cancer Agency*
Modelling Relative Risk of Cancer Incidence in a Cohort Study • Modélisation du risque relatif de l'incidence du cancer au moyen d'une étude de cohortes [MS-237]
- 16:00 Olaf BERKE, *University of Guelph and University of Veterinary Medicine, Hannover, Germany*
A Spatiotemporal Investigation of Echinococcus Multilocularis Infections Among Red Foxes in Lower Saxony, 1991 - 2005 • Une investigation spatiotemporelle des cas d'infections de Echinococcus Multilocularis chez les renards rouges de la Basse-Saxe, de 1991 à 2005 [MS-238]
- 16:15 Severien NKURUNZIZA & S. Ejaz AHMED, *University of Windsor*
Statistical Inference in Predator-prey System Modeled under Intrinsic Period Property • L'inférence statistique dans le système prédateur-proie modélisé sous la propriété intrinsèque de période [MS-239]

15:00–16:30 Session 12H**SSC3014****Modelling and Analysis in Sample Surveys****Modélisation et analyse pour l'échantillonnage**

Contributed Paper Session • Séance de communications libres

Chair • Président: Richard LOCKHART *Simon Fraser University*

- 15:00 Jack LOTHIAN, *Statistics Canada*
Heavy-tails and the Central Limit Theorem in Business Surveys • Ailes lourdes et le théorème limite centrale dans les enquêtes auprès des entreprises [MS-240]
- 15:15 Eddy REMPEL, *McMaster University*, Lehana THABANE, *Department of Clinical Epidemiology and Biostatistics, McMaster University, Centre for Evaluation of Medicines, St Joseph's Healthcare*
A Survey of Evangelical Mennonite Mission Conference of Churches: A Look at Factors that Impact Growth • Une étude de la conférence des églises de la mission évangélique mennonite et des facteurs en influençant la croissance [MS-241]
- 15:30 Floyd RUDMIN, *University of Tromsø, Norway*
Evaluating Likert Items Using Inter-Item θ • Estimation des items de Likert en utilisant l'inter-item θ [MS-242]

Abstracts • Résumés

Session 0 Dimanche 28 mai • Sunday, May 28, 19:00–21:00 SH - Great Hall

Contributed Poster Session

Présentation d'affiches

[MS-1]

Use of Contrasts in the Analysis of the Affymetrix Hypoalpha Dataset

L'emploi des contrastes dans l'analyse des données Affymetrix Hypolpha

Paul SUTCLIFFE & André DABROWSKI, *University of Ottawa*, Kathryn WILLIAMS, Bassam HAIDAR & Yves MARCEL, *Ottawa Heart Institute*

We explore the use of contrasts and robust standardization methods in analyzing an Affymetrix microarray data set related to cholesterol regulation. The data comprises readings on approximately 12,000 gene responses from each of thirty subjects partitioned in four groups. In an attempt to identify differentially expressed genes, we define three simple contrasts to reflect biological questions, standardize using 10 percent trimming, and plot the 12,000 points in 3D.

Nous explorons l'utilisation des contrastes et des méthodes robustes pour analyser un groupe de données microarray d'Affymetrix lié à la réglementation de cholestérol. Les données comportent des observations sur approximativement 12 000 gènes de trente sujets divisés dans quatre groupes. Afin d'essayer d'identifier l'expression différentielle de ces gènes, nous définissons trois contrastes simples pour refléter des questions biologiques, les normalisons en tronquant 10 pour cent, et traçons les 12 000 points dans 3D.

[MS-2]

A Robust Test for Asymptotic Independence in Multivariate Extreme Values

Un test robuste pour l'indépendance asymptotique de valeurs extrêmes multivariées

Yu-Ling TSAI, *University of Western Ontario*, Debbie DUPUIS, *HEC Montréal*, Duncan MURDOCH, *University of Western Ontario*

We start by examining three classical tests for asymptotic independence in multivariate extremes. We point out their sensitivities to outliers and the lack of accuracy of their asymptotic approximations. A robust bootstrapped test is proposed. The new test is made stable under contamination through a shrinking scheme. Bootstrapped p -values are used to resample the test statistic from the empirical distribution rather than from the estimated population distribution satisfying the null constraints. Applications to real and simulated data sets show the good performance of the new test.

Nous commençons par examiner trois tests classiques pour l'indépendance asymptotique pour des valeurs extrêmes multivariées. Nous précisons leurs sensibilités aux valeurs aberrantes et le manque d'exactitude de leurs approximations asymptotiques. Un test robuste Bootstrap est proposé. Le nouveau test est rendu stable sous une contamination par une méthode de resserrement. Les valeurs- p de Bootstrap sont utilisées pour reprélever le test statistique de la distribution empirique plutôt que de la distribution estimée de population satisfaisant les contraintes nulles. Des applications à des ensembles de données vrais et simulés montrent la bonne performance du nouveau test.

[MS-3]**Valid p -values for Stepwise Regression and Other Post-hoc Model Selection Methods****Valeurs- p valides pour la régression pas-à-pas et autres méthodes de sélection post-hoc d'un modèle**Richard HARSHMAN & Margaret E. LUNDY, *University of Western Ontario*

When a model modification is selected using post-hoc information (e.g., in stepwise regression) it is hard to determine the statistical significance of the resulting improvement. We describe a computer-intensive method that empirically approximates the improvement's null distribution using permutation-test methods combined with "null-set pruning" (elimination of cases inconsistent with prior step results or other dataset characteristics). Monte Carlo tests indicate that un-pruned randomization gives unbiased p -values for post-hoc step 1, but biased ones thereafter; pruned randomization gives apparently unbiased p -values for successive steps as well (within the confidence bounds of our Monte Carlo tests). Applications include complex sequential methods and stepwise canonical correlation/MANOVA.

Quand une modification d'un modèle est adoptée en utilisant l'information post-hoc (par exemple, dans la régression pas à pas) il est difficile de déterminer le niveau de signification statistique de l'amélioration résultante. Nous décrivons une méthode entièrement simulée par ordinateur avec laquelle on estime empiriquement la distribution de l'amélioration sous l'hypothèse nulle en employant des méthodes de tests par permutation combinées avec une technique d'« élagage » (élimination des cas contradictoires avec les résultats préliminaires ou d'autres caractéristiques des données). Les tests de Monte Carlo indiquent que la randomisation non élaguée semble mener à des valeurs- p non biaisées pour l'étape 1 post-hoc, mais biaisées par la suite ; la randomisation élaguée semble mener à des valeurs- p non biaisées pour toutes les étapes (à l'intérieur des limites de confiance de nos tests Monte-Carlo). Les applications incluent des méthodes séquentielles complexes et la corrélation canonique pas à pas/MANOVA.

[MS-4]**Screening for Risk Status and Statistical Power in Randomized Intervention Research****Criblage pour le statut de risque et puissance statistique dans la recherche d'intervention randomisée**Depeng JIANG & Debra PEPLER, *York University*

This study examined how screening for risk status affects statistical power in randomized intervention research. On the basis of Monte Carlo simulations, this study demonstrates how statistical power changes depending on available information on risk status and the strength of the intervention. It also examines how obtaining covariate information for the outcome, screening for risk status, and the predictive accuracy of risk status affect power. Using real prevention data as an example, this study illustrates how the estimation of intervention effects for diverse subgroups and the corresponding statistical power depend on the quality of classification of participants to subgroups.

Cette étude a examiné comment le criblage pour le statut de risque affecte la puissance statistique dans la recherche d'intervention randomisée. Sur la base des simulations de Monte Carlo, cette étude démontre comment la puissance statistique change selon l'information disponible sur le statut de risque et la force d'intervention. Elle examine également comment obtenir l'information sur la covariable pour le résultat, le criblage pour le statut de risque, et comment l'exactitude prédictive du statut de risque affecte la puissance. En utilisant de vraies données de prévention comme exemple, cette étude illustre comment l'estimation des effets d'intervention pour les divers sous-groupes et la puissance statistique correspondante dépendent de la qualité de la classification des participants aux sous-groupes.

[MS-5]**A New R Package for Biostatistics, Public Policy and Law****Une nouvelle logithèque R pour la biostatistique, les politiques d'intérêt public et le domaine judiciaire**

Wai Lok Wallace HUI & Yulia R. GEL, *University of Waterloo*, Joseph GASTWIRTH, *George Washington University*, Weiwen MIAO, *Macalester College*

We present a new R software package on statistical tests widely utilized in biostatistics, public policy and law. Along with well-known but not implemented tests on serial correlation such as the Bartels test, the package contains new distribution-free robust tests of symmetry; omnibus and directional tests of normality which are more sensitive to heavy tailed alternatives, and their graphical counterparts such as Robust QQ plot; measures of relative variability (Gini Index, Coefficient of Dispersion, Lorenze curve); tests of homogeneity of several sample proportions. All implemented tests and methods are illustrated by simulations and real-life examples from legal statistics, economics and biostatistics.

Nous présentons un nouveau progiciel de R sur les tests statistiques largement utilisés en biostatistiques, en politiques d'intérêt public et dans le domaine judiciaire. Avec les tests bien connus mais non mis en application sur l'autocorrélation tel que le test de Bartels, le progiciel contient de nouveaux tests robustes non paramétriques de la symétrie; tests composites et directionnels de la normalité qui sont plus sensibles aux solutions de rechange à aile lourde, et leurs contreparties graphiques telle que un QQ plot robuste; mesures de variabilité relative (indice de Gini, coefficient de dispersion, courbe de Lorenze); test d'homogénéité de plusieurs proportions d'échantillon. Tous les tests et les méthodes mis en application sont illustrés par des simulations et des exemples réels provenant des domaines des statistiques légales, des sciences économiques et de la biostatistique.

[MS-6]**Quality Assessment for Microarray Gene Expression Data****Évaluation de la qualité de données d'expression génétique mesurées par micro puces**

Julia BRETTSCHEIDER, *Queen's University*, François COLLIN, *Affymetrix Inc.*, Ben BOLSTAD, *University of California at Berkeley*, Terence SPEED, *University of California at Berkeley & Walter and Eliza Hall Institute, Australia*

Microarrays have become an established instrument for measuring tool for measuring the biochemical activity of thousands of genes simultaneously. Among the many applications are studies of the genetic components in the etiology of complex diseases such as cancer or schizophrenia. Despite technical improvement, the lack of accuracy and precision remains a characteristic of microarray data. Benchmarks for reproducibility are urgently needed. We introduce quality metrics and visual quality assessment tools for short oligonucleotide microarrays. They are tested on experimental data and compared them with the commercial quality report. Our assessment is based on byproducts obtained by statistical preprocessing by RMA/fitPLM.

Les micropuces sont devenus un instrument bien établi dans la mesure simultanée de l'activité biochimique de milliers de gènes. L'étude de la composante génétique dans l'étiologie de maladies complexes comme le cancer et la schizophrénie fait partie de ses multiples applications. Malgré une amélioration technique, l'absence d'exactitude et de précision demeure caractéristique des données provenant de micropuces. Il y a un besoin urgent de repères à fins de reproductibilité. Nous introduisons des métriques de qualité et des outils d'évaluation visuelle de la qualité pour les micropuces à oligonucléotides courts. Ils sont testés sur des données expérimentales et comparés au rapport de qualité commercial. Notre évaluation est basée sur les sous-produits obtenus lors du pré-traitement statistique par RMA/fitPLM.

[MS-7]**Generalizing Factor Analysis and Canonical Correlation to Three-way Arrays Increases their Ability to Disentangle Information****La généralisation de l'analyse factorielle et de corrélation canonique aux tableaux à trois niveaux augmente leur capacité de démêler l'information**Richard HARSHMAN, *University of Western Ontario*

Data analysis is often improved when stronger models are applied to stronger data. We discuss one way of strengthening both: generalizing outer-product models from two-way (matrix) to three-way (array) form, and then applying them to datasets such as objects \times variables \times occasions, or multiple covariance matrices. Often, the three-way decomposition is 'essentially unique' without imposing orthogonality (or any other) constraints, eliminating rotational indeterminacies. Consequently, with appropriate data, one can recover, and/or relate across two arrays, approximations of the source patterns that originally generated the covariation, enabling new scientific applications (e.g., see Google: parafac). Three-way PCA/factor analysis by PARAFAC and canonical analysis by PARACCON (still under development) are discussed.

L'analyse de données est souvent améliorée lorsque les modèles les plus fiables sont appliqués aux données les plus fiables. Nous discutons une façon de renforcer tous les deux : la généralisation des modèles outer-product d'une forme bidimensionnelle (matrice) à une forme tridimensionnelle (tableaux), et puis les appliquer aux jeux de données comme des objets \times variables \times occasions, ou des matrices de covariance multiple. Souvent, la décomposition tridimensionnelle est « essentiellement unique » sans imposer les contraintes d'orthogonalité (ou toute autre), élimination des indéterminations de rotation. En conséquence, avec des données appropriées, on peut récupérer, et/ou associer à travers deux tableaux, approximations des modèles de source qui ont à l'origine produit la covariation, permettant de nouvelles applications scientifiques (par exemple, voir Google : parafac). L'analyse à trois niveaux de PCA/factor par PARAFAC et l'analyse canonique par PARACCON (toujours en cours de développement) sont discutées.

[MS-8]**Some Ideas About Forestry Data Visualization****Quelques idées au sujet de la visualisation de données reliées à la foresterie**Lutong ZHOU & Willard BRAUN, *University of Western Ontario*

The grid package (P. Murrell, 2005) is a tool for visualizing forestry data. The viewport, a feature of grid, makes it possible to visually compare groups of plots arranged spatially. Several viewports can be used to construct multiple caveplots (Becker et al, 1994) to view lightning and fire ignition data at different locations. They can also be used to modify pairwise scatterplots for exploring relationships among fire weather variables and danger ratings. The editing of grid objects (grobs) makes animating graphics in R easier. As an example, we introduce the Fengche plot for visualizing wind data.

Le « grid package » (P. Murrell, 2005) est un outil pour visualiser les données reliées à la foresterie. La fenêtre d'affichage, un dispositif de grille, rend possible la comparaison visuelle des groupes de graphiques disposés dans l'espace. Plusieurs fenêtres d'affichage peuvent être employées pour construire les « caveplots » multiples (Becker et al, 1994) pour regarder la foudre et les données reliées à l'inflammation du feu à différents endroits. Ils peuvent également être employés pour modifier les nuages de points appariés pour explorer les rapports entre des variables de conditions d'alerte météorologiques et les indices de danger. L'édition des objets de grille (grobs) facilite les graphiques d'animation dans R. Comme exemple, nous présentons le graphique (Plot) de Fengche pour la visualisation des données reliées au vent.

[MS-9]**ANOVA Extensions for Mixed Data
Extension de l'ANOVA aux données mixtes**Yongtao ZHU & Alexander DE LEON, *University of Calgary*

We derive exact likelihood ratio tests for so-called multi-sample location hypotheses in the context of mixed continuous and discrete data modeled according to the general location model (GLOM). The tests can be viewed as extensions of conventional normal-theory ANOVA tests to the general multivariate mixed data setting. They provide a global approach to the mixed-data location problem which unifies, by testing only one hypothesis, results from a more traditional analysis via separate multiple tests. In addition, by modeling the mixed variables using the GLOM, the proposed tests are able to explicitly account for correlations between the continuous and discrete variables, resulting in a more powerful analysis. The size and power of the exact likelihood ratio tests are studied and compared, through extensive simulations, against asymptotic tests proposed earlier by Nakanishi (2003), Morales et al. (1998) and Bar-Hen & Daudin (1995). Mardia et al.'s (1979) academic prediction data and Krzanowski's (1975) advanced breast cancer data are used to illustrate the application of the methodology.

Nous obtenons les tests exacts du rapport de vraisemblance pour les hypothèses de localisations dites à échantillons multiples dans le contexte de données mixtes (continues et discrètes) modélisées à l'aide du modèle général de localisation (MGL). Les tests peuvent être vus comme une extension des tests d'ANOVA conventionnels au cadre général de données mixtes multivariées. Ils fournissent une approche globale au problème de localisation de données mixtes qui unifie, en ne testant qu'une hypothèse, les résultats d'une analyse plus traditionnelle via de multiples tests séparés. De plus, en modélisant les variables mixtes en utilisant le MGL, les tests proposés peuvent explicitement prendre en considération les corrélations entre variables continues et discrètes, ce qui résulte en une analyse plus puissante. Le niveau et la puissance des tests exacts du rapport de vraisemblance sont étudiés et comparés, par le biais de simulations, aux tests asymptotiques proposés antérieurement par Nakanishi (2003), Morales et al. (1998) et Bar-Hen & Daudin (1995). Les données de prédictions académiques de Mardia et al. (1979) et les données de cas avancés de cancer du sein de Krzanowski (1975) sont utilisées afin d'illustrer l'application de cette méthodologie.

[MS-10]**A Resampling Approach to Multilevel Models
Une approche de rééchantillonnage pour les modèles à plusieurs niveaux**Zilin WANG, *Wilfrid Laurier University*, Mary THOMPSON & Zhichang JIANG, *University of Waterloo*

With the presence of unequal probability sampling in a multilevel model, although the use of survey weights results in design consistent estimators of the parameters, the weight inflated estimators for variance components can be biased when the sample size at the first level is fixed. In this presentation, we propose a resampling method which serves as an alternative to weight scaling methods, and provides empirical distributions for the underlying estimators of the variances. The small sample properties of the estimators of parameters for a multilevel model from this resampling method are studied through simulations using artificial populations from the literature.

En présence d'échantillonnage à probabilités inégales dans un modèle à plusieurs niveaux, bien que l'utilisation de poids d'enquête permettent l'obtention d'estimateurs de paramètres cohérents avec le plan d'expérience, les estimateurs de composantes de variances gonflés par les poids peuvent être biaisés lorsque la taille de l'échantillon est fixée au premier niveau. Dans cette présentation, nous proposons une méthode de ré-échantillonnage servant d'alternative aux méthodes d'échelonnement pondéré, et nous en présentons les distributions empiriques des estimateurs de variance. Les propriétés des estimateurs de paramètres des petits échantillons pour un modèle à plusieurs niveaux issus de cette méthode de ré-échantillonnage sont étudiées par des simulations utilisant des populations artificielles tirées de la littérature.

Session 01A **Lundi 29 mai • Monday, May 29, 08:30–08:45** **SSC2050**
 Inaugural Session
 Séance inaugurale

Session 01B **Lundi 29 mai • Monday, May 29, 08:45–10:00** **SSC2050**
 SSC Presidential Invited Address
 Allocution de l'invité du président de la SSC

[MS-11]

MythMath: From Google to Wikis and Blogs

MythMath : De Google à wikis et blogues

Ivars PETERSON, *Science News*

Ponder the number of Eskimo words for snow, the tragic fate of lemmings, the mysterious pervasiveness of the golden ratio, and other myths of the modern media age. How do such myths become unquestioned truths in the realm of public discourse? Sorting out the truths of these stories and eradicating statistical bloopers can be a tricky business, and search engines, wikis, and blogs add to the confusion. Here's an insider's look at how old and new media handle science news (and data) and propagate misinformation.

Considérez le nombre de mots esquimaux pour la neige, le destin tragique des lemmings, la mystérieuse omniprésence du nombre d'or (golden ratio), et d'autres mythes de l'âge moderne des médias. Comment se fait-il que de tels mythes deviennent des vérités indiscutables dans le domaine du discours public? La distinction des vérités de ces histoires et la suppression des gaffes statistiques peuvent être épineuses, et les moteurs de recherche, wikis, et blogues ajoutent à la confusion. Voici le regard d'un initié à la façon dont les médias nouveaux et traditionnels manipulent les nouvelles scientifiques (et les données) et propage la désinformation.

Session 02A **Lundi 29 mai • Monday, May 29, 10:30–12:00** **SH3345**
 Current Issues in the Analysis of Longitudinal Data
 Problèmes courants en analyse de données longitudinales

[MS-12]

Estimation of Marginal Characteristics for Event History Processes

Évaluation des caractéristiques marginales pour des processus sur l'historique des événements

Jerry LAWLESS & Richard COOK, *University of Waterloo*

Event history analysis deals with the occurrence of events over time for individuals in some population. Sometimes we wish to develop a detailed probabilistic model for an event process, but often we may be content to examine marginal features such as the probability of r events in a stated time interval, or perhaps just the expected number of such events. In this talk we describe some approaches for estimating marginal features, and then compare them in some common scenarios involving recurrent events.

L'analyse historique des événements traite de l'occurrence des événements dans le temps pour des individus dans une certaine population. Parfois, nous souhaitons développer un modèle probabiliste détaillé pour un tel processus des événements, mais souvent nous pouvons être satisfaits d'examiner seulement les dispositifs marginaux tels que la probabilité d'observer r événements dans un intervalle déterminé, ou le nombre espéré d'événements dans cet intervalle. Dans cet entretien nous décrivons quelques approches pour estimer les dispositifs marginaux, et puis les comparons dans quelques scénarios communs impliquant des événements récurrents.

[MS-13]**Recent Developments in Joint Models for Longitudinal and Survival Data.****Développements récents de modèles conjoints pour des données longitudinales et de survie**Jeremy TAYLOR & Wen YE, *University of Michigan*, Xihong LIN, *Harvard University*

In this talk I will present selected recent advances of joint models for longitudinal biomarker data and event times. Joint models present a general way to describe such data. From these general models a variety of possible issues can be addressed, including inference about the parameters in the survival model, inference about the parameters in the longitudinal model. The model can also provide information about whether the longitudinal biomarker could be useful as a surrogate endpoint or auxiliary variable in a clinical trial. The models can also provide a basis for imputation of missing longitudinal data or event times. The typical form of the longitudinal model is as a parametric random effects model, and the typical form of the survival model is a proportional hazards model where the hazard depends on the “current true value” of the longitudinal variable. Estimation can be performed either in a 2-stage procedure or in a likelihood based way (either MLE or Bayesian). Recent advances include extensions of the model formulation and alternative estimation methods. I will discuss extending the standard parametric longitudinal model using a semi-parametric smoothing spline and investigate its properties in a 2-stage estimation procedure. The EM algorithm for finding the MLE’s can be very slow, requiring numerical integration in the E-step. I will discuss an alternative approach in which a Laplace approximation is used, leading to a penalised likelihood that eliminates the need to estimate the baseline hazard function.

Dans cet entretien, je présenterai des avancées récentes des modèles conjoints pour les données longitudinales de marqueurs biologiques et des temps d’événement. Les modèles conjoints présentent une manière générale de décrire de telles données. De ces modèles généraux une variété d’issues possibles peut être adressée, y compris l’inférence au sujet des paramètres dans le modèle de survie, ainsi que dans le modèle longitudinal. Le modèle peut également fournir des informations pour décider si le marqueur biologique longitudinal pourrait être utile comme proxy ou comme variable auxiliaire dans un essai clinique. Les modèles peuvent également fournir pour l’imputation des données manquantes ou des temps d’événement. La forme typique du modèle longitudinal est comme un modèle paramétrique à effet aléatoire, et la forme typique du modèle de survie est le modèle à risques proportionnels où le risque dépend de la « vraie valeur » de la variable longitudinale. L’estimation peut être effectuée dans une procédure à deux phases ou sur la base de la vraisemblance (EMV ou bayésien). Les avancées récentes incluent les extensions de la formulation du modèle et des méthodes alternatives d’estimation. Je discuterai l’extension du modèle longitudinal paramétrique standard en utilisant un lissage semi paramétrique spline et étudierai ses propriétés dans une procédure d’estimation à deux phases. L’algorithme EM pour trouver les EMV peut être très lent, exigeant l’intégration numérique dans l’étape-E. Je discuterai une approche alternative dans laquelle une approximation de Laplace est utilisée, menant à une vraisemblance pénalisée qui élimine la nécessité d’estimer la fonction de risque de référence.

[MS-14]**On Measurement Error Models and Complex Data Structures****Sur les modèles d’erreur de mesure et les structures de données complexes**Paul GUSTAFSON, *University of British Columbia*

One common problem with exposure-outcome data is poor measurement of exposure. Misleading inferences can arise from pretending that the poor measurements are accurate, and in general there is a large literature on methods which adjust for measurement error or misclassification in explanatory variables. I will discuss methods for such adjustment, with emphasis on their adaptation to complex

Un problème courant avec les données de type exposition-effet est la piètre qualité des mesures d’exposition. Des conclusions erronées peuvent résulter du fait de considérer ces mesures de mauvaise qualité comme précises. Il existe une vaste littérature sur les méthodes d’ajustement aux erreurs de mesure et à la mauvaise classification des variables explicatives. Je discuterai des méthodes permettant un tel ajustement en mettant l’accent sur leur adaptation aux struc-

data structures, such as longitudinal data. Other emphases include the debate on how ‘parametric’ one should be when adjusting for measurement error, and concern about the strength of assumptions needed for the sake of identifiability.

tures de données complexes, telles que les données longitudinales. La discussion portera, entre autres, sur les hypothèses requises pour l’identifiabilité et jusqu’à quel point on doit être « paramétrique » lorsqu’on ajuste pour la présence d’erreurs de mesure.

[MS-15]

Some Methods on Longitudinal Data Analysis

Quelques méthodes sur l’analyse de données longitudinales

Grace Yi, *University of Waterloo*

Longitudinal data analysis has attracted considerable research interest. Various methods, such as likelihood, imputation, and weighted estimating equations approaches, have been developed to address distinct features of longitudinal studies. Often, the interest lies in featuring correlation among repeated measurements and missingness of longitudinal data. However, measurement error in covariates is another typical feature. It is often the case that some covariates are subject to error due to their own nature or the mechanism of measurement. Common examples include biological measures such as blood pressure and serum cholesterol level. It is known that standard inferential procedures are biased if such measurement error is not taken into account. In this talk, I will briefly review existing inference methods, and then focus the discussion on some marginal methods that facilitate different features of longitudinal data.

L’analyse de données longitudinales a engendré un nombre considérable de recherches. Diverses méthodes, telles que la vraisemblance, l’imputation, et les approches d’équations d’estimations pondérées, ont été développées pour adresser les particularités distinctes des études longitudinales. Souvent, l’intérêt se situe les particularités de la corrélation parmi les mesures répétées et les données longitudinales manquantes. Cependant, l’erreur de mesure dans les covariables est une autre particularité typique. Il est fréquent que quelques covariables soient erronées à cause de leur nature propre ou à cause du mécanisme de la mesure. Des exemples incluent des mesures biologiques telles que le niveau de tension artérielle et le niveau du cholestérol dans le sérum. On sait que les procédures inférentielles usuelles sont biaisées si une telle erreur de mesure n’est pas prise en considération. Dans cet entretien, je réexamine brièvement des méthodes existantes et concentre la discussion sur quelques méthodes marginales qui facilitent différentes particularités des données longitudinales.

Session 02B

Lundi 29 mai • Monday, May 29, 10:30–12:00

SSC2050

The Statistics of Magnetic Resonance Imaging

Statistique pour l’imagerie par résonance magnétique

[MS-16]

Deformation Based Morphometry, Roy’s Maximum Root and Recent Advances in Random Fields

Morphométrie basée sur la déformation, la racine maximale de Roy, et des avancées récentes en champs aléatoires

Jonathan TAYLOR, *Stanford University*, Keith WORSLEY, *McGill University*, Robert ADLER, *Technion*

The starting point of our talk is a study of anatomical differences between controls and patients who have suffered non-missile trauma. We use a multivariate linear model at each location in space, using Hotelling’s T^2 to detect differences between cases and controls. If we include further covariates in the model, Roy’s maximum root is a natural general-

Le point de départ de notre exposé est une étude des différences anatomiques entre un groupe contrôle et des patients ayant souffert d’un traumatisme autre que par un projectile. Nous utilisons un modèle linéaire multivarié à chaque point dans l’espace, en utilisant le T^2 de Hotelling pour détecter les différences entre les patients et le groupe contrôle. Si nous ajoutons des covariables supplémentaires au modèle, la ra-

ization of Hotelling's T^2 . This leads to the Roy's maximum root random field, which includes many special types of random fields: Hotelling's T^2 , T , and F , so, in effect the Roy's maximum root random field "unifies" many different random fields. This leads to the recent advances in random fields. We describe some recent advances both in the "theory" and "application" of smooth random fields, particularly the behaviour of the maximum of a smooth random field and an integral-geometric "recipe" for using the EC approximation related to a "probability of tubes formula" rather than Weyl's better known "volume of tubes formula" which has been applied to similar problems.

cine maximale de Roy est une généralisation naturelle du T^2 de Hotelling. Ceci nous mène au champ aléatoire de la racine maximale de Roy, qui inclut plusieurs types de champs aléatoires : le T^2 , T et F de Hotelling. Le champ aléatoire de la racine maximale de Roy « unifie » donc plusieurs champs aléatoires différents.

Ceci nous mène aux récentes avancées en champs aléatoires. Nous décrivons quelques progrès dans la « théorie » et l'« application » des champs aléatoires lisses, particulièrement le comportement du maximum d'un champ aléatoire lisse et une « recette » intégrale-géométrique pour utiliser l'approximation de la caractéristique d'Euler liée à une « formule de probabilité des tubes » plutôt que la « formule de volume des tubes » de Weyl, plus connue, et qui fut appliquée à des problèmes similaires.

[MS-17]

Our entry in the Functional Imaging Analysis Contest

Notre participation dans le concours d'analyse d'imagerie par résonance magnétique fonctionnelle

Keith WORSLEY, *McGill University*, Jonathan TAYLOR, *Stanford University*

Last June the Organisation for Human Brain Mapping organised its first functional magnetic resonance imaging (fMRI) analysis contest. Six teams submitted a statistical analysis of an fMRI experiment on 16 subjects, where each subject listened to different sentences read by different speakers in a two-way factorial design. The object was to detect which brain regions were activated by the different speakers, different sentences, and an interaction between the two. We describe our contest entry, which used hierarchical mixed effects models for both magnitudes and delays, spatial smoothing to increase effective degrees of freedom, discrete local maxima, and random field theory.

En juin dernier, l'Organisation pour la Cartographie Cérébrale a organisé son premier concours pour l'analyse d'imagerie par résonance magnétique fonctionnelle (IRMf). Six équipes ont soumis une analyse statistique d'une expérience d'IRMf sur 16 sujets, où chaque sujet a écouté différentes phrases lues par différents orateurs dans une expérience factorielle à 2 facteurs. L'objet était de détecter quelles régions du cerveau ont été activées par les différents orateurs, les différentes phrases, et une interaction entre les deux. Nous décrivons notre entrée de concours, qui a employé les modèles à effets mixtes hiérarchiques pour amplitude et délai, le lissage spatial pour augmenter les degrés de liberté efficaces, les maxima locaux discrets, et la théorie aléatoire des champs.

Session 02C

Lundi 29 mai • Monday, May 29, 10:30–12:00

SSC2036

Social Statistics

Statistique sociale

[MS-18]

The Direct and Indirect Effects of Causal Systems with Longitudinal Data

Causalité directe et indirecte avec les données longitudinales

Rajulton FERNANDO, *University of Western Ontario*

Identifying causal processes is one of the major aims of collecting longitudinal data. In practice, however, little attention is being paid to this. This

La collecte de données longitudinales a pour but d'identifier les processus de causalité. En pratique, cependant, on accorde peu d'attention à cet objectif. Ainsi cet article présente

paper therefore outlines two major considerations in longitudinal studies: a) examining the structure of relationships not only between explanatory and criterion variables but also among the explanatory variables themselves—specially among categorical variables which are the most frequent type of variables collected by longitudinal surveys; and, b) formulating ways of identifying direct and indirect effects in a multivariate context. The paper will thus illustrate the importance of (chain) graphical models for longitudinal research.

deux réflexions majeures pour les études longitudinales : a) l'examen de la structure des relations, non seulement entre les variables explicatives et dépendantes, mais aussi entre les variables explicatives elles-mêmes, et plus spécifiquement entre les variables catégoriques, qui sont les plus fréquemment collectées dans les enquêtes longitudinales et b) la formulation de méthodes pour identifier les effets directs et indirects dans un contexte multivarié. Cet article va donc illustrer l'importance des modèles graphiques pour la recherche longitudinale.

[MS-19]

Tracking the Mental Health Consequences of Divorce: A Growth Curve Model Approach

Observation des conséquences du divorce sur la santé mentale : une approche utilisant un modèle avec courbe de croissance

Lisa STROHSCHN, *University of Alberta*

Although divorce has been conceptualized as a process, researchers have made little progress in distinguishing selection effects from the short and long term consequences of divorce. Using growth curve methods to analyze five waves of data from the 1994–2002 National Longitudinal Study of Children and Youth (NLSCY), this study prospectively tracks children living with two biological parents at initial interview in 1994, and compares the mental health trajectories of children from households that experience divorce during this time period with the trajectories of children whose households remain intact. In addition to identifying patterns of adjustment for children of divorce, I evaluate factors that predict whether levels of depression for children abate or persist in the years following divorce and discuss the implications of using growth curve models to address these issues.

Bien que le divorce ait été conceptualisé comme un processus, les chercheurs ont accompli peu de progrès dans la distinction des effets de sélection des conséquences à court et à long terme du divorce. En utilisant des méthodes de courbe de croissance pour analyser cinq vagues de données de l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ), cette étude observe prospectivement des enfants vivant avec deux parents biologiques lors de l'entrevue initiale en 1994, et compare l'évolution de la santé mentale des enfants des ménages qui vivront un divorce pendant cette période de temps avec l'évolution de la santé mentale des enfants dont les ménages demeurent intacts. En plus d'identifier des modèles d'ajustement pour les enfants de parents divorcés, j'évalue les facteurs qui prévoient si les niveaux de la dépression pour des enfants diminuent ou persistent durant les années suivant le divorce et je discute des implications d'employer des modèles de courbe de croissance pour aborder ces questions.

[MS-20]

Identifying Risk and Protective Factors Using Longitudinal Population Based Survey Data: Improving our Knowledge of Smoking Behaviour Development

Identification du risque et des facteurs protecteurs en utilisant des données longitudinales d'une population : amélioration de notre connaissance du développement du comportement des fumeurs

Bradley CORBETT, *Statistics Canada*

The study of human behaviour is a complex undertaking but remains as one of the most fascinating questions in society today: What influences someone to purchase product A over product B? Why do people engage in unhealthy behaviours in spite of

L'étude du comportement humain est une entreprise complexe mais ce sujet reste une des questions les plus fascinantes dans la société aujourd'hui : Qu'est ce qui influence quelqu'un à acheter le produit A plutôt que le produit B ? Pourquoi les gens s'engagent-ils dans des comportements

the inherent dangers? How does someone decide to participate in an activity that is contrary to their own principals and ethics? These and other similar questions can be found in nearly every academic discipline.

Recent improvements in analytical methods and behavioural theory have converged to provide a reasonable approach for developing models of human behaviour. This presentation will demonstrate how health behaviour theory and multi-level growth models can be used to improve the practice of smoking uptake prevention in Canadian adolescents. In particular, participants will learn how to separate risk and protective factors from concomitant or parallel factors in the study of human behaviour.

malsains malgré les dangers inhérents ? Comment est-ce que quelqu'un décide de participer à une activité qui est contraire à ses propres principes d'éthique ? Celles-ci et d'autres questions semblables peuvent être trouvées dans presque chaque discipline universitaire.

Les améliorations récentes des méthodes analytiques et de la théorie comportementale ont convergé pour fournir une approche raisonnable pour les développements des modèles du comportement humain. Cette présentation démontrera comment la théorie du comportement sanitaire et les modèles de croissance à plusieurs niveaux peuvent être utilisés pour améliorer la pratique de prévention de la reprise de tabagisme chez les adolescents canadiens. En particulier, les participants apprendront comment séparer le risque et les facteurs protecteurs des facteurs concomitants ou parallèles dans l'étude du comportement humain.

Session 02D

Lundi 29 mai • Monday, May 29, 10:30–12:00

SH3315

Parallel Computing in Statistics

Calcul parallèle en statistique

[MS-21]

Parallelism in Interface, Design, and Nonstationary Surrogate Modeling of Supercomputer Experiments
Parallélisme dans l'interface, conception, et modélisation avec suppléant non stationnaire des expériences sur super-ordinateurs

Robert GRAMACY, *University of Cambridge*, Herbert K. H. LEE, *University of California at Santa Cruz*

This talk explores nonstationary modeling methodologies for the surrogate modeling and the sequential design of computer experiments. The approach combines design techniques from Statistics and Machine Learning with treed partitioning and Bayesian inference. Parallelism is encountered as (1) a tool for efficient inference with MCMC methods, and (2) as a challenge inherent in interfacing complex statistical models with optimal design for experiments run on highly-parallel, asynchronous, supercomputers. The benefit using “opportunistic” rather than “optimal” sequential design strategies is illustrated through synthetic and real data, including a motivating example involving the computational fluid dynamics simulation of a NASA reentry vehicle.

Cet entretien explore les méthodologies de modélisation non stationnaire pour la modélisation suppléante et la conception séquentielle des expériences informatiques. L'approche présentée combine les techniques de conception de plans d'expérience utilisées en statistique et du compartimentage par arbres et l'inférence bayésienne de l'apprentissage automatique. Le parallélisme est utilisé comme (1) outil pour l'inférence efficace avec les méthodes MCMC, et (2) comme défi inhérent à l'établissement d'un lien entre les modèles statistiques complexes et la conception de plans optimaux pour les expériences faites sur des super ordinateurs hautement parallèles et asynchrones. L'avantage d'utiliser des stratégies de conception de plans séquentiels « opportunistes » plutôt que « optimaux » est illustré à l'aide de données simulées et réelles, y compris un exemple comportant une simulation de la dynamique des fluides d'un véhicule de rentrée de la NASA.

[MS-22]**Parallel Simulations for Sea Ice Modelling****Simulations parallèles pour la modélisation de la glace de mer**Theodoro KOULIS, *McGill University*

The temporal variability of seasonal sea ice at the poles is complex and closely linked to the Earth's climate. The amount of sea ice can have a significant effect on the energies transferred between the atmosphere and the ocean. Modelling the dynamics of the Earth's seasonal sea ice processes is therefore of great interest to scientists. We use a simple two-parameter model on a grid to analyze sea ice concentrations derived from Earth-orbiting satellites. Despite the apparent simplicity of the model, calculations of the model probabilities are cumbersome. With the aid of parallel simulations, we show how one can obtain estimates of the model probabilities, which in turn help us estimate the model parameters.

La variabilité temporelle de la glace de mer saisonnière aux pôles est complexe et intimement liée au climat terrestre. La quantité de glace de mer peut avoir un effet significatif sur les transferts d'énergie entre l'atmosphère et l'océan. La modélisation de la dynamique des processus de glace de mer saisonnière de la Terre est donc d'un grand intérêt pour les scientifiques. Nous utilisons un simple modèle à deux paramètres sur une grille pour analyser les concentrations de glace de mer obtenues par satellites. Malgré l'apparente simplicité du modèle, le calcul des probabilités est difficile. À l'aide de simulations parallèles, nous montrons comment obtenir des estimés des probabilités du modèle, ce qui nous aide à estimer les paramètres du modèle.

[MS-23]**A Bias Reduction Technique for Monte Carlo Pricing of Early Exercise Options****Une technique de réduction de biais pour des estimateurs de prix de type Monte-Carlo pour les premières options calculées**Mark REESOR, TYSON WHITEHEAD & Matt DAVISON, *University of Western Ontario*

We present a method for reducing the bias in Monte Carlo estimators of the price of American-style contingent claims. At each exercise opportunity in a time discretization, we assume an unbiased estimator of the hold value of the claim is available. The central limit theorem allows one to derive an expression for the approximate bias. This expression is estimated from the simulation, allowing for computation of bias-reduced estimators of the claim value. The method is applicable to both biased-high and biased-low estimators and is computationally much more efficient than a bootstrap approach to reducing bias.

Nous présentons une méthode pour réduire le biais dans les estimateurs de Monte-Carlo du prix des réclamations contingentes en style américain. À chaque occasion d'exercice dans une discrétisation de temps, nous supposons qu'un estimateur sans biais de la valeur retenue de la réclamation est disponible. Le théorème limite central permet de dériver une expression approximative pour le biais. Cette expression est estimée à partir de la simulation, permettant d'évaluer les estimateurs à biais réduit de la valeur de réclamation. La méthode est applicable aux estimateurs à biais élevé et aux estimateurs à faible biais et elle est beaucoup plus efficace au niveau informatique qu'une approche de Bootstrap pour la réduction du biais.

Session 02E**Lundi 29 mai • Monday, May 29, 10:30–12:00****SSC2024****The Challenges of Being an Environmental Statistician****Défis pour le statisticien environnemental****[MS-24]****Statistical Data Assimilation for Environmental Prediction****Assimilation de données statistiques à des fins de prédictions environnementales**Michael DOWD *Dalhousie University*

In this presentation I overview some challenges in developing statistical approaches for combining observations with deterministic mathematical models in the environmental sciences (a problem known as data assimilation). Two main elements define these estimation problems: (i) complex data types often having an indirect relation to the state (e.g. satellite observations), and (ii) high dimensional, nonlinear models used as the basis for prediction systems (e.g. weather models). I review the current state of the art for data assimilation in the ocean and atmospheric sciences. New directions are then explored based on ensemble (Monte Carlo) methods. Illustrative examples are drawn from numerical weather prediction and biological oceanography.

Dans cette présentation, je survole quelques difficultés dans le développement d'approches statistiques pour combiner des observations et des modèles mathématiques déterministes dans les sciences environnementales (un problème nommé assimilation de données). Deux principaux éléments définissent ces problèmes d'estimation : (i) des types de données complexes ayant parfois un lien indirect avec l'état (par exemple, des observations satellites), et (ii) des modèles non linéaires à haute dimensionnalité, utilisés comme base pour des systèmes de prédiction (par exemple, les modèles météorologiques). Je révisé les méthodes de pointe pour l'assimilation de données en sciences atmosphériques et de l'océan. De nouvelles directions sont ensuite explorées basées sur des méthodes Monte-Carlo. Des exemples illustratifs sont tirés de la prévision météorologique numérique et de l'océanographie biologique.

[MS-25]**Some Challenges in Measuring and Communicating National Environmental Performance****Quelques défis associés à la mesure et à la communication de la performance environnementale nationale**Tanja SREBOTNJAK, *Yale University*

The Environmental Performance Measurement Project aims to shift environmental decision-making onto firmer analytic foundations using environmental indicators and statistics. The project produces the Environmental Sustainability Index that tracks countries on 21 sustainability indicators and the new Environmental Performance Index that assesses key environmental policy outcomes using policy targets linked to the Millennium Development Goals. This presentation addresses some of the challenges the project team experiences in its work, specifically the tightrope walk environmental statisticians must master in terms of the rigorous analysis of environmental data and decision-makers' requests for straightforward analytical approaches and clear policy messages.

Le Projet de mesure environnemental vise à promouvoir les prises de décisions concernant l'environnement à partir de bases analytiques plus fiables qui utilisent des indicateurs environnementaux et des statistiques. Le projet produit l'indice de développement durable environnemental qui classe des pays sur 21 indicateurs de durabilité et sur le nouvel indice de performance environnementale qui évalue les résultats clés d'une politique environnementale en utilisant les cibles de cette politique, liées aux objectifs de développement du Millénaire. Cette présentation discute certains des défis que l'équipe a éprouvés dans son travail, spécifiquement l'équilibre délicat que les statisticiens environnementaux doivent maintenir entre des analyses rigoureuses de données environnementales, les demandes des décideurs concernant l'utilisation d'approches analytiques directes et le besoin d'une politique clairement explicitée.

[MS-26]**Decisions, Uncertainty, and the Law: Statistical Considerations of Environmental Risk Management.
Les décisions, l'incertitude, et la loi : considérations statistiques en gestion de risque écologique**Peter GUTTORP, *University of Washington*

Air quality laws are aimed at protecting individuals from harmful exposures. We discuss some aspects of the US Clean Air Act and its implementation. From a statistical point of view one may regard the issue as one of testing a statistical hypothesis. We develop statistical properties of such a hypothesis test, and of the US EPA implementation of it. Extensions beyond the network of monitoring stations can be implemented, and used for what a statistician might view as a better approach to setting air quality standards.

Les lois portant sur la qualité de l'air visent à protéger les individus à l'exposition de substances nocives. Nous discutons de quelques aspects de l'« US Clean Air Act » et de son implantation. D'un point de vue statistique, nous pouvons considérer ce problème comme un test d'hypothèse. Nous développons les propriétés statistiques d'un tel test d'hypothèse, et de son implantation par l'EPA américaine. Des étendues au-delà du réseau des stations de surveillance peuvent être implantées et utilisées pour ce qu'un statisticien pourrait voir comme une meilleure approche à la détermination des standards concernant la qualité de l'air.

Session 02F**Lundi 29 mai • Monday, May 29, 10:30–12:00****SSC3026****Case Study I: Record Linkage and Case Study II: Obstructive Sleep Apnea****Étude de cas I : Couplage d'enregistrements et Étude de cas II : L'apnée obstructive du sommeil****[MS-27]****Case Study I: Record Linkage: Maintaining a Registry****Étude de cas I : Couplage d'enregistrements : maintenance d'un registre**Peggy NG, *York University*

Data linkage falls into two large categories: statistical matching and exact matching. In many research and administrative situations the need arises to link two or more data sources together. The student assignment is given in progressively complex steps. First the student will use record linkage techniques to 'update' a population registry. Next the student can estimate linkage and error rates.

Le couplage des données peut se diviser en deux grandes catégories : couplage statistique et couplage exact. Dans de nombreuses situations de recherche et administratives, il est nécessaire de coupler deux ou plusieurs sources de données. L'étudiant devra compléter l'exercice par des étapes de plus en plus complexes. Dans un premier temps, il utilisera des techniques de couplage d'enregistrements pour « mettre à jour » un registre de population. Ensuite, il pourra évaluer les taux de couplage et d'erreur.

[MS-28]**Case Study II: Obstructive Sleep Apnea****Étude de cas II : L'apnée obstructive du sommeil**Alison GIBBS, *University of Toronto*

Obstructive sleep apnea is characterized by the cessation of breathing in sleep lasting more than 10 seconds. If left untreated it has been shown to result in an elevated mortality rate as well as decreased quality of life. Recently the healthcare system has mandated that having two nights of diagnos-

L'apnée obstructive du sommeil se caractérise par un arrêt de la respiration de plus de 10 secondes pendant le sommeil. En l'absence de traitement, le patient risque une mortalité accrue et une qualité de vie moindre. Le système de soins de santé a récemment ordonné qu'une nuit de tests suffirait au diagnostic plutôt que deux. Il est donc devenu nécessaire

tic sleep studies should be discontinued in favour of a single night. It is thus necessary for physicians to supplement the information from sleep studies with a variety of questionnaires in order to minimize missed diagnoses of sleep apnea. The data for this case study were collected from a retrospective review of patient charts from the sleep clinic at Toronto Western Hospital. The purpose is to examine the effectiveness of the Berlin Questionnaire, in combination with other questionnaires and with sleep studies, in diagnosing sleep apnea.

pour les médecins de compléter les informations tirées de ces tests par une variété de questionnaires afin de minimiser les erreurs de diagnostic d'apnée du sommeil. Les données utilisées pour cette étude de cas proviennent d'une revue rétrospective de dossiers de patients de la clinique de sommeil de l'hôpital Western de Toronto. Le but est d'examiner l'efficacité du Questionnaire de Berlin, en association avec d'autres questionnaires et études du sommeil, pour diagnostiquer l'apnée du sommeil.

Session 02G**Lundi 29 mai • Monday, May 29, 10:30–12:00****SSC3006**

Applications of Statistics I

Applications de la statistique I

[MS-29]

Developing a Daily Agro-climate Interpolation Grid for Canada: Data Verification and Model Validation
Développement d'une grille d'interpolation quotidienne pour l'agro-climat du Canada : vérification de données et validation des modèles

Nathaniel NEWLANDS, *Agriculture and Agri-Food Canada*, Harvey HILL, *Climate Impacts and Adaptation, PFRA-Agriculture and Agri-Food Canada*, Allan HOWARD, *National Agroclimate Information Service, Agriculture and Agri-Food Canada*

The use of daily climate data in agriculture has increased considerably over the past two decades as a result of the rapid development of information technology and the need to assess risk of future extreme weather due to climate change. Such events have the potential to impact agricultural production and environmental sustainability affecting soil, air and water resources. We describe the construction of an agroclimatic database of high spatio-temporal resolution (i.e., daily, 10 km) with coverage across agricultural regions. This database is intended for use by a wide range of end users concerned with agricultural management, planning, and policy.

L'utilisation de données climatiques sur une base quotidienne a considérablement augmenté au fil des deux dernières décennies. Ceci est dû à un développement rapide de l'informatique et aussi d'une demande pour évaluer les risques de climat extrême due au changement climatique. Ces types de changements ont le potentiel d'avoir un impact sur l'agriculture et la viabilité de l'environnement affectant le sol, l'air et l'eau. Nous décrivons ici la mise en place d'une base de données agro-climatiques de haute résolution spatio-temporelle (c'est-à-dire, quotidiennement, 10 km) couvrant différents régions agricoles. Cette base de données sera utilisée par ceux concernés par la gestion et la planification agricole et aussi par ceux qui développent les plans d'actions.

[MS-30]**Statistical Inferences Problems in High Energy Physics****Problèmes statistiques d'inférences en physique de forte énergie**

Zi JIN & Zheng ZHENG, *University of Toronto*

We describe some specific problems of statistical inference in the presence of nuisance parameters that are motivated by problems arising in particular physics. Our approach is based on higher order likelihood inference to tail areas in the pres-

Nous décrivons quelques problèmes spécifiques d'inférence statistique en présence de paramètres de nuisance motivés par des problèmes venant de la physique des particules. Notre approche repose sur l'inférence d'ordre supérieur basée sur la vraisemblance en présence de paramètres de nui-

ence of nuisance parameters, using recently developed third order approximations. Orthogonal parameters and adjustments to profile likelihood are also discussed. Practical examples arising in particle physics are outlined, and simulation studies show that the third order likelihood based approximations provide nice solutions and are one of the most promising methods for the examples.

sance et utilise les approximations du troisième ordre récemment développées. L'orthogonalité des paramètres et les ajustements permettant de profiler la vraisemblance sont également abordés. Des exemples pratiques venant de la physique des particules sont décrits. Des études par simulation démontrent que les approximations du troisième ordre basées sur la vraisemblance fournissent des solutions intéressantes et font partie des méthodes les plus prometteuses pour les exemples considérés.

[MS-31]

Point Process Modeling of Choice Reaction Time Data

Modélisation de données du temps de réaction à choix multiples par un processus ponctuel

Jennifer ASIMIT & W. John BRAUN, *University of Western Ontario*, William SIMPSON, *Defence Research and Development Canada*

A parametric model based on point process operations is developed to aid in the analysis of choice reaction time (RT) data. In a choice RT experiment black and white flashes are presented to an observer, who hits one of two buttons after perception of a flash, depending on the type of flash perceived. Expressions for the first and second order intensity functions are derived. The theoretical curves and estimated intensity curves (obtained using kernel methods) can then be compared.

Un modèle paramétrique basé sur des opérations d'un processus ponctuel est développé pour faciliter l'analyse de données du temps de réaction (TR) à choix multiples. Dans une expérience sur les TR à choix multiples, des cartes noires et blanches sont présentées à un observateur qui doit peser sur le bouton approprié en fonction de la carte montrée. Des expressions pour les fonctions d'intensité du premier et du second ordre sont dérivées. Les courbes théoriques et les courbes d'intensité estimées (obtenues en utilisant des méthodes de noyau) peuvent alors être comparées.

[MS-32]

Statistical Behaviour of a Spatial Renewal Process

Comportement statistique d'un processus de renouvellement spatial

Katherine DAVIES, *University of Western Ontario*

The one-dimensional renewal process has been extensively researched and investigated, and its importance in applications is well-understood. Recently, a multi-dimensional analogue of the renewal process has been proposed by Ivanoff and Merzbach (2005). It has been suggested that the two-dimensional version of this process could be used to model the spread of forest fires. The purpose of the current talk is to discuss this model from a statistical point of view. Using simulation, we study the behaviour of various tests of complete spatial randomness under the hypothesis of the spatial renewal model. We also consider the appropriateness of this kind of model for forestry applications.

Le processus de renouvellement unidimensionnel a été largement étudié, et son importance dans les applications est bien connue. Récemment, une approche multidimensionnelle du processus de renouvellement a été proposée par Ivanoff et Merzbach (2005). On a suggéré que la version bidimensionnelle de ce processus pourrait être employée pour modéliser la propagation des feux de forêt. Le but de cet exposé est de discuter ce modèle d'un point de vue statistique. À l'aide de simulations, nous étudions le comportement statistique sous l'hypothèse d'un processus de renouvellement spatial. Nous discutons aussi la pertinence d'utiliser ce genre de modèle pour des applications reliées à la foresterie.

[MS-33]

Modelling Lightning and Lightning Caused Forest Fires
Modélisation de la foudre et des feux de forêt qu'elle cause

Douglas WOOLFORD, Reg KULPERGER & W. John BRAUN, *University of Western Ontario*

We wish to relate forest fire ignitions to lightning stroke occurrences through the analysis of Ontario lightning and fire data, supplied by the Ontario Ministry of Natural Resources. We propose a mode-seeking clustering algorithm based on a convergent form of data sharpening methods. The algorithm is used to both reduce the lightning data to local spatial-temporal modes, and to estimate the short-term spatial track of a lightning storm system. These estimated tracks, storm centers and corresponding clusters are used as the foundation for preliminary models of lightning strike and fire ignition patterns in North-Western Ontario.

Nous souhaitons mettre en relation l'allumage des feux de forêt et la foudre à travers l'analyse de données du ministère des ressources naturelles d'Ontario. Nous proposons un algorithme de classification détecteur de modes basé sur une forme convergente de méthodes d'affinage de données. L'algorithme permet de réduire les données concernant la foudre à des modes spatio-temporaux locaux, ainsi que d'estimer, à court terme, le parcours de la tempête. Ces parcours estimés, les centres des tempêtes et les grappes correspondantes servent de fondation aux modèles préliminaires de la distribution de la foudre et des feux de forêt au nord-ouest de l'Ontario.

[MS-34]

Zero-truncated and Zero-inflated Regression Models to Predict the ROR Crash Data
Modèles de régression zéro-tronqués et zéro-majorés pour la prédiction de données sur les accidents ROR

B. M. Golam KIBRIA, *Florida International University*

We have considered several models to fit the count data that we encounter in the field of Health Science, Transportation Engineering, and Social Sciences. Several regression models, for examples, Poisson (PO), Zero-Inflated Poisson (ZIP), Zero truncated poisson (ZTP), Negative Binomial (NB), Zero-Inflated Negative Binomial (ZINB) and Zero Truncated NB (ZTNB) regression models are fitted to predict the run-off-road (ROR) crash data which were collected on arterial roads in south west region (urban) of Florida State. STATA 9.0 software has been used to analyze the data. From the empirical study we concluded that ZINB, ZTNB and ZIP models performed better than PO, ZTP and NB models for the zero inflated and over dispersed count data.

Nous avons considéré plusieurs modèles fréquemment employés pour la modélisation de données discrètes telles que rencontrées dans les sciences de la santé, les technologies des transports et les sciences sociales. Plusieurs modèles de régression sont ici considérés : Poisson (PO), Poisson Zéro-Majoré (PZM), Poisson Zéro-Tronqué (PZT), Binomial Négatif (BN), Binomial Négatif Zéro-Majoré (BNZM) et Binomial Négatif Zéro-Tronqué (BNZT). Ces modèles sont employés pour la prévision des accidents de la route à l'aide de données sur les sorties de route amassées sur des routes principales de la région (urbaine) occidentale du sud de l'état de la Floride. Le logiciel STATA 9.0 a été utilisé pour analyser les données. De l'étude empirique, nous avons conclu que les modèles BNZM, BNZT et PZM semblent plus efficaces que les modèles PO, PZT et BN dans le contexte de données zéro-majorées et hyperdispersées.

Session 02H**Lundi 29 mai • Monday, May 29, 10:30–12:00****SSC3014**

Business and Industrial Statistics
Statistique industrielle et de gestion

[MS-35]**Sequential Monitoring of Process Capability Indices****Supervision séquentielle des indices de capacité d'un processus**Abdulkadir HUSSEIN, *University of Windsor*

Based on two process capability indices that are well accepted in the industrial process control literature, I derive sequential test procedures for testing whether or not an industrial process is capable. I derive test statistics and their large-sample properties both under the null hypothesis and the alternative of interest. Through numerical examples, I analyze the performance of the procedures in terms of power and average sample size. The procedures are shown to be economical in the sense that they require much less sample than equally powerful non-sequential testing procedure.

En me basant sur deux indices de capacité des processus bien acceptés dans la littérature de contrôle des processus industriels, j'obtiens des procédures de tests séquentiels testant si un processus industriel est capable ou non. J'obtiens les statistiques de test ainsi que leurs propriétés en présence d'un grand échantillon, sous l'hypothèse nulle et l'alternative d'intérêt. Sur la base d'exemples numériques, j'analyse la performance des procédures en termes de puissance et de taille moyenne d'échantillon. Les procédures sont économes dans le sens où elle requièrent des échantillons beaucoup moins grands, à puissance égale, que les tests non séquentiels.

[MS-36]**A Robust Estimation of Process Capability Index****Une estimation robuste de l'indice de capacité d'un processus**Razi GHORI & S. Ejaz AHMED, *University of Windsor*

Process Capability Indices (PCI) are useful tools for measuring the potential capability of an industrial process and C_p is a commonly used index in managerial decisions. The main assumption underlying this index is the normality of the process under investigation. However, this assumption is not realistic for many real applications. In this talk, we suggest a robust estimator of C_p , based on Gini's mean difference. The asymptotic distribution of the suggested index is derived. The relative performance of the proposed estimator with the benchmark estimator based on sample standard deviation, is assessed theoretically as well as numerically. In addition to asymptotic confidence interval, three kind of bootstrap confidence intervals are presented. The probability coverages of confidence intervals based on inferential quantities associated with these estimators are evaluated using Monte Carlo simulation for small, moderate and large sample sizes. The proposed methodology is extended to other indices such as C_{pm} .

Les Indices de Capacité d'un Processus (ICP) sont des outils utiles pour mesurer les capacités potentielles d'un processus industriel et le C_p est généralement un indice utilisé dans des décisions de gestion. L'hypothèse principale sous-jacente à cet indice est la normalité du processus à l'étude. Cependant, cette hypothèse n'est pas réaliste pour plusieurs applications réelles. Dans cet entretien, nous suggérons un estimateur robuste de C_p , basé sur la différence moyenne de Gini. La distribution asymptotique de l'indice suggéré est obtenue. La performance relative de l'estimateur proposé avec l'estimateur de référence basé sur l'écart type échantillonnaire est évaluée théoriquement aussi bien que numériquement. En plus de l'intervalle asymptotique de confiance, trois genres d'intervalles de confiance bootstrap sont présentés. Les probabilités de couverture des intervalles de confiance basés sur des quantités inférentielles liées à ces estimateurs sont évaluées en utilisant la simulation de Monte-Carlo pour des échantillons petits, moyens et grands. La méthodologie proposée est étendue à d'autres indices tels que le C_{pm} .

[MS-37]**Special Mixture Discriminant Model for Drug Discovery Data
Modèle de mélange discriminant en recherche de médicaments**

Xu Sunny WANG, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, William J. WELCH, *University of British Columbia*

In drug discovery, statistical models have played a very important role for predicting activity of compounds against biological targets. In this supervised learning problem, descriptors of molecular structure (e.g. atomic weight, types of bonds, many other exotic characteristics) are used to predict activity. Compounds often cluster in the descriptor space, making mixture discriminant analysis a natural supervised learning tool. Conventional mixture discriminant methods have difficulty finding the best model for the drug discovery data. Complex structure and high dimensionality lead to overparameterized models. It is believed that the biological activity of compounds only depends on several descriptors, so we introduce a special mixture model, which has fewer parameters, and seeks to predict using multiple subspaces (i.e., multiple mechanisms). A constrained EM algorithm is used to estimate parameters, in conjunction with carefully chosen initial values and some other tuning parameters. We present results using both a first-order model that captures “main effect” and a second-order model capable of identifying interactive effects between pairs of descriptors.

En recherche de médicaments, les modèles statistiques ont joué un rôle important dans la prédiction de l'activité de composés contre différentes cibles biologiques. Dans ce problème d'apprentissage supervisé, des descripteurs de structure moléculaire (par exemple : poids atomique, types de liens, plusieurs autres caractéristiques exotiques) sont utilisés pour prédire l'activité. Les composés se regroupent souvent dans l'espace descripteur, faisant de l'analyse discriminante de mélange un outil naturel d'apprentissage supervisé. Les méthodes de mélange discriminant conventionnelles ont de la difficulté à trouver le meilleur modèle pour des données de recherche de médicaments. Des structures complexes et une haute dimensionnalité amènent des modèles surparamétrisés. Une supposition courante est que l'activité biologique des composés ne dépend que de quelques descripteurs. Nous introduisons donc un modèle de mélange spécial, ayant moins de paramètres et cherchant à prédire en utilisant de multiples sous-espaces (c.-à-d. de multiples mécanismes). Un algorithme EM avec contraintes est utilisé pour estimer les paramètres. Les valeurs initiales et quelques autres paramètres d'ajustement sont choisis soigneusement. Nous présentons des résultats utilisant un modèle de premier ordre qui identifie « l'effet principal » et un modèle de second ordre capable d'identifier des effets interactifs entre les paires de descripteurs.

[MS-38]**Risk-reducing Stein-like Estimation in a Lifetime Censored Regression Model****Estimation de Stein réduisant le risque dans un modèle de régression pour le temps de survie censuré**

Shakhwat HOSSAIN & S. Ejaz AHMED, *University of Windsor*

The problem of estimation of a regression parameter (vector) in a Weibull regression model with censoring is considered when it is suspected that parameters may be restricted to a linear subspace. Under this uncertainty, we present a basis for optimally combining estimation problems. The objective is to provide natural adaptive estimators that are free of subjective choices. In the context of two competing Weibull regression models (full and reduced), we demonstrate a Stein-type estimator, that outperforms the classical maximum likelihood estimator asymptotically. A Monte Carlo simulation study is also conducted to assess the performance

Nous considérons le problème de l'estimation d'un paramètre (vectoriel) de régression dans un modèle de régression de Weibull avec censure où nous suspectons que les paramètres peuvent être restreints à un sous-espace linéaire. Sous cette incertitude, nous présentons une base pour combiner de façon optimale les problèmes d'estimation. L'objectif est de fournir des estimateurs adaptatifs naturels libres de choix subjectifs. Dans le contexte de deux modèles de régression de Weibull en compétition (complet et restreint), nous démontrons un estimateur de type Stein qui, asymptotiquement, performe mieux que l'estimateur du maximum de vraisemblance classique. Une étude par simulation de Monte-Carlo est aussi effectuée afin d'évaluer la perfor-

of the suggested estimators for moderate and large samples. An empirical example illustrates the usefulness of this inferential procedure in practice. Bootstrapping procedures are used to develop confidence sets and serve as a basis for inference.

mance de l'estimateur proposé en présence d'échantillons grands et moyens. Un exemple empirique illustre l'utilité de cette procédure inférentielle en pratique. Des procédures de bootstrap sont utilisées pour développer des intervalles de confiance et servent de base à l'inférence.

[MS-39]

Compliance Testing for Random Effects Models with Joint Acceptance Criterion

Tests de conformité pour modèles à effets aléatoires avec critère d'acceptation conjointe

Crystal LINKLETTER, Chunfang LIN, Wilson LU, Pritam RANJAN, Derek BINGHAM & Richard LOCKHART, *Simon Fraser University*, Tom LOUGHIN, *Kansas State University*

For consumer protection, many countries regulate the production of goods sold by weight or volume to ensure consistency between contents and labelling. Government inspections are randomly performed to verify, generally speaking, that (i) on average, products being sold have net content sufficiently close to the labelled content and (ii) very few or no products being sold are significantly defective. This talk will cover challenges in evaluating the probability of jointly satisfying both criterion, and implications for tuning parameters when the production process is assumed to be a random effects model, with potentially more than one source of variability.

Afin d'assurer la protection des consommateurs, plusieurs pays réglementent la production de biens vendus au poids ou au volume pour assurer la cohérence entre le contenu et l'étiquetage. Des inspections gouvernementales sont menées, aléatoirement, pour vérifier que (i) en moyenne, les produits vendus ont un contenu net suffisamment près du contenu affiché et (ii) très peu ou aucun des produits vendus ne sont significativement défectueux. Cet exposé couvre les défis rencontrés dans l'évaluation de la probabilité de satisfaire conjointement ces deux critères, et les implications pour les paramètres d'ajustement lorsque nous supposons que le processus de production est un modèle à effets aléatoires, avec potentiellement plus d'une source de variabilité.

[MS-40]

An Alternative Wald Type Test for Two Linear Restrictions with Applications to Non-linear Models

Un test alternatif de type Wald pour deux restrictions linéaires avec applications aux modèles non linéaires

Krishna SAHA, *Central Connecticut State University*, Maxwell KING, *Monash University*

This paper proposes a new test procedure called a rel or full or half quel test to resolve the problem of the small-sample local biasedness and the power non-monotonic behaviors of the Wald test for two linear restrictions caused by inaccuracy of the estimated covariance matrix of the estimator. This new test procedure, which does not need the covariance matrix of the estimator, involves finding the critical region based on constructing the rel (ellipse) or quel (full or half) utilizing bootstrap in order to obtain the desired size with good power properties. Simulation results indicate that the new test procedure, especially the full quel test performs rather well in both respects showing the better monotonic power performance by controlling the size in the small samples.

Cet article propose un test avec une nouvelle méthode appelée un test « rel » ou un test « quel » entier ou test « mi quel » pour résoudre le problème de biais local d'un petit échantillon et les comportements non-monotones de la puissance du test Wald pour deux restrictions linéaires provoquées par l'inexactitude de la matrice estimée de covariance de l'estimateur. Cette nouvelle méthode, qui n'a pas besoin de la matrice de covariance de l'estimateur, implique de trouver la région critique basée sur la construction du rel (ellipse) ou le quel (entier ou mi) utilisant le bootstrap afin d'obtenir le niveau désiré avec une bonne puissance. Les résultats de simulation indiquent que ce test, particulièrement le quel entier est plutôt efficace pour les deux problèmes et montre une meilleure puissance monotone en contrôlant le niveau pour les petits échantillons.

Session 02I**Lundi 29 mai • Monday, May 29, 10:30–12:00****SSC3018****Design and Analysis of Clinical Trials
Planification et analyse d'essais cliniques****[MS-41]****New Allocation Rule for Repeated Measurement Designs****Nouvelles règles de répartition pour plans à mesures répétées**Yuanyuan LIANG & Keumhee Chough CARRIÈRE, *University of Alberta*

In response-adaptive designs, we modify the trial on the basis of outcomes in order to achieve a specific goal. Optimal designs for clinical trials are usually constructed under a single optimality criterion. We improve the current strategies of building response-adaptive designs in two directions: (1) construct optimal multiple-objective designs to increase both the estimation precision and the proportion of patients assigned to a better treatment, and (2) use a more general model. In simulations we demonstrate that the designs constructed under the new proposed allocation rule are more efficient than fixed optimal designs in terms of the mean squared error.

Dans des plans adaptatifs à la réponse, nous modifions l'essai sur la base des résultats afin d'atteindre un but spécifique. Les plans optimaux pour des essais cliniques sont généralement construits avec un seul critère d'optimalité. Nous améliorons les stratégies présentes de construction de plan adaptatifs à la réponse de deux façons : (1) en construisant des plans à objectifs multiples optimaux afin d'augmenter à la fois la précision d'estimation et la proportion de patients assignés à un meilleur traitement, et (2) en utilisant un modèle plus général. Par des simulations, nous montrons que les plans construits sous ces nouvelles propositions de règles d'allocation sont plus efficaces que les plans optimaux fixes en termes d'erreur quadratique moyenne.

[MS-42]**Testing Treatment Effects of Binary Response Adaptive Designs****Test sur les effets de traitement des plans adaptatifs de réponse binaire**Yanqing YI & Xikui WANG, *University of Manitoba*

The dependency of data in response adaptive designs induces the difficulties of statistical inference such as testing treatment effects. This paper discusses the consistency and asymptotic normality of maximum estimator of the parameters for a wide class of response adaptive designs. Although the samples are dependent in the designs, it is proved that chi square statistic is still valid under certain conditions of an allocation rule.

La dépendance des données dans des plans adaptatifs de réponse induit des difficultés dans l'inférence statistique telles que les tests sur les effets d'un traitement. Cet article discute la convergence et la normalité asymptotique de l'estimateur maximum des paramètres pour une large classe de plans adaptatifs de réponse. Bien que les échantillons soient dépendants dans les plans, on démontre que la statistique khi-carré est encore valable sous certaines conditions sur la règle d'attribution.

[MS-43]**Improving Efficiency in Clinical Trials Using a Markov Chain Transition Rate Model with Conditional Power****Augmentation de l'efficacité des essais cliniques par l'utilisation de chaînes de Markov avec puissance conditionnelle**Gregory POND & Lillian SIU, *Princess Margaret Hospital/University of Toronto*

Optimising efficiency in oncology clinical trials is critical as the high costs, financially and ethically, to run these trials limit the availability of

Les coûts éthiques et financiers élevés des essais cliniques en oncologie limitant la disponibilité des patients, le besoin d'une efficacité optimale lors de ces essais est cri-

patients. Incorrect trial decisions are even more costly in terms of future patients and finances. A Markov chain transition rate model utilising conditional power methods is proposed to increase the information available and improve decision making. Repeated measurements on a single patient are incorporated through the Markov chain, and conditional power allows for a group-sequential design, thus, early termination of trials with futile treatments. Illustrative examples will be presented of trials performed at Princess Margaret Hospital.

tique. De mauvaises conclusions suite à ces essais sont particulièrement coûteuses, tant financièrement qu'en terme de futurs patients. Un modèle basé sur la construction d'une chaîne de Markov et utilisant des méthodes de puissance conditionnelle est ici proposé afin d'augmenter la quantité d'information disponible et d'améliorer la prise de décision. La chaîne de Markov permet d'incorporer des mesures répétées provenant d'un même patient. Les méthodes de puissance conditionnelle permettent l'utilisation d'un plan séquentiel par groupes et donc, l'interruption rapide des essais liés à des traitements inutiles. Des exemples issus d'essais cliniques effectués à l'hôpital Princess Margaret seront présentés à titre illustratif.

[MS-44]

A Robust Two-Group Step-Down Test for Multiple Outcomes

Un test descendant robuste à deux groupes pour des résultats multiples

Lisa LIX & Kathleen DEERING, *University of Manitoba*

In clinical trials, it is often of interest to test the difference between treatment and control groups on multiple outcomes when the treatment is expected to have a positive effect on all outcomes. Few tests of directional multivariate alternative hypotheses have been proposed for the case when covariances are heterogeneous and the distribution is heavy-tailed. We propose a composite approximate degrees of freedom test with trimmed estimators based on Roy's (1954) step-down analysis. The accuracy of the approximation is verified by simulations, and comparisons are made with an approximate likelihood ratio test and a non-parametric test for the Behrens-Fisher problem.

Dans les essais cliniques, il est souvent d'intérêt d'examiner la différence entre le groupe de traitement et le groupe de contrôle sur des résultats multiples quand on s'attend à ce que le traitement ait un effet positif sur tous les résultats. Peu de tests d'hypothèses alternatives multivariées directionnelles ont été proposés pour le cas où les covariances sont hétérogènes et la distribution a une queue lourde. Nous proposons un test composé de degrés de liberté approximatifs avec les estimateurs tronqués basés sur les analyses de méthodes « descendante » de Roy (1954). L'exactitude de l'approximation est vérifiée par des simulations, et les comparaisons sont faites avec un test de ratio de vraisemblance approximative et un test non paramétrique pour le problème de Behrens-Fisher.

[MS-45]

A Maximum Likelihood Estimator for the Analysis of Randomized Trials with Treatment Noncompliance and Subsequent Nonresponse

Un estimateur du maximum de vraisemblance pour l'analyse d'essais cliniques aléatoires en présence de non-observance et de non-réponse

Yang ZHAO, *University of Regina*

Treatment non-compliance and subsequent nonresponse are common problems in the analysis of randomized trials. In this research I consider maximum likelihood methods for the estimation of the causal effect of treatment assignment for a two-armed randomized trial. I will describe an iterative re-weighted algorithm to obtain a maximum likelihood estimator of the causal effect. Closed form

La non-observance de traitement et la non-réponse subséquente sont des problèmes courants dans l'analyse d'essais randomisés. Dans cette recherche, je considère des méthodes d'estimation de l'effet causal de l'assignation de traitement pour un essai clinique randomisé à deux groupes basées sur le maximum de vraisemblance. Je décris un algorithme itératif re-pondéré pour obtenir un estimateur de l'effet causal. Des expressions analytiques permettant de mettre à jour les

expressions for updating the estimates at each iteration step will be provided.

estimations à chaque itération sont présentées.

[MS-46]

Semi-parametric Models and Efficiency in Mean Cost Estimation

Modèles semi-paramétriques et efficacité dans l'évaluation du coût moyen

Eleanor PULLENAYEGUM, *University of Waterloo*, Andrew WILLAN, *University of Toronto*

Cost-effectiveness analysis from censored clinical trial data requires an estimate of the between-treatment difference in mean cost. While estimation can be non-parametric, there are advantages to semi-parametric regression. Firstly, incorporating baseline covariates may explain some of the residual variance. Secondly, costs may be collected at several time points. Multivariate regression allows such longitudinal data to be pooled across time intervals. The drawback of regression is that inverse-probability weighting, the usual estimation procedure, is inefficient. Existing results on semi-parametric efficiency are used to derive efficient estimators. The benefits of these new estimators and models are illustrated using two recent clinical trials.

L'analyse de rentabilité des données d'essais cliniques censurées exige une évaluation de la différence, en coût moyen, entre les traitements. L'évaluation peut être non paramétrique mais il y a des avantages à utiliser la régression semi-paramétrique. Premièrement, les covariables de base peuvent expliquer une partie de la variance résiduelle. Deuxièmement, des coûts peuvent être épargnés à plusieurs points dans le temps durant le processus. La régression multivariée permet de telles données longitudinales d'être mises en commun à travers des intervalles de temps. L'inconvénient de la régression est que la pondération selon la probabilité inverse, la procédure habituelle d'estimation, est inefficace. Des résultats existants sur l'efficacité des méthodes semi-paramétriques sont employés pour dériver des estimateurs efficaces. Les avantages de ces nouveaux estimateurs et modèles sont illustrés en utilisant deux essais cliniques récents.

Session 03A

Lundi 29 mai • Monday, May 29, 13:30–14:30

SSC2050

Canadian Journal of Statistics Award Address

Allocution du récipiendaire du Prix de la Revue canadienne de statistique

Session 03B

Lundi 29 mai • Monday, May 29, 13:30–15:00

SH3345

Statistical Issues in Designing and Monitoring Clinical Trials

Problèmes statistiques dans la planification et la surveillance d'essais cliniques

[MS-47]

Challenges in the Design, Implementation, and Monitoring of HIV Treatment Trials

Défis dans la conception, l'implantation, et la surveillance d'essais de traitement du VIH

James NEATON, *University of Minnesota*

An HIV trial called SMART (Strategies for Management of Antiretroviral Therapy) illustrates a number of challenges faced in designing and carrying out an international HIV treatment trial. Features of this trial that will be discussed: 1) Sample size estimation in the absence of critical data; 2) implementation of the trial in 33 countries; 3) monitoring guidelines which consider not only the primary outcome, progression to AIDS or death,

Un essai clinique relié au VIH appelé SMART (Strategies for Management of Antiretroviral Therapy) illustre un certain nombre de défis relevés en planifiant et en effectuant un essai international de traitement du VIH. Quelques thèmes de cet essai seront discutés : 1) Estimation de la taille de l'échantillon en absence de données critiques ; 2) Implantation de l'essai dans 33 pays ; 3) Lignes de conduite qui considèrent non seulement le résultat primaire, la progression au SIDA ou la mort, mais également un résultat « de soutien »

but also another “supportive” outcome – morbidity and mortality due to cardiovascular, renal or liver disease.

différent – morbidité et mortalité dues aux maladies cardiovasculaires, rénales ou du foie.

[MS-48]

Adaptivity in Group Sequential Designs
L’adaptabilité dans les plans séquentiels par groupes

Christopher JENNISON, *University of Bath, UK*

Although there has recently been heightened interest in adaptive and flexible trial designs, the ideas of adaptive design go back a long way. This talk focuses on hypothesis testing, in particular:

- efficient sequential stopping rules,
- adapting to unpredictable information,
- adapting to nuisance parameters,
- adapting to external factors.

We shall see how the first three topics can be dealt with in traditional approaches — which for group sequential tests means methods introduced in the 1980s. We shall discuss the added value offered by more recent proposals for adaptive designs and point out areas where caution may be needed.

Malgré la hausse récente de l’intérêt porté à la planification d’essais adaptatifs et flexibles, les idées de la planification adaptative sont plus anciennes. Cette présentation porte sur les tests d’hypothèses, plus précisément :

- les règles d’arrêt séquentielles efficaces,
- l’adaptation à de l’information imprévisible,
- l’adaptation aux paramètres de nuisance,
- l’adaptation aux facteurs externes.

Nous allons voir comment les trois premiers sujets peuvent être traités avec les approches traditionnelles — ce qui signifie, pour les tests séquentiels par groupes, des méthodes introduites dans les années 1980. Nous allons discuter de la valeur ajoutée offerte par des propositions plus récentes pour des plans adaptatifs et signaler des zones où plus de prudence est de rigueur.

Session 03C Lundi 29 mai • Monday, May 29, 13:30–15:00 SH3315

Change-point Detection
Détection de point de changement

[MS-49]

The Probable Error of a Mean and That of a Change in a Mean
L’erreur probable d’une moyenne et celle d’un changement de moyenne

Miklós Csörgő & Barbara Szyszkowicz, *Carleton University*, Qiying WANG, *University of Sydney*

Weighted approximations in probability of self-normalized and Studentized partial sums processes will be reviewed and applied to studying the problem of change in the mean of random variables in the domain of attraction of the normal law. The talk will be based on our joint works as reviewed and further developed in our exposition: “On weighted approximations and strong limit theorems for self-normalized partial sums processes,” in: *Asymptotic Methods in Stochastics—Festschrift for Miklós Csörgő* (Lajos Horváth and Barbara Szyszkowicz, eds.), Fields Institute Communications, Volume 44, Amer. Math. Soc., Providence, RI, 2004, pp. 489–521.

Des approximations pondérées en probabilité de processus de sommes partielles auto-normalisées et studentisées seront passées en revue et appliquées à l’étude du problème du changement de moyenne de variables aléatoires dans le domaine d’attraction de la loi normale. Cet exposé est basé sur nos travaux conjoints présentés et développés dans notre exposition : « On weighted approximations and strong limit theorems for self-normalized partial sums processes », dans : *Asymptotic Methods in Stochastics – Festschrift for Miklós Csörgő* (Lajos Horváth and Barbara Szyszkowicz, edit.), Fields Institute Communications, volume 44, Amer. Math. Soc., Providence, RI, 2004, p. 489-521.

[MS-50]**Change-point Detection in a General Regression Model****Détection de changement structurel dans un modèle de régression général**Murray BURKE, *University of Calgary*

In statistics, model building is a very important task. When considering bivariate observations taken over time, we may wish to establish a relationship between the X and the Y components by some regression model, either parametric or semi-parametric, linear or nonlinear. However, if, after some point, there is a change in the relationship between X and Y, then the single regression model will fit the data poorly. Thus, it is important to determine whether a change has occurred before the particular form of the model is postulated. Tests for a change-point under fairly general assumptions will be proposed.

L'élaboration de modèles est une tâche très importante en statistique. Lorsque nous considérons des données bivariées observées à travers le temps, il est possible d'établir un lien entre les composantes X et Y grâce à un modèle de régression paramétrique ou semi-paramétrique, linéaire ou non-linéaire. Cependant, si, à un moment donné, un changement survient dans la relation entre X et Y, alors un seul modèle de régression ajustera mal les données. Il est donc important de déterminer si un tel changement s'est produit avant d'émettre une conjecture sur la forme particulière du modèle. Des tests permettant de détecter un point de changement sous des hypothèses relativement peu restrictives seront présentés.

[MS-51]**Monitoring Parameter Change in Time Series Models****Surveillance du changement de paramètre dans une série chronologique**Edit GOMBAY & Daniel SERBAN, *University of Alberta*

Sequential tests, that are generalizations of Page's CUSUM tests, are proposed to detect an abrupt change in any parameter(s) of an autoregressive time series model. The tests accommodate nuisance parameters. They are based on large sample approximations of the efficient score vector under the null hypothesis of no change, and under the alternative. The empirical power of the tests is evaluated in a simulation study. The new method performs better than the existing ones found in the literature, if the criterion is the type I error probability, which can be unacceptably high for methods that minimize the expected value of reaction time.

Les tests séquentiels, qui sont une généralisation des tests CUSUM de Page, sont proposés pour détecter un changement brusque parmi un ou plusieurs paramètres d'un modèle de séries chronologiques autorégressif. Ces tests prennent en considération les paramètres de nuisance. Ils sont basés sur des approximations de grande taille d'échantillons du score sous l'hypothèse nulle de l'absence de changement, et sous l'hypothèse alternative. La puissance empirique des tests est évaluée via simulations. La nouvelle méthode performe mieux que les méthodes existantes retrouvées dans la littérature si le critère est la probabilité d'une erreur de type I, qui peut être élevée de façon inacceptable pour les méthodes qui minimisent l'espérance du temps de réaction.

Session 03D Lundi 29 mai • Monday, May 29, 13:30–15:00 SSC2036**Design and Analysis of Experiments for Complex Computer Models
Planification et analyse d'expériences pour des modèles d'ordinateur complexes****[MS-52]****Uncertainty in Kriging Predictions with and Without Random Error
Incertitude dans la prévision par krigeage avec et sans erreur aléatoire**Bela NAGY, JASON LOEPPKY & William J. WELCH, *University of British Columbia*

The standard Kriging prediction variance formula used in geostatistics and for computer experiments underestimates the true variance because it doesn't incorporate the variability due to estimating the covariance parameters. Our results show that the difference can be quite substantial both with and without random error in the model, especially for small sample sizes. Corrections have been proposed, but they tend to be numerically cumbersome. Alternatives are explored and evaluated by simulations on performance and computational feasibility.

La formule habituelle de variance pour la prévision par krigeage utilisée en géostatistique et pour les expériences par ordinateur sous-estime la vraie variance parce qu'elle ignore la variabilité causée par l'estimation de la covariance. Nos résultats montrent que cette différence peut être substantielle, surtout pour des petits échantillons, que le modèle inclue ou non une erreur aléatoire. Des correctifs ont été proposés, mais ils requièrent généralement de lourds calculs numériques. Différentes alternatives sont étudiées et évaluées à l'aide de simulations afin de déterminer leur performance et leur faisabilité numérique.

[MS-53]**Quantifying Uncertainty in the Biospheric Carbon Flux for England and Wales
Quantifier l'incertitude dans le flux du carbone biosphérique en Angleterre et au pays de Galles**Marc KENNEDY, Clive ANDERSON, Anthony O'HAGAN, Mark LOMAS & Ian WOODWARD, *University of Sheffield, UK*,
Andreas HEINEMEYER, *University of York, UK*

A crucial issue in the global warming debate is the effect of vegetation and soils on atmospheric CO_2 concentrations. SDGVM is a deterministic code simulating the extraction of CO_2 through photosynthesis, and other processes which return it to the atmosphere. We estimate the total net CO_2 flux for England and Wales in 2000, involving inference for the sum over multiple outputs (different sites and vegetation types) with uncertain and correlated inputs. A series of maps are produced, illustrating the spatial variations in uncertainty for each code output. Detailed information about the relative sources of the overall uncertainty is also produced.

Un problème crucial dans le débat du réchauffement climatique est l'effet de la végétation et des sols sur les concentrations atmosphériques de CO_2 . Le modèle SDGVM est un programme déterministe simulant l'extraction de CO_2 par la photosynthèse et autres processus le renvoyant dans l'atmosphère. Nous estimons le flux total net de CO_2 pour l'Angleterre et le pays de Galles en 2000, ce qui implique d'inférer la somme de différentes sorties (différents sites et types de végétation) avec des entrées incertaines et corrélées. Une série de cartes est produite, illustrant les variations spatiales des incertitudes pour chaque sortie. Des informations détaillées quant aux sources relatives de l'incertitude globale sont aussi produites.

[MS-54]**Design of Experiments for Integrated Computer and Physical Experiments
Planification d'expériences intégrant l'expérimentation physique et informatique**Shane REESE, *Brigham Young University*, Derek BINGHAM, *Simon Fraser University*

Physical experimentation is an expensive endeavor. Computer experiments, which are often a less expensive alternative, provide a reasonably accu-

L'expérimentation physique est généralement très coûteuse. Les expériences informatiques, souvent moins dispendieuses, nous fournissent une représentation raisonna-

rate representation of physical experimental results. This work proposes an approach to designing experiments when the researcher has a limited amount of data collected on each of the physical and computer experiments and wishes to allocate future experiments to be tested at inputs for the physical experiments or for the computer experiments, or perhaps some experiments to be run at combinations of both physical and computer models. This hybrid experimental situation relies on Gaussian process formulations for the modeling of the experimental outputs as functions of the experimental inputs. We demonstrate the methodology with both simulated and actual experimental situations.

blement précise des résultats d'expériences physiques. Ce travail propose une approche à la planification d'expérience où le chercheur dispose d'une quantité limitée de données obtenues par expérimentations physique et informatique et souhaite tester à l'entrée si l'expérience devrait être informatique, physique, ou peut-être une combinaison du modèle physique et informatique. Cette situation expérimentale hybride s'appuie sur un processus gaussien de formulations pour la modélisation des sorties expérimentales comme fonction des entrées expérimentales. Nous démontrons cette méthodologie avec des situations expérimentales simulées et réelles.

Session 04A**Lundi 29 mai • Monday, May 29, 15:30–16:30****SSC2050****Record Linkage and Counterterrorism****Couplage d'enregistrements et la lutte au terrorisme****[MS-55]****Record Linkage, Nondisclosure, Counterterrorism, and Statistics****Couplage d'enregistrements, non-divulgarion, contre-terrorisme et statistiques**Michael LARSEN, *Iowa State University*

Governments collect information for many purposes, including preventing terrorism. Record linkage is the process of merging databases containing common units and has many applications. Both false and missed links between records have potentially severe consequences in the context of counterterrorism. Use of record linkage methods for counterterrorism could affect other data collection efforts in terms of increasing nonresponse and decreasing data quality. Most government data collection efforts include guarantees of privacy, confidentiality, and nondisclosure. The focus of this talk is the relationship between advances in record linkage, disclosure risk, and counterterrorism efforts involving large databases.

Les gouvernements recueillent de l'information à plusieurs fins, incluant la prévention du terrorisme. L'appariement des fichiers permet la combinaison de bases de données contenant des unités communes et a plusieurs applications. Des faux liens ou des liens manquants entre les enregistrements peuvent avoir des conséquences graves dans le contexte du contre-terrorisme. L'utilisation de méthodes d'appariement des fichiers à des fins de contre-terrorisme pourrait affecter d'autres cueillettes d'information en termes d'augmentation de la non-réponse et de diminution de la qualité des données. La plupart des cueillettes de données gouvernementales incluent des garanties concernant la vie privée, la confidentialité et la non-divulgarion des données. Cet exposé se concentre sur la relation entre les progrès dans le domaine de l'appariement de fichiers, les risques de divulgation et les activités de contre-terrorisme impliquant l'utilisation de grandes bases de données.

Session 04B**Lundi 29 mai • Monday, May 29, 15:30–17:00****SSC2036**
Statistical Applications in Forestry
Applications statistiques en foresterie
[MS-56]
Forest Fire Management—Practical Problems of Uncertainty in Decision Support
Gestion des feux de forêts : problèmes pratiques d'incertitude dans le support décisionnel

 Al TITHECOTT & Rob McALPINE, *Ontario Ministry of Natural Resources*

Forest fire management is achieved through a series of nested decisions completed in a dynamic natural environment. Analytical decision support requires an understanding of the context, scale, and sources of uncertainty of both natural and human systems. Ontario's forest fire management program has a long history of collaborative decision analysis and information systems development. This presentation will describe the hierarchical context of fire management decision-making and some practical problems faced by managers that may benefit from statistical analysis. Examples of decision analysis completed by the authors and others in Ontario will be used to highlight successes and opportunities for improvement.

La gestion des feux de forêt est faite grâce à une série de décisions imbriquées, prises dans un environnement naturel dynamique. Le support décisionnel analytique requiert une compréhension du contexte, de l'échelle et des sources de l'incertitude des systèmes naturels et humains. Le programme de gestion des feux de forêts de l'Ontario a une longue histoire d'analyse décisionnelle collaborative et de développement de systèmes d'information. Cette présentation décrit le contexte hiérarchique de la prise de décision en gestion des feux et quelques problèmes pratiques rencontrés par les gestionnaires pouvant bénéficier d'une analyse statistique. Des exemples d'analyses décisionnelles effectuées, entre autres par les auteurs, sont utilisées afin de souligner les réussites et opportunités de développement.

[MS-57]
Statistics in Forest Ecology
La statistique en écologie forestière

 Edward JOHNSON, *University of Calgary*

Traditionally, forest ecology has used statistics to describe data, evaluate the strength of evidence, and design sampling and plot (experimental) studies. Outside of these areas, forest ecology has used statistics to model the distribution of forest species and the dynamics of the forest. Multivariate techniques have been used to describe and understand the spatial distribution of plants and the environmental correlation to these patterns. Multivariate models, being used mostly in a descriptive manner, have not proved to be useful in understanding the underlying processes that give rise to the patterns. However, in the last few years the use of a more climatic-hydrological-ecological distributed modeling approach seems to offer more promise in a more mechanistic understanding of community organization and distribution. The study of forest dynamics has also seen a change from the ear-

Traditionnellement, l'écologie forestière s'est servie de statistiques pour décrire les données observées, évaluer la robustesse des preuves scientifiques, ainsi que pour la planification d'échantillons et d'études expérimentales. En outre, l'écologie forestière s'est servie de statistiques pour modéliser la distribution d'espèces forestières et les dynamiques forestières. Des techniques d'analyses multivariées ont été utilisées pour décrire et comprendre la répartition spatiale des plantes et les corrélations environnementales avec cette répartition. Les modèles à plusieurs variables, surtout utilisés de façon descriptive, ne se sont pas montrés utiles à la compréhension des processus sous-jacents à cette répartition. Cependant, l'utilisation au cours des dernières années d'une approche de modélisation climatique, hydrologique et écologique semble très prometteuse quant à une compréhension plus mécanique de l'organisation et de la distribution des communautés. Cette étude de la dynamique forestière se démarque aussi des approches précédentes, plus descrip-

lier, largely descriptive approach of stand dynamics to recent population process approaches which include both local population dynamics and regional population dynamics (metapopulations).

tives, pour se tourner vers l'approche des processus de population. Celle-ci inclut à la fois les dynamiques de populations locales et régionales, ou métapopulations.

[MS-58]

Mixture Models for Spatio-temporal Multi-state Processes

Modèles de mélange pour des processus spatio-temporels multi-état

Farouk NATHOO & Charmaine DEAN, *Simon Fraser University*

Multi-state models can be useful in longitudinal studies where at any point in time, an individual may be said to occupy one of a discrete set of states and interest centers on the process governing state-to-state transitions. For example, states may refer to the number of recurrences of an event, or the stages of a disease. Methodology for the analysis of multi-state models is well developed for the health context, where typically, individuals may be considered to be independent. In forestry, statistical methodology for the analysis of this type of longitudinal data needs to recognize the added feature of incorporating spatial correlation. For example, how the rates of transitions over states differ spatially over a region may be of interest. Here, joint modeling of both spatial correlation as well as correlation between different transition rates is required and a multivariate spatial approach is employed. Spatial random effects are considered in a special case: the two-state mover-stayer model. Methodological developments will be presented in the context of a study examining recurrent weevil infestation in British Columbia forests.

Les modèles à plusieurs états peuvent être utiles pour des études longitudinales où à un moment quelconque dans le temps, on peut dire qu'un individu occupe un état parmi l'ensemble discret des états et où l'intérêt est centré sur le processus régissant les transitions entre les états. Par exemple, les états peuvent se rapporter au nombre de répétitions d'un événement, ou aux étapes d'une maladie. La méthodologie pour l'analyse des modèles à plusieurs états est bien développée dans le contexte de la santé, où typiquement, les individus peuvent être considérés comme indépendants. Dans le domaine forestier, la méthodologie statistique pour l'analyse de ce type de données longitudinales doit identifier le dispositif supplémentaire d'incorporation de la corrélation spatiale. Par exemple, comment les taux de transitions entre les états diffèrent spatialement d'une région peut être d'intérêt. Des effets aléatoires spatiaux sont considérés dans un cas spécial : le modèle mover-stayer à deux états. Notre exemple est une étude de l'infestation récurrente de charançon dans des forêts de la Colombie-Britannique. Cette étude de sept ans a été entreprise par le Ministère de la foresterie de la Colombie-Britannique. L'intérêt primaire était de décrire la forme de l'infestation de charançon dans tout le secteur au cours de la période de l'observation.

Session 04C

Lundi 29 mai • Monday, May 29, 15:30–17:00

SH3345

Statistics in Public Health

Statistique en santé publique

[MS-59]

Weighted Likelihood for Semiparametric Models and Two Phase Stratified Samples, with Application to Cox Regression

Vraisemblance pondérée pour modèles semiparamétriques et échantillons stratifiés à deux niveaux, avec application à la régression de Cox

Norman BRESLOW & Jon WELLNER, *University of Washington*

We consider inverse probability weighted (IPW) likelihood equations for both Euclidean and non-parametric parameters in semiparametric models

Nous considérons les équations de vraisemblance pondérées par probabilités inverse (PPI) pour des paramètres euclidiens et non-paramétriques dans des modèles semi-paramétriques

SH = Somerville House

SSC = Social Science Centre

UC = University College

fitted to two phase stratified samples. By proving weak convergence of the IPW empirical process, and borrowing results on weighted bootstrap processes, we derive asymptotics for finite population stratified sampling. Several of our key results have already been derived for the special case of Cox regression with stratified case-cohort studies, other complex survey designs and missing data problems more generally. This paper serves to interpret this prior work and to pave the way for other applications.

ajustés à des échantillons stratifiés à deux niveaux. En prouvant la convergence faible du processus PPI empirique, et en empruntant des résultats sur les processus de bootstrap pondérés, nous dérivons des résultats asymptotiques pour l'échantillonnage fini de populations stratifiées. Plusieurs de nos résultats-clés ont déjà été dérivés pour le cas particulier de la régression de Cox sous une étude de cas-cohorte stratifiée, et, plus généralement, d'autres plans de sondages complexes et problèmes de données manquantes. Cet article a pour but d'interpréter ce travail antérieur et d'ouvrir la voie à d'autres applications.

[MS-60]

Analysis of Observational Studies in the Presence of Treatment Selection Bias: Effects of Invasive Cardiac Management on AMI Survival

Analyse des études d'observation en présence d'un biais de sélection d'un traitement : effets de gestion cardiaque invasive sur la survie IMA

Thérèse STUKEL, *Institute for Clinical Evaluative Sciences*

Comparisons of treated and untreated patients in observational studies may be biased due to differences in patient prognosis. We collected chart and health administrative data on 158,831 heart attack patients. The relative mortality rate among those who received invasive versus conservative therapy was estimated using propensity-based matching and instrumental variable (IV) analyses. Invasive management was associated with a 50% decrease in mortality using propensity-based risk adjustment. This is unrealistic based on randomized trials, possibly due to residual bias from unmeasured prognostic variables. IV analyses showed a 3% survival benefit, which is valid but has a different interpretation and answers policy questions.

Les comparaisons des patients ayant ou non subi un traitement dans les études par observation peuvent être biaisées étant donné les différences dans le pronostic émis pour le patient. Nous avons recueilli les diagrammes et les données administratives de santé sur 158 831 patients ayant souffert d'une crise cardiaque. Le taux relatif de mortalité parmi ceux qui ont reçu une thérapie invasive versus une thérapie modérée a été estimé en utilisant l'appariement basé sur la propension et les analyses de la variable instrumentale (VI). L'administration invasive a été associée à une réduction de 50% dans la mortalité en utilisant l'ajustement de risque basé sur la propension. C'est peu réaliste basé sur des épreuves randomisées, probablement dû au biais résiduel des variables pronostiques non mesurées. Les analyses VI ont montré 3% de survie additionnelle, ce qui est valide mais a une interprétation différente et répond à des questions politiques.

[MS-61]

The Consequences of Misspecified Random Effects Distributions on Predictions of Random Effects

Les conséquences d'une mauvaise spécification de la distribution des effets aléatoires lors de la prédiction d'effets aléatoires

John NEUHAUS & Charles McCULLOCH, *University of California at San Francisco*

Predictions of quality of care in hospitals or physician performance often use predicted random effects from generalized linear mixed models. Standard applications of this approach require specification of a distribution for the random effects, which will be misspecified to some degree. We examine the effects of this misspecification on pre-

Les effets aléatoires dérivés à partir de modèles linéaires généralisés mixtes sont souvent utilisés dans les prédictions de la qualité des soins en hôpitaux ou de la performance des médecins. Certaines applications standard de cette approche nécessitent que la distribution des effets aléatoires soit spécifiée, avec toutefois un certain niveau d'erreur de la spécification. Nous examinons les effets de cette erreur de spécifica-

diction of the random effects. Using analytic results, simulation studies and fits to example data, we show that misspecifying the random effects distribution can lead to inaccurate predicted random slopes and intercepts. These results contrast with those in other settings where random effects distribution misspecification typically produces little bias.

tion sur la prédiction des effets aléatoires. Nous utilisons des résultats analytiques, des études de simulation, et des ajustements à des données pour montrer que l'erreur de spécification de la distribution des effets aléatoires peut produire des erreurs au niveau des ordonnées à l'origine et pentes aléatoires prédites. Ces résultats sont en contraste avec d'autres situations où une erreur de spécification de la distribution des effets aléatoires typiquement produit peu de biais.

Session 04D**Lundi 29 mai • Monday, May 29, 15:30–17:00****SH3315**
**Statistical Applications in Marketing
Applications statistiques en marketing**
[MS-62]**Customer Value Segmentation—Informing Better Marketing Decisions****Segmentation de la valeur client – informer à propos de meilleures décisions de marketing**Alison BURNHAM, *Digital Cement*

Many of today's companies have built impressive data warehouses that contain a goldmine of customer information. Information on transactions, response to marketing campaigns, opinion survey results, interactions with the company (call centre, sales representatives), where they live, which stores they frequent and any program memberships (loyalty programs, newsletter communications) they have signed up for. The next step for many is to dig into this data to better inform their marketing decisions. The idea of finding segments within the database that can be profiled and understood is well entrenched.

At the same time marketing is under increasing pressure to show return on investment. What makes sense at this point is to not only identify interesting customer segments but to make sure that what is interesting about them is directly tied to the value (profit) they bring to the company.

In this talk I will discuss the process of building customer value segments and also address some of the challenges businesses face when doing such projects.

De nos jours, la plupart des entreprises ont construits d'importantes banques de données contenant une mine d'or d'informations sur les clients. De l'information sur leurs transactions, leur réponse aux campagnes de promotion, les résultats d'enquêtes d'opinion, leurs interactions avec l'entreprise (centre d'appel, représentants de commerce), où habitent-ils, quels magasins fréquentent-ils et à quels programmes ont-ils adhéré (programme de fidélité, bulletins d'informations). Pour plusieurs, la prochaine étape est de fouiller ces données afin de prendre des décisions de marketing plus éclairées. L'idée de trouver des segments dans la base de donnée pouvant être profilés et explicités est bien ancrée.

En même temps, le marketing est sous la pression grandissante de montrer un rendement sur le capital investi. Il est donc logique non seulement d'identifier des segments intéressants mais aussi de s'assurer que ce qu'ils ont d'intéressant est directement lié à la valeur (profit) qu'ils amènent à l'entreprise.

Dans cet exposé, je parle du processus de construction de segment de valeur-client ainsi que de certains des défis auxquels font face les entreprises se lançant dans de tels projets.

[MS-63]**National Automated Valuation Models in Canada
Modèles nationaux d'évaluation automatisée au Canada**Anthony PERCACCIO, *Municipal Property Assessment Corporation*

The Municipal Property Assessment Corporation (MPAC) has developed an Automated Valuation Model (AVM) product in response to a growing need by the financial services industry for more accurate and cost-effective 'real-time' estimates of market value. The AVM product draws on MPAC's state-of-the-art automated property valuation database covering more than 3.5 million residential properties in the province of Ontario representing over \$ 950 billion. The AVM product is statistically derived from extensive market & time series analyses and is updated monthly to reflect current market conditions. The talk will discuss the AVM suite of products and how they provide a reliable and accurate estimate of market value for all residential properties in Ontario, Newfoundland & Labrador and beyond.

La Société d'évaluation foncière des municipalités (MPAC) a développé un modèle d'évaluation automatisée (MEV) en réponse aux besoins croissants de l'industrie des services financiers pour des estimations de la valeur foncière plus précises, rentables et en « temps réel ». Le MEV s'appuie sur la base de données d'évaluation automatisée des propriétés, qui couvre plus de 3,5 millions de propriétés résidentielles en Ontario représentant plus de 950 milliards de dollars. Le MEV est statistiquement obtenu à partir d'études de marché et de séries chronologiques approfondies et est mis à jour mensuellement afin de refléter les conditions présentes du marché. Cet exposé porte sur la suite de produits MEV et la précision et fiabilité des estimations fournies pour toute propriété résidentielle en Ontario, Terre-Neuve & Labrador et plus.

Session 04E**Lundi 29 mai • Monday, May 29, 15:30–17:00****SSC2024****Limit Theorems with Applications in Statistics****Théorèmes limites avec applications en statistique****[MS-64]****Hermite Wavelet Estimation of Probability Density****Estimation de la densité de probabilité à l'aide d'ondelettes de Hermite.**Mei Ling HUANG & RON KERMAN, *Brock University*

The paper studies a Hermite wavelet estimator of an unknown density function $f(x)$ using a sequence X_1, X_2, \dots, X_n of i.i.d. random variables. The convergence rate and other properties of this estimator are given.

L'article étudie l'estimation, à l'aide d'ondelettes de Hermite, d'une fonction de densité inconnue $f(x)$ à partir d'une suite de variables aléatoires i.i.d. X_1, X_2, \dots, X_n . Diverses propriétés de l'estimateur résultant (incluant le taux de convergence) sont présentées.

[MS-65]**Some Recent Results on the Consistency of Kernel-Type Estimators****Résultats récents sur la convergence d'estimateurs à noyau**Uwe EINMAHL, *Vrije Universiteit Brussel, Belgium*

We give a survey of some recent results dealing with uniform consistency of kernel-type estimators. Our method is based on some general results from the theory of abstract empirical processes. Among other things, we present results on the pre-

Nous faisons une revue de résultats récents au sujet de la convergence uniforme d'estimateurs à noyaux. Notre théorie se base sur des résultats généraux de la théorie des processus empiriques abstraits. Nous présentons, entre autres, des résultats sur l'ordre de convergence précis de l'estimateur de

cise order of the convergence speed for the classical kernel type density estimator (with respect to (weighted) sup-norms) and also for the Nadaraya-Watson estimator (uniform over locations from a compact interval). A special feature of our results is that they also hold uniformly over certain ranges for the bandwidth. This allows us to prove consistency results when the bandwidth is chosen depending on the data and/ or location.

densité à noyau classique (par rapport aux normes-sup pondérées) et de l'estimateur Nadaraya-Watson (uniforme sur un intervalle compact). Nos résultats sont également valides uniformément sur certaines étendues pour la largeur de fenêtre. Ceci nous permet de prouver la convergence lorsque la fenêtre est choisie d'après les données et/ou la position.

[MS-66]

How Many Entries of a Typical Orthogonal Matrix can be Approximated by Independent Normals? Combien d'entrées d'une matrice orthogonale typique peuvent être approximées par des variables normales indépendantes ?

Tiefeng JIANG, *University of Minnesota*

I will present my solution to the open problem of Diaconis: what are the largest orders of the upper left block of a random matrix which is uniformly distributed on the orthogonal groups, can be approximated by independent standard normals? This problem is solved by two different approximation methods: the variation norm and a weak norm.

The history of the problem since 1906 will be reviewed; applications and future problems will also be given in this talk.

Je présenterai ma solution au problème non résolu de Diaconis : quels sont les ordres les plus élevés du bloc gauche supérieur d'une matrice aléatoire, qui est uniformément distribuée sur les groupes orthogonaux, qui peuvent être approximés par des variables normales typiques indépendantes ? Ce problème est résolu par deux méthodes différentes d'approximation : la norme de la variation et la norme faible.

L'historique du problème depuis 1906 sera passée en revue ; des applications et des problèmes futurs seront également abordés dans cette présentation.

[MS-67]

On the Golden ratio, Strong Law, and First Passage Problem Sur le nombre d'or, la loi forte, et le problème du premier passage

Andrei VOLODIN, *University of Regina*, Andrew ROSALSKY, *University of Florida*

For a sequence of correlated square integrable random variables $\{X_n, n \geq 1\}$, conditions are provided for the strong law of large numbers $\lim_{n \rightarrow \infty} (S_n - ES_n)/n = 0$ almost surely to hold where $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. The hypotheses stipulate that two series converge, the terms of the first series involve both the Golden Ratio $\varphi = \frac{1}{2}(1 + \sqrt{5})$ and bounds on $\text{Var}(X_n)$ and the terms of the second series involve both φ and bounds on $\text{Cov}(X_n, X_{n+m})$. An application to first passage times is provided.

Pour une séquence de variables aléatoires corrélées de carré intégrables $\{x_n, n \geq 1\}$, des conditions sont présentées pour que la loi forte des grands nombres soit satisfaite presque sûrement, $\lim_{n \rightarrow \infty} (S_n - ES_n)/n = 0$ où $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. Les hypothèses stipulent que deux séries convergent, les termes de la première série impliquent le nombre d'or $\varphi = \frac{1}{2}(1 + \sqrt{5})$ et des bornes sur $\text{Var}(X_n)$ et les termes de la deuxième série impliquent φ et des bornes sur $\text{Cov}(X_n, X_{n+m})$. Une application aux premiers temps de passage est fournie.

Session 04F**Lundi 29 mai • Monday, May 29, 15:30–17:00****SSC3026**

Bayesian Statistics

Statistique bayésienne

[MS-68]**A Quasi-Bayesian Change Point Detection with Random Exchangeable Weights****Détection quasi bayésienne de changement structurel avec poids aléatoires échangeables**Mahmoud ZAREPOUR, *University of Ottawa*, Reza HABIBI, *Shiraz University*

A quasi-Bayesian change point test statistic is studied under a new scenario of weights which are asymptotically close to an important subclass of Poisson-Dirichlet weights. This approach helps to get closer to a full nonparametric Bayesian machinery for the time of change. With detecting change at each point, the prior can be updated for the future change point detection. Asymptotic theory for the quasi-Bayesian test is developed under the null and alternative hypothesis and results are represented in terms of Itô stochastic integrals.

Une statistique de test quasi bayésienne pour la détection d'un point de changement structurel est étudiée sous un nouveau scénario de poids asymptotiquement proches d'une importante sous-classe de poids de type Poisson-Dirichlet. Cette approche permet de se rapprocher d'un traitement bayésien non paramétrique de l'étude du point de changement. En détectant le changement à chaque point, l'information *a priori* peut être mise à jour pour la détection de futurs changements. La théorie asymptotique liée au test quasi bayésien est développée sous les hypothèses nulle et alternative et mène à des résultats présentés en termes d'intégrales stochastiques d'Itô.

[MS-69]**Comparing Methods for Estimating Parameters in Dynamic Systems****Comparaison de méthodes pour estimer des paramètres dans des systèmes dynamiques**David CAMPBELL, James O. RAMSAY, Giles HOOKER & Jiguo CAO, *McGill University*

Systems of differential equations model the behaviour dynamics between variables. In linear systems the differential equations can often be solved and parameters estimated through nonlinear regression techniques. However with nonlinear systems typically there is no closed form solution. Furthermore the likelihood surface is often full of local extrema, making direct likelihood maximization techniques slow or converge to non-optimal solutions. In this talk I will contrast the Bayesian method with the profile estimation method for overcoming these hurdles while highlighting challenges from both methods using simulated and real data examples.

Les systèmes d'équations différentielles peuvent modéliser la dynamique des comportements entre les variables. Dans les systèmes linéaires, les équations différentielles peuvent souvent être résolues et les paramètres sont estimés par des techniques de régression non linéaires. Cependant avec les systèmes non linéaires, en général il n'y a aucune résolution analytique. En outre, la surface de vraisemblance contient souvent des maximums locaux. Ainsi les techniques directes de maximisation de vraisemblance peuvent être moins performantes ou converger vers des solutions non optimales. Dans cette présentation, je vais discuter de l'utilisation de la méthode bayésienne et de la méthode d'évaluation de profil pour surmonter ces obstacles. Je vais mettre l'accent sur les défis des deux méthodes en utilisant des exemples de données simulées et des vraies données.

[MS-70]**Semi-parametric Bayesian Logistic Regression
Régression logistique bayésienne semi-paramétrique**Sohee KANG & Michael ESCOBAR, *University of Toronto*

Non-parametric and semi-parametric extensions of logistic regression have been studied by different approaches from Bayesian and non-Bayesian perspectives. In this talk, we propose another semi-parametric Bayesian way of estimating the binary regression curve based on the density estimation such that the fitted regression curve is estimated as the k mixtures of logistic regression functions weighted by functions of marginal distribution of covariates.

This Bayesian model provides an easy-to-implement simulation environment. The Dirichlet process mixture model is used to develop an appropriately flexible class of mixture distributions and a Gibbs sampling is used for the computations. The performance of this method is illustrated by using both simulated and real data sets.

The exciting feature of our proposed method is that it is applicable to generalized linear models without restriction to the binary data. We also apply our method to the Poisson regression and demonstrate the performance with simulated data sets.

Des extensions non paramétriques et semi-paramétriques de la régression logistique ont été étudiées par différentes approches de perspectives Bayésiennes et non bayésiennes. Dans cet entretien, nous proposons une autre manière Bayésienne semi-paramétrique d'estimer la courbe de régression binaire basée sur l'estimation de densité telle que la courbe de régression adaptée est estimée comme étant un mélange de k fonctions de régression logistiques pondérées par des fonctions de la distribution marginale des covariables.

Ce modèle bayésien fournit un environnement de simulation facile à implanter. Le modèle de mélange du processus de Dirichlet est employé pour développer une classe convenablement flexible des distributions de mélange et l'échantillonnage de Gibbs est employé pour les calculs. La performance de cette méthode est illustrée en utilisant à la fois des jeux de données simulées ainsi que réelles.

La particularité passionnante de la méthode proposée est qu'elle est applicable aux modèles linéaires généralisés sans restriction aux données binaires. Nous appliquons également notre méthode à la régression de Poisson et démontrons la performance avec des jeux de données simulées.

[MS-71]**Checking for Prior-data Conflict with Hierarchically Specified Priors****Vérification de l'existence d'un conflit dans l'*a priori* avant les observations avec des lois *a priori* hiérarchiques**Hadas MOSHONOV & Mike EVANS, *University of Toronto*

Many statistical models include few parameter components. In the Bayesian context priors are often specified component wise. This may entail placing independent priors on parameter components or specifying the priors in a sequential or hierarchical fashion. We examine methods for checking for the source of any prior-data conflict in the individually specified components of the prior.

Plusieurs modèles statistiques incluent peu de composantes paramétriques. Dans le contexte bayésien, les distributions *a priori* sont souvent spécifiées paramètre par paramètre. Ceci peut impliquer que les paramètres suivent des lois *a priori* indépendantes, ou que les *a priori* sont spécifiés de façon séquentielle ou hiérarchique. Nous examinons des méthodes pour déterminer la source de conflits dans les *a priori* avant même d'observer les données parmi les composantes spécifiées individuellement de la distribution *a priori*.

[MS-72]

Bayesian LAGO for Statistical Detection Problems**LAGO bayésienne pour des problèmes statistiques de détection**Wanhua SU, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*, Mu ZHU, *University of Waterloo*

In statistical detection problems, we have data $\{Y_i, X_i\}$, where X_i is the predictor vector of the i th observation and $Y_i = 0 \setminus 1$ is its class label. The objective of a statistical detection problem is to identify class-1 observations, which are extremely rare. Our previous work (LAGO) was demonstrated to be superior to methods such as KNN and SVM in a real drug discovery problem. One drawback of the existing LAGO is that it only provides a point estimate of a test point's possibility of being class 1, ignoring the uncertainty of the model. In this talk, we present a probabilistic framework for LAGO, referred as BLAGO. Uncertainty can be accommodated by assigning prior distributions to the parameters. MCMC is used to obtain the posterior distributions. BLAGO provides proper probabilistic predictions that have support on $(0, 1)$ and capture uncertainty of the predictions as well.

Dans les problèmes statistiques de détection, nous avons des données $\{Y_i, X_i\}$, où X_i est le vecteur prédictif de la i^e observation et $Y_i = 0 \setminus 1$ est son étiquette de classe. L'objectif d'un problème statistique de détection est d'identifier les observations de la classe-1, qui sont extrêmement rares. Notre travail précédent sur le LAGO a été démontré qu'il peut être supérieur aux méthodes telles que KNN et SVM dans un problème réel de découverte d'un médicament. Un désavantage présent du LAGO est qu'il fournit seulement une estimation ponctuelle qu'une observation test puisse être dans la classe 1, ignorant l'incertitude du modèle. Dans cet entretien, nous présentons un cadre probabiliste pour LAGO, référé par BLAGO. L'incertitude peut être adaptée en assignant des distributions aux paramètres. MCMC est utilisé pour obtenir *a priori* les distributions postérieures. BLAGO fournit les prévisions probabilistes appropriées de support sur $(0, 1)$ et détecte aussi l'incertitude des prévisions.

[MS-73]

Profile-monitoring Curve Data**Contrôle de profils pour des données fonctionnelles**Sofia MOSESOVA, *University of Waterloo*, Hugh CHIPMAN, *Acadia University*

In this talk we generalize some of the traditional process monitoring techniques to functional (curve) data, with emphasis on a problem taken from the automotive industry. Two types of models involving B -spline smoothing and differential equations are proposed for describing the dynamics of such data. Estimated coefficients obtained from fitting these models are subsequently monitored using a Hotelling T^2 statistic. Simulated data is used to validate the results.

Dans cette présentation, nous généralisons certaines des techniques traditionnelles de contrôle des procédés à des données fonctionnelles (courbes) en mettant l'accent sur un problème de l'industrie automobile. Deux types de modèles basés sur le lissage de B -splines et des équations différentielles sont proposés pour décrire la dynamique de telles données. Les coefficients estimés obtenus à partir de ces modèles sont par la suite contrôlés à l'aide de la statistique T^2 de Hotelling. On utilise des données de simulation pour valider les résultats.

Session 04G**Lundi 29 mai • Monday, May 29, 15:30–17:00****SSC3006**

Methodology for Time-to-event Data

Méthodologie pour des données de temps d'attente

[MS-74]**Efficient Semiparametric Estimation in a Class of Conditional Single-index Survival Models****Estimation semi paramétrique efficace dans une classe de modèles de survie conditionnels à indice simple**Xuewen LU & Jie SUN, *University of Calgary*

In this article, we investigate a class of conditional single-index survival models. Our models assume that the hazard functions depend on the covariates through a single-index but not otherwise restrict the shape of the hazard function. Therefore, our models generalize many existing parametric or semiparametric survival models in the literature, including proportional hazards models, accelerated failure time models, in more general, generalized linear models and linear transformation models. A pseudo-likelihood-based inference is developed to estimate the single-index coefficients: the resulting estimator is strongly consistent, asymptotically normal and achieves the semiparametric efficiency bound. A simulation study is carried out to examine the finite sample behaviours of the estimator.

Dans cet article, nous nous intéressons à une classe de modèles de survie conditionnels à indice simple. Nos modèles supposent que la fonction de risque dépend des covariables à travers un indice simple mais ne restreignent pas autrement la forme de la fonction de risque. Notre modèle généralise donc plusieurs modèles de survie paramétriques et semiparamétriques existant dans la littérature, incluant les modèles à risques proportionnels, les modèles à temps accéléré, les modèles linéaires généralisés et les modèles de transformation linéaire. Une inférence basée sur le maximum de la pseudo-vraisemblance est développée pour estimer les coefficients de l'indice simple : l'estimateur résultant converge fortement, est asymptotiquement normal et atteint la borne d'efficacité semi-paramétrique. Une étude de simulation est effectuée afin d'examiner le comportement de l'estimateur sous un échantillon fini.

[MS-75]**Marginal Hazards Model for Multivariate Failure Times with Auxiliary Covariates****Modèle des risques marginaux pour des temps de ruine multivariés avec des covariables auxiliaires**Zhaozhi FAN, *Memorial University of Newfoundland*

We consider marginal hazards model of multivariate failure times with continuous auxiliary covariates. We extend the kernel smoothing procedure of Zhou and Wang (2000) to correlated failure time data. We estimate the marginal partial likelihood function with a semiparametric procedure. We present asymptotic properties of the induced estimator and provide some simulation results.

Nous considérons le modèle des risques marginaux pour des temps de ruine multivariés avec des covariables auxiliaires multivariés. Nous adaptons la procédure de lissage par le noyau de Zhou et Wang (2000) aux données corrélées des temps de ruine. Nous estimons la fonction de vraisemblance partielle marginale avec une procédure semi paramétrique. Nous présentons les propriétés asymptotiques de l'estimateur par induction et fournissons quelques résultats de simulation.

[MS-76]**A Marginal Regression Model for Multivariate Failure Time Data with a Surviving Fraction
Un modèle de régression marginal pour données multivariées de temps de défaillance avec une proportion qui survit**

Yingwei PENG, *University of Michigan & Memorial University of Newfoundland*, Jeremy TAYLOR, *University of Michigan*, Binbing YU, *Information Management Services, Inc., USA*

In this work, we consider a semiparametric marginal proportional hazard mixture cure model for clustered survival data. Unlike the clustered data in previous work, the latent binary cure statuses of patients in one cluster tend to be correlated in addition to the possible correlated failure times among the patients in the cluster who are not cured. The complexity of specifying appropriate correlation structures for the data becomes even worse if the potential correlation between cure statuses and the failure times in the cluster has to be considered. We formulate a semiparametric marginal proportional hazards mixture cure model and propose an estimation method. Simulation studies are conducted to assess finite sample properties of the proposed model and method. The marginal model is illustrated with a real cancer data set.

Nous considérons ici un modèle marginal semi paramétrique à risques proportionnels pour des données de survie en grappes. Contrairement aux données en grappes des travaux précédents, les états latents de guérison binaire des patients au sein d'une grappe tendent à être corrélés en plus d'être possiblement corrélés avec les temps de défaillance des patients de la grappe qui ne sont pas guéris. La complexité de la spécification de la structure de corrélation entre les états de guérison et les temps de défaillance est augmentée lorsque les corrélations potentielles entre les états de guérison et les temps de défaillance à l'intérieur de la grappe doivent être considérées. Nous formulons un modèle marginal semi paramétrique à risques proportionnels et proposons une méthode d'estimation. Des simulations sont effectuées afin d'évaluer les propriétés du modèle et de la méthode proposés sous un échantillon de taille finie. Le modèle marginal est illustré à l'aide de données sur le cancer.

[MS-77]**On the Asymptotics of NPMLE of a Distribution Function Based on Bivariate Interval-censored Data
Propriétés asymptotiques de l'EVMNP d'une distribution basé sur des données bidimensionnelles censurées par intervalles**

Dianliang DENG, *University of Regina*

The nonparametric maximum likelihood estimation of a bivariate distribution function based on interval-censored data is studied and a self-consistent equation is derived for the estimator. Further, the strong consistency and convergence rate for NPMLE are obtained for the specified cases.

L'estimation par maximum de vraisemblance non paramétrique (EVMNP) d'une distribution bidimensionnelle basée sur des données censurées par intervalles est étudiée et une équation auto-cohérente satisfaite par l'estimateur est obtenue. De plus, la convergence presque sûre et le taux de convergence de l'EVMNP sont obtenus pour des cas spécifiques.

[MS-78]**Low-risk Dose Estimation for Environmental Toxic Agents—Weibull Hazard Model
Estimation de la dose à faible risque d'agents toxiques environnementaux – le modèle de Weibull**

Abdul Sattar AL-KHALIDI, *Insurance Bureau of Canada*

This paper concerns with modeling and analysis of toxicity data in which the time to advert response is recorded or estimated. The low risk dose estimator is derived from a parametric probability distribution for the time to toxic response to provide ex-

Cet article porte sur la modélisation et l'analyse de données de toxicité où le temps de réponse est enregistré ou estimé. L'estimateur de la dose à faible risque est dérivé d'une distribution paramétrique pour le temps de réponse toxique afin de fournir la durée de l'effet d'exposition et les niveaux de dose.

posure duration effect, as well as dose levels. The derived probability distribution stems from multi-hit assumption in which a Weibull hazard function is induced from the maximum response time. In the area of toxicology and environmental health, the damaging effects of environmental chemical (such as gaseous toxins) or other stimuli are often studied in animal and microbial systems. Data from such experiments will be analyzed via the model developed in this work. The procedure developed is illustrated on data from carcinogenicity assays.

La distribution dérivée découle du présupposé de frappes multiples où une fonction de risque de Weibull est induite du temps de réponse maximal. Dans le domaine de la toxicologie et de la santé environnementale, les effets dommageables d'agents chimiques présents dans l'environnement (comme les toxines gazeuses) ou autres stimuli sont souvent étudiés dans des systèmes animaux et microbiens. Les données de telles expériences seront analysées avec le modèle développé ici. La procédure développée est illustrée sur des données de dosage de cancérogénéité.

Session 04H

Lundi 29 mai • Monday, May 29, 15:30–17:00

SSC3014

Nonparametric Methods

Méthodes non paramétriques

[MS-79]

Two-dimensional Survival Analysis Under General Censoring

Analyse de survie bidimensionnelle sous censure générale

Alberto CARABARIN & B. Gail IVANOFF, *University of Ottawa*

The filtering model in two dimensions based on 'clouds' and 'anti-clouds' is reviewed, and it is shown to have a fundamental problem with the observability of the Nelson-Aalen estimator for the integrated hazard function. A solution is proposed based on filtering by a more general type of random set. A functional central limit theorem is given for the Nelson-Aalen estimator obtained under the new model.

Nous passons en revue le modèle de filtration à deux dimensions basé sur des « nuages » et des « anti-nuages », et montrons qu'il comporte un problème fondamental avec l'observabilité de l'estimateur de Nelson-Aalen pour la fonction de risque cumulatif. Une solution basée sur la filtration par un ensemble aléatoire de type plus général est proposée. Un théorème central limite fonctionnel est donné pour l'estimateur de Nelson-Aalen obtenu sous le nouveau modèle.

[MS-80]

Nonparametric Tests of Hypotheses for Umbrella Alternatives

Tests d'hypothèse non paramétriques pour des alternatives de type parapluie

Mayer ALVO & Vladislav BRION, *University of Ottawa*

We consider the problem of testing for the hypothesis of no treatment effect against the alternative that there is a monotone treatment effect subject to change in direction. We propose a general procedure based on the ranks of the observations. Two sets of rankings are defined: one labeled the extremal set corresponding to the alternative and the other defined by the observations. A metric between any two rankings is selected. The test statistic is then based on the average overall distance between pairs of rankings, one selected from each set. We illustrate the approach using the Spearman metric.

Nous considérons le problème de tester l'hypothèse qu'il n'y ait aucun effet de traitement contre l'alternative qu'il y a un effet de traitement monotone et sujet à un changement de direction. Nous proposons une procédure générale basée sur le rang des observations. Deux ensembles de rangs sont définis : l'ensemble dit externe et correspondant à l'alternative et l'ensemble défini par les observations. Une métrique entre n'importe quelle paire de rangs est choisie. La statistique de test est alors basée sur la distance moyenne entre les paires de rangs où un rang est choisi parmi chaque ensemble. Nous illustrons cette approche en utilisant la métrique de Spearman.

[MS-81]**Projection-directed Invariant Tests for the Multivariate Multi-sample Problem
Tests invariants par projection pour le problème multivarié à plusieurs échantillons**Xiaobin YUAN & Cheng CHENG, *St. Jude Children's Research Hospital*

A class of simple tests combining projection and univariate nonparametric tests is proposed for the multivariate multi-sample problem. The direction for projection is based on principal components of the pooled sample. For the projected data at each direction, we can choose proper univariate tests for different alternatives. Our proposed tests have little asymptotic efficiency loss for testing multivariate Normal means compared to the optimal T-square test. This method can be modified for right-censored bivariate data.

Une classe de tests simples combinant des tests non paramétriques univariés et des projections est proposée pour le problème multivarié à plusieurs échantillons. La direction de la projection est basée sur les composantes principales de l'échantillon groupé. Pour les données projetées dans chaque direction, nous pouvons choisir le test univarié approprié pour différentes alternatives. La perte d'efficacité asymptotique des tests que nous proposons pour tester les moyennes d'une Normale multivariée, par rapport au test optimal du T-carré, est très faible. Cette méthode peut être modifiée pour des données bivariées censurées à droite.

[MS-82]**Nonparametric Likelihood Weights
Poids non paramétriques pour la vraisemblance pondérée**Jean-François PLANTE & James V. ZIDEK, *University of British Columbia*

Suppose samples are available from M similar populations, one of which is of inferential interest. The weighted likelihood described by Wang (2001) uses exponential weights to include all data for inference. Determining the weights is challenging. We present a data-based nonparametric criterion for choosing the weights and justify its construction intuitively. We then explore the properties and performances that result from using these adaptive weights.

Vous disposez de M échantillons provenant de populations semblables dont une sur laquelle porte l'inférence. La vraisemblance pondérée telle que décrite par Wang (2001) utilise des poids exponentiels afin d'inclure toutes les données dans le processus inférentiel. Un grand défi consiste à déterminer ces poids. Nous présentons et justifions intuitivement un critère échantillonnal non paramétrique pour choisir les poids. Nous étudions ensuite les propriétés et les performances qui découlent de ce choix.

[MS-83]**Comparison of Two Non Parametric Regression Methods with Application to a Mammary Density Study
Comparaison de deux méthodes de régression non paramétrique multivariées par une application à une étude portant sur la densité mammaire**Nathalie VANDAL, *Université Laval*

Non parametric regression is a statistical tool to describe the relationship between a dependent variable and one or more predictors without specifying an exact form for this relationship. In this presentation, a short description of GAM and polyMARS estimation methods will be given, followed by a comparison of results obtained when applying these two methods to a dataset from a mammary density study.

La régression non paramétrique est un outil statistique permettant de décrire la relation entre une variable dépendante et une ou plusieurs variables explicatives, sans spécifier de forme stricte pour cette relation. Dans cette présentation, une brève description des méthodes GAM et polyMARS sera exposée, suivi d'une comparaison des résultats obtenus lors de l'application de ces deux méthodes à des données provenant d'une étude portant sur la densité mammaire.

[MS-84]**Science, Mathematics, and Statistics
Science, mathématiques, et statistique**G. ARTHUR MIHRAM, *Princeton, NJ*, DANIELLE MIHRAM, *University of Southern California*

We review our statistical literature dealing particularly with our relationship to Science. Science is that human activity devoted to the search for the very explanation for (i.e., for the truth about) some particular naturally occurring phenomenon. The Scientific Method is a six-stage model-building process. Statistical model-building often is supportive of the scientist's effort; yet, we conclude with a probabilistic statement regarding any designed experiment, meaning that we necessarily fall short of the goal of Science (viz., providing the very explanation). Are there shortcomings of mathematics (applied, then pure) to Science? Our conclusion: Mathematics is neither necessary nor sufficient for Science.

Nous examinons notre littérature de statistique qui traite en particulier de notre rapport avec la science. La science est cette activité humaine consacrée à la recherche de l'explication exacte (c.-à-d., la vérité au sujet) d'un phénomène naturel en particulier. La Méthode Scientifique est un processus de six étapes pour la construction d'un modèle. La construction d'un modèle statistique soutient souvent l'effort de l'homme de science ; néanmoins, nous concluons avec une déclaration probabilistique au sujet de n'importe quelle expérience bien conçue, ce qui sous-entend que nécessairement nous faillons à notre obligation envers le but de la science (à savoir, fournir l'explication exacte). Y a-t-il des imperfections en mathématiques (appliquées, puis pures) par rapport à la Science ? Notre conclusion : Les mathématiques ne sont ni nécessaires ni suffisantes pour la Science.

Session 05A**Mardi 30 mai • Tuesday, May 30, 08:30–09:45****SSC2050****Gold Medal Address****Allocution de la médaille d'or****[MS-85]****Statistical Analysis for Public Policy
L'analyse statistique au service de la politique publique**David ANDREWS, *University of Toronto*

Public policy is increasingly informed by the statistical analysis of studies and administrative data. Policy is guided by measures of the magnitude of problems and interventions. To be effective, these measures have to be understood and accepted. The largest problem is that of bias.

In this lecture, symbolic computation in R is used to show surprising new and some forgotten results about common statistical procedures. From these results follow general principles for statistical analysis for public policy. These principles are illustrated with recent health studies.

La politique publique est de plus en plus informée par l'analyse statistique d'études et de données administratives. La politique est guidée par des mesures démontrant l'ampleur des problèmes et des interventions. Pour être efficaces, ces mesures doivent être comprises et acceptées. Le plus grand problème est alors celui du biais.

Dans cet exposé, le calcul symbolique avec R est utilisé pour montrer des résultats surprenants, nouveaux et oubliés, sur quelques procédures statistiques communes. De ces résultats découlent des principes généraux concernant l'analyse statistique en politique publique. Ces principes sont illustrés avec de récentes études sur la santé.

Session 06A Mardi 30 mai • Tuesday, May 30, 10:15–11:45 SSC3018

**Some Current Developments in the Analysis of Complex Survey Data
Développements récents en analyse de données d'enquêtes complexes**

[MS-86]

Analysis of Longitudinal Survey Data with Missing Observations: An Application to the National Longitudinal Survey of Children and Youth (NLSCY)

Analyse de données longitudinales en présence d'observations manquantes : une application à l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ)

Ivan CARRILLO-GARCIA, *University of Waterloo*, Milorad KOVACEVIC, *Statistics Canada*, Changbao WU, *University of Waterloo*

All surveys, either cross-sectional or longitudinal, contain nonresponse. Under these circumstances one must make assumptions about the response mechanism. There is extensive literature about nonresponse for cross-sectional surveys. For longitudinal surveys, the missing data problem, the properties of different response mechanisms, and their impacts on variances are much less known. Assuming MAR, we apply the Weighted GEE modeling, following the lines of earlier research for longitudinal studies under non-survey settings (Robins et al., 1995), to estimate the regression coefficients and their joint randomization variance when either dropouts or intermittent nonresponse is present. We use the NLSCY for our analyses.

Toutes les enquêtes, qu'elles soient transversales ou longitudinales, sont sujettes à la non-réponse. Dans ces conditions, il est nécessaire d'établir des postulats quant au mécanisme de réponse. La littérature au sujet de la non-réponse dans les enquêtes transversales est abondante. Cependant, le problème des données manquantes, les propriétés des différents mécanismes de réponse et leurs impacts sont moins bien connus dans le cas d'études longitudinales. Sous l'hypothèse de données manquantes aléatoirement, nous appliquons la modélisation par EEG pondérées, suivant l'approche de recherche antérieure en études longitudinales (Robins et al., 1995), afin d'estimer les coefficients de régression et leur variance conjointe lorsque des abandons ou de la non-réponse intermittente sont présents. Nous utilisons l'ELNEJ pour nos analyses.

[MS-87]

Treatments for the Problem of Identification of Links in Indirect Sampling

Traitements pour le problème de l'identification des liens en échantillonnage indirect

Xiaojuan XU, *University of Alberta*, Pierre LAVALLÉE, *Statistics Canada*

In the context of indirect sampling, we address the problem of overestimation caused by link nonresponse when using generalized weight share method (GWSM). A few adjustment methods, with and without use of auxiliary variables, have been constructed. The properties of these methods were studied in a simulation study based on the longitudinal survey data. The simulation results show that these adjusted using the GWSM perform well in reducing both, the estimation bias and the variance. The advancement in bias reduction is significant.

Dans le contexte de l'échantillonnage indirect, nous nous intéressons au problème de la surestimation causée par la non-réponse de lien lorsque la méthode de la part des poids généralisée (MPPG) est utilisée. Quelques méthodes d'ajustement, utilisant ou non des variables auxiliaires, ont été construites. Les propriétés de ces méthodes ont été étudiées dans une étude de simulation basée sur des données longitudinales. Les résultats de simulation montrent que les ajustements utilisant la MPPG performant bien en réduisant à la fois le biais et la variance. Le progrès en réduction de biais est significatif.

[MS-88]

Resampling Methods for Two-stage Sampling in Complex Surveys**Méthodes de ré-échantillonnage pour l'échantillonnage à deux phases dans les enquêtes à plans complexes**Chunfang (Devon) LIN & Randy SITTER, *Simon Fraser University*

In two-stage sampling in complex surveys, replication-based variance estimation is often applied to the first-stage sampling units. Theoretical justification is based on the assumption that the first-stage sampling fraction is negligible. Motivated by some surveys where this assumption is not met, we develop adaptations of the method of balanced repeated replications and bootstrapping that do not require this assumption, explore the asymptotic properties of the derived variance estimators and conduct a simulation study to evaluate the performance of the proposed methods with finite sample size.

Dans l'échantillonnage à deux phases dans les enquêtes aux plans complexes, l'estimation de la variance basée sur la réplication est souvent appliquée aux unités sélectionnées à la première phase. La justification théorique est fondée sur l'hypothèse que la fraction de l'échantillonnage de la première phase est négligeable. Motivé par l'existence de quelques enquêtes où cette hypothèse n'est pas satisfaite, nous développons des adaptations de la méthode de réplifications répétées équilibrées et de la méthode de ré-échantillonnage qui n'exigent pas cette hypothèse. Nous explorons les propriétés asymptotiques des estimateurs dérivés de la variance et réalisons une étude de simulation pour évaluer la performance des méthodes proposées avec une taille d'échantillon finie.

Session 06B**Mardi 30 mai • Tuesday, May 30, 10:15–11:45****SSC2024****Regression Models: Recent Advances and Applications****Modèles de régression : Développements récents et applications**

[MS-89]

Modelling Dependent Bivariate Time-to-event Data Using Accelerated Life Regression**Modélisation des données dépendantes bidimensionnelles en utilisant une régression accélérée de la survie**Yun Hee CHOI, *Samuel Lunenfeld Research Institute at the Mount Sinai Hospital*, E. David MATTHEWS & Y. Yi YI, *University of Waterloo*

To analyze bivariate time-to-event data from matched or naturally paired study designs, we introduce a bivariate accelerated life regression model that uses shared frailties, and describe a flexible computational framework for fitting this model. We investigate the statistical properties of the model, and develop diagnostic tools for assessing model fit using the bivariate probability integral transformation of univariate residuals. Our proposed residuals plots will help confirm the choice of an appropriate frailty distribution. Through simulation studies and graphical displays, we characterize the sampling behaviour of these residuals and demonstrate how well these tools address questions of model fit.

Pour les analyses de survie pour des données bidimensionnelles des plans d'étude appariés ou naturellement jumelés, nous présentons un modèle accéléré bidimensionnel de régression de la survie qui emploie des faiblesses communes, et décrivons un cadre informatique flexible pour ajuster ce modèle. Nous étudions les propriétés statistiques du modèle, et développons des outils diagnostiques pour évaluer l'ajustement de modèle en utilisant la transformation intégrale de la probabilité bidimensionnelle des résidus unidimensionnels. Nos graphiques de résidus proposés aideront à confirmer le choix d'une distribution appropriée de faiblesse. Par des simulations et des représentations graphiques, nous caractérisons le comportement de prélèvement de ces résidus et démontrons à quel point ces outils ajustent le modèle.

[MS-90]**Regression Models for Longitudinal Data and its Applications****Les modèles de régression en présence de données longitudinales et leurs applications**Tulay KORU-SENGUL, *McMaster University*

Longitudinal data analysis has been extensively used in many fields especially in medicine, epidemiology and social sciences to study the same group of subjects over time. There are different modeling approaches for analyzing longitudinal studies which may differ substantially from each other and the choice of a model depends, of course, on the desired goal of the study. This talk will introduce longitudinal designs and investigate different modeling approaches by outlining the similarities and dissimilarities. Applications of the models will be given for several clinical studies. Recent advances for analyzing longitudinal studies such as joint models will also be discussed.

L'analyse de données longitudinales est largement utilisée dans plusieurs domaines, en particulier en médecine, en épidémiologie et en sciences sociales, pour étudier le même groupe de sujet à travers le temps. Il existe différentes approches à la modélisation de données longitudinales différant substantiellement l'une de l'autre et le choix du modèle utilisé dépend, bien sûr, de l'objectif de l'étude. Cet exposé présente des plans longitudinaux et examine différentes approches à la modélisation en soulignant leur similitudes et dissimilitudes. Des applications des modèles à des études cliniques seront présentées, ainsi que certaines avancées récentes en analyse longitudinales, telles les modèles conjoints.

Session 06C**Mardi 30 mai • Tuesday, May 30, 10:15–11:45****SSC2050**

Bioinformatics

Bioinformatique

[MS-91]**Applications of Computational Intelligence to Pattern Discovery in Biological Systems****Applications de l'intelligence informatique à la découverte de patrons reliés aux systèmes biologiques**Gary FOGEL, *Natural Selection, Inc.*

Biological systems are inherently nonlinear. One method of understanding biological systems is to first reduce this nonlinearity to linear relationships and apply standard statistical methods for model development. When the underlying mechanics are largely linear, these methods work very well. However, when the biology is vastly nonlinear, novel computational methods can be employed that avoid reduction to first principles. This lecture describes efforts to use computational intelligence methods for pattern recognition and pattern discovery in an assortment of biological problems which contain nonlinearity in feature space and/or have vast number of possible solutions.

Les systèmes biologiques sont intrinsèquement non linéaires. Une méthode pour comprendre les systèmes biologiques est de réduire cette non-linéarité à des relations linéaires pour ensuite appliquer les méthodes statistiques usuelles de développement de modèles. Lorsque les mécaniques sous-jacentes sont largement linéaires, ces méthodes fonctionnent très bien. Cependant, lorsque les variables en biologie sont largement non linéaires, de nouvelles méthodes informatiques peuvent être utilisées pour éviter la réduction aux premiers principes. Cet exposé décrit des tentatives d'utilisation d'intelligence informatique pour la reconnaissance et la découverte de formes dans un assortiment de problèmes biologiques dont l'espace de paramètres contient de la non-linéarité et/ou qui ont un grand nombre de solutions possibles.

[MS-92]**Operons and the Quality of Predictions of Functional Interactions Using Phylogenetic Profiles****Les opérons et la qualité des prévisions des interactions fonctionnelles en utilisant des profils phylogénétiques**Gabriel MORENO-HAGELSIEB, *Wilfrid Laurier University*, Sarath JANGA, *CCG-UNAM*

Operons are composed of functionally-related genes. Phylogenetic profiles are vectors describing the presence/absence of a gene across a genome dataset. If two genes work together, they should be both either absent or present across organisms. Thus, their phylogenetic profiles should correlate. We devised a confidence value for operon predictions using phylogenetic profiles valid in any Prokaryote. A non-redundant genome dataset increases the number of predictions at a confidence value of 0.90 by an average ratio of 1.36. The results show the power and importance of positive and negative gold standards for the improvement of predictions using genomic context.

Les opérons se composent de gènes qui fonctionnent d'une façon reliée. Les profils phylogénétiques sont des vecteurs décrivant la présence ou l'absence d'un gène à travers d'un ensemble de données de génome. Si deux gènes fonctionnent ensemble, ils devraient être absents ou présents à travers des organismes. Par conséquent, leurs profils phylogénétiques devraient être corrélés. Nous avons déterminé une valeur de confiance pour des prévisions d'opéron en utilisant des profils phylogénétiques valides dans n'importe quel Prokaryote. Un ensemble de données non redondant de génomes augmente le nombre de prévisions à un niveau de confiance de 0,90 par un ratio moyen de 1,36. Les résultats montrent la puissance et l'importance des étalons positifs et négatifs pour l'amélioration des prévisions en utilisant le contexte génomique.

[MS-93]

Multiclustering, Avoiding the Shape of Underlying Metrics
Éviter la forme des métriques sous-jacentes en classification

Daniel ASHLOCK, *University of Guelph*

Any clustering technique depends strongly on the underlying distance or similarity measure used to compare the objects being clustered. Methods for estimating the "correct" number of clusters are dependent on human judgment or may require multiple runs of an algorithm followed by statistical analysis of the clusters produced. This talk introduces multiclustering which combines the output of hundreds or thousands of clustering runs in a manner that both avoids some kinds of bias introduced by the underlying metric and also gives an advisory, called a cut plot, that advises the user as to the structure of the data including the natural number of clusters if one exists.

Toute méthode de classification dépend fortement de la mesure de distance ou de similarité utilisée pour comparer les objets à regrouper. Les méthodes estimant le nombre « correct » de groupes dépendent du jugement humain ou requièrent de multiples itérations d'un algorithme suivi d'une analyse statistique des groupes produits. Cet exposé présente une méthode de multiclassification qui combine les résultats de centaines ou milliers d'itérations tout en évitant le biais introduit par la métrique sous-jacente et fournit un graphe de coupe conseillant l'utilisateur quant à la structure des données, incluant le nombre naturel de groupes, s'il existe.

Session 06D**Mardi 30 mai • Tuesday, May 30, 10:15–11:45****SH3345**

Meta-analysis of Randomized Controlled Trials
Méta-analyse d'essais contrôlés randomisés

[MS-94]

Combining Studies with Continuous and Binary Outcomes
Combinaison d'études avec résultats continus et binaires

Andrew WILLAN & Joseph BEYENE, *SickKids Research Institute/University of Toronto*

In recent years, the medical literature has witnessed an exponential growth in research that uses meta-analytic methods. The analytical methods are well-established for combining studies involving effect measures reported using uniform scales for the outcome of interest, for example scenarios where all studies that need to be combined use a continuous scale. A challenge arises in situations where a combination of measurement scales is reported for the same end point. We present an overview of methods that can be used for combining effect sizes where some studies are reporting summary measures on a continuous scale and others on binary scales. The performances of the various methods are compared using simulations and real data sets are used for illustration.

Ces dernières années, la littérature médicale a été témoin d'une croissance exponentielle de la recherche employant des méthodes méta-analytiques. Les méthodes analytiques sont bien établies pour combiner des études impliquant l'effet de mesures rapportées en utilisant des échelles uniformes pour les résultats d'intérêt, comme par exemple, les scénarios où toutes les études devant être combinées utilisent une échelle continue. Un défi surgit dans les situations où des études employant des échelles de mesure complètement différentes sont combinées dans un même but. Nous présentons une revue des méthodes qui peuvent être employées pour combiner les résultats provenant d'études rapportant des mesures sommaires sur des échelles continues et binaires. Les performances des diverses méthodes sont comparées à l'aide de simulations et de données réelles.

[MS-95]**Use of Bayesian Methodology in Meta-analysis****Utilisation de méthodologie bayésienne pour la méta-analyse**Anna SAWKA, *University of Toronto*

An advantage of using a Bayesian approach in meta-analysis is that prior knowledge, beliefs, or assumptions can be systematically incorporated in a hierarchical model. For example, prior knowledge about the treatment effect of an intervention in a different population than the one being studied may be incorporated in a model. In another example, prior knowledge about a design effect may be incorporated in pooling data from cluster randomized trials. This presentation will discuss the applications and benefits of Bayesian methods in pooling data from randomized controlled trials with complex designs.

Un avantage d'employer une approche bayésienne pour la méta-analyse vient du fait que les connaissances antérieures, croyances, ou hypothèses peuvent être systématiquement incorporées à un modèle hiérarchique. Par exemple, une connaissance antérieure quant à l'effet d'une intervention dans une population différente de celle étant étudiée peut être incorporée dans un modèle. Dans un autre exemple, une connaissance antérieure au sujet de l'effet dû au plan d'expérience peut être incorporée en regroupant des données produites lors d'essais randomisés par grappes. Cette présentation présentera l'application et les avantages des méthodes bayésiennes permettant le regroupement des données provenant d'essais contrôlés randomisés à plans d'expérience complexes.

[MS-96]**A Generalized Meta-analysis Model: The Combination of Observations and Studies****Un modèle généralisé de méta-analyse : la combinaison des observations et des études**Keith O'ROURKE, *Ottawa Health Research Institute*

Most analysis techniques default to addressing a single study and place restrictions on the distributions and parameterizations—i.e., Linear, Generalized Linear, Generalized Additive, and Proportional Hazards models. Meta-analysis defaults to the situation of multiple studies and given independent groups of independent observations a wide range of distributions with common and non-common parameters and various parameterizations—can be accommodated. Importance sampling based likelihoods are used for various reported summaries to enable direct Bayesian or Classical inference. Numerical integration with guaranteed bounds allows various random effect models and simulated modified profile likelihood may provide some assurance that Classical confidence intervals are valid.

La plupart des techniques d'analyse prennent en considération une seule étude et imposent des restrictions sur les distributions et les paramétrisations – par exemple, les modèles linéaires, linéaires généralisés, additifs généralisés et à défaillance proportionnelle. La méta-analyse prend en considération de multiples études et, en présence de groupes indépendants d'observations indépendantes, une large gamme de distributions – avec ou sans paramètres communs – et de paramétrisations peuvent être accommodées. Les vraisemblances basées sur l'échantillonnage d'importance sont utilisées pour produire divers sommaires permettant directement une inférence bayésienne ou classique. L'intégration numérique avec bornes garanties permet plusieurs modèles à effets aléatoires et la vraisemblance de profil modifiée simulée peut fournir une certaine assurance relative à la validité des intervalles de confiance classiques.

Session 06E**Mardi 30 mai • Tuesday, May 30, 10:15–11:45****UC85**

Advances in Markov Processes

Développements dans les processus de Markov

[MS-97]**A Mean Field Analysis of Decentralized Random Access Protocols****Analyse par une approximation champ moyen d'un protocole d'accès décentralisé**David McDONALD, *University of Ottawa*, Charles BORDENAVE & Alexandre PROUTIERE, *École Normale Supérieure*

The stability and performance of decentralized random access protocols for Wireless LANs and Ad-Hoc networks is complicated due to inherent interaction between sources. One can circumvent this difficulty by decoupling the sources; i.e., by assuming that the (re)-transmission processes of the different sources are mutually independent. This assumption allows one to derive explicit estimates of the performance. This approach was for example applied by Bianchi to analyze the IEEE 802.11 Decentralized Coordination Function (DCF) algorithm and since has been widely used to predict the performance of similar protocols.

Using mean field techniques, we prove that, for a wide range of random back-off algorithms, the decoupling assumption is asymptotically exact as the number of sources grows. In the specific case of exponential back-off algorithm (the DCF is based on this algorithm), the mean field analysis provides the transient and stationary distributions of the (re)-transmission processes.

La stabilité et la performance des protocoles d'accès aléatoire décentralisé pour les LANs sans fil et les réseaux ad hoc sont compliquées par la présence d'interaction inhérente entre les sources. On peut éviter cette difficulté en découplant les sources; c.-à-d. en supposant que les processus de (re)-transmission des différentes sources sont mutuellement indépendants. Cette hypothèse permet de dériver des estimations explicites de la performance. Cette approche a été appliquée, par exemple, par Bianchi pour analyser l'algorithme de la Fonction de Coordination Décentralisée IEEE 802.11 (FCD). Depuis, son utilisation est largement répandue pour la prévision de la performance de protocoles semblables.

À l'aide des techniques de champ moyen, nous montrons que, pour un grand nombre d'algorithmes d'accès aléatoire de reprise, l'hypothèse de découplage est asymptotiquement exacte lorsque le nombre de sources augmente. Dans le cas spécifique de l'algorithme exponentiel à reprise (le FCD est basé sur cet algorithme), l'analyse du champ moyen fournit les distributions transitoires et stationnaires des processus de (re)-transmission.

[MS-98]**Reversibility for diffusions via Quasi-invariance****La réversibilité pour un processus de diffusion par le biais de la quasi-invariance**Byron SCHMULAND, *University of Alberta*

My topic is the regularity of reversible distributions for diffusion processes. After discussing the problem in Euclidean space, I turn to the more complex situation of the Fleming-Viot (measure-valued) process. We show that reversible distributions are quasi-invariant under a certain flow, forcing the mutation operator to satisfy a cocycle identity.

Mon sujet est la régularité de distributions réversibles pour processus de diffusion. Après une discussion du problème dans l'espace euclidien, je me tourne vers la situation plus complexe du processus de Fleming-Viot. Nous montrons que les distributions réversibles son quasi-invariantes sous un certain écoulement, forçant l'opérateur de mutation à satisfaire une identité de cocycle.

[MS-99]**Catalytic Branching Networks
Réseaux avec branchement catalytique**Donald DAWSON, *Carleton University*

We consider a class of diffusion processes that model catalytic branching, that is, a branching process in which the branching rate is enhanced by the presence of a catalyst. A catalytic branching network corresponds to a multitype system in which some types serve as catalysts for other types. Catalytic branching networks in which there are closed cycles of catalytic types and catalytic systems distributed in space pose a number of challenging mathematical problems that will be discussed. Parts of this work are joint with E.A. Perkins, T. Cox and A. Greven.

Nous considérons une classe de processus de diffusion modélisant le branchement catalytique, c'est-à-dire un processus de branchement dans lequel le taux de branchement augmente en présence d'un catalyseur. Un réseau de branchement catalytique correspond à un système à plusieurs types où certains types servent de catalyseur pour d'autres. Les réseaux de branchement catalytique où se trouvent des cycles fermés de types catalytiques et des systèmes catalytiques distribués dans l'espace posent plusieurs problèmes mathématiques difficiles dont nous discuterons. Ce travail fut en partie réalisé conjointement avec E.A. Perkins, T. Cox et A. Greven.

Session 06F**Mardi 30 mai • Tuesday, May 30, 10:15–11:45****SSC3006****Classical Statistics
Statistique classique****[MS-100]****A Weighted Multivariate Sign Test for Cluster Correlated Data
Un test pondéré des signes multivarié pour des données corrélées en grappes**Denis LAROCQUE, *HEC Montréal*, Jaakko NEVALAINEN & Hannu OJA, *University of Tampere*

We consider the multivariate location problem with cluster correlated data. A family of multivariate weighted sign tests are introduced for which observations from different clusters can receive different weights. The asymptotic distribution of the test statistic is obtained under the null hypothesis and also for a local alternative model under multivariate normality. Optimal weights maximizing Pitman asymptotic efficiency are provided. Several approaches for estimating these weights are presented. Using Pitman asymptotic efficiency, it is shown that appropriate weighting can increase substantially the efficiency compared to a test that gives the same weight to each cluster.

Nous considérons le problème de position multivarié avec des données corrélées en grappes. Une famille de tests des signes pondérés multivariés est proposée. La loi asymptotique des statistiques de test est obtenue sous l'hypothèse nulle et sous une suite de contre-hypothèses locales multivariées. Les poids optimaux maximisant l'efficacité asymptotique de Pitman sont obtenus. Plusieurs approches pour estimer ces poids sont présentées. Nous démontrons que l'utilisation de poids judicieux peut améliorer de manière notable l'efficacité par rapport au test donnant un poids égal à chaque observation.

[MS-101]**Inference of the Normal Correlation Coefficient****Inférence à propos du coefficient de corrélation de variables normales**Ye SUN & Augustine WONG, *York University*

Inference concerning the correlation coefficient of two random variables from the bivariate normal distribution has been investigated by many authors. In particular, Fisher derived various forms of the density for the sample correlation coefficient and Fisher's z -Transformation. Hotelling suggested several modifications of Fisher's methods to obtain the confidence interval for the correlation coefficient. However, inference for the correlation coefficient is sometimes biased. In this paper, a likelihood based higher order asymptotic method is proposed to obtain confidence interval for the correlation coefficient. Simulation results indicated that the proposed method is very accurate even when the sample size is small.

L'inférence liée au coefficient de corrélation de deux variables aléatoires provenant d'une distribution normale bidimensionnelle a été étudiée par de nombreux auteurs. En particulier, Fisher a trouvé différentes formes de la densité du coefficient de corrélation échantillonnal et a défini ce qui est maintenant connu comme la transformation z de Fisher. Hotelling a suggéré plusieurs modifications aux méthodes de Fisher afin d'obtenir un intervalle de confiance pour le coefficient de corrélation. Cependant, l'inférence liée au coefficient de corrélation est parfois biaisée. Dans cet article, nous proposons une méthode asymptotique d'ordre supérieur basée sur la vraisemblance permettant d'obtenir un intervalle de confiance pour le coefficient de corrélation. Des résultats de simulation indiquent que la méthode proposée est très précise, même en présence d'échantillons de petite taille.

[MS-102]**A Moment-based Percentile Approximation Methodology****Une méthodologie d'approximation des centiles basée sur les moments**Hyung-Tae HA & Serge PROVOST, *University of Western Ontario*

The exact distributions of test statistics encountered for instance in multivariate analysis are often either intractable or difficult to derive in closed forms. However, various distributional approximations can be obtained from their moments or cumulants, which can usually be determined with relative ease. The proposed methodology which essentially consists in expressing an approximate density function as the product of an appropriate base density and a polynomial whose coefficients are obtained by matching moments, will be applied to several test criteria. It appears to be easier to implement than those currently in use while providing percentiles that are at least as accurate.

La distribution exacte des statistiques de test rencontrées, par exemple, en analyse multivariée sont souvent soit insolubles, soit difficile à obtenir sous forme fermée. Néanmoins, plusieurs approximations distributionnelles peuvent être obtenues à partir de leurs moments ou cumulants, qui peuvent souvent être facilement obtenus. La méthodologie proposée, qui consiste essentiellement à exprimer une densité approximative sous forme d'un produit de la densité de base appropriée et d'un polynôme dont les coefficients sont obtenus par égalisation des moments, est appliquée à plusieurs critères-tests. Elle semble être plus aisée à mettre en œuvre que celles présentement utilisées, tout en fournissant des percentiles au moins aussi précis.

[MS-103]**Statistical Inference for the Difference Between the Best Treatment Mean and a Control Mean****Inférence statistique sur la différence entre la moyenne du groupe avec le meilleur traitement et celle du groupe contrôle**

Chu-In LEE, *Memorial University of Newfoundland*, Jianan PENG, *Acadia University*, Lin LIU, *University of California at San Diego*

In many experiments researchers are interested in comparing several treatment means with a control mean. When there are some treatments significantly better than the control, it is often of interest to evaluate the difference between the best treatment mean and the control mean and to identify the best treatment. In this talk we derive lower confidence bounds for the aforementioned difference for the case that treatments are at least as effective as the control and for the case that no restriction is placed on the treatment means and the control mean. Two efficient computation algorithms are proposed. Expected lower confidence bounds of the two procedures are compared with that of Dunnett's procedure. An application to a real life data is included.

Dans plusieurs expérimentations, les chercheurs sont intéressés à comparer plusieurs moyennes de traitement à une moyenne de contrôle. Lorsque certains traitements sont significativement meilleurs que le contrôle, il est souvent d'intérêt d'estimer la différence entre la meilleure moyenne de traitement et la moyenne de contrôle et d'identifier le meilleur traitement. Dans cette présentation, nous dérivons les limites inférieures de confiance pour la différence mentionnée ci-dessus pour le cas où les traitements sont au moins aussi efficaces que le contrôle et pour le cas où aucune restriction n'est placée sur les moyennes de traitement et la moyenne de contrôle. Deux algorithmes efficaces de calcul sont proposés. Des limites inférieures de confiance espérées des deux procédures sont comparées à celles utilisées dans la procédure de Dunnett. Une application à des vraies données est incluse.

[MS-104]**Characterization of Distributions through Conditional Expectation of Functions of Generalized Order Statistics****Caractérisation de distributions par le biais d'espérances conditionnelles de fonctions de statistiques d'ordre généralisées**

Abdul KHAN, *Aligarh Muslim University, India*

Generalized order statistics (GOS) introduced by Kamps (J. Statist. Plan. Infer., 1995), contain order statistics, records, sequential order statistics, etc. as its particular cases. In this paper, a general form of continuous distribution is obtained by considering conditional expectation of GOS conditioned on non-adjacent GOS. Various examples and deductions on characterizations of distributions are discussed, thus unifying earlier known results on order statistics, records and GOS.

Les statistiques d'ordre généralisées (SOG) introduites par Kamps (J. Statist. Plan. Infer., 1995) incluent les statistiques d'ordre, les records et les statistiques d'ordre séquentielles comme cas particuliers. Dans cet article, une forme générale de distribution continue est obtenue en considérant l'espérance conditionnelle de SOG, en conditionnant sur des SOG non-adjacentes. Des exemples variés et des conséquences quant aux caractérisations des distributions sont présentés, unifiant ainsi des résultats déjà connus relatifs aux statistiques d'ordre, records et SOG.

Session 06G**Mardi 30 mai • Tuesday, May 30, 10:15–11:45****SSC3010**

Biostatistics in Action

Biostatistique en action

[MS-105]**An Index for Cultivar Resistance Based on Disease Progress Curves****Un indice pour la résistance du cultivar basé sur des courbes de progrès de maladie**Kenneth McRAE & H.W. (Bud) PLATT, *Agriculture & Agri-Food Canada*

In breeding programs for disease susceptible crops, the degree of resistance in new cultivars is compared with the standard(s) in field trials. The spread of a foliar disease depends on the inocula, temperature and humidity which can change dramatically during the growing period and from year-to-year. A robust index for cultivar resistance is proposed. Estimation of the proportionality and onset parameters for individual cultivars are based on weighted areas under an empirical, standardized disease progress curve. Censoring of the disease record is considered. Originally developed for late blight in potato, the statistical method has potential for other crop-disease combinations.

Dans les programmes d'amélioration génétique pour les récoltes susceptibles à la maladie, suite à des tests sur le terrain, le degré de résistance de nouveaux cultivars peut être comparé au(x) standard(s) obtenus. La propagation d'une maladie foliaire dépend des inoculants, de la température et de l'humidité qui peuvent changer nettement pendant la période de croissance, et d'année en année. On propose un indice robuste pour mesurer la résistance du cultivar. L'estimation de la proportionnalité et des paramètres de départ pour différents cultivars sont basées sur des surfaces pondérées sous une courbe empirique et normalisée de progrès de maladie. La censure sur le relevé de la maladie est considérée. Cette technique développée à l'origine pour la pomme de terre atteinte du mildiou, peut potentiellement être appliquée à d'autres combinaisons de récolte-maladie.

[MS-106]**Modeling Prostate Specific Antigen (PSA) Kinetics in Prostate Cancer Patients during Surveillance****Modélisation de la cinétique de l'antigène spécifique de prostate (PSA) chez des patients sous surveillance souffrant de cancer de prostate**Liyang ZHANG, Andrew LOBLAW & Laurence KLOTZ, *Sunnybrook & Women's College Health Sciences Centre, University of Toronto*

A prospective single-arm cohort study was carried out on 231 prostate cancer patients. PSA doubling time (DT) was used to stratify patients into low and high risks of progression. The PSA kinetics in the groups were estimated by general linear mixed model.

Given the baseline status of individuals, high and low risk average curves were derived with corresponding average PSA DTs of 2.97 and 6.54 years, respectively. By applying the dynamic prognostic rule in combination with serial biopsy, a decision for intervention based on the risk of progression could be optimally recommended about 2.3 years after the initiated surveillance.

Une étude prospective de cohorte à une branche a été effectuée sur 231 patients souffrant du cancer de prostate. Le PSA à temps double (TD) a été utilisé pour stratifier les patients à petit et grand risque de progression. Les cinétiques PSA dans les groupes ont été estimés par le modèle linéaire général mixte.

Étant donné l'état des individus, des courbes pour un grand et un petit risque moyen ont été tracées avec un PSA et un TD moyen correspondant à 2,97 et 6,54 ans, respectivement. En appliquant la règle pronostique dynamique en combinaison avec la biopsie périodique, une décision pour l'intervention basée sur le risque de progression a pu être recommandée de façon optimale environ 2,3 ans après l'observation initiale.

[MS-107]**Modeling Lung Disease and its Progression in a Chronic Obstructive Pulmonary Disease Cohort
Modélisation d'une maladie pulmonaire et sa progression au sein d'une cohorte obstructive chronique**

Liqin XU & Marek SMIEJA, *McMaster University*, Lehana THABANE, *McMaster University, Centre for Evaluation of Medicines, St Joseph's Healthcare*

Chronic obstructive pulmonary disease (COPD) is a common, irreversible, slowly progressive lung disease, usually associated with smoking and with respiratory infections. The objective of the study was to investigate the association between lung function and several patient characteristics that include age, gender and height. This presentation will discuss the results of analysis to model the relationship between lung function decline and age, gender, height, smoking and Chlamydia pneumoniae infection. We compare the results based on generalized estimating equations, random effects and robust regression models. We will also discuss some results based on different imputation methods to handle missing data.

La maladie pulmonaire obstructive chronique (MPOC) est une affection pulmonaire fréquente, irréversible et lentement progressive, habituellement liée au tabagisme et aux infections respiratoires. L'objectif de l'étude était d'étudier l'association entre le fonctionnement des poumons et plusieurs caractéristiques des patients qui incluent l'âge, le sexe et la taille. Cette présentation discutera les résultats de l'analyse pour modéliser la relation entre le déclin du fonctionnement des poumons et l'âge, le sexe, la taille, le tabagisme et l'infection Chlamydia pneumoniae. Nous comparons les résultats basés sur des équations d'estimations généralisées, des effets aléatoires et des modèles robustes de régression. Nous discuterons également quelques résultats basés sur différentes méthodes d'imputation pour manipuler des données manquantes.

[MS-108]**A Meta-analysis on Bleeding and Cardiovascular Outcomes
Une méta-analyse sur le saignement et les résultats cardiovasculaires**

Ji CHENG, Jack SUN & Richard WHITLOCK, *McMaster University*, Lehana THABANE, *Centre for Evaluation of Medicines, St. Joseph Healthcare & McMaster University*

Despite guidelines recommending that aspirin be discontinued seven days prior to coronary bypass surgery, many cardiac surgeons continue to keep their patients on aspirin until the day of surgery. This may be due to conflicting evidence surrounding its effects. The objective of this study was to perform a systematic review and to summarize the evidence on the effect of preoperative ASA on bleeding and transfusion requirements in coronary bypass surgery. In this presentation, we discuss the results of the meta-analysis of both RCT and non-RCT studies using both fixed and random-effects approaches to determine the relevant clinical endpoints.

En dépit des directives recommandant que l'absorption d'aspirine soit cessée pendant sept jours avant un pontage coronarien, beaucoup de chirurgiens cardiaques continuent à garder leurs patients sur l'aspirine jusqu'au jour de la chirurgie. Ceci peut être dû aux preuves contradictoires entourant ses effets. L'objectif de cette étude était d'accomplir une révision systématique et de récapituler la preuve sur l'effet de l'aspirine avant l'opération sur les conditions de saignement et de transfusion durant la chirurgie. Dans cette présentation, nous discutons les résultats de la méta-analyse des études RCT et non-RCT en employant des approches à effets aléatoires et des approches à effets fixes pour déterminer les critères d'évaluations cliniques pertinents.

[MS-109]**Analyzing Long Term Hospitalization Records From an Administrative Database
Analyse des dossiers d'hospitalisation d'une base de données administratives**Shih-Wa YING & Gurbakhshash SINGH, *Simon Fraser University*, Mary McBRIDE, *BC Cancer Agency*

Health insurance administrative databases provide person-based health information that can be used to assess the health care utilization of patients and to identify risk factors and evaluate their effects. This talk presents analysis of hospital utilization records from Health Insurance BC (Medical Service Plan) of childhood cancer survivors. We explore several models to analyze the data, regarding the time to the first hospital admission, the admission frequency over time and the duration in hospital.

Les bases de données administratives d'assurance-santé fournissent de l'information sur la santé des individus pouvant être utilisée pour évaluer l'utilisation de soins de santé par les patients et pour identifier les facteurs de risque et évaluer leurs effets. Cet exposé présente une analyse de l'utilisation de dossiers d'hospitalisation d'enfants ayant survécu au cancer de la « Health Insurance BC ». Nous explorons plusieurs modèles pour analyser les dossiers en ce qui a trait à la date d'hospitalisation, la fréquence d'hospitalisation et la durée des hospitalisations.

Session 06H**Mardi 30 mai • Tuesday, May 30, 10:15–11:45****SSC3014**

Censored Data

Données censurées

[MS-110]**Smooth Nonparametric Quantile Regression with a Censored Response
Régression non paramétrique lisse des percentiles avec réponse censurée**Valérie ROY, Thierry DUCHESNE & Belkacem ABDOUS, *Université Laval*

Parametric models are often used in conditional quantile or distribution estimation. However, when parametric assumptions are not verified, one might prefer to use nonparametric inference. In some applications, it also happens that the response is censored.

In this talk, we are interested in the smooth nonparametric estimation of regression quantiles when the response variable is subject to right-censoring. We consider three estimators: one based on the generalized Kaplan-Meier estimator, one that optimizes a loss function with weights proposed by Stute and one proposed by Bowman and Wright. We investigate the properties of these estimators by simulation and we illustrate the methods with a data analysis example.

Les modèles paramétriques sont très souvent utilisés en estimation de percentiles conditionnels. Or, comme les postulats ne sont pas toujours vérifiés, le choix d'un modèle non paramétrique est parfois préférable. Il peut également arriver que la variable réponse soit censurée.

Dans cet exposé, nous nous intéressons aux méthodes d'estimation non paramétrique lisse de régression des percentiles dans le cas où la variable réponse est censurée à droite. Nous considérons trois estimateurs : un employant l'estimateur de Kaplan-Meier généralisé, un employant une optimisation pondérée par les poids Stute et un proposé par Bowman & Wright. Les propriétés de ces estimateurs sont étudiées par voie de simulation et nous illustrons ces méthodes en les appliquant à un jeu de données.

[MS-111]**Improving the Estimation of Kendall's tau when Censoring Affects Only One of the Variables****Amélioration de l'estimation du tau de Kendall lorsque la censure n'affecte qu'une seule des deux variables**David BEAUDOIN, Thierry DUCHESNE & Christian GENEST, *Université Laval*

This session considers the estimation of Kendall's tau for bivariate data (X, Y) when only Y is subject to right-censoring. The estimators proposed by Brown et al. (1974), Weier & Basu (1980) and Oakes (1982), which are standard in this context, fail to be consistent when tau is different from zero because they only use information from the marginal distributions. Wang & Wells (2000) were the first to propose an estimator which accounts for joint information. Four more are developed here to account for information provided by X . Simulations show that the new estimators perform significantly better than the existing ones mentioned above.

Cette présentation considère l'estimation du tau de Kendall en présence de données bivariées (X, Y) où seulement Y est sujette à la censure à droite. Les estimateurs proposés par Brown et al. (1974), Weier & Basu (1980) et Oakes (1982) ne sont pas convergents lorsque le tau de Kendall est différent de zéro puisqu'ils n'utilisent que l'information contenue dans les lois marginales. Wang & Wells (2000) ont été les premiers à proposer un estimateur qui tient compte de l'information conjointe. Quatre nouveaux estimateurs utilisant l'information fournie par la variable X sont proposés dans cet exposé. Des simulations démontrent que les nouveaux estimateurs sont supérieurs à ceux mentionnés ci-haut.

[MS-112]**A Nonparametric Two Sample Test for Bivariate Right Censored Data****Un test non paramétrique pour données bidimensionnelles censurées à droite**Shojaeddin CHENOURI, *University of Waterloo*

In this talk, we will propose an asymptotically distribution free test for testing the equality of two bivariate time-to-event distributions based on randomly right censored observations. The test statistic is defined by using the notion of data depth. Data depth is a measure of how deep or central a given point is with respect to a given data cloud or distribution, and it introduces a center outward ordering of data.

Dans cet exposé, nous proposons un test asymptotiquement non paramétrique pour tester l'égalité de deux distributions de temps d'événements bidimensionnels basé sur des observations aléatoirement censurées à droite. La statistique de test est définie en utilisant la notion de profondeur des données. La profondeur est une mesure de la centralité d'un point par rapport à un nuage de points ou une distribution, et elle introduit un classement partant du centre des données.

Session 07A**Mardi 30 mai • Tuesday, May 30, 13:15–14:15****SH3345**

Isobel Loutit Invited Address on Business and Industrial Statistics

Allocution Isobel Loutit sur la statistique industrielle et de gestion

[MS-113]**The Challenges of Statistical Leadership****Les défis du leadership en statistique**Geoff VINING, *Virginia Tech*

This talk discusses what we mean by statistical leadership, the several forms of statistical leadership, and the challenges that we face, particularly within industrial statistics. Much of this talk is

Cet exposé traite de ce que nous entendons par « leadership statistique », des différentes formes de leadership statistique, et des défis auxquels nous devons faire face, particulièrement dans le domaine de la statistique industrielle. Cet exposé est

based on several interviews of senior corporate executives. It concludes with several concrete recommendations for statisticians interested in leadership roles.

basé en grande partie sur plusieurs entrevues de dirigeants d'entreprise chevronnés, et se terminera par plusieurs recommandations concrètes à l'intention des statisticiens intéressés par des positions de leadership.

Session 07B **Mardi 30 mai • Tuesday, May 30, 13:15–14:45** **SSC2024**

Spatial/Temporal Modelling for Marine Ecological Systems

Modèles spatiaux/temporels dans les systèmes écologiques marins

[MS-114]

Data Needs and Challenges for Providing Ecosystem-based Advice for the Marine Environment

Besoin en données et défis reliés à l'établissement de conseils concernant l'écosystème de l'environnement marin

Stephen SMITH, *Bedford Institute of Oceanography*

The Department of Fisheries and Oceans (DFO) is in the process of refocusing its scientific program to concentrate on providing ecosystem-based advice on the impact of human activities. This work will involve the collection, analysis and understanding of complex data sets consisting ocean characteristics from satellites and submersible vehicles, seabed characteristics based upon multi-beam sonar and biological sampling not only from traditional sampling gear but also from video and other types of images. These data sets all share the common attributes of consisting of high volumes of temporal and spatial measurements of large numbers of variables with potential for missing data and measurement errors. The Department's change to ecosystem research and the complexity of the data sets being collected has already placed heavy demands upon our research staff for the development of and training on quantitative methods. In this talk I outline some of the challenges that will need to be faced in terms of data management, survey design and analysis for these types of data.

Le ministère des Pêches et des Océans (MPO) est présentement à repenser son programme scientifique afin de se concentrer sur l'offre de services de consultation concernant l'écosystème et les l'impacts sur des activités humaines. Ce travail impliquera la collecte, l'analyse et la compréhension des données complexes sur les océans obtenues à partir des satellites, des véhicules submersibles, des caractéristiques de fonds marins obtenues par un sonar multifaisceaux, et par le prélèvement d'échantillons biologiques, non seulement par l'engin du pêche traditionnel, mais également par le truchement de film vidéo et d'autres types d'images. Ces ensembles de données partagent tous les attributs communs suivants : des volumes importants de données temporelles et spatiales, un grand nombre de variables, et potentiellement de données manquantes et d'erreurs de mesure. Ce changement de cap du Ministère vers des projets de recherche sur les écosystèmes, ainsi que la complexité des données collectées, ont déjà engendrés des investissements importants de la part de notre personnel de recherche pour le développement des méthodes quantitatives et pour la formation. Dans cette présentation, j'élabore sur certains des défis qui devront être relevés en termes de gestion des données, plans d'enquête et plans d'analyse pour ces types de données.

[MS-115]

State Space Approaches for Nonlinear Ecological Models

Approches d'espace d'état pour les modèles écologiques non linéaires

Michael DOWD, *Dalhousie University*

Marine ecological data are characterized by complex dependence structure, as well as sparseness with respect to the time and space scales that must be resolved. Statistical techniques for analysis of-

Des données écologiques marines sont caractérisées par une structure d'une dépendance complexe, aussi bien que par la rareté en ce qui concerne les échelles de temps et d'espace qui doivent être résolues. Les techniques statistiques pour

ten must incorporate population or ecosystem models governed by nonlinear differential equations. In this presentation, I will discuss time dependent estimation of the state and parameters for such stochastic dynamic systems. The general nonlinear and non-Gaussian state space model provides the framework, with sequential Monte Carlo methods yielding sampling based solutions. Problems in marine ecological prediction are discussed.

l'analyse doivent souvent incorporer la des modèles de population ou d'écosystème régis par des équations différentielles non linéaires. Dans cette présentation, je discuterai de l'estimation dépendante du temps de l'état et des paramètres pour de tels systèmes dynamiques stochastiques. Le modèle non linéaire et non gaussien général de l'espace d'état fournit le cadre, avec des méthodes Monte-Carlo séquentielles rapportant des solutions basées sur l'échantillonnage. Des problèmes dans la prévision écologique marine sont discutés.

[MS-116]

Using Robust Hierarchical State Space Switching Models to Elucidate the Behaviour of Seals and Turtles at Sea

Utilisation des modèles à changements de régime hiérarchiques d'espace d'état robustes pour élucider le comportement des phoques et des tortues marines

Ransom MYERS, Ian JONSEN, Greg BREED, Mike JAMES & Joanna FLEMMING, *Dalhousie University*

Even the most casual observer of animals will notice that they have distinct behavioural modes, i.e., they forage, they search for food, they rest, they migrate. These behavioural modes change over time in response to environmental variation. The purpose of this talk is to provide a modeling framework to analyze such behaviour. Specifically, we model data from satellite tracks using a robust hierarchical state space framework to implement switching models. With the recent advent of sophisticated tracking methods large amounts of data that clearly describe such behaviour in time and space are becoming available. We show how state-space switching models can be used to identify behaviours of seals and marine turtles from satellite linked tags that is crucial for the conservation of the species. This approach allows questions to be addressed that were previously difficult for ecologist to formulate using conventional approaches with such data.

Même l'observateur le plus occasionnel des animaux notera qu'ils ont des modes de comportement distincts, c.-à-d. ils vont paître, ils cherchent de la nourriture, ils se reposent, et ils émigrent. Ces modes de comportements changent avec le temps en réponse à des variations dans l'environnement. Le but de cette présentation est de fournir un cadre analytique à l'aide de modèles afin d'analyser ces comportements. Spécifiquement, nous modélisons des données obtenues par satellite qui permettent d'identifier le passage des animaux. Nous utilisons un cadre d'espace d'état hiérarchique robuste pour mettre en application des modèles de commutation. Avec l'arrivée récente de procédés de poursuites sophistiqués, les grandes quantités de données qui décrivent clairement un tel comportement dans le temps et dans l'espace deviennent disponibles. Nous montrons comment les modèles de commutation espace-état peuvent être utilisés pour identifier les comportements des phoques et des tortues marines par des repères liés au satellite, ce qui est crucial pour la conservation de ces espèces. Cette approche permet aux écologistes d'étudier certaines questions qui pouvaient difficilement l'être avec les approches conventionnelles.

Session 07C**Mardi 30 mai • Tuesday, May 30, 13:15–14:45****SSC2050**

Statistical Methods in Computational Genetics

Méthodes statistiques en calculs génétiques

[MS-117]**Inferring Gene Regulation From Integrated Bayesian Analysis of Genomic Sequence and Expression Data
Inférer la régulation des gènes à partir de l'analyse bayésienne intégrée de données de séquençage et d'expression génomique**Mayetri GUPTA, Joseph IBRAHIM & Pingping QU, *University of North Carolina at Chapel Hill*

The behavior of multiple genes under different experimental conditions is generally analyzed by clustering mRNA expression data from microarrays, ignoring genomic sequence features. Computational discovery of regulatory motifs (binding sites) in promoter sequences of clustered genes is typically considered a post-processing step. This stepwise approach often involves the construction of biologically irrelevant clusters, in turn leading to errors in motif discovery. We present a novel Bayesian framework and methodology for simultaneously addressing the problems of discovering clusters of co-regulated genes, and motifs that are involved in their regulation. We also extend this approach to temporal gene expression profiling studies.

Le comportement de plusieurs gènes sous différentes conditions expérimentales est généralement analysé en groupant les données d'expression d'ARNm de micropuces en ignorant les caractéristiques de séquence génomique. La découverte informatique de motifs régulateurs (sites de liaison) dans les séquences promotrices de gènes regroupés est typiquement considérée comme une étude post-traitement. Cette approche pas-à-pas implique souvent la construction de groupes sans justification biologique, ce qui mène à des erreurs en découverte de motifs. Nous présentons un nouveau cadre bayésien et une méthodologie pour simultanément régler le problème de la découverte de groupes de gènes co-régulés et de motifs impliqués dans leur régulation. Nous étendons aussi cette approche à des études de profilage d'expression temporelle des gènes.

[MS-118]**Bayesian Analysis of ChIP-chip Experiments: Data, Questions and Partial Answers
Analyse bayésienne d'expériences ChIP-chip : données, questions et réponses partielles**Raphael GOTTARDO, *University of British Columbia*

The advent of microarray technology has enabled biomedical researchers to monitor changes in the expression levels of thousands of genes. Until recently, however, the mechanisms driving these changes have been harder to study in a similarly high-throughput level. A recent technological innovation, chromatin immuno-precipitation (ChIP) coupled with microarray (chip) analysis, hence the name ChIP-chip, now makes it possible for researchers to identify regions of a given genome that are bound by transcription factors. During this talk, I will present a flexible and robust Bayesian hierarchical model for the analysis ChIP-chip data that incorporates the spatial structure (probe dependence) while borrowing strength from all the probes.

L'avènement des micropuces a permis aux chercheurs biomédicaux d'observer les changements du niveau d'expression de milliers de gènes. Jusqu'à récemment, cependant, il était difficile d'étudier les mécanismes causant ces changements dans un tel contexte de données à très haute fréquence. Une innovation technologique récente, l'immuno-précipitation de chromatine (ChIP) associée à une analyse de micropuces (chip) permet aux chercheurs d'identifier les régions d'un génome liées à des facteurs de transcription. Au cours de cet exposé, je présenterai un modèle hiérarchique bayésien flexible et robuste permettant l'analyse de données ChIP-chip. Ce modèle incorpore la structure spatiale (dépendance des sondes) du problème tout en tirant sa force de toutes les sondes.

[MS-119]**Separating the Wheat from the Chaff: Statistical Methods for False Discovery Control
Méthodes statistiques pour le contrôle des fausses découvertes**

Radu CRAIU, *University of Toronto*, Shelley BULL, *Samuel Lunenfeld Research Institute*, Andrew PATERSON, *Hospital for Sick Children, Toronto*, Lei SUN, *Departments of Public Health and Statistics, Hospital for Sick Children, Toronto*

The multiplicity problem is increasingly relevant and important in many scientific studies. With the diminishing power of using the traditional control of family-wise error rate (FWER), the use of false discovery rate (FDR) has become common in many applications. In this talk, I will investigate the performance of false discovery control after a set of p -values has been stratified based on auxiliary information. Under the fixed rejection region framework, I will demonstrate that the aggregated FDR is a weighted average of the stratified FDR. Under the fixed FDR framework, I will provide the condition under which the expected total number of correct rejections from the stratified FDR approach is greater than that from the aggregated FDR approach. I will illustrate the ideas with analyses of microarray gene-expression data, pedigree data and genome-wide association data.

Le problème des tests multiples est de plus en plus d'actualité et important pour les recherches scientifiques. Dans le contexte où le pouvoir des méthodes traditionnelles est diminué, la méthode de contrôle de la proportion de fausses découvertes (PFD) représente un choix alternatif généralement utilisé. Je vais discuter la performance de la méthode de contrôle de la PFD quand on utilise l'information auxiliaire pour stratifier les valeurs- p . Je prouverai que la PFD agrégée est une moyenne pondérée des PDF obtenues dans chaque strate. Dans le contexte de la PDF fixée, je fournirai la condition pour que le nombre total espéré de vrais rejets pour l'approche PFD stratifiée soit plus grand que pour celle agrégée. Les idées seront illustrées avec des données génétiques.

Session 07D**Mardi 30 mai • Tuesday, May 30, 13:15–14:45****SSC3006****Probability in High Dimensions and its Applications****Probabilités en dimensions élevées et leurs applications****[MS-120]****Local U -statistic Processes and the Estimation of Densities of Functions of Several Sample Variables: Rates of Approximation in Probability****Processus local de la statistique- U et estimation de densités de fonctions de plusieurs variables indépendantes : taux d'approximation en probabilité**

Evarist GINÉ, *University of Connecticut*, David M. MASON, *University of Delaware*

It is shown that the density of a function of more than one sample can be approximated in probability at the parametric rate of the square root of n , when one measures the discrepancy in the sup norm or in any of the L^p norms, p between 1 and infinity. Several moment and exponential inequalities for U -statistics are developed in order to prove the stated results. This work builds on previous work by Frees (1994) and by Schick and Wefelmeyer (2004).

Il est montré que la densité d'une fonction de plus d'un échantillon peut être approximée en probabilité au taux paramétrique de la racine carrée de n , quand l'écart est mesuré à l'aide de la norme sup ou de n'importe laquelle des normes L^p , p entre 1 et l'infini. Plusieurs moments et inégalités exponentielles pour les statistiques- U sont développés afin de prouver les résultats indiqués. Ce travail est basé sur les travaux de Frees (1994) et de Schick et Wefelmeyer (2004).

[MS-121]**LIL for the Local U -Statistic Process with Application to the Strong Consistency of a Class of Kernel Density Estimators****LLI pour le processus local de la statistique- U avec application à la convergence forte d'une classe d'estimateurs de densité par le noyau**David MASON, *University of Delaware*, Evarist GINÉ, *University of Connecticut*

The local U -statistic process is based upon an estimator of the density of a function of several i.i.d. variables proposed by Frees (1994). Laws of the iterated logarithm for the local U -statistic process are discussed and are shown to be directly applicable to the derivation of exact rates of uniform in bandwidth consistency in the sup and in the L^p norms for these density estimators. The probability inequalities and moment bounds for U -processes that were developed to establish these laws will also be described and should be of separate interest.

Le processus local de la statistique- U est basé sur un estimateur de la densité d'une fonction de plusieurs variables i.i.d. proposé par Frees (1994). Des lois du logarithme itéré pour le processus local de la statistique- U sont discutées et il s'avère qu'elles sont directement applicables à la détermination des taux exacts de la convergence uniforme des tailles de fenêtre dans les normes sup et L^p pour ces estimateurs de densité. Les inégalités de probabilité et les limites de moment pour les processus- U qui ont été développés pour établir ces lois seront également décrites.

[MS-122]**Bahadur-Kiefer Theory for Weakly and Strongly Dependent Sequences****Théorie de Bahadur-Kiefer pour séquences faiblement et fortement dépendantes**Rafal KULIK, *Carleton University*

I would like to discuss recent developments concerning Bahadur-Kiefer theory for sample quantiles in case of weakly and strongly dependent linear processes. In the first case, we will obtain the Bahadur-Kiefer representation (with optimal rates) for a large class of linear processes. In the latter case, the optimal rates are established (as far as I know) for the first time. An idea is based on so called reduction principle for empirical processes (see Ho and Hsing, *Ann. Probab.* 1996). In particular, with help of the Bahadur-Kiefer theory I will show that for quantile processes we do not have a reduction principle which mirrors the one for the empirical processes. It provides the negative answer to the question posed in Ho and Hsing paper.

J'aimerais parler de développements récents concernant la théorie Bahadur-Kiefer sur les quantiles échantillonnaires en présence de processus linéaires faiblement ou fortement dépendants. Dans le premier cas, nous obtenons la représentation de Bahadur-Kiefer (avec taux optimaux) pour une grande classe de processus linéaires. Dans le second cas, les taux optimaux sont établis (d'après ce que nous savons) pour la première fois. Une idée est basée sur le principe dit de réduction pour processus empiriques (voir Ho et Hsing, *Ann. Probab.* 1996). En particulier, avec l'aide de la théorie de Bahadur-Kiefer, je montre que pour les processus de quantiles nous n'avons pas de principe de réduction similaire à celui pour les processus empiriques. Ceci nous apporte une réponse négative à la question posée dans l'article de Ho et Hsing.

Session 07E**Mardi 30 mai • Tuesday, May 30, 13:15–14:45****SSC3018****Innovative Teaching in Probability and Statistics****Enseignement innovateur en probabilités et statistique****[MS-123]****Use of Simulation in the Teaching of Non-specialists****Utilisation de simulations dans l'enseignement pour non-spécialistes**Paul COREY, *University of Toronto*

Simulation has a long history with an early famous example being the resampling carried out by William Seally Gossett in his famous 1908 paper introducing Student's t ratio for comparing means. He re-sampled from a collection of prisoner finger lengths. Simulation in applied statistical research is used to verify whether optimal asymptotic properties of estimators exist for small samples. This talk demonstrates the value of using simulation for teaching biostatistics to the non-specialist. One example involves the comparison of small sample properties of Student's t test and the randomization test in the comparison of two independent sample means.

La simulation a une longue histoire et a débuté avec un exemple célèbre, celui du rééchantillonnage réalisé par William Seally Gossett dans son célèbre article (1908) introduisant le ratio t de Student pour comparer des moyennes. Il a ré-échantillonné des longueurs à partir d'une collection de longueurs de doigts de prisonniers. La simulation dans la recherche statistique appliquée est utilisée pour vérifier si les propriétés asymptotiques optimales des estimateurs existent pour de petits échantillons. Cette présentation démontre la valeur d'employer la simulation pour l'enseignement de la biostatistique au non-spécialiste. Un exemple implique la comparaison des propriétés du test t de Student d'un petit échantillon et le test de randomisation dans la comparaison des moyennes de deux échantillons indépendants.

[MS-124]**The Education of Us****L'éducation des nôtres**Jim STALLARD, *University of Calgary*

Recent advances in Statistical Education have spawned a variety of innovative and exciting pedagogical approaches that can be implemented into any undergraduate course in statistics—spanning first-year service courses to fourth-year courses in probability theory. Many of these methods are implemented in a student-centered learning environment, focusing on experiential learning and authentic assessment. In this presentation, I will briefly outline some of these teaching models and methods that have been recently (and successfully) adopted by members of the Statistics Division at the University of Calgary.

Des progrès récents dans le domaine de l'éducation statistique ont permis l'apparition de toute une variété de nouvelles approches pédagogiques pouvant être mises en place dans n'importe quel cours sous-gradué en statistique – allant des cours appliqués de première année aux cours de théorie des probabilités de quatrième année. Plusieurs de ces méthodes ont été mises en place dans un environnement d'apprentissage centré sur l'étudiant, mettant l'accent sur l'apprentissage expérimental et l'évaluation authentique. Dans cette présentation, je survole brièvement certains de ces modèles d'enseignement et des méthodes, utilisés récemment (et avec succès!), adoptés par les membres de la « Statistics Division » de l'Université de Calgary.

[MS-125]**An Integrated Computer Aided Approach to Teaching Introductory Statistics****Une approche intégrée assistée par ordinateur pour enseigner la statistique de base**Paul CABILLO, *Acadia University*

The challenge in teaching introductory statistics and probability courses is to find ways to engage students more fully in the learning process. The hope is that the outcomes of such engagement will result in a deeper understanding of the fundamental ideas of statistics. In this talk I will present some of the evolving strategies that we have employed at Acadia over the last decade to achieve this end. At the core of our approach lies the integration of class presentations and studio labs in teaching. The computer plays an important role in the delivery of both components.

Le défi dans l'enseignement des cours de statistiques de base et de probabilité est de trouver des moyens d'engager les étudiants plus entièrement dans le processus d'apprentissage. Nous espérons que les résultats d'un tel engagement auront comme conséquence une compréhension plus profonde des idées fondamentales de la statistique. Dans cette présentation, je présenterai les différentes stratégies que nous avons utilisées à Acadia pendant la dernière décennie pour atteindre ce but. Au cœur de notre approche se trouve l'intégration des présentations en classe et des laboratoires d'enseignement. L'informatique joue un rôle important dans la livraison des deux composants.

Session 07F**Mardi 30 mai • Tuesday, May 30, 13:15–14:45****UC85**

Joint Modelling of Longitudinal and Survival Data and Applications in Cancer Patients
Modélisation conjointe de données longitudinales et de survie et applications aux patients atteints de cancer

[MS-126]

Multivariate Modeling of Longitudinal Data and Survival Data with Application to Pediatric Cancer Study
Modélisation multivariée de données longitudinales et de survie, avec une application à l'étude du cancer pédiatrique.

Deo Kumar SRIVASTAVA, Shesh N. RAI, Lijun ZHANG, James M. BOYETT & Sean PHIPPS, *St. Jude Children's Research Hospital*

We briefly describe a longitudinal study of pediatric cancer patients involving bone marrow transplants who are followed for quality of life measures (response process) during and immediately after the cancer treatment. In this study longitudinal follow-ups are incomplete on a significant proportion of patients due to many reasons. For example, data related to response process are missing due to scheduling conflict or due to death. Data missing due to death (observation process) cannot be treated as random. We model these two processes jointly. We study the properties of our method through a simulation study and also apply this to the pediatric cancer study.

Nous décrivons brièvement une enquête longitudinale sur le cancer pédiatrique avec transplantation de moelle osseuse où la qualité de vie (processus de réponse) est mesurée pendant et immédiatement après le traitement. Dans cette enquête, le suivi longitudinal est incomplet chez une proportion significative des patients pour différentes raisons. Par exemple, les données liées au processus de réponse sont manquantes pour cause de conflit d'horaire ou de mort du patient. Les données manquantes dues à la mort (processus d'observation) ne peuvent être traitées comme aléatoires. Nous modélisons donc ces deux processus conjointement. Les propriétés de notre méthode sont analysées par le biais de simulations et nous l'appliquons pour une enquête sur le cancer pédiatrique.

[MS-127]**Regression Analysis of Longitudinal Data with Informative Observation and Follow-up Times****Analyse par régression de données longitudinales en présence d'observations informatives et des temps de suivi**Jianguo (Tony) SUN, Xingwei TONG & Xin HE, *University of Missouri*

This talk considers statistical analysis of longitudinal studies. In this case, three processes are usually involved and they are the longitudinal process of interest, a survival process representing the stop or dropout time of study subject, and a point process representing observation time points. Many methods have been proposed for the analysis of these studies when the three processes are independent of each other given covariates or the first two are independent of the third process given covariates. In this talk, we will focus on the situation where all three processes may be related and some analysis approaches are discussed.

Cet exposé considère l'analyse statistique d'enquêtes longitudinales. Dans ce cas, trois processus sont généralement impliqués : le processus longitudinal d'intérêt, un processus de survie représentant le temps d'arrêt ou d'abandon des sujets de l'enquête, et un processus ponctuel représentant les moments où le processus est observé. Plusieurs méthodes ont été proposées pour l'analyse d'enquêtes où, sachant des covariables, les trois processus sont indépendants, ou les deux premiers sont indépendants du troisième. Dans cet exposé, nous nous concentrons sur la situation où les trois processus peuvent être liés et nous proposons quelques approches d'analyse.

[MS-128]**Gap Time Analysis of Longitudinal Data with the Application to Cancer Studies****Analyse des temps d'intervalle pour des données longitudinales avec une application aux études sur le cancer**Do-Hwan PARK, *University of Nevada-Reno*, Liuquan SUN, *Chinese Academy of Sciences*

This talk discusses the analysis of a special type of longitudinal data in which the response variable is the count of occurrences of a recurrent event. Specifically, it is assumed that one is interested in the gap time between the occurrences, but only longitudinal data are available. For the analysis, the use of the additive hazards model is investigated and an estimating equation approach is presented for estimation of regression parameters. The asymptotic and finite sample properties of the proposed parameter estimates are established and the approach is applied to a cancer study.

Cette présentation porte sur l'analyse d'un type particulier de données longitudinales où la variable réponse est le nombre de réalisations d'un événement récurrent. Plus spécifiquement, nous supposons que la variable d'intérêt est le temps entre chacune des réalisations, mais que nous ne disposons que des données longitudinales. Pour l'analyse, nous nous intéressons au modèle de risques additifs et nous présentons une approche par équation estimatrice pour l'estimation des paramètres de régression. Nous établissons les propriétés, asymptotiques et avec échantillons finis, des estimations paramétriques proposées et appliquons cette approche à une étude sur le cancer.

Session 07G Mardi 30 mai • Tuesday, May 30, 13:15–14:45 SSC3014

**Tail Behaviour and Elliptically Contoured Distributions
Comportement des ailes et distributions à contours elliptiques**

[MS-129]

**A Simulation Study of the GMM and the Second-order Least Squares Estimators in Nonlinear Models
Une étude de simulation pour les estimateurs MMG et des moindres carrés du second ordre sous des modèles non linéaires**

Taraneh ABARIN, *University of Manitoba*

The Generalized Method of Moments (GMM) is a widely applied estimation technique in nonlinear regression models and many other setups. Wang and Leblanc (2005) extended the ordinary least squares estimation method by including in the criterion function the distance of the squared response variable to its second conditional moment. This so-called second-order least squares estimator (SLSE) has been shown to be asymptotically more efficient than the ordinary least squares estimator in general. This simulation study compares the SLSE with the GMM estimator for various nonlinear regression models.

La méthode des moments généralisée (MMG) est une technique d'estimation largement appliquée en présence, entre autres, de modèles de régression non linéaire. Wang et Lablanc (2005) ont étendu la méthode d'estimation par moindres carrés ordinaire en incluant, dans la fonction critère, la distance entre le carré de la variable réponse et son deuxième moment conditionnel. Il a été montré que cet estimateur, appelé estimateur des moindres carrés de second ordre (EMCS), est généralement asymptotiquement plus efficace que l'estimateur des moindres carrés ordinaire. Cette étude de simulation compare l'EMCS et l'estimateur MMG sous différents modèles de régression non linéaire.

[MS-130]

**Robust Prediction Error Criterion for Pareto Modeling of Upper Tails
Un critère robuste de prévision pour la modélisation Pareto des ailes supérieures**

Debbie DUPUIS, *HEC Montréal*, Maria-Pia VICTORIA-FESER, *Université de Genève*

Estimation of the Pareto tail index from extreme order statistics is an important problem in many settings: income distributions, finance, and insurance. The upper tail of the distribution, where data are sparse, is typically fitted with a model such as the Pareto model from which probabilities associated with extreme events are deduced. The success of this procedure relies heavily not only on the choice of the estimator for the Pareto tail index but also on the procedure used to determine the number of extreme order statistics used for the estimation. We develop and investigate a robust prediction error criterion.

L'estimation de l'indice de Pareto à partir des statistiques d'ordre extrêmes est un problème important dans plusieurs disciplines : les distributions de revenu, la finance et l'assurance. L'aile supérieure de la distribution, où les données sont rares, est souvent ajustée par un modèle tel que le modèle de Pareto. Les probabilités des événements extrêmes sont ensuite calculées à partir de ce dernier. Le succès de cette procédure dépend beaucoup non seulement du choix de l'estimateur de l'indice, mais aussi de la procédure utilisée pour déterminer le nombre de statistiques d'ordre extrêmes incluses dans l'estimation. Nous développons et examinons un critère robuste de prévision.

[MS-131]**On the Extreme Behaviour of Continuous Bivariate Elliptical Distributions
Le comportement extrême des distributions elliptiques bidimensionnelles continues**Alexandru ASIMIT & Bruce JONES, *University of Western Ontario*

The elliptical class of distribution functions has been widely studied recently with applications in actuarial science, economics and finance. Under some regularity conditions, the continuous bivariate elliptical distributions allow some explicit computations of the indexes of regular variation for the marginal distributions, which are sufficient for extreme behavior analysis. The study is motivated by Hult, H. and Lindskog, F. (2002) "Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions," *Adv. Appl. Prob.*, 2002, 34.

La classe des distributions elliptiques a été récemment étudiée à travers diverses applications en actuariat, en économie et en finance. Sous certaines conditions de régularité, les distributions elliptiques bidimensionnelles continues permettent le calcul explicite des indices de variation régulière pour les distributions marginales, ce qui est suffisant pour une analyse des comportements extrêmes. Cette étude est motivée par Hult, H. et Lindskog, F. (2002) « Multivariate Extremes, Aggregation and Dependence in Elliptical Distributions », *Adv. Appl. Prob.*, 2002, 34.

[MS-132]**Estimation of Parameters of the Simple Multivariate Linear Model with Student-*t* Errors
Estimation des paramètres du modèle linéaire multivarié simple avec des erreurs *t* de Student**Shahjahan KHAN, *University of Southern Queensland*, A K Md Ehsanes SALEH, *Carleton University*

Estimation of the slope and intercept parameters of the multivariate linear model with Student-*t* errors in the presence of uncertain prior information on the value of the slope vector is considered. The unrestricted maximum likelihood, restricted, preliminary test, shrinkage, and positive-rule shrinkage estimators are defined for the parameters of the model. The expressions for the bias, quadratic bias, quadratic risk and mean squared errors (mse) of the estimators are derived. The performance of the estimators under different criteria and conditions are covered. The uniform dominance of the positive-rule shrinkage estimator over the other estimators is noted under the quadratic risk criteria.

L'estimation de la constante et des pentes du modèle linéaire multivarié avec des erreurs *t* de Student en présence d'information *a priori* incertaine sur la valeur du vecteur des pentes est considérée. Les estimateurs de maximum de vraisemblance sans contraintes, avec contraintes, test préliminaire, rétrécissement et rétrécissement à valeur positive sont définis pour les paramètres du modèle. Les expressions du biais, biais quadratique, risque quadratique et le carré moyen des erreurs (cme) des estimateurs sont calculés. L'efficacité des estimateurs sous différents critères et conditions sont couvertes. La dominance uniforme de l'estimateur par rétrécissement à valeur positive sur les autres estimateurs est notée sous le critère de risque quadratique.

[MS-133]**On Small Area Estimation Under a Sub-area Level Model
L'estimation de petites régions à l'aide d'un modèle défini pour les sous-régions**Mahmoud TORABI & J.N.K. RAO, *Carleton University*

An extension of the Fay and Herriot (1979) area level model to sub-area level is introduced. Empirical best linear unbiased prediction (EBLUP) of sub-area level and area level means are obtained by replacing the unknown model variances in the BLUP by suitable estimators, using an iterative method of estimating the model variances based

Une extension du modèle de Fay et Herriot (1979) pour les domaines est présentée pour des sous-domaines. Les meilleures prévisions linéaires empiriques sans biais (MPLESB) des moyennes au niveau des domaines et des sous-domaines sont obtenues en remplaçant les variances inconnues du modèle dans le MPLESB par les estimateurs appropriés, en utilisant une méthode itérative pour estimer

on weighted residual sum of squares. We obtain second order approximations to the mean squared error (MSE) of the model-based small area estimators and then derive MSE estimators unbiased to second order. A small simulation on the performance of the proposed estimators is also provided. Key Words: Best linear unbiased prediction; Fay-Herriot model; Linear mixed models; Mean squared error; Variance components

les variances du modèle basée sur la somme des carrés résiduelles pondérées. Nous obtenons des approximations de second ordre à l'erreur quadratique moyenne (EQM) des estimateurs du modèle basé sur un petit domaine et puis dérivons des estimateurs sans biais de l'EQM au deuxième ordre. Une petite simulation sur la performance des estimateurs proposés est également fournie.

Mots clés : La meilleure prévision linéaire sans biais ; modèle Fay-Herriot ; Modèles mixtes linéaires ; erreur quadratique moyenne ; Composants de la variance

Session 07H

Mardi 30 mai • Tuesday, May 30, 13:15–14:45

UC222

Dependence, Scores and Applications Dépendance, cotes et applications

[MS-134]

Estimating Intraclass Correlation Coefficients from the Class of Elliptically Contoured Distributions Estimation des coefficients de corrélation intraclasse pour la famille des distributions à courbes de niveau elliptiques

Sohail KHAN & S. Ejaz AHMED, *University of Windsor*

In this talk, we showcase inference tools for the intraclass correlation coefficients from several elliptical distributions. The class of elliptical distributions, which includes the more familiar normal distribution, provides flexibility in modeling the thickness of tails or extremes that are becoming commonly important in many fields. Further, besides this flexibility, it preserves several properties of the normal distribution.

We suggest a general result that allows us to combine data from various different sources to improve the efficiency of the estimators within the context of intraclass correlation analysis. Asymptotic properties of the proposed estimators are investigated both analytically and computationally. A Monte Carlo simulation study is also conducted to assess the performance of the suggested estimators for moderate and large samples. An empirical example illustrates the usefulness of this inferential procedure in practice.

Dans cet entretien, nous présentons des outils d'inférence pour les coefficients de corrélation intraclasse de plusieurs distributions elliptiques. La classe des distributions elliptiques, qui inclut la distribution normale, fournit de la flexibilité en modélisant l'épaisseur des ailes ou les valeurs extrêmes qui deviennent de plus en plus importantes dans beaucoup de domaines. En plus de cette flexibilité, elle préserve plusieurs propriétés de la distribution normale.

Nous présentons un résultat général permettant de combiner des données de sources différentes dans le but d'améliorer l'efficacité des estimateurs dans le contexte de l'analyse de corrélation intraclasse. Les propriétés asymptotiques des estimateurs proposés sont étudiées analytiquement et numériquement. Une étude de simulation Monte Carlo est également menée pour évaluer la performance des estimateurs suggérés dans le cas de moyens et grands échantillons. Un exemple illustre l'utilité de la procédure en pratique.

[MS-135]**Estimating Kendall's tau Under Bivariate Censoring
Estimation du tau de Kendall à partir de données censurées**Lajmi LAKHAL-CHAIEB, *University of Waterloo*, Louis-Paul RIVEST & David BEAUDOIN, *Université Laval*

We investigate the nonparametric estimation of Kendall's tau for measuring the association between two random variables under bivariate censoring. The proposed estimator is a modification of the estimator proposed by Oakes (1982), using an Horvitz-Thompson-type correction for the orderable pairs. Under bivariate censoring, a pair is orderable if one can establish the concordance/discordance status of the original pair using the observed data. Our estimator is shown to be consistent and asymptotically normally distributed. A simulation study shows that the proposed estimator performs well compared with competing alternatives. The various methods are illustrated with two real data sets.

On considère l'estimation non paramétrique du tau de Kendall mesurant l'association entre deux variables aléatoires censurées. L'estimateur proposé est une modification de celui proposé par Oakes (1982) en utilisant une correction du type Horvitz-Thompson pour les paires orderables. Pour les données variées censurées, une paire de points est dite orderable si on peut déterminer son statut de concordance/discordance à partir des données observées. On montre que notre estimateur est convergent et possède une distribution asymptotique normale. Les simulations montrent que les performances de l'estimateur proposé sont meilleures que celles des estimateurs existants, sous des conditions de censure simplifiées. On illustre les différentes méthodes avec l'analyse de deux jeux de données réels.

[MS-136]**Analysis of Quality of Life Data (QoL) from Cancer Clinical Trials based on Item Response Theory
Analyse de mesures de qualité de vie (QdV) d'essais cliniques basées sur la théorie de réponse aux items**Miao WANG, Dongsheng TU & Deb FELDMAN-STEWART, *Queen's University*

QoL is an important endpoint in the comparison of cancer treatments in clinical trials and also useful for helping patients determine their treatment preferences. Traditionally, the classical test theory (CTT) is adopted to calculate scores for a given QoL scale but CTT scores are sample dependent and, therefore, the results derived from these scores may not be generalizable across different patient populations. In this presentation, models based on the item response theory are applied to analyze the QoL data from two clinical trials conducted by the NCIC Clinical Trials Group and the results are compared from those based on CTT.

La QdV est un important critère d'évaluation dans la comparaison de traitements pour le cancer en essais cliniques. Elle est aussi utile pour aider les patients à déterminer leurs préférences relatives aux traitements. Traditionnellement, la théorie classique des tests (TCT) est adoptée pour calculer les scores d'une échelle donnée de QdV, mais les scores TCT sont dépendants de l'échantillon, rendant difficile la généralisation des résultats obtenus à partir de ces scores à différentes populations de patients. Dans cette présentation, des modèles basés sur la théorie de réponse dans les items sont appliqués à l'analyse de données de QdV provenant de deux essais cliniques conduits par le NCIC Clinical Trials Group et les résultats sont comparés à ceux basés sur la TCT.

[MS-137]**Self Reported Risk for Drug Related Problems in Ambulatory Elderly Patients Referred for Family Practice Pharmacist Assessment****Le risque auto rapporté des problèmes reliés à la médication de patients âgés référés pour une évaluation d'un pharmacien de famille praticien**

Mahbubul HAQ, *Centre for the Evaluation of Medicines (Hamilton, Ontario)*, Lisa DOLOVICH, *Centre for the Evaluation of Medicines, Leslie Dan Faculty of Pharmacy (University of Toronto), Pharmacy Department (St. Joseph's Healthcare, Hamilton, Ontario), Department of Clinical Epidemiology and Biostatistics (McMaster University, Hamilton, Ontario), Department of Family Medicine (McMaster University, Hamilton, Ontario)*, Lisa MCCARTHY, *Centre for the Evaluation of Medicines (Hamilton, Ontario)*, Janusz KACZOROWSKI, *Centre for the Evaluation of Medicines (Hamilton, Ontario), Department of Clinical Epidemiology and Biostatistics (McMaster University, Hamilton, Ontario), Department of Family Medicine (McMaster University, Hamilton, Ontario)*, Lehana THABANE, *Centre for the Evaluation of Medicines (Hamilton, Ontario), Department of Clinical Epidemiology and Biostatistics (McMaster University, Hamilton, Ontario)*

Mechanisms to help clinicians and patients identify medication related risk can reduce drug related problems. The objectives of this study were to determine the risk for drug related problems based on patient self-report using a structured tool and to identify patient characteristics associated with this risk in elderly ambulatory patients. In this presentation we discuss a cross sectional analysis that was conducted with consecutive patients referred for pharmaceutical care assessments who self-completed the previously validated 10-item Medication Risk Questionnaire (MRQ, scored 0–10). Multivariable regression analyses used MRQ score as the dependent variable and age, gender, SF-12, number of medications and medical conditions as independent variables.

Les mécanismes pour aider des cliniciens et des patients à identifier le risque relié à la médication peuvent réduire les problèmes reliés aux médicaments. Les objectifs de cette étude étaient de déterminer le risque pour des problèmes reliés au médicament basé sur l'auto-évaluation du patient à l'aide d'un outil structuré et d'identifier des caractéristiques du patient liées à ce risque pour des patients âgés. Dans cette présentation nous discutons une analyse transversale qui a été conduite avec les patients référés pour des évaluations de soins pharmaceutiques, patients qui ont complété eux-mêmes le Questionnaire de Risques de Médicaments composé de 10 items (QRM). Des analyses de régression multiple, avec les scores QRM utilisés comme variable dépendante et l'âge, le genre, le SF-12, le nombre de médicaments et les conditions médicales en tant que variables indépendantes.

[MS-138]**Sleepiness in Adolescence: A Survey of 3200 Canadian Students****Somnolence chez les adolescents : une enquête sur 3200 étudiants canadiens**

Mi YAN, *University of Waterloo*, Edward GIBSON, Lehana THABANE & A.C. Peter POWLES, *McMaster University*, Susan O'BRIEN, *Ancaster High School*, Danielle Sirriani MOLNAR, *Brock University*

Sleepiness is a serious problem in adolescents, which can negatively impact teenage students' behaviours, impede their cognitions and reduce their activities; however, neither education nor health care systems in Canada are aware of this issue. More research is needed to further understand the prevalence of the problem in order to study proper interventions. A total of 3200 high school students in Ontario were surveyed to measure sleepiness using Epworth Sleepiness Score (ESS). We used generalized estimating equations to identify the factors associated with sleepiness (ESS > 10).

La somnolence est un problème sérieux pendant l'adolescence, qui peut affecter négativement les étudiants adolescents, leurs comportements, réduire leur capacité d'apprendre et réduire leurs activités ; cependant, ni le système d'éducation ni le système de santé au Canada ne se rendent compte de ce problème. Davantage de recherche est nécessaire afin de mieux comprendre l'importance du problème dans le but d'étudier des interventions appropriées. Un total de 3 200 étudiants du secondaire en Ontario a été examiné pour mesurer la somnolence en utilisant les Scores de somnolence d'Epworth (SSE). Nous avons utilisé des équations d'estimation généralisée pour identifier les facteurs liés

This presentation will report the association between sleepiness and both academic performance and extra-curricular activities.

à la somnolence (SSE > 10). Cette présentation rapportera l'association entre la somnolence et le rendement scolaire et les activités parascolaires.

Session 07I

Mardi 30 mai • Tuesday, May 30, 13:15–14:45

SSC3010

Business Surveys

Enquêtes sur les entreprises

[MS-139]

The Use of The Goods and Services Tax by the Monthly Restaurants, Caterers and Taverns Survey

L'utilisation de la taxe sur les produits et services pour l'Enquête mensuelle sur les restaurants, traiteurs et tavernes

Zachary PRITCHARD & Charles TARDIF, *Statistics Canada*

The Monthly Restaurants, Caterers and Taverns Survey (MRCTS) collects sales data from a sample of restaurants, caterers and Taverns in Canada. The sales estimates, an input to the System of National Accounts, are used by various agencies for planning purposes. Since May 2004, data from selected companies have been modelled using Goods and Services Tax (GST) data provided by the Canada Revenue Agency (CRA). We will discuss the usage of GST data by the MRCTS and its effects on the survey in terms of data quality and response burden.

Key words: Data quality, response burden, tax data.

L'Enquête mensuelle sur les restaurants, traiteurs et tavernes (EMRTT) recueille les ventes d'un échantillon représentatif de restaurants, traiteurs et tavernes au Canada. Les estimations de l'EMRTT sont partie intégrante des comptes nationaux et sont utilisées par divers membres de l'industrie pour fins de planification. Depuis mai 2004, les données de la Taxe sur les produits et services (TPS) provenant de l'Agence de revenu du Canada (ARC) sont utilisées pour modéliser les ventes de compagnies sélectionnées. Nous discuterons de l'utilisation de la TPS par l'EMRTT et de son impact sur l'enquête au niveau de la qualité des données et du fardeau de réponse.

Mots clés : Les données TPS, le fardeau de réponse, la qualité des données.

[MS-140]

Comparison between Two Data Sources: Tax Data and Respondent Data of the Unified Enterprise Survey of Statistics Canada

Comparaisons entre deux sources de données : données fiscales et données des répondants de l'enquête unifiée auprès des entreprises de Statistique Canada

Éric PELLETIER, Wisner JOCELYN & Lenka MACH, *Statistique Canada*

Since 2002, several Statistics Canada enterprise surveys are making more use of tax data in lieu of survey data to reduce response burden and costs. This is feasible because of an agreement with Canada Revenue Agency. In 2005, the Data Integration Project was launched in order to determine the optimal methods for surveys' use of tax data. The first step of the project was to compare the concepts and definitions used for tax data and the data collected via a survey. In this presentation, the Unified Enterprise Survey (UES) data will be used to compare the two data sources.

Depuis 2002, plusieurs enquêtes entreprises à Statistique Canada utilisent davantage de données fiscales comme données de remplacement dans le but de réduire le fardeau de réponse et le coût des enquêtes. Ceci a été rendu possible grâce à une entente avec l'Agence du revenu du Canada. En 2005, le Projet d'intégration des données a été développé dans le but d'étudier comment les données fiscales peuvent être utilisées de façon optimale dans les enquêtes. Une des premières étapes de ce projet a été de comparer les concepts et les définitions utilisées pour les données fiscales et les données recueillies auprès des entreprises. Les données provenant de l'Enquête unifiée auprès des entreprises (EUE) seront utili-

sées lors de la présentation pour comparer les deux sources de données.

[MS-141]

**Review of the Collection Strategy of Statistics Canada's Unified Enterprise Survey
Revue des stratégies de collecte de l'enquête unifiée auprès des entreprises de Statistique Canada**

Rose-Carline EVRA & Sylvie DEBLOIS, *Statistique Canada*

The Unified Enterprise Survey (UES) is a major annual survey conducted by Statistics Canada since 1998. Data collection for the survey is challenging due, in part, to time and cost constraints. In order to determine which enterprises should be assigned a high priority follow-up in cases of non-response, a score function was introduced in 2002. This function is currently being reviewed in order to make it more efficient. The presentation will focus on the re-evaluation of the score function and on follow-up process for enterprises with specific characteristics.

Key words: collection, enterprise, score function

L'Enquête unifiée auprès des entreprises (EUE) est une enquête annuelle majeure de Statistique Canada, en cours depuis 1998. La collecte de données pour cette enquête pose certains défis dus, entre autres, aux contraintes de temps et de coûts. Dans le but de pouvoir déterminer à quelles entreprises un suivi prioritaire devrait être accordé dans le cas de non-réponse, une fonction de score fut introduite en 2002. Celle-ci est présentement revue afin de la rendre plus efficace. La présentation portera sur la réévaluation de cette fonction et sur le processus de suivi d'entreprises ayant des caractéristiques particulières.

Mots-clés : collecte, entreprises, fonction de score

[MS-142]

Dealing with the Problem of Combined Reports at the Sampling Design Stage for the Workplace and Employee Survey

Le traitement du problème des rapports combinés à l'étape du plan de sondage à l'Enquête sur le milieu de travail et les employés

Cynthia BOCCI & Jean-François BEAUMONT, *Statistique Canada*

The Workplace and Employee Survey is a longitudinal survey that collects information on employers (at the location level) and on employees working in the selected locations. Enterprises having more than one location, referred to as multi-location employers, are often unfortunately not able to report for a chosen location and report combined information which is often simply thrown out or imputed. These combined reports have the effect of introducing measurement errors and thus reducing data quality. We discuss various options to select fewer multi-locations in the sample and investigate how much increase of variance is acceptable in order to reduce the number of combined reports and increase data quality.

L'Enquête sur le milieu de travail et les employés est une enquête longitudinale qui recueille des informations auprès des employeurs (au niveau de l'emplacement) et des employés travaillant dans les emplacements sélectionnés. Les entreprises qui ont plus d'un emplacement, ce qu'on appelle des employeurs multi-emplacements, ont malheureusement souvent de la difficulté à rapporter les données pour un emplacement choisi et donc ils fournissent un rapport combiné qui est souvent simplement jeté ou imputé. Ces rapports combinés ont pour effet d'introduire des erreurs de mesure et ainsi de réduire la qualité des données. On discute plusieurs options pour sélectionner moins de multi-emplacements dans l'échantillon. On étudie également à quel point une augmentation de la variance est acceptable afin de réduire le nombre de rapports combinés et d'accroître la qualité des données.

[MS-143]**Recent Developments in the Redesign of Statistics Canada's Business Register
Développements récents dans le remaniement du registre des entreprises de Statistique Canada**Stuart PURSEY, Paul HUNSBERGER, Zeeshan ZEESHAN & Stuart PURSEY, *Statistics Canada*

The Business Register of Statistics Canada was developed in the mid-1980s. Its role is to provide a frame for more than 90 recurring business surveys. Statistics Canada has undertaken to completely revise and update the infrastructure and approaches of the current Business Register. Methodological issues in survey design are discussed with emphasis on the proposed methods used to derive sampling units from the structure of a business. The new approaches in survey design are compared and contrasted with the current approaches.

Le Registre des entreprises de Statistique Canada a été développé au milieu des années 80. Il constitue une base de sondage utilisée de façon récurrente par plus de 90 enquêtes auprès des entreprises. Statistique Canada a entrepris une révision complète du Registre des entreprises actuel tout en mettant à jour son infrastructure et ses façons de faire. On discutera de questions méthodologiques reliées à la conception d'enquêtes en mettant l'accent sur de nouvelles méthodes proposées pour dériver les unités d'échantillonnage à partir de la structure des entreprises. Ces nouvelles approches seront comparées et on fera ressortir les contrastes entre celles-ci et les approches actuelles.

Session 08A Mardi 30 mai • Tuesday, May 30, 15:15–16:45 SSC2050**Data Mining Complex Structures
Structures complexes de forage de données****[MS-144]****Haplotype Inference Using a Hidden Markov Model with Efficient Markov Chain Sampling
Inférence sur le haplotype en utilisant un modèle de Markov caché avec un échantillonnage efficace par chaînes de Markov**Shuying SUN, Celia GREENWOOD & Radford NEAL, *University of Toronto*

This talk will introduce a new population-based haplotype inference method which uses a Hidden Markov Model (HMM) for the source of the ancestral haplotype segments. This HMM-based method includes the genotyping error rate, the mutation rate and recombination rate in the model and infers them using Markov Chain Monte Carlo (MCMC) methods. It has been tested using data sets from Daly et al. 2001 and the HAPMAP Project. It produces the results that are quite close to family-based reconstructions, and comparable to the PHASE program of Stephens et al. 2001.

Nous présentons une nouvelle méthode d'inférence sur le haplotype pour une population, méthode qui emploie un modèle de Markov caché (MMC) pour la source des segments héréditaires de haplotype. Cette méthode inclut le taux d'erreur dans le génotypage, le taux de mutation et le taux de recombinaison dans le modèle et les infère en employant les méthodes de chaînes de Markov Monte-Carlo (CMCMC). Elle a été testée en utilisant des jeux de données de Daly et al. (2001) et le projet de HAPMAP. Elle produit des résultats qui sont près des reconstructions basées sur le lien de famille, et comparable au programme PHASE de Stephens et al. (2001).

[MS-145]**Visualizing Binary Pairwise Similarity Data
Visualisation de données binaires de similarité deux-à-deux**Sam ROWEIS, *University of Toronto*, Amir GLOBERSON, *MIT*

Binary pairwise similarity data is available in many domains where quantifying the similarity/difference between objects is extremely difficult

Les données binaires de similarité deux-à-deux sont disponibles dans plusieurs domaines où quantifier la similarité/différence entre objets est extrêmement difficile ou im-

or impossible. Nonetheless, it is often desirable to obtain insight into such data by associating each object (record) with a point in some abstract feature space – for visualization purposes this space is often two or three dimensional. We present an algorithm for visualizing such similarity data, which delivers an embedding of each object such that similar objects are always closer in the embedding space than dissimilar ones. Many such mappings may exist, and our method selects amongst them the one in which the mean distance between embedded points is as large as possible. This has the effect of stretching the mapping and, interestingly, favoring embeddings with low effective dimensionality.

We study both the parametric and non-parametric variants of the problem, showing that they both result in convex Semidefinite Programs (SDP). In the non-parametric version, input points may be mapped to any point in space, whereas the parametric version assumes that the mapping is given by some function (e.g. a linear or kernel mapping) of the input. This allows us to generalize the embedding to points not used in the training procedure.

[MS-146]

Robustness in Data Mining

Robustesse en forage de données

David BANKS, *Duke University*

Robustness is a long-standing issue in data mining, and people have developed a number of strategies for addressing it. This talk describes an approach in which structures in the data are iteratively found and removed. The ideas are illustrated in the context of regression, multidimensional scaling, and cluster analysis. Some of the benefits of the method include better interpretability, high-breakdown points, and some relaxation of the statistical need for random sampling.

possible. Néanmoins, il est souvent souhaitable d'obtenir une compréhension de telles données en associant chaque objet à un point dans un espace de caractéristiques abstrait – à des fins de visualisation, cet espace est souvent à deux ou trois dimensions. Nous présentons un algorithme pour visualiser de telles données de similarité, fournissant une fonction de chaque objet de façon à ce que des objets similaires soient toujours plus rapprochés dans l'espace d'arrivée que des objets dissimilaires. Plusieurs transformations de ce type existent, et notre méthode sélectionne celle où la distance moyenne entre les points transformés est la plus grande possible. Ceci a pour effet d'étirer la transformation et favorise les fonctions à faible dimension effective.

Nous étudions les variantes paramétrique et non paramétrique du problème, et montrons que nous obtenons toujours des programmes semi-définis convexes. Dans le cas non paramétrique, des points en entrée peuvent être transformés en n'importe quel point dans l'espace, alors que le cas paramétrique présuppose que la transformation est donnée par une certaine fonction (par exemple, une transformation linéaire ou à noyau) des entrées. Ceci nous permet de généraliser la transformation à des points non utilisés lors de la procédure d'entraînement.

La robustesse est une question importante en exploration de données et différentes stratégies furent élaborées pour y répondre. Cet exposé décrit une approche où les structures présentes dans les données sont trouvées et retirées de façon itérative. Ces idées sont illustrées dans le contexte de la régression, de l'échelonnement multidimensionnel et de l'analyse de groupement. Les avantages de la méthode sont, entre autres, une meilleure interprétabilité, des points de rupture élevés et un certain relâchement des besoins statistiques d'échantillonnage aléatoire.

Session 08B**Mardi 30 mai • Tuesday, May 30, 15:15–16:45****SH3345****Genomic Assessment Implications of Biologic Heterogeneity****Implications sur l'évaluation génomique de la variabilité biologique****[MS-147]****Implications of Tumor Heterogeneity on Assessment of Differential Expression using Genomic Data****Conséquences de l'hétérogénéité des tumeurs pour déterminer les différences d'expression en utilisant des données génomiques**Debashis GHOSH, *University of Michigan*

In expression studies, differential expression between cancer and non-cancerous samples is routinely studied. A major issue unaddressed in these settings is the fact that the tumor itself is quite heterogeneous. We discuss the nature of this problem and causes as well as some statistical methods for incorporating tumor heterogeneity into the analysis.

Dans des expérimentations d'expression génétique, la différence d'expression entre des échantillons cancéreux et non-cancéreux est couramment utilisée. Un problème majeur qui n'est pas pris en considération dans ce contexte est la grande hétérogénéité de la tumeur. Nous discutons de la nature de ce problème et de ses causes et décrivons quelques méthodes statistiques permettant d'incorporer l'hétérogénéité à l'analyse.

[MS-148]**Design and Analysis of Genetical Genomics Studies Involving Complex Traits****Conception et analyse des études des génomes génétiques impliquant des traits complexes**Guilherme ROSA, *Michigan State University*

Modern techniques being used to unravel the genetic mechanisms governing variation in complex traits combine information on phenotypic traits, molecular markers, and gene expression. These methods are generally referred to as genetical genomics approaches. Such studies, however, require careful planning and refined quantitative methods for their success. In this presentation we will discuss some issues related to the design and analysis of genetical genomics studies. Current methods available in the literature will be reviewed, and some directions for future research will be discussed.

Les techniques modernes étant employées pour démêler les mécanismes génétiques régissant la variation des traits complexes combinent l'information sur des traits phénotypiques, des marqueurs moléculaires, et l'expression de gène. Ces méthodes sont généralement mentionnées comme des approches de génomiques génétiques. Cependant, de telles études exigent une planification soignée et des méthodes quantitatives raffinées pour leur succès. Dans cette présentation, nous discuterons quelques questions liées à la conception et à l'analyse des études génétiques de génomique. Des méthodes courantes disponibles dans la littérature seront passées en revue, et quelques directions pour la future recherche seront discutées.

[MS-149]**Clinical Implications of Tumor Heterogeneity Demonstrated Pathologically and with Image Analysis****Implications cliniques de l'hétérogénéité des tumeurs démontrées pathologiquement et par analyse d'image**Judy-Anne CHAPMAN, *NCIC Clinical Trials Group*, David AXELROD, *Rutgers University*, Naomi MILLER, *University of Toronto*, Jin QIAN, *University of Waterloo*, William CHRISTENS-BARRY, *Equipoise Imaging LLC*, Yan YUAN, *University of Waterloo*, Yuejiao FU, *York University*

The current trend in assessment of malignancies is towards evaluation of ever smaller samples, containing smaller numbers of tumour cells. The va-

La tendance actuelle dans l'évaluation des tumeurs malignes se dirige vers l'évaluation d'échantillons toujours plus petits, contenant moins de cellules tumorales. La va-

lidity of small sample assessment may be limited by tumour heterogeneity. A cohort study of non-invasive breast DCIS was designed to assess heterogeneity in nuclear grade and its apparent effects on clinical outcome. The results demonstrate that assessment of very small samples of tumour containing only small numbers of cells may be insufficient to characterize the tumour for diagnostic or prognostic purposes.

lidité des évaluations basées sur de petits échantillons peut être limitée par l'hétérogénéité des tumeurs. Une étude de cohortes de DCIS du sein non invasive a été établie afin d'évaluer l'hétérogénéité au niveau nucléaire et ses effets apparents sur les résultats cliniques. Les résultats démontrent que l'évaluation de très petits échantillons de tumeurs contenant très peu de cellules peut être insuffisante pour caractériser la tumeur à des fins de diagnostic ou de pronostic.

Session 08C

Mardi 30 mai • Tuesday, May 30, 15:15–16:45

SSC2024

Probability and Finance

Probabilités et finance

[MS-150]

Stochastic Methods and Statistical Technique for Equity-linked Life Insurance

Méthodes stochastiques et technique statistique pour l'assurance-vie liée à des indices boursiers

Alexander MELNIKOV & Yulia ROMANIUK, *University of Alberta*

Equity-linked life insurance contracts are agreements in which the payoff depends on the performance of one or more risky assets as well as on the client's mortality. The presence of both financial and insurance risk elements complicates the pricing of these agreements and calls for appropriate hedging methods that would optimize the risk management strategies available to the firm selling the contracts. We will focus on quantile and efficient hedging, explain the idea behind the methods and the related statistical concepts. Theoretical results will be illustrated with numerical examples based on real-world financial and mortality data.

Les contrats d'assurance-vie reliés aux valeurs boursières sont des accords dans lesquels le remboursement dépend d'un ou de plusieurs capitaux de risque aussi bien que de la mortalité du client. La présence des éléments financiers et d'assurance à risque complique l'évaluation de ces accords et demande des méthodes appropriées de couverture qui optimiseraient les stratégies de gestion des risques disponibles à la société vendant les contrats. Nous nous concentrerons sur le quantile et les couvertures efficaces, nous expliquerons l'idée derrière les méthodes et les concepts statistiques reliés à ces méthodes. Des résultats théoriques seront illustrés avec des exemples numériques basés sur des données financières réelles et des données de survie.

[MS-151]

Extensions of Levy-Khintchine Formula and Beurling-Deny Formula in Semi-Dirichlet Forms Setting

Extensions de la formule de Levy-Khintchine et de la formule Beurling-Deny dans le contexte des formes semi-Dirichlet

Wei SUN, *Concordia University*, Ze-Chun HU, *Nanjing University, China*, Zhi-Ming MA, *Institute of Applied Mathematics, Chinese Academy of Sciences*

The Levy-Khintchine formula characterizes the infinitesimal generators of Levy processes. For more general Markov processes, the formula that comes closest to such a characterization is the Beurling-Deny formula for symmetric Dirichlet forms. In this talk, we extend these celebrated structure results to include general right continuous strong Markov processes, which are supposed to be as-

La formule de Levy-Khintchine caractérise les générateurs infinitésimaux des processus de Levy. Pour des processus plus généraux de Markov, la formule la plus proche d'une telle caractérisation est la formule de Beurling-Deny pour les formes symétriques de Dirichlet. Dans cette présentation, nous extrapolons ces célèbres résultats de structure pour inclure les processus continus de Markov forts et continus à droite, qui sont censés être associés aux formes semi-

sociated with semi-Dirichlet forms. Moreover, we extend LeJan’s formula in non-symmetric Dirichlet forms setting, which characterizes the diffusion part of a Dirichlet form.

Dirichlet. D’ailleurs, nous extrapolons la formule de LeJan aux formes dissymétriques de Dirichlet, qui caractérise la partie de diffusion d’une forme Dirichlet.

[MS-152]

Optimal Dividends Under a Ruin Probability Constraint
Dividendes optimaux sous une contrainte de probabilité de ruine

Steve DREKIC, *University of Waterloo*

We consider a surplus process modified by the payment of dividends when the insurer’s surplus exceeds a certain threshold. We use a probabilistic argument to obtain general expressions for the expected present value of dividend payments. We then consider the question of maximizing the expected present value of dividend payments subject to a constraint on the insurer’s ruin probability.

Nous considérons un processus de surplus modifié par le paiement des dividendes quand le surplus de l’assureur excède un certain seuil. Nous employons un argument probabiliste afin d’obtenir des expressions générales pour la valeur présente espérée des paiements de dividendes. Nous considérons ensuite le problème de la maximisation de la valeur présente espérée des paiements de dividendes sous une contrainte liée à la probabilité de ruine de l’assureur.

Session 08D

Mardi 30 mai • Tuesday, May 30, 15:15–16:45

SSC3018

The History of Statistics and Probability

L’histoire de la statistique et des probabilités

[MS-153]

André-Michel Guerry and the Rise of Moral Statistics
André-Michel Guerry et la naissance de la statistique morale

Michael FRIENDLY, *York University*

André-Michel Guerry’s (1833) *Essai sur la statistique morale de la France* was one of the foundation studies of modern social science. Guerry assembled data on crimes, suicides, literacy, and other “moral statistics” to shed light on social issues in France, particularly those used to understand factors affecting crime. He used tables and maps to analyze these issues and was likely the first to try to use such data to answer questions about the relations among multiple social variables.

In this talk I trace the development of the study of moral statistics in Guerry’s work. I also illustrate some recent attempts to ask what light modern methods of statistical graphics and thematic cartography can shed on Guerry’s questions and data.

L’« *Essai sur la statistique morale de la France* » d’André-Michel Guerry (1833) fut l’une des études fondatrices de la science sociale moderne. Guerry avait rassemblé des données sur les crimes, les suicides, l’alphabétisation et autres « statistiques morales » afin de faire la lumière sur les grandes questions sociales en France, particulièrement celles utilisées pour comprendre les facteurs liés au crime. Il utilisa des tableaux et des cartes pour analyser ces questions et fut vraisemblablement le premier à utiliser de telles données pour tenter de répondre à des questions sur les relations entre de multiples variables sociales.

Dans cet exposé je suis la trace du développement de l’étude des statistiques morales dans le travail de Guerry. J’illustre également quelques tentatives récentes de déterminer comment les méthodes modernes de graphiques statistiques et de cartographie thématique peuvent éclairer les questions et données de Guerry.

[MS-154]**Abraham De Moivre: Genius in Exile
Abraham de Moivre : génie en exil**Christian GENEST, *Université Laval*, David R. BELLHOUSE, *University of Western Ontario*

November 27, 2004, marked the 250th anniversary of the death of Abraham De Moivre, whose famous approximation to the binomial distribution inspired the Central Limit Theorem. De Moivre, a French Huguenot refugee living in England, was one of the great pioneers of classical probability theory. He also made seminal contributions in analytic geometry, complex analysis and the theory of annuities. A summary of his life and works will be presented within the social and scientific context of the time. This talk is based on joint work with David R. Bellhouse of the University of Western Ontario.

Le 27 novembre 2004 a marqué le 250^e anniversaire du décès d'Abraham De Moivre, dont la célèbre approximation de la loi binomiale a inspiré le théorème limite central. Calviniste français exilé en Angleterre, de Moivre fut l'un des grands pionniers de la théorie classique des probabilités. Il contribua en outre de façon significative à la géométrie analytique, à l'analyse complexe et au calcul des rentes. Un résumé de sa vie et de son œuvre sera présenté dans le contexte social et scientifique de l'époque. Cet exposé s'appuie sur des travaux réalisés en collaboration avec David R. Bellhouse, de l'Université de l'Ontario occidental.

[MS-155]**Some Comments About Issai Schur (1875–1941) and The Early History of Schur Complements
Quelques commentaires sur Issai Schur (1875-1941) et le développement des compléments de Schur**George STYAN, *McGill University*, Simo PUNTANEN, *University of Tampere*

After some biographical remarks about Issai Schur (1875–1941), we note that he attended the Nicolai Gymnasium in Libau (Kurland) and suggest that Nicolai is used here in honour of Saint Nicholas of Bari, Bishop of Myra, who is widely associated with Christmas and after whom Santa Claus is named. The term “Schur complement” was introduced in 1968 by Emilie Virginia Haynsworth (1916–1985) in view of a 1917 lemma by Schur. We present some early results involving Schur complements by A.C. Aitken, Tadeusz Banachiewicz, W.J. Duncan, Louis Guttman, and Haynsworth, with some statistical applications. Several pictures illustrate this talk.

Suivant quelques remarques biographiques au sujet d'Issai Schur (1875-1941), nous notons qu'il fréquenta le Gymnasium Nicolai à Libau (Courlande) et suggérons que le nom « Nicolai » soit utilisé ici en honneur à Saint Nicolas de Bari, évêque de Myre, largement associé avec Noël. Le terme « complément de Schur » fut introduit en 1968 par Emilie Virginia Haynsworth (1916-1985) d'après un lemme de 1917 de Schur. Nous présentons quelques résultats préliminaires impliquant les compléments de Schur par A.C. Aitken, Tadeusz Banachiewicz, W.J. Duncan, Louis Guttman et Haynsworth, avec quelques applications statistiques. Plusieurs images illustrent cet exposé.

Session 08E**Mardi 30 mai • Tuesday, May 30, 15:15–16:45****UC85****Estimating Function Methodology
Méthodologie d'estimation de fonctions****[MS-156]****Information Matrix: From Fisher to Godambe
Matrices d'information : de Fisher à Godambe**Peter SONG & Qian ZHOU, *University of Waterloo*

In the theory of estimating functions, Godambe information matrix is the essence in the description of asymptotic normality for an estimator. In this talk I will present some empirical studies to demonstrate merits of Godambe information matrix over the Fisher information matrix in the maximum likelihood (ML) inference. Although Godambe information matrix reduces to Fisher Information matrix in the ML setting, its unique sandwich-type construction gives rise to some desirable robustness against influential data points and model misspecification. Results from several simulation studies will be given in the presentation.

En théorie de l'estimation fonctionnelle, la matrice d'information de Godambe est au cœur de la description de la normalité asymptotique d'un estimateur. Lors de cet exposé, je présenterai des études empiriques afin de démontrer les mérites de la matrice d'information de Godambe par rapport à la matrice d'information de Fisher dans le cadre de l'inférence basée sur le maximum de vraisemblance (MV). Bien que la matrice d'information de Godambe se réduise à la matrice d'information de Fisher dans le contexte MV, sa construction unique de type sandwich permet une meilleure robustesse par rapport aux données influentes et à une mauvaise spécification du modèle. Les résultats de plusieurs études de simulation seront exposés lors de la présentation.

[MS-157]**Estimating Functions and Analysis of Missing Data
Fonctions d'estimation et analyse des données manquantes**Bent JØRGENSEN, *University of Southern Denmark*

We first review basic results for optimality of estimating functions, such as Godambe and Crowder optimality of the score and quasi-score functions, respectively. We also review some basic notions of parameter orthogonality, and introduce the idea of nuisance parameter insensitivity (Jørgensen and Knudsen, *Scand. J. Statist.* 2004). These results are then applied to the analysis of missing data in the context of estimating functions, where we discuss inverse probability weighted GEE, and the use of the BLUP for imputation of missing data.

Nous passons en revue des résultats de base concernant l'optimalité de fonctions estimatrices, tels l'optimalité de Godambe et de Crowder pour les fonctions de cotes et les quasi-fonctions de cotes, respectivement. Nous révisons également quelques notions d'orthogonalité des paramètres, et introduisons l'idée d'insensibilité aux paramètres de nuisance (Jørgensen et Knudsen, *Scand. J. Statist.* 2004). Ces résultats sont alors appliqués à l'analyse de données manquantes dans le contexte des fonctions d'estimation, où nous discuterons d'EEG pondérées par probabilités inverses et de l'utilisation du « BLUP » pour l'imputation de données manquantes.

[MS-158]**Estimating Functions for Correlated Recurrent and Terminal Events
Fonctions d'estimation pour événements récurrents et terminaux corrélés**Jack KALBFLEISCH, Yining YE & Douglas SCHAUBEL, *University of Michigan*

In clinical and observational studies, recurrent event data (e.g., hospitalizations) with a terminal event (e.g., death) are often encountered. In many

Dans les études cliniques et observationnelles, des événements récurrents (par exemple, des hospitalisations) et terminaux (par exemple, la mort) sont fréquemment rencon-

instances, the terminal and recurrent event rates are strongly correlated which we model with a shared random effect for the recurrent event rate and the terminal event hazard function. The analysis is based on estimating functions that arise from average event rates for the terminal and recurrent event processes, and a Poisson likelihood is used to estimate the distribution of the random effect. The approach is compared with others in the literature and illustrated on some hospitalization data from a Canada/US study on peritoneal dialysis.

trés. Dans plusieurs cas, les événements récurrents et terminaux sont fortement dépendants, ce que nous modélisons à l'aide d'un effet aléatoire partagé entre la fréquence de l'évènement récurrent et la fonction de défaillance associée à l'évènement terminal. Cette analyse est basée sur des fonctions estimatrices obtenues à partir des fréquences moyennes des événements pour les processus d'évènements récurrents et terminaux, et une vraisemblance de Poisson est utilisée pour estimer la distribution de l'effet aléatoire. Cette approche est comparée à d'autres tirées de la littérature et est illustrée à l'aide de données d'hospitalisation d'une étude canado-américaine de la dialyse péritonéale.

Session 08F **Mardi 30 mai • Tuesday, May 30, 15:15–16:45** **SSC3006**

Computational Methods in Statistics
Méthodes computationnelles en statistique

[MS-159]

An Introduction to Linear Grouping Analysis
Introduction à l'analyse linéaire de groupes

Justin HARRINGTON, *University of British Columbia*

The popular “*k*-means” clustering algorithm assumes that the data form distinct spheres around *k* centers. In this talk, we will introduce a new clustering method, first presented in Van Aelst et al (2005), called Linear Grouping Analysis (LGA) that clusters the data around hyperplanes. Some important challenges that arise are: the misclassification effect produced by hyperplanes that intersect other data groups; and issues of scaling and the “curse of dimensionality.” In this talk we will outline a parametric solution to the former issue, and describe a parallel computing implementation that partly addresses the latter.

Le populaire algorithme des centres de groupes suppose que les données forment des sphères centrées en *k* centres. Dans cet exposé, nous présentons une nouvelle méthode de classification proposée par Van Aelst et al (2005), l'analyse linéaire de groupes, qui classe les données selon leur proximité avec des hyperplans. Quelques défis de taille se présentent à nous : les mauvaises classifications causées par l'intersection d'un plan avec un groupe de données ainsi que les problèmes liés à la complexité du temps de calcul lorsque le nombre de dimensions augmente. Nous proposons une solution paramétrique au problème de classifications erronées et décrivons une programmation parallèle de l'algorithme qui permet de compenser pour la lenteur des calculs en plusieurs dimensions.

[MS-160]

The EM Algorithm for Finite Mixtures with Subsampling
Un algorithme EM pour des mélanges finis de distributions avec sous-échantillonnage

Peter MACDONALD, *McMaster University*, Ruochu GAO, *University of Western Ontario*

A finite mixture distribution can be formulated as a missing data problem. There is an observable *y* and a finite set of states. We are interested in the conditional distribution of *y* given its associated state. Complete data would be a sample from the joint distribution of *y* and state. The states are

Un mélange fini de distributions peut être formulée comme un problème de données manquantes. Il y a un *y* observable et un ensemble fini d'états. Nous sommes intéressés par la distribution conditionnelle de *y* étant donné son état associé. Les données complètes seraient un échantillon provenant de la distribution à deux variables de *y* et d'état. Les états sont

unobservable, however, so we have only the mixture distribution, that is, the marginal distribution of y . The EM algorithm provides a natural analysis, estimating an indicator vector for state on the E-step, then the conditional distribution of y given state on the M-step, repeating these two steps to convergence. We have extended the EM algorithm to include samples from the conditional distribution of state given y . This is of great practical importance in fisheries applications, where the length of a fish is easy to measure but its age is not. A mixed sample of fish length can be supplemented by aging a few fish from selected length classes, and this additional information leads to much better estimates than can be obtained from the mixed sample alone.

inobservables, cependant, ainsi nous avons seulement la distribution de mélange, c.-à-d., la distribution marginale de y . L'algorithme EM fournit une analyse convenable, estimant un vecteur d'indicateur pour l'état sur l'étape-E, puis la distribution conditionnelle de l'état donné y sur l'étape-M, répétant ces deux étapes jusqu'à la convergence. Nous avons développé l'algorithme EM pour inclure des échantillons provenant de la distribution conditionnelle de y étant donné l'état. C'est de grande importance pratique dans une application de pêche, où il est facile de mesurer la longueur d'un poisson mais son âge n'est pas facilement connu. Un échantillon mélangé de longueur de poissons peut être augmenté en déterminant l'âge de quelques poissons provenant des classes de longueur choisies, et cette information additionnelle mène à de bien meilleures estimations que ce qui peut être obtenu à partir de l'échantillon mélangé seul.

[MS-161]

Context Sensitive Regression Models with Binary Predictors Modèles de régression contextuels avec prédicteurs binaires

Johan VAN HOREBEEK, *CIMAT*, Hugh CHIPMAN, *Acadia University*

We introduce a class of linear regression models in which the effect of a predictor on the response can depend on the level of another predictor. Similar to Graphical Models, by representing such relations as a directed graph, a compact and interpretable modelling tool is developed. Both frequentist and Bayesian methods for identifying these models are discussed.

Nous présentons une classe de modèles de régression linéaire pour lesquels l'effet du prédicteur sur la réponse peut dépendre du niveau d'un autre prédicteur. Similairement aux modèles graphiques, en représentant ces relations comme un graphe orienté, un outil de modélisation compact et interprétable est développé. Des méthodes bayésiennes et fréquentistes permettant d'identifier ces modèles seront exposées.

[MS-162]

Estimating Prediction Error in Linear Regression by Cross Validation Estimer l'erreur de prévision en régression linéaire par validation croisée

Hui SHEN & William J. WELCH, *University of British Columbia*

Cross validation is commonly used to estimate prediction error. In this paper, we study the properties of the cross validation estimator of mean squared prediction error (CVMSE) in linear regression. We compare the prediction performance for different numbers of folds, ν , in cross validation. We also propose a correction method to reduce the bias. Our correction can reduce the bias of CVMSE significantly when the number of parameters in the model is not small relative to the number of observations in the data set. Moreover, we find that the bias correction can help in model selection.

La validation croisée est fréquemment utilisée pour estimer l'erreur de prévision. Dans cet article, nous étudions les propriétés de l'estimateur par validation croisée de l'erreur quadratique moyenne de prévision (VCEQM) en régression linéaire. Nous comparons les performances de prévision pour différents nombres de groupes, ν , en validation croisée. Nous proposons aussi une méthode de correction pour réduire le biais. Notre correction peut réduire le biais du VCEQM de façon significative lorsque le nombre de paramètres du modèle n'est pas petit relativement au nombre d'observations. De plus, nous trouvons que la correction pour le biais peut aider à la sélection du modèle.

[MS-163]**Bivariate Exponential Distributions and its Application on Matched Pair Data****Distributions exponentielles bidimensionnelles et leur application dans un contexte de données appariées**Yu ZHANG, *University of Regina*

In Clinical Research, we have data which came from the paired organs of patients. For example, eyes, ears, kidneys. The lifetime of these two organs are of course not independent, simply because the organs are exposed to the same environments. Also, we sometimes couldn't get the complete data in medical research. For example, some patients died before the organ died, or some patients were withdrawn from the experiment. Under such cases, the data are called censored data. We have many types of censored data, they are random type I censoring, fixed type I censoring, type II censoring and hybrid censoring.

Lors d'une recherche clinique, nous avons obtenu des données traitant d'organes appariés, et donc provenant d'un même patient. Par exemple, les yeux, les oreilles, les reins. Les durées de vie de ces deux organes sont naturellement non indépendantes, simplement parce que les deux organes sont exposés au même environnement. De plus, il a été quelque fois impossible d'obtenir des données complètes. Par exemple, quelques patients sont morts avant que l'organe considéré soit mort et quelques patients ont été tout simplement retirés de l'étude. Dans de tels cas, les données sont dites censurées. Nous avons été confrontés à plusieurs types de données censurées : censure aléatoire et fixe de type I, censure de type II et censure hybride.

[MS-164]**Robust Designs for Logistic Model with Overdispersion****Plans robustes pour un modèle logistique avec hyperdispersion**Adeniyi ADEWALE & Douglas P. WIENS, *University of Alberta*

We have developed criteria that generate robust designs that insure against possible misspecifications in logistic models. We construct designs that protect against possible fixed effects misspecification and overdispersion. We addressed the problems of misspecification of the systematic effects by some form of averaging of the loss—an approach somewhat akin to a Bayesian treatment. The problem of extra-binomial variation is addressed by incorporating the variance function prescribed by a superior model similar to a logistic mixed model. Examples for various misspecification scenarios are considered, including an example on the toxicity of ethylene oxide to grain beetles.

Nous avons développé des critères permettant de construire des plans d'expérience robustes quant aux spécifications erronées possibles dans les modèles logistiques. Nous construisons des plans offrant une protection contre une spécification erronée des effets fixes et d'hyperdispersion. Nous considérons les problèmes liés à la spécification erronée des effets systématiques par une certaine forme d'étalement de la perte – une approche quelque peu apparentée à un traitement bayésien. Le problème de la variation extra-binomiale est abordé en incorporant la fonction de variance prescrite par un modèle supérieur semblable à un modèle logistique mixte. Des exemples traitant différents scénarios de spécification erronée, y compris un exemple sur la toxicité de l'oxyde d'éthylène pour les cucujides, sont présentés.

Session 08G**Mardi 30 mai • Tuesday, May 30, 15:15–16:45****SSC3010**

Methodology for Complex Data Structures

Méthodologie pour des données à structure complexe

[MS-165]**Joint Analysis of Longitudinal and Survival Data with Application to Cystic Fibrosis****Analyse conjointe de données longitudinales et de survie, avec une application à la fibrose kystique**Eshetu ATENAFU, *University of Toronto*

It is common in many longitudinal studies that both repeated measures and time-to-event data are collected, along with treatment information and patient characteristics at baseline. Since the longitudinal process and survival times are often related, simultaneous modeling of these outcomes has received considerable attention. In this talk, we introduce the transition model approach, which is based on stratification of the longitudinal measures and survival endpoints through transition states. We will discuss parameter estimation procedures, illustrate the method through application to a cystic fibrosis study, and draw comparisons to other methods such as selection models, pattern-mixture models and shared-parameter models.

Dans plusieurs études longitudinales, des mesures répétées et les temps d'événements sont recueillis, en plus d'information sur le traitement et des caractéristiques des patients. Puisque le processus longitudinal et le temps de survie sont souvent corrélés, la modélisation simultanée de ces résultats a fait l'objet de beaucoup d'attention. Dans cet exposé, nous introduisons l'approche par modèle de transition, basée sur la stratification des mesures longitudinales et des temps d'arrêt de survie à travers des états de transition. Nous discutons de procédures d'estimation paramétrique, illustrons cette méthode par l'application à une étude sur la fibrose kystique et établissons des comparaisons avec d'autres méthodes telles les modèles de sélection, les modèles de mélanges de formes et les modèles à paramètres partagés.

[MS-166]**Flexible Parametric Models for Truncated and Interval-censored Multi-state Data****Modèles paramétriques flexibles pour des données tronquées, à intervalle censuré et multi-état**David TOLUSSO & Richard COOK, *University of Waterloo*

The evolution of many disease processes can be conveniently characterized through multistate models. Data available are often subject to both interval-censoring and truncation. In this talk, the appeal of flexible parametric models will be highlighted for analysis of this kind of data. Attention will be directed at robust methods of estimation. An application to data from a trial of patients with metastatic cancer will be given for illustration.

L'évolution de plusieurs processus d'une maladie peut être pratiquement caractérisée par des modèles multiples. Les données disponibles sont sujettes souvent à être censurées par intervalle et par troncation. Dans cet entretien, la pertinence des modèles paramétriques flexibles sera mise en évidence pour l'analyse de ce genre de données. On portera une attention particulière aux méthodes robustes d'estimation. Une application aux données d'un essai sur des patients avec le cancer métastatique sera donnée comme illustration.

[MS-167]**Fitting HIV Mixed-effects Dynamic Models to Clinical Trials****Ajustement de modèles dynamiques à effets mixtes pour des essais cliniques pour le VIH**Jiguo CAO, James O. RAMSAY & Giles HOOKER, *McGill University*

HIV dynamic models describe the rate of population change of uninfected cells, infected cells and virus as a function of their populations and inter-

Les modèles dynamiques de VIH décrivent le taux de changement de population des cellules non infectées, des cellules infectées et du virus en fonction de leurs populations et des

actions. They have significantly contributed to our understanding of HIV infection and the development of antiviral drug therapy. The talk will introduce a set of differential equations to model long-term HIV dynamical system. A generalized profiling method will be applied to fit HIV dynamic models to clinical data, with a random patient effect taken into account. The idea is to reduce the dimensionality of parameter space by treating the nuisance parameters as functions of structural parameters.

interactions. Ils ont significativement contribué à notre compréhension de l'infection par le VIH et du développement du médicament antiviral thérapeutique. Dans cette présentation nous discuterons d'un ensemble d'équations différentielles pour modéliser le système dynamique à long terme du VIH. Une méthode de profilage généralisée sera appliquée aux modèles dynamiques de VIH pour l'ajuster aux données cliniques, avec un effet « patient » aléatoire pris en considération. L'idée est de réduire la dimensionnalité de l'espace de paramètres en traitant les paramètres nuisibles comme fonctions des paramètres structuraux.

[MS-168]

Dynamic Variable Selection in SNP Classification Based on an APEX Microarray Platform

Sélection de variable dynamique en classification de SNP basée sur une plate-forme de puces APEX

Mohua PODDER, William J. WELCH & Ruben H. ZAMAR, *University of British Columbia*

We develop a classification algorithm based on simple Linear Discriminant Analysis (LDA) using dynamic variable selection for an APEX microarray platform of SNP genotyping. Using a set of 32 Coriell sample as a training data set, we develop a supervised learning algorithm which is robust and fully automated and also achieve an accuracy level of 98 in a new data set of 270 SIRS sample based on the same APEX microarray platform for SNP genotyping. This model-based algorithm captures the redundancy in the system considering all the underlying information for a SNP automatically down weighting the 'bad data'.

Nous développons un algorithme de classification basé sur l'analyse discriminante linéaire (ADL) utilisant la sélection de variables dynamiques pour une plate-forme de biopuces APEX de génotypage SNP. En utilisant un ensemble de 32 échantillons de Coriell comme jeu de données d'entraînement, nous développons un algorithme d'apprentissage supervisé robuste et complètement automatisé et obtenons un niveau de précision de 98 dans un nouveau jeu de données de 270 SIRS obtenu à partir de la même plate-forme de biopuces pour génotypage SNP. Cet algorithme basé sur des modèles utilise la redondance dans le système en considérant toute l'information sous-jacente à un SNP, accordant ainsi moins d'importance aux « mauvaises données ».

[MS-169]

Sequential Learning of SVMs for Target Identification in Drug Discovery

Apprentissage séquentielle de SVMs pour l'identification de cible dans la découverte de médicament

Qian ZHOU, *University of Waterloo*, S. Stanley YOUNG, *National Institute of Statistical Sciences*, Mu ZHU, *University of Waterloo*

An important problem in drug discovery is to build a model (e.g., an SVM) to screen a large database and identify a few active chemical compounds. To build an SVM, we must test some compounds first in order to construct a training set. Instead of randomly selecting these initial compounds, we use active learning to construct the initial training set sequentially. We investigate several different sequential strategies. Using a randomized block design, we conduct a number of experiments to statistically assess these different strategies. Our main

Un problème important dans la découverte de médicament est d'établir un modèle (par exemple, un SVM) pour examiner une grande base de données et pour identifier quelques composés chimiques actifs. Pour construire un SVM, nous devons examiner quelques composés d'abord afin de construire un ensemble d'entraînement. Au lieu de choisir aléatoirement ces composés initiaux, nous employons l'apprentissage actif pour construire l'ensemble d'entraînement initial séquentiellement. Nous étudions plusieurs différentes stratégies séquentielles. En utilisant un plan d'expérience en blocs aléatoires, nous mènerons un

conclusion is: an SVM using a sequentially constructed training set is significantly more effective.

certain nombre d'expériences pour évaluer statistiquement ces différentes stratégies. Notre conclusion principale est la suivante : un SVM employant un ensemble d'entraînement séquentiellement construit est significativement plus efficace

[MS-170]

Classification of Longitudinal Data with Missing Values

Classification de données longitudinales avec des valeurs manquantes

Julie HORROCKS, Azim BHAMANI, Hoki PONG, Marianne VAN DEN HEUVEL, FRANCIS TEKPEY & B. Anne CROY,
University of Guelph

In this talk, we compare methods for classifying individuals into groups, based on longitudinal data with missing values. We consider methods that require imputation of missing values and methods that model the longitudinal data under the assumption that data are missing at random (MAR). As an application, women are classified according to their response to in-vitro fertilization, using longitudinal data on adhesion of certain blood lymphocytes in the days prior to ovulation.

Dans cette présentation, nous comparons des méthodes pour classifier des individus en groupes, basées sur des données longitudinales avec des valeurs manquantes. Nous considérons les méthodes qui exigent l'imputation des valeurs manquantes et les méthodes qui modélisent les données longitudinales sous l'hypothèse que les données sont manquantes d'une façon aléatoire. Comme application, des femmes sont classifiées selon leur réponse à la fertilisation in vitro, en utilisant des données longitudinales sur l'adhérence de certains lymphocytes de sang les jours avant l'ovulation.

Session 08H

Mardi 30 mai • Tuesday, May 30, 15:15–16:45

SSC3014

Survey Analysis and Operations

Analyse d'enquêtes et opérations

[MS-171]

Methodological Innovations for the Canadian Labour Force Survey

Innovations méthodologiques dans l'Enquête canadienne sur la population active

Harold MANTEL & Edward J. CHEN, *Statistics Canada*

Statistics Canada's Labour Force Survey (LFS) is a monthly rotating panel household survey that produces official measurements of employment characteristics. The LFS sample and frame are also used by many other Statistics Canada household surveys. In this talk we will review methodological innovations introduced in a recent redesign of the survey including: use of more detailed geographical information in the clustering and stratification; introduction of special strata to target aboriginals and immigrants; sample reduction in high cost areas; use of the Address Register for listing dwellings in some clusters; and the use of telephone for first contact whenever possible.

L'Enquête sur la population active (EPA) de Statistique Canada est une enquête mensuelle sur les ménages qui utilise une rotation de panels et qui produit des mesures officielles des caractéristiques d'emploi. L'échantillon et la base de sondage de l'EPA sont aussi utilisés par plusieurs autres enquêtes sur les ménages de Statistique Canada. Dans cet exposé, nous passons en revue les innovations méthodologiques introduites lors du remaniement récent de l'enquête incluant : l'utilisation d'information géographique plus détaillée lors de la segmentation et de la stratification ; l'introduction de strates particulières pour cibler les autochtones et les immigrants ; la réduction de l'échantillon dans les zones à coûts élevés ; l'utilisation du registre des adresses pour lister les foyers dans certains segments ; et l'utilisation du téléphone pour le premier contact lorsque possible.

[MS-172]**A Methodological Approach to Controlling Inter-monitor Variability in Computer Assisted Telephone Interviewing Operations****Une approche méthodologique du contrôle de la variabilité entre les surveillants dans les opérations d'entrevues téléphoniques assistées par ordinateur**Hansheng XIE & Walter MUDRYK, *Statistics Canada*

Quality Control (QC) monitoring has been implemented in Computer Assisted Telephone Interviewing operations of most social surveys at Statistics Canada. The QC monitoring has been demonstrated to be instrumental in lowering interviewer errors. However, the consistency and reliability of the monitors who evaluate and assess the interviewers remains a general concern of this technique. In this paper, we propose two methods to objectively evaluate inter-monitor variability over time and determine monitors' consistency in assessing interviewers.

Statistique Canada a exécuté la surveillance du Contrôle de la Qualité dans les opérations de l'Interview Téléphonique Assistée par Ordinateur de la plupart des enquêtes sociales. Cet exercice a contribué à diminuer les erreurs commises par les interviewers. Cependant, cette technique cause une certaine inquiétude quant à la cohérence et la fiabilité des surveillants qui évaluent les interviewers. Dans notre article, nous proposons deux méthodes pour évaluer objectivement la variabilité entre les surveillants au cours du temps ainsi que pour déterminer objectivement la cohérence des surveillants lors de leur évaluation des interviewers.

[MS-173]**What to do when Survey Weights Weigh you Down****Quoi faire quand les pondérations (les poids) d'une enquête pèsent sur vous**Pat NEWCOMBE, *Statistics Canada*

Researchers using Statistics Canada survey data sometimes are frustrated when the application of the survey weights produces model results notably different from those obtained without the use of weights. Although Statistics Canada strongly encourages researchers to use the weights to account for the sampling design, it is a useful diagnostic technique to compare the weighted and unweighted results. However, when there is a difference, it is not always obvious how to "get to the bottom of it". This paper discusses approaches to this problem, including some fairly simple graphical methods which may be carried out in SAS.

Des chercheurs, employant des données d'enquête de Statistique Canada, sont parfois frustrés quand l'utilisation des poids d'enquête dans les modèles mène à des résultats différents de ceux obtenus sans l'utilisation des poids. Bien que Statistique Canada encourage fortement les chercheurs à employer les poids afin de tenir compte du plan de sondage, la comparaison des résultats pondérés et non pondérés demeure un outil diagnostique intéressant. Cependant, quand il y a des différences, il n'est pas toujours facile de les justifier. Cet article discute de différentes approches face à ce problème et de l'utilisation de méthodes graphiques assez simples qui peuvent être effectuées dans SAS.

Session 09A Mercredi 31 mai • Wednesday, May 31, 08:30–09:30 SSC2050

Pierre Robillard Award Address

Allocution du récipiendaire du prix Pierre-Robillard

Session 09B Mercredi 31 mai • Wednesday, May 31, 08:30–10:00 SSC2024**Applied Probability
Probabilités appliquées****[MS-174]****Three Months Journeying of a Hawaiian Monk Seal
Le voyage de trois mois d'un phoque moine hawaïen**David BRILLINGER, *University of California at Berkeley*, Brent STEWART, *Hubbs-Sea World Research Institute*, Charles LITTNAN, *Pacific Islands Fisheries Science Center, NOAA Fisheries*

Hawaiian monk seals are endemic to the Hawaiian Islands and are the most endangered species of marine mammal that lives entirely within the jurisdiction of the United States. Data have been collected recently on their foraging habitats, movements, and behaviours throughout the main Hawaiian Islands. Our work is directed to exploring a data set from a relatively shallow offshore submerged bank (Penguin Banks) southwest of Molokai. The work ends by fitting a stochastic differential equation that mimics some aspects of the behaviour of one female seal by working with location data.

Le phoque moine d'Hawaii est endémique aux îles hawaïennes et est l'espèce la plus menacée d'extinction parmi les mammifères marins vivants entièrement sur le territoire des États-Unis. Nous avons récemment obtenu des données sur son habitat de broutage, ses mouvements et son comportement à travers les îles hawaïennes principales. Notre travail porte sur l'exploration de données d'un banc submergé (Penguin Banks) peu profond au sud-ouest de Molokai. Notre travail se termine par l'ajustement d'une équation différentielle stochastique imitant certains aspects du comportement d'une femelle phoque en travaillant avec des données de positionnement géographique.

[MS-175]**Return Period Models for High Sea Levels
Modèles de la période de retour pour les niveaux élevés de la mer**Bruce SMITH, *Dalhousie University*

Flood risk is a concern of individuals living in coastal areas. One sometimes hears statements like "the 20 year flood level here is 12 metres", which is interpreted to mean that the expected number of sea level exceedances above 12 metres over a 20 year time period is one. But how does one determine that 12 metres is the appropriate 20 year level? Some probabilistic results for point processes will be reviewed, followed by an examination of some empirical features of sea level records, which will be incorporated to develop models of the intensity function of sea level exceedance processes.

Le risque d'inondation est une préoccupation pour les individus vivant dans les secteurs côtiers. Nous entendons parfois des énoncés tels que « le niveau d'inondation aux 20 ans à cet endroit est de 12 mètres », ce qui veut dire que, sur une période de 20 ans, le niveau dépassera celui de la mer de 12 mètres en moyenne une fois. Mais comment détermine-t-on que 12 mètres est le niveau approprié pour 20 ans ? Quelques résultats probabilistes relatifs aux processus ponctuels seront passés en revue ainsi que quelques propriétés empiriques du niveau de la mer. Le tout sera utilisé pour développer des modèles pour la fonction d'intensité des processus de dépassement du niveau de la mer.

[MS-176]**Finite Memory Polya Urns
Urnes de Polya à mémoire finie**Glen TAKAHARA, *Queen's University*, Chris LEITH, *Royal Bank of Canada*

The well-known Polya urn model and its generalizations are characterized by infinite memory in the process of ball draws: balls placed in the urn

Le célèbre modèle d'urne de Polya et ses généralisations sont caractérisées par une mémoire infinie dans le processus du tirage des boules : les boules placées dans l'urne ont

have a diminishing but never-ending influence on future ball draws. However, infinite memory is superfluous or inappropriate for certain modeling situations. In this talk we describe a class of finite memory Polya urn models and compare them to their infinite memory counterparts in terms of applications and asymptotics. Applications include communication channels with memory and online evolutionary algorithms. The asymptotics are under various limiting regimes one obtains as time and/or memory get large.

une influence décroissante mais sans fin sur les futurs tirages. Cependant, la mémoire infinie est superflue ou inappropriée pour certaines modélisations. Dans cet exposé, nous décrivons une classe de modèles d'urnes de Polya à mémoire finie et les comparons à leurs équivalents infinis en termes d'applications et de propriétés asymptotiques. Les applications étudiées incluent les canaux de communication avec mémoire et les algorithmes évolutifs en ligne. Les propriétés asymptotiques sont obtenues sous différents régimes de limites lorsque le temps et/ou la mémoire devient grande.

Session 09C Mercredi 31 mai • Wednesday, May 31, 08:30–10:00 SSC2036

Analysis of Longitudinal Complex Survey Data Using Marginal Modelling and GEE: Theory and Practice

Analyse de données longitudinales d'enquêtes complexes utilisant la modélisation marginale et les GEE : théorie et pratique

[MS-177]

Goodness-of-fit Tests of Marginal Logistic Models for Longitudinal Complex Survey Data

Tests d'adéquation de modèles logistiques marginaux pour des données longitudinales d'enquêtes complexes

Georgia ROBERTS, *Statistics Canada*, J.N.K. RAO, *Carleton University & Statistics Canada*, Qunshu REN, *Carleton University*

Due to survey design features, longitudinal survey data exhibit complexities not present in traditional longitudinal data. Ren and Roberts (2005) studied marginal logistic regression models for binary responses using design-weighted Generalized Estimating Equations to estimate model parameters, odds ratios for association between pairs of binary responses of the same subject, and a one-step estimating function bootstrap for variance estimation. In this paper, we develop formal tests to assess the goodness-of-fit of the model mean specification, accounting for the survey design, including quasi-score, adjusted chi-squared, normal-approximation-based and simulation-based methods. Proposed tests are applied to longitudinal data from Statistics Canada's NPHS.

Étant donné les caractéristiques du plan de sondages, les données de sondage longitudinales présentent des complexités absentes pour des données longitudinales traditionnelles. Ren et Roberts (2005) ont étudié les modèles de régression logistique marginaux pour réponses binaires en utilisant des équations estimatrices généralisées pondérées par le plan d'expérience pour estimer les paramètres du modèle, les rapports de cotes pour l'association entre les paires de réponses binaires pour le même sujet et un bootstrap à une étape de la fonction estimatrice pour estimer la variance. Dans cet article, nous développons des tests formels pour évaluer l'ajustement de la spécification de la moyenne du modèle en prenant en considération le plan de sondage, incluant les méthodes de quasi-cote, de khi-deux ajusté, d'approximation normale et de simulation. Les tests proposés sont appliqués à des données longitudinales de l'ENSP de Statistique Canada.

[MS-178]**Efficiency Comparisons of GEE versus IEE for Longitudinal Complex Survey Data with Ordinal Responses
Comparaisons de l'efficacité de la méthode des EEG par rapport à la méthode des EEI pour des données
longitudinales ordinales d'enquêtes complexes**Abdelnasser SAIDI, *Statistics Canada*, Diane STUKEL, *UNESCO*, Susana RUBIN-BLEUER, *Statistics Canada*

We use the proportional odds marginal model and the “independence estimating equations” (IEE) and the “generalized estimating equations” (GEE) approaches for modeling longitudinal ordinal complex survey data. The IEE assumes that observations on a subject over time are independent, whereas the GEE assumes a working correlation structure over time. Our objective is to determine “best practices” under this set-up: we compare through simulation the efficiency of GEE versus IEE under models having differing longitudinal correlation structures of both the responses and the covariates. To measure the efficiency, we consider two approaches to obtaining variance estimates: the Liang-Zeger Taylor (“sandwich”) variance and the One-Step Jackknife.

Nous avons recours au modèle marginal à cotes proportionnelles de même qu'aux méthodes des « équations d'estimation indépendantes » (EEI) et des « équations d'estimation généralisées » (EEG) pour modéliser les données longitudinales ordinales d'enquêtes complexes. Dans le cadre de la méthode des EEI, on suppose que les observations faites sur un sujet au fil du temps sont indépendantes, tandis qu'on suppose une structure de corrélation au fil du temps dans le cadre de la méthode EEG. Notre objectif est d'établir des « pratiques exemplaires » dans ce contexte : nous exécutons une simulation comparant les deux méthodes sous l'angle de l'efficacité de modèles présentant des structures de corrélation longitudinale de réponses et de covariables différentes. Pour mesurer l'efficacité, nous examinons deux méthodes de calcul d'estimation de la variance, soit les méthodes Liang-Zeger Taylor (« sandwich ») et la méthode Jackknife à une étape.

[MS-179]**Modelling of Longitudinal Polytomous Outcomes From Complex Survey Data: Using GEE and Missing Data Analysis****Modélisation des résultats polychotomiques longitudinaux pour des données complexes d'une enquête : utilisation du EEG et de l'analyse de données manquantes**Punam PAHWA, *Dept. of Community Health and Epidemiology*, Chandima KARUNANAYAKE & Helen McDUFFIE, *Inst. of Agricultural Rural and Environmental Health*

The statistical methods were used to fit ordered logistic regression models for polytomous outcomes which account for the complexity of a multi-stage survey-design and also incorporate the subject-level clustering (as a result of repeated observations on a subject). The first set of models were based on the design-weighted generalized estimating equations (GEE) approach assuming missing observations were missing completely at random. The second set of models were fitted using GEE to investigate how missing data patterns influence the results obtained from the previous analysis. The analyses were applied to mental health longitudinal data available from Statistics Canada's NPHS.

Les méthodes statistiques ont été utilisées pour adapter les modèles de régressions logistiques ordonnés pour les résultats polychotomiques qui expliquent la complexité d'un plan d'enquête à plusieurs degrés et incorporent également le regroupement sujet-niveau (en raison des observations répétées sur un sujet). Le premier ensemble de modèles a été basé sur l'approche d'équations d'estimation généralisées (EEG) à plan pondéré en supposant que les observations manquantes étaient absentes complètement au hasard. Le deuxième ensemble de modèles a été adapté en utilisant EEG pour étudier comment la structure des données manquantes influence les résultats obtenus à partir de l'analyse précédente. Les analyses ont été appliquées aux données longitudinales de santé mentale disponibles à partir de ENSP de Statistique Canada.

[MS-180]**Design-based Versus Model-based Methods: A Comparative Study Using Longitudinal Survey Data.****Les méthodes basées sur le plan de sondage versus les méthodes basées sur un modèle : étude comparative en utilisant des données longitudinales d'une enquête.**Sunita GHOSH & Punam PAHWA, *University of Saskatchewan*, Geert MOLENBERGHS, *Hasselt University*

Survey data analysis using complex sampling designs ought to account for clustering, stratification and unequal probability of selection. Design-based and model-based methods are two commonly used routes taken to account for such survey designs. Several cross-sectional survey design studies have shown that these two approaches provide similar results when the model fits the data well. The present paper aims at comparing these two approaches using NPHS dataset. A marginal modeling approach proposed by Rao (1998) (design based) and Generalized Estimating Equation method, proposed by Liang and Zeger (1986) (model-based), will be used. Results obtained from these methods will be compared.

L'analyse de données d'un sondage en employant des plans d'échantillonnage complexes doit tenir compte du regroupement des unités, de la stratification et de la probabilité inégale de sélection. La méthode basée sur le plan et celle basée sur le modèle sont deux méthodes généralement utilisées pour considérer de tels plans de sondage. Plusieurs études de plan de sondage transversales ont prouvé que ces deux approches fournissent des résultats semblables quand le modèle représente bien les données. Ce papier vise à comparer ces deux approches en utilisant les données de l'ENSP. Un modèle utilisant les marginales proposé par Rao (1998) (basée sur le plan) et la méthode généralisée d'estimation d'équation, proposée par Liang et Zeger (1986) (basé sur le modèle), seront utilisés. Des résultats obtenus à partir de ces méthodes seront comparés.

Session 10A Mercredi 31 mai • Wednesday, May 31, 10:30–11:30 SSC2050**Special Invited Session of the Biostatistics Section****Allocution de l'invité d'honneur du Groupe de biostatistique****[MS-181]****Weaknesses and Strengths of a Bayesian Approach in Clinical Research****Faiblesses et forces d'une approche bayésienne en recherche clinique**Donald BERRY, *The University of Texas M.D. Anderson Cancer Center*

I argue that almost all statistical analyses are wrong, regardless of their philosophical underpinnings! Bayesian analyses are especially susceptible to erroneous conclusions. A rigorous frequentist approach is immune. But it comes with heavy baggage that slows progress. In attempting to lighten the load, I turn to Bayesian innovations in clinical trials, especially design. I describe some of these benefits and relate them to modern attitudes in drug and medical device development, and to attitudes in cancer cooperative groups and at my institution. Of special importance are the uses of (i) flexible, adaptive designs, (ii) predictive probabilities, and (iii) hierarchical modeling.

Je soutiens que presque toutes les analyses statistiques, indépendamment de leurs fondements philosophiques, sont erronées ! Les analyses bayésiennes sont particulièrement susceptibles de mener à des conclusions erronées. Une approche rigoureuse fréquentiste y est immunisée. Cependant, une telle approche vient avec une lourde charge rendant difficile le traitement des problèmes. Dans le but d'alléger cette charge, je me tourne vers les innovations bayésiennes dans les essais cliniques, particulièrement vers la conception de plans d'expérience. Je décris certains des avantages et j'établis un rapport avec les attitudes modernes vis-à-vis le développement de médicaments et d'appareils médicaux, et les attitudes prônées par les groupes coopératifs de cancer et par mon établissement. Une importance particulière est accordée à l'utilisation (i) de plans d'expérience adaptatifs et flexibles, (ii) de probabilités prédictives, (iii) d'une modélisation hiérarchique.

Session 10B Mercredi 31 mai • Wednesday, May 31, 10:30–12:00 SSC2036**Recent Advances in SPC and Data Driven Statistics****Développements récents en CSP et statistiques guidées par les données****[MS-182]****Statistical Process Control and the Problem of Unknown Parameters****Difficultés causées par la présence de paramètres inconnus en contrôle des procédés.**Douglas HAWKINS, *University of Minnesota*

Until recently, statistical process control (SPC) has been reminiscent of statistics pre-Student. Parameter estimates would be plugged in where exact parameters were required, and the resulting tools used as if these estimates were the true parameters. It is now appreciated that estimated parameters cause much more severe distortion in SPC tools than in classical statistics. Some researchers, paralleling Student, are studying the impact that parameter estimation has on existing SPC tools. Others are devising new tools that do not involve parameters that need to be estimated. The talk sketches some ideas and results from both threads of research.

Jusqu'à récemment, la statistique du contrôle des procédés (SCP) rappelait la statistique de la période pré-Student. Des valeurs estimées des paramètres sont souvent substituées aux vraies valeurs de ces paramètres, les outils résultants les utilisant sans distinction. On se rend compte maintenant que l'estimation des paramètres amène un biais beaucoup plus grave dans l'utilisation des outils de SCP que dans la statistique classique. Quelques chercheurs, imitant Student, étudient l'impact de l'estimation des paramètres sur les outils usuels de SCP. D'autres conçoivent des nouveaux outils n'impliquant pas de paramètres de nuisance. Cette présentation propose quelques idées et résultats des deux écoles de pensées.

[MS-183]**Shewhart Control Charts in New Perspective****Une nouvelle perspective pour les cartes de contrôle Shewhart**Willem ALBERS & Wilbert C.M. KALLENBERG, *University of Twente*

Parameter estimation and/or lack of normality strongly affect standard control charts. Corrections are presented, as well as extensions to parametric and nonparametric charts. A flexible parametric choice is based on the normal power family. If this is still not adequate, a nonparametric chart is called for. Finally, a data driven procedure can choose between the three different charts. Once the nonparametric option is selected, many Phase I observations should be available. If that is not the case, the individual chart should be replaced by a grouped one. A newly proposed 'minimum chart' is then recommended.

L'estimation des paramètres et/ou l'absence de normalité affectent fortement les cartes de contrôle standard. Des corrections sont présentées, aussi bien qu'une extension aux cartes paramétriques et non paramétriques. Un choix paramétrique flexible est basé sur la famille normale de puissance. Si ce n'est toujours pas adéquat, une carte non paramétrique est favorisée. Finalement, un procédé guidé par les données mène à un choix entre les trois cartes différentes. Une fois que l'option non paramétrique est choisie, plusieurs observations en phase I devraient être disponibles. Si ce n'est pas le cas, la carte individuelle devrait être remplacée par une autre en groupe. Une carte minimale nouvellement proposée est alors recommandée.

[MS-184]**On t and EWMA t Charts for Monitoring Changes in the Process Mean****Sur des cartes de contrôle t et EWMA t pour étudier le changement dans la moyenne du processus**Lingyun ZHANG, *Massey University*, Gemai CHEN, *University of Calgary*

The performance of \bar{X} chart is usually and naturally studied under the assumption that the process variance is well estimated or does not change. This is, of course, not always the case in practice. We find that in the No case, \bar{X} charts are not robust against errors in variance estimation or changing variance. In this talk we discuss the use of a t chart and an Exponentially Weighted Moving Average (EWMA) t chart to monitor the process mean. We show that the EWMA t chart has the desired robustness property when the process variance is poorly estimated or is changing, and performs similarly to the EWMA \bar{X} chart when the variance is well estimated or does not change.

La performance du diagramme \bar{X} est habituellement et naturellement étudiée sous l'hypothèse que la variance du processus est bien estimée ou ne change pas. Ce n'est, naturellement, pas toujours le cas dans la pratique. Nous constatons que dans un nombre de cas, les diagrammes \bar{X} ne sont pas robustes contre des erreurs d'estimation de la variance ou du changement de la variance. Dans cet entretien nous discutons l'utilisation d'un diagramme t et d'un diagramme t moyenne mobile exponentiellement pondérée (MMEP) pour contrôler la moyenne du processus. Nous démontrons que le diagramme t MMEP possède la propriété désirée de robustesse quand la variance du processus est mal estimée ou instable, et performe aussi bien que le diagramme \bar{X} MMEP quand la variance est bien estimée ou n'est pas instable.

Session 10C Mercredi 31 mai • Wednesday, May 31, 10:30–12:00 SH3345**Analyzing Data from International Surveys****Analyse de données d'enquêtes internationales****[MS-185]****Examples From Large-scale Skill Assessment****Exemples à partir d'évaluations d'habiletés à grande échelle**Scott MURRAY, *UNESCO Institute for Statistics*

The large-scale assessment of proficiency is one of the most rapidly expanding areas of statistics. This paper draws upon examples from three international comparative studies—the OECD PISA study, Statistics Canada's and NCES's International Adult Literacy Survey (IALS) and UNESCO OREALC's Latin American Laboratorio for the Evaluation of Educational Quality (LLECE)—to illustrate the analytic techniques that are being applied to extract information value from the assessments for use by policy makers. The paper will also provide some information on the nature of measurement error that characterizes such assessments and what steps are being taken to control them to manageable levels.

L'évaluation à grande échelle de la compétence est l'un des secteurs statistiques dont l'expansion est la plus rapide. Cet article utilise les exemples de trois études comparatives internationales – l'étude OCDE PISE, l'Enquête internationale sur l'alphabétisation et les compétences des adultes (EIACA) de Statistique Canada et de « NCES » et « l'UNESCO OREALC Latin American Laboratorio for the Evaluation of Educational Quality (LLECE) » – pour illustrer les techniques analytiques qui sont appliquées pour extraire la valeur de l'information à partir des évaluations à l'usage des décideurs. L'article fournira également quelques informations sur la nature de l'erreur de mesure qui caractérise de telles évaluations et quelles mesures sont prises pour les contrôler à des niveaux gérables.

[MS-186]**Child Poverty and Changes in Child Poverty****La pauvreté chez l'enfant et les changements dans la pauvreté chez l'enfant**Miles CORAK, *Statistics Canada*

This paper documents levels and changes in child poverty rates in 12 OECD countries using data from the Luxembourg Income Study project, and focusing upon an analysis of the reasons for changes over the 1990s. The objective is to uncover the relative role of income transfers from the state in determining the magnitude and direction of change in child poverty rates, holding other demographic and labour market factors constant. As such the paper offers a cross-country overview of child poverty, changes in child poverty, and the impact of public policy in North America and Europe.

Cet article documente les niveaux et les changements dans les taux de pauvreté chez l'enfant dans 12 pays de l'OCDE en utilisant les données du projet Luxembourg Income Study project, et se focalisant sur une analyse des raisons des changements au cours des années 90. L'objectif est de découvrir le rôle relatif des transferts de revenu de l'état en déterminant la grandeur et la direction du changement des taux de pauvreté chez l'enfant, en fixant les autres facteurs démographiques et le marché du travail. Ainsi, cet article offre une vue d'ensemble transnationale de la pauvreté chez l'enfant, changement dans la pauvreté chez l'enfant, et l'impact des politiques d'intérêt public en Amérique du nord et l'Europe.

[MS-187]**Analysis and Availability of Data Collected Via International Household Surveys Programs****Analyse et disponibilité des données provenant de programmes d'enquêtes internationales sur les ménages**Edilberto LOAIZA, *UNICEF*

Since the 1970s a couple of household surveys programs have been provided, for developing countries. They needed data to develop and monitor policies, programs and interventions in the areas of health, nutrition, family planning and population. The Demographic and Health Surveys (DHS) program sponsored by USAID and UNICEF's Multiple Indicators Cluster Survey (MICS) are analyzed in this presentation with the objective of presenting the richness of these two data sources as well as the potentials for further analysis and comparability across time. Some examples are presented in the context of the work developed by UNICEF in the countries where these programs have taken place. Of particular importance is the reference to the role these two data sources have played in monitoring country's commitments towards the 2015 Millennium Development Goals.

Depuis les années 1970, deux programmes d'enquêtes sur les ménages ont été fournis pour répondre aux besoins des pays en voie de développement. Ils avaient besoin de disposer de données afin de pouvoir développer des politiques et faire des suivis dans les domaines de la santé, de la nutrition, de la planification familiale et de la population. Le programme « Demographic and Health Surveys (DHS) », sous l'égide de l'USAID, ainsi que le « Multiple Indicators Cluster Survey » (MICS) d'UNICEF sont analysés dans cette présentation, avec pour objectif de présenter la richesse de ces deux sources d'information et le potentiel en données pour faire des analyses plus approfondies et une compatibilité à travers le temps. Certains exemples sont présentés dans le contexte du travail développé par UNICEF dans les pays où ces programmes sont utilisés. La référence au rôle de ces deux sources de données dans la surveillance des engagements pris par les pays envers les « Objectifs de développement du millénaire 2015 » est d'une importance particulière.

Session 10D Mercredi 31 mai • Wednesday, May 31, 10:30–12:00 SSC3018**Statistical Analysis of Non-precise Data
Analyse statistique des données imprécises****[MS-188]****Fuzzy Information and Statistics
L'information floue et la statistique**Reinhard VIERTL, *Vienna University of Technology*

Data used in statistics are frequently not precise numbers but more or less non-precise. Especially all measurement results from continuous quantities are non-precise. This kind of data uncertainty is different from errors and can be best described by fuzzy models. Now statistical methods have to be adapted for that kind of data. This is possible for descriptive statistics as well as for statistical inference by applying the concept of propagation of imprecision from fuzzy set theory. In Bayesian inference there is another kind of non-precise information: non-precise a-priori information. The mathematical model for that are so-called fuzzy probability distributions.

Il est fréquent que les données utilisées en statistique ne soient pas des nombres précis mais soient plus ou moins imprécis. En particulier, toutes les mesures de quantités continues sont imprécises. Cette sorte d'incertitude relative aux données est différente des erreurs et les modèles flous conviennent mieux pour la décrire. Les méthodes statistiques doivent être adaptées pour ce type de données. En utilisant le concept de propagation de l'imprécision de la théorie des ensembles flous, cela est possible à la fois pour les statistiques descriptives et l'inférence statistique. En inférence bayésienne, il faut prendre en considération un autre type d'information imprécise : l'information a priori imprécise, modélisée par les distributions dites floues.

[MS-189]**Classification of Non-precise Data
Classification de données non précises**François THÉBERGE & Mayer ALVO, *University of Ottawa*

Non-precise data arise when it is the imprecision of the observation itself that is of interest rather than the uncertainty due to statistical variation. We present a framework to address the problem of multi-population classification for non-precise data. This is achieved by adapting the usual classification rule to take into account the characterizing functions of the observations. We also compare a new likelihood score function for non-precise quantities to a straightforward extension from probability density functions and show that the former is always a better choice under some global error criterion. Results are illustrated via examples.

Les données non précises surgissent lorsque c'est l'imprécision de l'observation elle-même qui est d'intérêt plutôt que de l'incertitude due à la variation statistique. Nous présentons un cadre pour adresser le problème de la classification de multi-population pour des données imprécises. Ceci est réalisé en adaptant la règle habituelle de classification pour tenir compte des fonctions caractérisantes des observations. Également, nous comparons une nouvelle fonction de cotes de vraisemblance pour des quantités imprécises à une extension directe des fonctions de densité de probabilité et nous prouvons que la première est toujours un meilleur choix sous un certain critère d'erreur global. Des résultats sont illustrés par des exemples.

[MS-190]**A Classical Goodness-of-fit Test Based on Fuzzy Random Variables
Un test classique d'ajustement basé sur des variables aléatoires floues**Maria GIL, Gil GONZALEZ-RODRIGUEZ & Ana COLUBI, *Universidad de Oviedo, Spain*

In a previous paper a class of characterizing fuzzy representations of a random variable has been introduced. Each fuzzy representation corresponds to a fuzzy random variable (in Puri and Ralescu's sense) whose (fuzzy) expected value characterizes the distribution of the original random variable.

The problem of the one-sample testing of the two-sided hypothesis about the mean of a fuzzy random variable has been also discussed in depth in another paper.

By combining both tools, classical goodness-of-fit tests for real-valued random variables can be developed. The first preliminary empirical studies show a suitable behavior when compared with well-known methods.

Dans un article précédent, nous avons introduit des représentations floues d'une variable aléatoire. Une représentation floue est une variable aléatoire floue (au sens de Puri et Ralescu) dont la valeur de l'espérance (floue) caractérise la distribution de la variable aléatoire originale.

Le problème du test pour un échantillon des hypothèses bilatérales sur l'espérance d'une variable aléatoire floue a été traité en détail dans un autre article.

La combinaison de ces deux instruments nous conduit à l'obtention de tests d'ajustement pour variables aléatoires à valeurs réelles. Les premières études empiriques ont montré des résultats encourageants lorsqu'on les compare à ceux des tests d'ajustement habituels.

Session 10E Mercredi 31 mai • Wednesday, May 31, 10:30–12:00 SSC2024**Case Study II: Obstructive Sleep Apnea****Étude de cas II : L'apnée obstructive du sommeil****Session 10F Mercredi 31 mai • Wednesday, May 31, 10:30–12:00 SSC3026****Time Series****Séries chronologiques****[MS-191]****Variable Selection in Generalized Linear Models by Empirical Likelihood
Sélection de variables par la vraisemblance empirique dans un modèle linéaire généralisé**ASOKAN MULAYATH VARIYATH, BOVAS ABRAHAM & JIAHUA CHEN, *University of Waterloo*

Variable selection is an important topic widely discussed in the statistical literature. The main objective of variable selection is to identify the least number of covariates that explain the response variable appropriately. There are many variable selection procedures such as AIC, BIC, cross validation. The information based criteria such as AIC and BIC often need a well defined parametric likelihood. In generalized linear models, however, the likelihood is usually not well defined. In this context, we propose to use the empirical likelihood based AIC and BIC for the purpose of model selection. A sub-model in GLM is defined by setting

La sélection de variables est un sujet important très présent dans la littérature statistique. L'objectif principal de la sélection de variables est d'identifier le plus petit nombre de covariables expliquant adéquatement la variable réponse. Il existe plusieurs procédures de sélection de variables, comme le AIC, le BIC et la validation croisée. Les critères AIC et BIC, basés sur l'information, nécessitent souvent une vraisemblance paramétrique bien définie. Sous des modèles linéaires généralisés, cependant, la vraisemblance n'est généralement pas bien définie. Dans ce contexte, nous proposons l'utilisation du AIC et BIC basés sur la vraisemblance empirique afin de sélectionner le modèle. En MLG, un sous-modèle est défini en fixant certains coefficients de régres-

some regression coefficients to 0. Its profile empirical likelihood is computed under the constraints of the equations specified by the full GLM, with these regression coefficients equating to 0. We propose a method to ensure the existence of solutions. The new procedure has similar properties to that of AIC and BIC under parametric model assumptions. The simulations results show that the new method works well.

sion à 0. Sa vraisemblance empirique de profil est calculée sous les contraintes sur les équations spécifiées par le MLG complet, avec ces coefficients de régression égaux à 0. Nous proposons une méthode pour nous assurer de l'existence de solutions. La nouvelle procédure a des propriétés similaires au AIC et BIC sous l'hypothèse d'un modèle paramétrique. Les résultats de simulation démontrent que la nouvelle méthode fonctionne bien.

[MS-192]

**Normal Mixtures GARCH Models and Option Pricing
Modèles GARCH de mélanges normaux et évaluation d'option**

Alex BADESCU & Reg KULPERGER, *University of Western Ontario*, Emese LAZAR, *University of Reading*

The skewness and excess kurtosis of the asset returns were recently explained in the GARCH framework by using different distributional assumptions on the driving noise. However derivatives valuation was rarely addressed. We analyze the pricing performance of an asymmetric normal mixture GARCH model with two variance components. We compute option prices on an Index stock for different GARCH specifications under different risk neutral measures.

L'asymétrie et l'aplatissement des excès de retours d'actifs ont été récemment expliqués dans le cadre d'un modèle GARCH en employant différentes hypothèses sur la distribution du bruit. Cependant l'évaluation des dérivés a été rarement abordée. Nous analysons l'évaluation de la performance d'un modèle GARCH de mélange normal asymétrique avec une variance à deux composantes. Nous calculons différentes options de prix pour un indice d'action pour différentes caractéristiques de GARCH sous différentes mesures indifférentes au risque.

[MS-193]

**Quasi-empirical Bayes Methods of Estimation in ARMA models
Méthodes quasi bayésiennes dans l'estimation des modèles ARMA**

Mohamedou OULD HAYE & A. K. Md. E. SALEH, *Carleton University*

This talk investigates the asymptotic properties of various estimators of coefficients in an ARMA(p, q) model when vague non-sampling information on the moving average part is available. In particular, we discuss the usual MLE (called unrestricted estimator), restricted MLE (based on vague information), preliminary test estimator, shrinkage estimator and the positive rule estimator of the autocorrelations. It is shown that near the prior information on the MA-parameters, the restricted, preliminary test and shrinkage estimators perform better than the unrestricted estimator, while their superiority changes as MA-parameters divert from the prior informations. The analysis is based on the asymptotic properties of the estimators under contiguous alternatives.

Dans cet expose, nous explorons les propriétés asymptotiques de plusieurs estimateurs des coefficients d'un processus ARMA(p, q) sous l'hypothèse de vague information sur la composante moyenne mobile MA. En particulier, nous comparons l'estimateur classique de maximum de vraisemblance (MLE), que nous appelons ici l'estimateur non restreint, à différents types d'estimateurs tels que l'estimateur de test préliminaire (PTE), le maximum de vraisemblance restreint, l'estimateur de Stein, ainsi que l'estimateur dit de règle positive. Nous montrons qu'en présence d'information sur la composante moyenne mobile, et en termes de minimisation de la variance, tous ces estimateurs donnent de meilleurs résultats que le classique maximum de vraisemblance. Nous comparons aussi l'efficacité de ces estimateurs sous des hypothèses alternatives contiguës.

[MS-194]**Algorithm of Maximum Likelihood Parameters Estimation for First-order Superdiagonal Bilinear Time Series****Algorithme d'estimation par maximum de vraisemblance des paramètres du modèle bilinéaire superdiagonal d'ordre un**

Khadija BOUZAACHANE, *Ecole Mohammadia d'Ingénieurs, Maroc*, Youssef BENGHABRIT, *Ecole Nationale Supérieure des Arts et Métiers, Meknès*, Mostaf HARTI, *Université Sidi Mohammed Ben Abdellah, Maroc*

Bilinear models represent a simple and powerful class of nonlinear models. Their analysis and particularly the estimation of their parameters is of central interest. Our work deals with the design of a new algorithm for estimating the parameters of the first-order superdiagonal bilinear time series model. This iterative algorithm is based on maximum likelihood method and the widely used Kalman filter.

To demonstrate the usefulness of our algorithm series of Monte Carlo simulations were performed.

Les modèles bilinéaires représentent une classe particulière des modèles non linéaires de séries chronologiques. L'estimation des paramètres de ces modèles a été peu développée et demeure un axe de recherche intéressant. Dans notre travail nous avons établi un nouvel algorithme d'estimation des paramètres des modèles bilinéaires. Cet algorithme est fondé sur la méthode du maximum de vraisemblance et le filtre de Kalman. Pour démontrer l'utilité et la bonne performance de notre algorithme nous avons généré des séries de simulations par la méthode de Monte Carlo.

[MS-195]**Improved Estimation for Dynamic Linear Regression Model****Estimation améliorée pour un modèle de régression linéaire dynamique**

Zahirul HOQUE, *University of Newcastle, Australia*, Richard GERLACH, *University of Sydney, Australia*

This study investigates the preliminary test and shrinkage estimators of linear state space regression model via Kalman filtering. The performance of the estimators, with respect to mean squared error, has been investigated. It has been revealed that under certain conditions both preliminary test and shrinkage estimators outperform Kalman filter but shrinkage estimator is superior to preliminary test estimator. Hence, the result presented in this paper invalidates the minimum mean square error property of Kalman filter that is widely used by the engineers for estimation of the parameters of linear state space models.

Cette étude porte sur le test préliminaire et les estimateurs par rétrécissement d'un modèle de régression spatiale à état linéaire avec filtrage de Kalman. La performance des estimateurs a été étudiée en ce qui a trait à l'erreur quadratique moyenne. Sous certaines conditions, le test préliminaire et l'estimateur par rétrécissement sont plus performants que le filtre de Kalman, mais l'estimateur par rétrécissement s'avère supérieur à l'estimateur du test préliminaire. Les résultats présentés dans cet article invalident donc la propriété d'erreur quadratique minimale du filtre de Kalman, largement utilisé en ingénierie pour l'estimation des paramètres de modèles spatiaux à état linéaire.

[MS-196]**Subset Autoregression: A New Approach****Une nouvelle approche pour modèles autorégressifs sur sous-ensembles**

Ian McLEOD, *University of Western Ontario*, Ying ZHANG, *Acadia*

A new family of subset autoregressive models are introduced and a comprehensive approach to model identification, estimation and diagnostic checking is developed for these models. Also a new version of the partial autocorrelation plot is introduced.

Une nouvelle famille de modèles à sous-ensembles autorégressifs est introduite et une approche compréhensive à l'identification de modèle, à l'estimation et à la vérification diagnostique est développée pour ces modèles. Une nouvelle version du graphe d'autocorrélation partielle est aussi pré-

These new models are better suited to efficient model building of high-order autoregressions with long time series. Several illustrative examples are given.

An R package implementation is available.

sentée.

Ces nouveaux modèles sont plus adaptés à la construction de modèles efficaces d'autorégression d'ordre élevé avec longues séries chronologiques. Plusieurs exemples illustratifs sont présentés.

Une implémentation sous R est disponible.

Session 10G Mercredi 31 mai • Wednesday, May 31, 10:30–12:00 SSC3006

Probability and Mathematical Statistics Probabilité et statistique mathématique

[MS-197]

Forward-backward Stochastic Differential Equations and Transforms for Affine Diffusions

Les équations différentielles stochastiques directes et rétrogrades et les transformations pour diffusions affines

Cody HYNDMAN, *University of Calgary*

For an affine diffusion we show how forward-backward stochastic differential equations (FBSDEs) can be employed to characterize general transforms, including the Fourier transform as a special case, in terms of the solution of Riccati ordinary differential equations. The transforms can be employed to calculate prices of derivative securities where the underlying asset price is driven by the affine diffusion. Numerical techniques for the solution of the associated FBSDE are explored in the context of simulation techniques not requiring the solution of the Riccati equations.

Pour une diffusion affine nous montrons comment des équations différentielles stochastiques directes et rétrogrades (EDSDR) peuvent être utilisées pour caractériser les transformations générales, y compris la transformation de Fourier comme cas spécial, en termes de solution des équations différentielles ordinaires de Riccati. Les transformations peuvent être utilisées pour calculer les prix des titres dérivés où le prix des actifs sous-jacents est conduit par la diffusion affine. Des techniques numériques pour la solution des EDSDR associées sont explorées dans le contexte des techniques de simulation n'exigeant pas la solution des équations de Riccati.

[MS-198]

Maximizing Expected Utility Through Forward-backward Stochastic Differential Equations

Optimisation de l'utilité espérée par le biais d'équations différentielles stochastiques directes et rétrogrades

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

Utility portfolio selection is concerned for example with developing dynamic strategies which maximize the expected utility from terminal wealth. In the literature explicit optimal portfolios are mainly given when the interest rate of the bond, the appreciation rate and volatility of the stocks are taken as deterministic. We derive here an optimal portfolio for a continuous-time model with random coefficients and express the optimal portfolio in terms of the solution of forward-backward stochastic differential equations (solutions of these can be computed explicitly in the case of some classes of utility functions).

En sélection de portefeuille par utilité, nous nous intéressons, par exemple, à développer des stratégies dynamiques maximisant l'utilité espérée de la richesse terminale. Dans la littérature, des portefeuilles optimaux explicites sont surtout donnés lorsque les coefficients du marché sont considérés comme fixes. Nous obtenons ici un portefeuille optimal pour un modèle à temps continu avec coefficients du marché aléatoires. Nous montrons que des solutions à ce problème peuvent être obtenus par l'étude d'équations différentielles stochastiques directes et rétrogrades, dont les solutions peuvent être calculées explicitement dans le cas de certaines classes de fonctions d'utilité.

[MS-199]**A Two-sample Copula Process****Un processus de copules pour deux échantillons**Adriana JORDAN & Gail IVANOFF, *University of Ottawa*

We define a two-sample version of the empirical copula process; this can be viewed as a type of two-dimensional p - p plot. We prove a functional limit theorem which can be applied to tests of stochastic order in higher dimensions, including precedence tests, when only partial data is available.

Nous définissons une version à deux échantillons du processus empirique de copule ; ceci peut être regardé comme un graphique bidimensionnel de type p - p . Nous prouvons un théorème de limite fonctionnel qui peut être appliqué aux tests d'ordre stochastique dans des dimensions plus élevées, y compris des tests de priorité, quand seulement des données partielles sont disponibles.

[MS-200]**Empirical Central Limit Theorems under Model Misspecification****Versions empiriques du théorème central limite pour un modèle mal spécifié**Hanna JANKOWSKI, *University of Washington*

Asymptotic behaviour of infinite-dimensional data may be studied by considering the limit of the empirical measure and the associated fluctuations. We discuss empirical central limit theorems for stochastic models under perturbations. In particular, a $n^{-1/2}$ perturbation in the parameters describing the modelling process results in a non-zero mean for the limiting distribution. We derive an expression for the mean operator in terms of Malliavin derivatives. The perturbations we consider can be seen as random contiguous alternatives.

Le comportement asymptotique de données en dimension infinie peut être déterminé en étudiant la limite de la mesure empirique et des fluctuations qui y sont associées. Nous discutons des versions empiriques du théorème central limite pour des modèles stochastiques contenant des perturbations. En particulier, une perturbation $n^{-1/2}$ des paramètres décrivant le processus de modélisation résulte en une distribution limite possédant une moyenne non-nulle. Nous dérivons une expression pour l'espérance en terme des dérivées de Malliavin. Les perturbations considérées sont des limites sous des alternatives contiguës aléatoire

[MS-201]**Strong Approximation for Mixing Sequences with Infinite Variance****Approximation forte pour des suites « mélangeantes » avec variance infinie**Raluca BALAN, *University of Ottawa*, Ingrid-Mona ZAMFIRESCU, *Baruch College, City University of New York*

In this talk we will present a strong approximation result for a mixing sequence with infinite variance and logarithmic decay rate of the mixing coefficient. The result is proved under the assumption that the distribution is symmetric and lies in the domain of attraction of the normal law. Moreover the truncated variance function is supposed to be slowly varying with log-log type remainder.

Dans cet exposé, nous présenterons le résultat d'une approximation pour des suites mélangeantes avec variance infinie et avec un taux logarithmique de décroissance des coefficients. Nous travaillons sous l'hypothèse que la distribution est symétrique et se situe dans le domaine d'attraction de la loi normale. De plus, nous supposons que la fonction de la variance tronquée varie lentement avec un memento de type bilogarithmique.

[MS-202]**New Central Limit Theorems via Studentization in Linear Structural Error-in-variables Models****Nouveaux théorèmes limites centraux basés sur la studentisation dans les modèles linéaires avec erreurs de mesure**Yuliya MARTSYNYUK, *Carleton University*

Linear structural error-in-variables models are revisited for studying least squares estimators of slope and intercept. New central limit theorems (CLT's) are established for these estimators under the least restrictive ever model conditions in this regard. Moreover, due to Studentization, these CLT's are invariant in form within the introduced class of explanatory and error variables, and free of unknown parameters of their joint distribution. In contrast, in related CLT's in the literature so far, variances of the limiting normal distributions, in general, are complicated and depend on various, typically unknown, hard-to-estimate moments of error and explanatory variables.

Les modèles linéaires structuraux avec erreur de mesure sont revus pour l'étude des estimateurs des moindres carrés de la pente et de l'ordonnée à l'origine. De nouveaux théorèmes centraux limite (TCL) sont établis pour ces estimateurs avec les pré-supposés les moins restrictifs employés jusqu'ici à cet égard. De plus, grâce à la studentisation, ces TCL sont invariants en forme à l'intérieur de la classe introduite des variables explicatives et d'erreur, et libres de paramètres inconnus liés à leur distribution conjointe. Ceci contraste avec les TCL similaires retrouvés dans la littérature, pour lesquels les variances des distributions normales limites sont généralement compliquées et dépendent de divers moments des variables d'erreur et explicatives, souvent inconnus et difficiles à estimer.

Session 10H Mercredi 31 mai • Wednesday, May 31, 10:30–12:00 SSC3014**Statistical Genetics and Genomics****Génétique statistique et génomique****[MS-203]****Group Sequential Approach for Penetrance Estimation of Genes Involved in Complex Diseases Based on Weighted Log-rank Statistics****Estimation de la pénétrance des gènes impliqués dans des maladies complexes : une approche séquentielle par groupes basée sur des statistiques de (log) rang pondérées**M. Tariqul HASAN & LAURENT BRIOLLAIS, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital.*

Advances in the identification and treatment of genetically transmitted diseases have led to an increased need for reliable estimates of genetic susceptibility risk. These estimates are used to identify individuals at risk of being a disease allele carrier and to define the probability of developing a disease in carriers (penetrance). In this work, we are interested of family-based group sequential design that could include one or more interim analyses. The sample size necessary to estimate the penetrance in mutation carriers with accurate precision or to conclude to a significant difference between two groups, e.g. mutation carrier vs. non-carriers is updated at each phase. The experiment stops when enough families have been collected. Our method is based on weighted log-rank statistics for cluster survival data.

Des avancées dans l'identification et le traitement de maladies transmises génétiquement ont conduit à un besoin accru d'estimations fiables du risque de susceptibilité génétique. Ces estimés sont utilisés pour identifier les individus à risque d'être porteur de l'allèle de la maladie et pour définir, chez les porteurs, la probabilité que se développe la maladie (la pénétrance). Dans ce travail, nous sommes intéressés par un plan séquentiel par groupes basé sur la famille pouvant inclure au moins une analyse intérimaire. La taille de l'échantillon requis pour estimer la pénétrance chez les porteurs de l'allèle mutant avec précision, ou pour conclure à une différence significative entre deux groupes (par exemple, porteurs et non-porteurs), est mise à jour à chaque phase. L'étude prend fin lorsque suffisamment de données de familles sont recueillies. Notre méthode est basée sur des statistiques de log-rang pondérées pour des données de survie en grappes.

[MS-204]**Estimating the Lifetime Risk Associated with Mutated Genes Involved in Complex Diseases
Estimation du risque associé aux gènes mutants impliqués dans des maladies complexes**Yun Hee CHOI & Laurent BRIOLLAIS, *Samuel Lunenfeld Research Institute, Mount Sinai Hospital*

Estimating the lifetime risk (penetrance) associated with genetic abnormalities has major implications for developing effective prevention strategies for susceptible individuals. We develop new statistical approaches to estimate the penetrance of mutated genes involved in complex diseases by using families who carry the mutated gene. Various study designs and statistical methods are considered for estimating the age-specific risk of disease among carriers and non-carriers of the identified mutation. The efficiency of several study designs—population- and clinic-based—is compared via simulation studies using different likelihood-based methods—prospective, retrospective and joint likelihoods—to account for specific ascertainment of the family.

L'estimation du risque à vie (pénétrance) associé aux anomalies génétiques a des implications majeures sur le développement de stratégies de prévention efficaces pour les individus susceptibles. Nous développons de nouvelles approches statistiques pour estimer la pénétrance de gènes mutants impliqués dans des maladies complexes à l'aide de familles porteuses du gène mutant. Différents plans d'expérience et méthodes statistiques sont considérés pour étudier le risque lié à l'âge de maladie chez les porteurs et non-porteurs de la mutation identifiée. L'efficacité de plusieurs plans d'expérience est comparée par des études de simulation utilisant différentes méthodes basées sur la vraisemblance – vraisemblance prospective, rétrospective et conjointe – pour considérer des découvertes spécifiques à la famille.

[MS-205]**Genetic Effect Estimation via Bootstrap in Linkage Analysis of Quantitative Trait Loci
Estimation d'effet génétique par bootstrap dans l'analyse de liaison du trait quantitatif du loci (emplacement des gènes)**Longyang WU, *Samuel Lunenfeld Research Int.*, LongYang WU, *SLRI*, Lei SUN, *University of Toronto*, Bull SHELLEY, *SLRI*

Bias occurs in locus-specific effect-size estimation in genome-wide linkage analysis of quantitative trait loci (QTL) when the original data are used for both hypothesis testing and parameter estimation. We propose bootstrap resampling estimators for bias reduction in the situation in which linkage peaks at more than one QTL are of interest. We conducted simulation studies in nuclear families with 0 to 5 QTLs and applied the methods in a genome-wide analysis of a blood pressure phenotype in extended pedigrees from the Framingham Heart Study (FHS). The new estimators substantially reduce the upward bias in genetic effect estimation.

Le biais dans l'estimation de la taille de l'effet d'un locus spécifique dans l'analyse de liaison sur tout le génome du Trait Quantitatif du Loci (TQL) se produit quand les données originales sont utilisées à la fois pour les tests d'hypothèses et l'estimation de paramètres. Nous proposons ses estimateurs par rééchantillonnage « bootstrap » pour la réduction du biais dans la situation où la liaison est importante à plus d'un QTL. Nous avons réalisé des études par simulation dans les familles nucléaires avec 0 à 5 TQLs et avons appliqué les méthodes dans une analyse de liaison sur tout le génome d'un phénotype de tension artérielle dans des familles généalogiques à partir de l'étude « Framingham Heart Study (FHS) ». Les nouveaux estimateurs réduisent sensiblement le biais positif dans l'estimation d'effet génétique.

[MS-206]**A Method for Analysis of CGH Microarray Data
Une méthode pour l'analyse de données de micropuces CGH**Wenqing HE & Ian McLEOD, *University of Western Ontario*

Genomic DNA copy number alterations are important features for the development of human diseases. CGH microarray is a powerful technique that enables us to search genome-wide for possible regions with DNA copy number alterations. The DNA copy number may be viewed as a sequence along the whole genome, and the alteration regions correspond to the sequence changes. We propose to use a Bayesian change-point approach to model this sequence and alteration regions may then be identified. Simulation studies are conducted to evaluate the performance of the proposed method and a real CGH data set is analyzed.

Les amplifications d'ADN de génomes sont des dispositifs importants pour le développement dans le domaine de maladies humaines. Les micropuces CGH sont une technique puissante nous permettant de chercher les régions possibles avec des amplifications d'ADN sur tout le génome. L'amplification d'ADN peut être considérée comme une séquence sur le génome entier, et les régions de changement correspondent aux changements de la séquence. Nous proposons d'utiliser une approche bayésienne de point de rupture pour modéliser cette séquence, et les régions de changement peuvent alors être identifiées. Des études de simulation sont entreprises pour évaluer la performance de la méthode proposée et de vraies données CGH sont analysées.

[MS-207]**Improved Estimation for Differential Gene Expression by Shrinking Variance Components
Estimation améliorée de l'expression de gènes par le rétrécissement des composantes de variance**Lihua AN & S. Ejaz AHMED, *University of Windsor*

The difficulty is to know what problems to combine together... why should not all our estimation problems be lumped together on one grand meele? George Barnard (1962).

A major concern with microarray experiments is that they have little replication. The variance estimates obtained from an individual gene may be very imprecise, and therefore the inferences reached may not be trustworthy due to the weak variance estimates. An appealing idea for improving inferences from microarray experiments would be combining information across genes. In this communication we construct several estimators based on pretest and shrinkage rules. The risks of the proposed estimators are studied and compared with the gene-based unbiased estimator under quadratic loss and entropy loss functions. It is demonstrated that the shrinkage estimators are superior to the baseline estimators. Interestingly, the relative performance of the proposed estimators is significantly different under the two loss functions.

La difficulté est de savoir quels problèmes combiner.... pourquoi tous nos problèmes d'estimation ne devraient-ils pas être empilés ensemble en une grande mêlée? George Barnard (1962).

Un problème majeur avec les expériences à micropuces est qu'elles ont peu de répétitions. Les estimations de variance obtenue pour un gène donné peuvent être très imprécises; les inférences obtenues peuvent donc ne pas être sûres à cause de la faiblesse des estimations des variances. Une idée intéressante pour améliorer la qualité des inférences des expériences à micropuces serait de combiner l'information de plusieurs gènes. Dans cette communication, nous construisons plusieurs estimateurs basés sur des règles du pré-test et de réduction. Les risques des estimateurs proposés sont étudiés et comparés avec l'estimateur sans biais basé sur le gène, sous les fonctions de perte quadratique et entropique. Il est intéressant de noter que la performance relative des deux estimateurs proposés est significativement différente sous les deux fonctions de perte.

[MS-208]**Improving the Relevance of Microarray Data Clustering through Integration with Gene Ontology
Amélioration de la pertinence du regroupement des données de micropuces par l'intégration avec l'ontologie du gène**Adam ZAGDANSKI & Rafał KUSTRA, *University of Toronto*

The Gene Ontology (GO) offers structured, controlled vocabulary for describing the molecular function, biological process and cellular component characteristics. It has already been shown that GO may facilitate genome-wide studies, e.g. large-scale predictive tasks in functional genomics or knowledge-based analysis of microarray gene expression data.

We propose general clustering framework to integrate microarray expression data with GO annotations through combination of corresponding (dis)similarity measures. This integration improves both biological relevance and stability of clustering results. Our approach can be easily extended to incorporate other genomic data sources (e.g. protein-protein interactions or transcription factors binding sites).

L'ontologie du gène (OG) offre le vocabulaire structuré et contrôlé pour décrire la fonction moléculaire, le processus biologique et les caractéristiques de composante cellulaire. Il a été déjà montré que l'OG peut faciliter les études à la grandeur du génome, par exemple, les tâches de prédiction à grande échelle dans la génomique fonctionnelle ou l'analyse basée sur la connaissance des données de l'expression du gène.

Nous proposons un cadre général de groupement pour intégrer les données d'expression avec les annotations OG par la combinaison des mesures de (dis)similarité correspondantes. Cette intégration améliore la pertinence biologique et la stabilité des résultats en groupement. Notre approche peut être facilement étendue pour incorporer d'autres sources de données génomiques (par exemple des interactions protéine-protéine ou sites de liaison facteurs de transcription).

Session 10I Mercredi 31 mai • Wednesday, May 31, 12:00–16:30 SSC - 2nd floor hallway

Posters for Case Study II

Affiches pour l'étude de cas II

Session 11A Mercredi 31 mai • Wednesday, May 31, 13:30–14:30 SSC2050

Address of the winner of the Centre de recherches mathématiques/Statistical Society of Canada Prize

Allocution du lauréat du Prix du Centre de recherches mathématiques/Société statistique du Canada

Session 11B Mercredi 31 mai • Wednesday, May 31, 13:00–14:30 SSC2024**Risk Theory I
Théorie du risque I****[MS-209]****Erlangized Fluid Queues and their Relation to Finite-time Ruin Probabilities****Les files d'attente de fluides erlangisées et leur relation aux probabilités de ruine en temps fini**David STANFORD, *University of Western Ontario*, Vaidyanathan RAMASWAMI, *ATT Research*, Douglas WOOLFORD, *University of Western Ontario*

In recent years the techniques of fluid queues have been used to address a variety of problems in risk theory. Uses have included the time of ruin, the surplus prior to ruin and/or deficit upon ruin for correlated claims processes (and their relevant joint distributions), and the Erlangization approach for approximating finite-time ruin probabilities for renewal risk processes. In this paper, we present the theoretical results for the Erlangization approach as applied to the fluid queue model in general. We show how the resulting approximations can be used to answer risk-theoretic questions such as the approximation of finite time ruin probabilities for some correlated risk processes, and the containment of forest fires through the construction of fire lines.

Au cours des dernières années, les techniques liées aux files d'attente de fluides ont été utilisées dans l'étude de plusieurs problèmes en théorie du risque. Cette utilisation inclut le temps de ruine, le surplus antérieur à la ruine et/ou le déficit au moment de la ruine pour des processus de réclamations corrélés (et leurs distributions conjointes pertinentes), ainsi que l'approche d'erlangisation permettant d'approximer les probabilités de ruine en temps fini des processus de renouvellement du risque. Dans cet article, nous présentons les résultats théoriques liés à l'approche d'erlangisation telle qu'appliquée aux modèles de files d'attente de fluides en général. Nous montrons comment les approximations résultantes peuvent être utilisées afin de résoudre des problèmes reliés à la théorie du risque tels l'approximation de la probabilité de ruine en temps fini pour des processus de risques corrélés, ainsi que le confinement des feux de forêts via la construction de tranchées pare-feu.

[MS-210]**Risk Models with Constant Dividend Barrier****Modèles de risque avec barrière de dividende constante**Andrei BADESCU, *University of Waterloo*, Vaidyanathan RAMASWAMI, *AT&T Labs*, Soohan AHN, *University of Seoul*

Finite buffer fluid flow models are useful in analyzing the performance of high speed communications systems. A possible application of this class of models is to insurance risk models with a dividend barrier. Based on the matrix-analytic methods of Ramaswami for fluid flows, here we develop an efficient time dependent analysis for a general Markov modulated fluid flow model with a finite buffer and an arbitrary initial fluid level at time 0. We apply these to an insurance risk model with a dividend barrier and a general Markovian arrival process of claims with possible dependencies in successive inter-claim intervals and claim sizes. We demonstrate the implementation and accuracy of our algorithms through a set of numerical examples.

Les modèles d'écoulement des fluides à mémoire finie sont utiles dans l'analyse de performance de systèmes de communications à haute vitesse. Une application possible de cette classe de modèles sont les modèles de risque en assurances avec une barrière de dividende. À partir des méthodes matricielles analytiques de Ramaswami pour l'écoulement des fluides, nous développons ici une analyse efficace dépendante sur le temps pour un modèle général de Markov modulé d'écoulement des fluides avec mémoire finie et un niveau initial de fluide arbitraire. Nous appliquons ensuite ceci à un modèle de risque en assurances avec barrière au dividende et un processus markovien général d'arrivée des réclamations avec possibles dépendances sur les intervalles entre réclamations successives ainsi que sur les tailles des réclamations. Nous démontrons la mise en œuvre et la précision de nos algorithmes à l'aide d'exemples numériques.

[MS-211]

Constant Dividend Barrier in a Risk Model With Time-dependent Claim Sizes**Barrière de dividende constante dans un modèle de risque où le montant des réclamations dépend du temps**David LANDRIAULT, *University of Waterloo*

The risk model with time-dependent claim sizes proposed in Boudreault et al. (2006) is studied in the presence of a constant dividend barrier. An integro-differential equation for the Gerber-Shiu discounted penalty function is derived. We show that its solution can be expressed as the solution to the Gerber-Shiu discounted penalty function in the same risk model with the absence of a barrier plus a combination of two linearly independent solutions to the associated homogeneous integro-differential equation. Finally, we analyze the expected present value of dividend payments up to the time of ruin in the same risk model. A homogeneous integro-differential equation is derived and then solved. Its solution can be expressed as a different combination of the two linearly independent solutions to the homogeneous integro-differential equation associated to the Gerber-Shiu discounted penalty function.

Le modèle du risque avec montant des réclamations dépendant du temps proposé dans Boudreault et al. (2006) est étudié en présence d'une barrière de dividende constante. Une équation intégro-différentielle pour la fonction de pénalité escomptée de Gerber-Shiu est dérivée. Nous montrons que sa solution peut être exprimée comme la solution de la fonction de pénalité escomptée de Gerber-Shiu du même modèle en absence de barrière plus une combinaison de deux solutions linéairement indépendantes à l'équation intégro-différentielle homogène associée. Finalement, nous analysons l'espérance de la valeur présente des paiements de dividendes jusqu'au temps de ruine sous le même modèle de risque. Une équation intégro-différentielle homogène est obtenue et résolue. Sa solution peut être exprimée comme une combinaison différente des deux solutions linéairement indépendantes à l'équation intégro-différentielle homogène associée à la fonction de pénalité escomptée de Gerber-Shiu.

Session 12A Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SSC2050**Industrial Experimental Design****Planification d'expériences industrielles**

[MS-212]

Second-order Split-plot Design Construction and Evaluation**Plan d'expérience à parcelles subdivisées du 2^e ordre : construction et évaluation**Peter PARKER, *National Aeronautics and Space Administration*

When designing industrial response surface experiments, we often encounter practical or economic constraints that restrict complete randomization and require a split-plot design. While statistical efficiency is well-defined for complete randomization, overall experimental efficiency of a split-plot design requires a consideration of the experimental apparatus. Two aerospace research case studies are used to illustrate second-order split-plot design construction and evaluation. Competing designs are evaluated from a statistical and practical perspective considering their utilization of experimental resources, clarity of structure, and simplicity of analysis. Subjective compromises among design

Lors de la planification d'expériences portant sur une surface de réponse dans le domaine industriel, nous rencontrons souvent des contraintes pratiques ou économiques limitant la randomisation complète et exigeant l'emploi d'un plan à parcelles subdivisées. Alors que l'efficacité statistique est bien définie pour la randomisation complète, l'efficacité de la méthode des parcelles subdivisées exige une considération de l'appareillage expérimental. Deux études de cas tirées de la recherche aérospatiale sont utilisées pour illustrer la conception et l'évaluation de plans à parcelles subdivisées du second ordre. Des plans concurrents sont évalués d'un point de vue statistique et pratique en considérant leur utilisation des ressources expérimentales, simplicité de structure, et facilité d'analyse. Une importance particulière est accordée à

criteria are emphasized to develop an effective experimental strategy to accommodate restrictions on randomization.

certain compromis subjectifs entre les différents critères de planification afin de développer une stratégie expérimentale efficace permettant une adaptation aux contraintes de randomisation.

[MS-213]

Optimal Foldover Plans for Two-level Fractional Factorial Split-plot Designs

Plans de repliement optimaux pour plans factoriels fractionnels à deux niveaux à parcelles subdivisées

Rob McLEOD, *University of Winnipeg*, John BREWSTER, *University of Manitoba*

We consider the construction of foldovers of two-level fractional factorial split-plot designs. The foldover technique is a useful follow-up strategy when the objective is to de-alias effects of interest after the initial fractional factorial split-plot design has been run. Moreover, if the runs of the initial and foldover fractional factorial split-plot designs cannot be conducted under homogeneous conditions, the resulting combined design is a blocked fractional factorial split-plot design. The impact of blocking on the selection of optimal foldover plans is considered. Using various optimality criteria, we provide a catalog of optimal foldover plans for initial minimum aberration fractional factorial split-plot designs consisting of 16 and 32 runs.

Nous considérons la construction de repliements de plans d'expérience fractionnels factoriels à deux niveaux à parcelles subdivisées. La technique de repliement est une stratégie de relance utile lorsque l'objectif est de départir les effets d'intérêt une fois le plan initial fractionnel factoriel à subdivision de parcelles exécuté. De plus, si l'exécution des plans initiaux et de repliement fractionnels factoriels à subdivision des parcelles ne peuvent être effectués sous des conditions homogènes, le plan combiné résultant est un plan fractionnel factoriel à subdivision des parcelles par blocs. L'impact de cette division par blocs sur la sélection de plans de repliement optimaux est considéré. En utilisant divers critères d'optimalité, nous fournissons un catalogue de plans de repliements optimaux pour plans initiaux fractionnels factoriels à subdivision de parcelles à minimum d'aberration et comportant entre 16 ou 32 exécutions.

[MS-214]

Sequential Experiment Design for Contour Estimation from Complex Computer Codes.

Plan d'expérience séquentiel pour l'estimation d'une courbe de niveau à partir d'un programme informatique complexe

Pritam RANJAN & Derek BINGHAM, *Simon Fraser University*, George MICHAILIDIS, *University of Michigan*

In many engineering applications, one is interested in identifying the values of the inputs in computer experiments that leads to a response above a pre-specified threshold or, in a region of interest. In this talk we introduce statistical methodology that identifies the desired contours (iso-surfaces) in the input space. Stochastic model is used to approximate the response surface which is followed by a sequential approach for the selection of trials which is directed towards improvement of the estimation of the contours. Finally, extraction of the contours from the estimated surface will be outlined. Application also includes multiple contour estimation and highest posterior density region.

Dans plusieurs applications en ingénierie, on est intéressé à identifier les valeurs d'entrées dans les expériences informatiques qui mène à une réponse au-dessus d'un seuil pré-spécifié ou, dans une région d'intérêt. Dans cette présentation, nous discuterons de la méthodologie statistique qui identifie les courbes de niveaux désirées (iso-surfaces) dans l'espace des entrées. Un modèle stochastique est employé pour approximer la surface de réponse qui est suivie d'une approche séquentielle pour le choix des essais qui sont orientés vers l'amélioration de l'estimation des courbes de niveaux. En conclusion, l'extraction des courbes de niveaux sera décrite. L'application inclut également l'estimation d'une courbe de niveau multiple et la région de densité à posteriori la plus élevée.

Session 12B Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SSC2036**Latent Variable and Multilevel Modelling Using Survey Data****Variables latentes et modélisation multiniveaux pour des données d'enquêtes**

[MS-215]

Pseudo Maximum Likelihood Estimation of Multilevel and Latent Variable Models**Estimation par pseudo-maximum de vraisemblance de modèles à niveaux multiples et à variables latentes**Sophia RABE-HESKETH, *University of California, Berkeley*, Anders SKRONDAL, *London School of Economics*

A common approach to analyzing complex surveys that involve multistage sampling, unequal sampling probabilities, and stratification is pseudo maximum likelihood estimation. This approach is available for multilevel, item response, latent class and structural equation models in the Stata program *gllamm*. Simulation results will be presented to show that, for multilevel logistic models, inclusion of weights at level 1 can lead to bias not only in the variance components but also in the regression coefficients. For latent variable models, weights at level 1 (the item level) are rarely needed so that the issue of bias does not arise.

Une approche courante pour analyser des études complexes impliquant de l'échantillonnage à plusieurs niveaux, des probabilités d'inclusion inégales et de la stratification est l'estimation par pseudo-maximum de vraisemblance. Cette approche est disponible pour les modèles à plusieurs niveaux, de réponse dans les items, de classe latente et d'équations structurelles dans le programme Stata *gllamm*. Des résultats de simulation sont présentés afin de montrer que, pour les modèles logistiques à plusieurs niveaux, l'inclusion de poids au premier niveau peut entraîner un biais non seulement dans les composantes de variance mais aussi dans les coefficients de régression. Pour les modèles à variable latente, les poids au premier niveau (le niveau des items) sont rarement requis ; le problème du biais ne s'y présente donc pas.

[MS-216]

Modeling Change in Longitudinal Studies With Individual Trajectories: Parametric and Nonparametric Multilevel Models**Méthodes d'analyse du changement fondées sur la trajectoire de développement individuelle : modèles multiniveaux paramétriques et non paramétriques**Eric LACOURSE & Véronique DUPÉRÉ, *Université de Montréal*

Multilevel models encompass a variety of modern longitudinal analytic approaches sharing some common features. The purpose of this presentation is to describe and illustrate, in a general and non-technical way, the similarities and differences between two types of multilevel models relying on parametric and non-parametric techniques for estimating population heterogeneity. Fully parametric models assume a normal distribution of the "latent" variable whereas non-parametric models rely on a discrete distribution for approximating an unspecified, possibly continuous distribution of the random effects. Each model is fit for answering different kinds of research questions.

Les modèles multiniveaux regroupent une variété de méthodes modernes d'analyse du changement. La présentation qui sera faite décrit et illustre, de façon générale, les similitudes et différences entre deux types de modèles multiniveaux s'appuyant sur des méthodes paramétriques et non paramétriques d'estimation de l'hétérogénéité d'une population. Les modèles complètement paramétriques présupposent que la variable « latente » suive une distribution normale, alors que les modèles non paramétriques s'appuient sur une distribution discrète pour approximer la distribution non spécifiée et possiblement continue des effets aléatoires. Chacun des modèles a été conçu pour répondre à des questions différentes.

[MS-217]

Fitting Multi-level Models to Survey Data**Modélisation de données d'enquête complexes à l'aide de modèles à multi-niveaux**Milorad KOVACEVIC & Rong HUANG, *Statistics Canada*

The importance of consistent estimation of variance components for the estimation of the fixed parameters of a multi-level model fit to complex survey data is examined. Beginning with a brief review of available methods, the discussion concentrates on conditions under which many of traditional methods for variance components estimation when applied to survey data may lead to biased estimation of the fixed effects parameters. Results of a small simulation study based on the Canadian Workplace and Employee Survey are presented.

L'importance de la convergence des estimateurs des composantes de la variance pour l'estimation des paramètres fixes du modèle d'ajustement multi-niveaux appliquée à des données d'enquête complexe est examinée. Commencé par une brève revue des méthodes disponibles, la discussion se concentre sur les conditions sous lesquelles beaucoup de méthodes d'estimation des composantes de variance donnent lieu à des estimations biaisées des paramètres à effets fixes. Les résultats d'une petite étude de simulation fondée sur l'Enquête canadienne sur le milieu du travail et les employés sont présentés.

Session 12C Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SSC2024**Risk Theory II****Théorie du risque II**

[MS-218]

On the Discounted Penalty Function in the Renewal Risk Model with Arbitrary Interclaim Times**Sur la fonction de pénalité escomptée dans le modèle de risque de renouvellement avec des temps d'inter-réclamation arbitraires**Gordon WILLMOT, *University of Waterloo*

The defective renewal equation satisfied by the Gerber-Shiu discounted penalty function in the renewal risk model with arbitrary interclaim times is analyzed. The ladder height distribution is shown to be a mixture of residual lifetime claim severity distributions, resulting in an invariance property satisfied by many claim size distributions. When claims are exponentially distributed, a simple result follows when the penalty function only involves the deficit, and the Laplace transform of the density of the surplus prior to ruin is obtained. Finally, an expression for the moments of the discounted deficit is obtained for arbitrary claim sizes.

L'équation défectueuse de renouvellement satisfaite par la fonction de pénalité pondérée Gerber-Shiu dans le modèle de risque de renouvellement avec des temps d'inter-réclamation arbitraires est analysée. La distribution en échelle de la hauteur s'avère un mélange des lois de survie résiduelle de sévérité de réclamation, ayant pour résultat une propriété d'invariance satisfaite par beaucoup de distributions de grandeurs de réclamation. Quand les réclamations sont exponentiellement distribuées, un résultat simple suit quand la fonction de pénalité implique seulement le déficit, et la transformation Laplace de la densité de l'excédent avant la ruine est obtenue. En conclusion, une expression pour les moments du déficit escompté est obtenue pour des tailles de réclamation arbitraires.

[MS-219]**The Gerber-Shiu Function for Risk Models Based on Lévy Processes****La fonction de Gerber-Shiu pour des modèles de risque basés sur des processus de Lévy**José GARRIDO, *Concordia University*, Manuel MORALES, *Université de Montréal*

Dufresne et al. (1991) introduced a general risk model defined as the limit of compound Poisson processes. Such a model is either a compound Poisson process itself or a process with an infinite number of small jumps. They work with the classical and the perturbed risk models and hint that their results can be extended to gamma and inverse Gaussian risk processes.

The construction of Dufresne et al. (1991) is based on a non-negative, non-increasing function that governs the jumps of the process. This function, it turns out, is the tail of the Lévy measure of the process. We extend their work to a generalized risk model driven by an increasing Lévy process, that is when the aggregate claims process is a subordinator. Embedded in this wide family of risk models we find the gamma, inverse Gaussian and generalized inverse Gaussian processes.

Dufresne et al. (1991) présentent un modèle général du risque défini comme la limite de processus de Poisson composés. Leur modèle est soit lui-même un processus de Poisson composé ou un processus avec un nombre infini de petits sauts. Ils travaillent avec le modèle de risque classique ou le modèle perturbé et font la conjecture que leurs résultats peuvent s'étendre au processus de risque gamma et inverse gaussien.

La construction de Dufresne et al. (1991) est basée sur une fonction non négative, non croissante qui gouverne les sauts du processus. Cette fonction est en réalité l'aile de la mesure de Lévy du processus. Nous étendons leur travaux à un modèle de risque généralisé issu d'un processus de Lévy croissant, c'est à dire que les réclamations totales forment un processus qui est un subordonateur. Dans cette ample famille de modèles de risque nous retrouvons les processus gamma, inverse Gaussien et inverse Gaussien généralisé.

[MS-220]**The Compound Poisson Risk Model with Multiple Thresholds****Le modèle de risque composé de Poisson à seuils multiples**Kristina SENDOVA, *University of Western Ontario*, X. Sheldon LIN, *University of Toronto*

In this paper we consider a multi-threshold compound Poisson risk model. We further solve a general integro-differential equation with potential applications to ruin theory problems. Consequently, an important particular case is discussed. Namely, the Gerber-Shiu discounted penalty function under the compound Poisson model with multiple thresholds. Finally, examples are considered to illustrate the applicability of the main result.

Dans cet article, nous considérons un modèle de risque composé de Poisson à seuils multiples. Nous résolvons une équation intégro-différentielle générale ayant des applications possibles en théorie de la ruine. Nous discutons également d'un cas particulier important, soit la fonction de pénalité escomptée de Gerber-Shiu sous le modèle composé de Poisson avec seuils multiples. Finalement, nous présentons des exemples illustrant l'applicabilité du résultat principal.

Session 12D Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SH3345**Geometry, Topology and Statistics
Géométrie, topologie et statistique****[MS-221]****Reconstruction of 3D Scenes and Projective Shape Analysis****Reconstruction de scènes en trois dimensions et analyse projective des formes**VICTOR PATRANGENARU & Samanmalee SUGATHADASA, *Texas Tech University, USA*

A classical problem in computer vision is the reconstruction of a 3D scene in absence of obstructions, from multiple views obtained by a noncalibrated camera. We show that since the reconstruction is unique up to a projective transformation, the actual object recovered is the projective shape of the scene, which can be regarded as a point on a product of a finite number of copies of projective spaces. Many problems in pattern recognition from images, can be thus answered using inference in multivariate directional data analysis.

Un problème classique en vision artificielle est la reconstruction d'une scène tridimensionnelle en absence des obstructions, à partir de plusieurs images obtenues par une caméra non calibrée. Nous montrons que puisque la scène est unique jusqu'à une transformation projective, l'objectif est de reconstruire la forme projective de la scène, qui peut être regardée comme point dans un produit d'un nombre fini des copies de l'espace projectif. Plusieurs problèmes en reconnaissance des formes à partir des images, peuvent ainsi être résolus en utilisant l'inférence dans l'analyse des données directionnelles multivariées.

[MS-222]**Exploratory Data Analysis for Human Motion Data****Analyse exploratoire du mouvement humain**Theodore CHANG, *University of Virginia*

This talk will discuss data that is collected by placing sensors on a subject's body as he walks a treadmill. In this case we will consider exploratory data analysis for the knee joint. Considering the upper and lower legs as rigid bodies, and ignoring the translational component, the motion of one relative to the other can be described by a path in the rotation group in three dimensions $SO(3)$. Our approaches to exploratory data analysis will be deeply rooted in the geometry of $SO(3)$.

Cet exposé porte sur des données obtenues en plaçant des capteurs sur le corps d'un sujet alors qu'il marche sur un tapis roulant. Dans ce cas, nous considérons une analyse exploratoire des données de l'articulation du genou. En considérant la jambe supérieure et inférieure comme des corps rigides, et en ignorant la composante translationnelle, le mouvement de l'une par rapport à l'autre peut être décrit comme un trajet dans le groupe des rotations en trois dimensions $SO(3)$. Nos approches à l'analyse exploratoire des données sont profondément enracinées dans la géométrie de $SO(3)$.

[MS-223]**A Statistical Approach to Persistent Homology****Une approche statistique sur l'homologie persistante**Peter BUBENIK, *Cleveland State University*, Peter KIM, *University of Guelph*

We attempt to lay down a statistical foundation to some recent advances in computational topology. We are interested in problems, such as that of manifold learning, in which one is interested in non-linear properties of potentially high-dimensional finite samples. The technique of persistent homol-

Nous essayons d'établir la justification statistique des progrès récents dans la topologie informatique. Nous sommes intéressés par des problèmes, tels que ceux reliés à l'apprentissage, où l'intérêt se porte sur les propriétés non linéaires des échantillons finis potentiellement de dimension d'ordre élevé. La technique de l'homologie persistante peut être

ogy can be used to identify topological features of data sets. We assume that the data has been sampled from a probability density, for which we define persistent homology in a way that is compatible with the persistent homology of finite samples. We estimate the underlying density from a given sample, and use the persistent homology of the density to determine the topological features of the data. In the present study we focus on parametric cases.

utilisée pour identifier les caractéristiques topologiques des jeux de données. Nous supposons que les données ont été prélevées d'une densité de probabilité, pour laquelle nous définissons l'homologie persistante d'une manière dont elle est compatible avec l'homologie persistante des échantillons finis. Nous estimons la densité sous-jacente d'un échantillon donné, et utilisons l'homologie persistante de la densité pour déterminer les caractéristiques topologiques des données. Dans la présente étude nous nous concentrons sur des cas paramétriques.

Session 12E Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SSC3018

Applications of Statistics II Applications de la statistique II

[MS-224]

Exploratory Analysis of Sphere Packings By Shape-based Statistics

Analyse exploratoire d'empilements de sphères avec des statistiques basées sur les formes

Jeffrey PICKA, *University of New Brunswick*

Given a random point pattern, many aspects of its structure are summarized by its Delaunay triangulation. The shape of each triangular cell can be mapped into a point on a sector of the unit disc. Analysis of the distribution of these points on the sector can be used to characterize realizations of different algorithms, and to identify unusual features of the realization. Comparisons are given between this approach and other statistics which can be derived from the triangulation.

La triangulation de Delaunay permet de présenter plusieurs aspects importants de la structure d'un nuage de points aléatoires. La forme de chaque cellule triangulaire peut être associée à un point dans un secteur du disque unitaire. Une analyse de la distribution de ces points dans le secteur peut être utilisée pour caractériser les réalisations de différents algorithmes, et pour identifier les éléments inhabituels de ces réalisations. Des comparaisons entre cette approche et d'autres statistiques pouvant être obtenues à partir de la triangulation sont présentées.

[MS-225]

A New Statistical Model for Random Unit Vectors

Un nouveau modèle statistique pour des vecteurs unitaires aléatoires

Karim OUALKACHA & Louis-Paul RIVEST, *Université Laval*

In this presentation, we propose a new symmetrical model for unit vectors, as being an alternative to the exponential model of Bingham. This model has interesting properties, for instance the normalizing constant and the moments have explicit and simple forms and the simulation of data following the proposed distribution is easy. In dimension 4, the new distribution can be used to model quaternions associated to 3×3 rotation matrices for describing the posture of experimental subjects. We present the moment estimators for the parameters of this model and we investigate their sampling distribu-

Dans cette présentation, nous proposons un nouveau modèle symétrique pour des vecteurs unitaires, comme étant une alternative du modèle exponentiel de Bingham. Quelques propriétés de ce modèle sont, entre autres que, la constante de normalisation et les moments ont des formes explicites et simples et que la simulation des données provenant de la distribution proposée est facile. Ce modèle s'applique en particulier, en dimension 4, pour modéliser des quaternions associés à des matrices de rotation 3×3 qui décrivent la posture de sujets expérimentaux. Nous présentons les estimateurs des paramètres du modèle par la méthode des moments, ainsi que leurs propriétés asymptotiques. Une analyse de données

tions. A data analysis of the right hand posture of subjects handling a drill is also presented to illustrate the proposed methodology.

sur la posture de la main droite de sujets manipulant une perceuse est aussi présentée afin d'illustrer la méthodologie proposée.

[MS-226]

Modified Likelihood Ratio Test for Homogeneity in a Mixture of von Mises Distributions

Test du rapport de vraisemblance modifié pour l'homogénéité en présence d'un mélange de distributions von Mises

Pengfei LI, *University of Waterloo*, Yuejiao FU, *York University*, Jiahua CHEN, *University of Waterloo*

Recent studies suggest that a proportion of circadian genes have systematically different phase/peak expressions in two different tissues. The statistical problem of interest is to test homogeneity in a mixture of von Mises distributions. The modified likelihood ratio test (MLRT) is proposed to test whether the circular sample is drawn from a single von Mises distribution or a mixture of two von Mises distributions with common unknown concentration parameter. The MLRT statistic is shown to have a simple chi-squared null limiting distribution. The result is extended to general parametric kernels. For illustration, the MLRT is applied to a real example.

De récentes études suggèrent qu'une proportion des gènes circadiens possède des expressions phase/pic totalement différentes dans deux tissus différents. Le problème statistique d'intérêt est de tester l'homogénéité dans un mélange de distributions von Mises. Le test du rapport de vraisemblance modifié (TRVM) est proposé pour tester si un échantillon circulaire est tiré d'une seule distribution von Mises ou bien d'un mélange de deux distributions von Mises avec un paramètre de concentration commun mais inconnu. Nous montrons que la loi limite nulle de la statistique du TRVM est une loi du khi-deux. Le résultat est ensuite étendu aux noyaux paramétriques généraux. À des fins d'illustration, le TRVM est appliqué à un exemple réel.

[MS-227]

Optimal Tolerance Regions for Future Regression Vector and Residual Sum of Squares of Multiple Regression Model with Multivariate Spherically Contoured Errors

Régions de tolérance optimales pour un futur vecteur de régression et sommes de carrés résiduels sous un modèle de régression multiple avec erreurs multivariées aux contours sphériques

Shahjahan KHAN, *University of Southern Queensland, Australia*

This paper considers multiple regression model with multivariate spherically symmetric errors to determine optimal beta-expectation tolerance regions for the future regression vector (FRV) and future residual sum of squares (FRSS) by using the prediction distributions of some appropriate functions of future responses. The prediction distribution of the FRV, conditional on the observed responses, is multivariate Student- t distribution. Similarly, the prediction distribution of the FRSS is a beta distribution. Optimal beta-expectation tolerance regions for the FRV and FRSS are based on the F -distribution and beta distribution respectively. The results in this paper generalise those for the multiple regression model with normal and Student- t errors.

Cet article considère le modèle de régression multiple avec des erreurs multivariées sphériquement symétriques pour déterminer les régions de tolérance optimales de la bêta-espérance du vecteur de régression future (VRF) et de la somme de carrés résiduels futurs (SCRF) en utilisant les distributions de prévision de fonctions de réponse future appropriées. La distribution de prévision du VRF, conditionnel aux réponses observées, est une distribution de Student multivariée. De façon similaire, la distribution de prévision de la SCRf est la distribution bêta. Les régions de tolérance optimales de la bêta-espérance pour le VRF et la SCRf sont basées sur la distribution F et la distribution bêta, respectivement. Les résultats de cet article généralisent ceux obtenus pour un modèle de régression multiple avec erreurs normales et suivant une loi de Student.

[MS-228]

Theories in Linear Regression Model
Théorie en régression linéaire

Anwar SAGR, *University of Al-Jabal Al-Gharbi*

Theory 1. This theory for analysis unexplained variance (SSE)

$$\begin{aligned} \text{SSE} &= \text{SSE}_y + \text{SSE}_x \\ \text{SST} &= \text{SSR} + \text{SSE}_y + \text{SSE}_x \end{aligned}$$

the (SST) is total variance, (SSR) is explained variance, (SSE_y) is sum of square error in the model, (SSE_x) is sum of square of measurement error in the independent variable.

Theory 2. If $(X_i, Y_i), \dots, (X_n, Y_n)$ is random sample and Y is dependent variable and X is independent variable and the relation between X and Y is linear relation, then: the average of X and Y it will equal the average of X and Y after reflection on the line of regression, so the reflection does not change the averages in the regression linear model.

Théorie 1. Cette théorie concerne l'analyse de variance non expliquée (SSE)

$$\begin{aligned} \text{SSE} &= \text{SSE}_y + \text{SSE}_x \\ \text{SST} &= \text{SSR} + \text{SSE}_y + \text{SSE}_x \end{aligned}$$

(SST) est la variance totale, (SSR) est la variance expliquée par le modèle, (SSE_y) est la somme des carrés des erreurs dans le modèle, (SSE_x) est la somme des carrés des erreurs de mesure dans la variable indépendante.

Théorie 2. Si $(X_i, Y_i), \dots, (X_n, Y_n)$ est un échantillon aléatoire, Y est la variable dépendante et X la variable indépendante, et la relation entre X et Y est linéaire, alors : la moyenne de X et Y est égale à la moyenne de X et Y après une réflexion par rapport à la droite de régression, c'est-à-dire que la réflexion ne change pas les moyennes dans le modèle de régression linéaire.

Session 12F Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SSC3026

Multivariate Time Series
Séries chronologiques multivariées

[MS-229]

Empirical Bayes Analysis for a Hierarchical Negative Binomial Generalized Linear Model
Analyse bayésienne empirique sous un modèle binomial négatif hiérarchique linéaire généralisé

Mohanad AL-KHASAWNEH & S. Ejaz AHMED, *University of Windsor*

In this talk, we present a parametric hierarchical empirical Bayes approach for estimating the mean proportion response from generalized linear regression models (GLiM's) based on a negative binomial distribution. Following Piegorsch and Casella (1996, JABES, 231–247), GLiM's are extended via parametric families of link functions that include the log-link function as a special case. We estimate the hyper-parameters from the beta-negative binomial marginal distribution. A simulation study is carried out to investigate the performance of the proposed estimators in a practical setting. An empirical example is also given to illustrate the usefulness of this inferential procedure in practice.

Dans cet exposé, nous présentons une approche paramétrique hiérarchique bayésienne empirique pour estimer la réponse moyenne de modèles de régression linéaires généralisés (MliGs) basés sur une distribution binomiale négative. Suivant Piegorsch et Casella (1996, JABES, 231–247), les MliGs sont généralisés par le biais de familles paramétriques de fonctions de liens incluant la fonction de lien logarithmique comme cas particulier. Nous estimons les hyperparamètres de la distribution marginale bêta-binomiale négative. Une étude de simulation est effectuée afin d'évaluer la performance des estimateurs proposés dans un contexte pratique. Un exemple empirique est aussi donné afin d'illustrer l'utilité de cette procédure inférentielle en pratique.

[MS-230]**Fixed Point Clustering in Robust Mixture Discriminant Analysis
Regroupement point fixe dans l'analyse discriminante de mélange robuste**Shaheena BASHIR & Edward CARTER, *University of Guelph*

In classification problems, in the presence of masking outliers, the approach of Bashir and Carter (2005) is useful when the data are mixtures of multivariate normal. In this paper an improvement in the robust mixture discriminant analysis of Bashir and Carter (2005) is made, using fixed point clustering of Hennig (2003) instead of k -means clustering to get initial estimates. This approach is shown through simulation studies to be more effective. The efficiency of the resulting model is compared with the standard mixture discriminant analysis approach using probability of misclassification criterion. This comparison showed better results producing smaller errors of classification.

Dans les problèmes de classification, en présence d'observations aberrantes ayant un effet masquant, l'approche de Bashir et Carter (2005) est utile quand les données suivent un mélange de distributions normales multivariées. Dans cette présentation, nous présentons une version améliorée de l'analyse discriminante robuste de Bashir et de Carter (2005) quand les données suivent un mélange de distributions normales multivariées. Nous utilisons le regroupement de point fixe de Hennig (2003) au lieu de l'analyse à classification automatique à k moyennes pour obtenir des estimations initiales. Nos simulations ont démontré que cette approche est plus efficace. L'efficacité du modèle résultant est comparée à l'approche standard d'analyse discriminante en utilisant comme critère la probabilité de classification erronée. Cette comparaison a montré que de meilleurs résultats sont obtenus avec la nouvelle méthode et qu'il en résulte moins d'erreurs de classification.

[MS-231]**On Robust Forecasting in Dynamic Vector Time Series Models
Sur les prévisions robustes dans les modèles de séries chronologiques vectoriels dynamiques**Christian GAGNÉ & Pierre DUCHESNE, *Université de Montréal*

Robust estimation and prediction in multivariate autoregressive models with exogenous variables (VARX) are considered. The conditional least squares estimators (CLS) are known to be non robust when outliers occur. To obtain robust estimators, the method introduced in Duchesne (2005) and Bou Hamad and Duchesne (2005) is generalized for VARX models. The asymptotic distribution of the new estimators is studied and from this is obtained in particular the asymptotic covariance matrix of the robust estimators. Classical conditional prediction intervals normally rely on estimators such as the usual non robust CLS estimators. In the presence of outliers, such as additive outliers, these classical predictions can be severely biased. More generally, the occurrence of outliers may invalidate the usual conditional prediction intervals. Consequently, the new robust methodology is used to develop robust conditional prediction intervals which take into account parameter estimation uncertainty. In a simulation study, we investigate the finite sample properties of the robust prediction in-

L'estimation et la prévision robustes sont considérées dans les modèles autorégressifs avec variables exogènes multivariées (VARX). Les estimateurs par moindres carrés conditionnels (CLS) ne sont pas robustes quand des valeurs aberrantes surviennent. Afin d'obtenir des estimateurs robustes, la méthode introduite par Duchesne (2005) et Bou Hamad et Duchesne (2005) est généralisée pour les modèles VARX. La distribution asymptotique des nouveaux estimateurs est étudiée et en particulier est obtenue de ce résultat la matrice de covariance asymptotique des estimateurs robustes. Les intervalles de prévision conditionnels classiques reposent normalement sur des estimateurs comme les estimateurs CLS non robustes. En présence de valeurs aberrantes, telles les valeurs aberrantes additives, ces prévisions classiques risquent d'être sévèrement biaisées. Plus généralement, la présence de valeurs aberrantes peut invalider les intervalles de prévision conditionnels usuels. Conséquemment, la nouvelle méthodologie robuste est utilisée afin de développer des intervalles de prévision conditionnels robustes qui prennent en compte l'incertitude entourant l'estimation des paramètres. Dans une étude de simulation, nous étudions les propriétés en échantillons finis des intervalles de prévision robustes en

tervals under several scenarios for the occurrence of the outliers, and the new intervals are compared to non-robust intervals based on classical CLS estimators.

considérant plusieurs scénarios pour l'apparition des valeurs aberrantes, et les nouveaux intervalles sont comparés aux intervalles non robustes basés sur les estimateurs CLS classiques.

[MS-232]

Robust Optimal Tests for Non-causality in Multivariate Time Series

Tests optimaux et robustes de non-causalité pour des séries chronologiques multivariées

Abdessamad SAIDI & Roch ROY, *Université de Montréal*

We are deriving optimal rank-based tests for Granger non-causality in multivariate time series. Assuming that the global process admits a joint vector autoregressive (VAR) representation with elliptic innovation density, both no feedback and one direction non-causality hypotheses are tested. These tests are based on multivariate residual ranks and signs and are shown to be asymptotically distribution free under elliptically symmetric innovation densities and invariant with respect to some affine transformations. Local powers and asymptotic relative efficiencies are derived. Simulations confirm theoretical results about the size and power of the proposed rank-based methods, and establish their good robustness properties.

Nous développons des tests optimaux de non-causalité au sens de Granger pour des séries chronologiques multivariées. En supposant que le processus global admet une représentation autorégressive vectorielle à symétrie elliptique, l'hypothèse de non-causalité unidirectionnelle et de rétroaction sont testées. Ces tests sont basés sur les signes et les rangs multivariés des résidus. Ils sont valides sous la classe des distributions à symétrie elliptique et sont invariants sous l'action de certains groupes de transformations affines. Les puissances locales et les efficacités asymptotiques relatives sont calculées. Les simulations confirment les résultats théoriques sur le niveau et la puissance des tests proposés, et montrent la robustesse de ces derniers.

[MS-233]

Empirical Study of Portmanteau Test Statistics Designed for Multivariate Time Series

Étude empirique de tests de type portemanteau pour séries chronologiques multivariées

Jennifer POULIN & Pierre DUCHESNE, *Université de Montréal*

In the analysis of multivariate time series, Duchesne and Roy (2004) introduced portmanteau test statistics based on kernel-based multivariate spectral density estimators. The test statistics converge in distribution toward a standard normal distribution under the null hypothesis of no correlation in the error term. However, empirical studies reveal that the exact levels of the spectral-based test statistics appear to be relatively far from the nominal empirical levels when they are calculated using the asymptotic critical values. In this talk, we discuss a modification of the Duchesne and Roy (2004) test statistics using the Chen and Deo (2003) transformation. The purpose of this presentation is to compare by simulations the modified test statistics and the original spectral tests.

Dans l'analyse des séries chronologiques multivariées, Duchesne et Roy (JMVA 2004) ont introduit des statistiques de test de type portemanteau qui reposent sur des estimateurs de la densité spectrale multivariée. Les statistiques de test convergent en distribution vers une loi normale centrée réduite sous l'hypothèse nulle d'absence de corrélation dans le terme d'erreur. Cependant, des études empiriques révèlent que les niveaux exacts des tests spectraux sont relativement loin des niveaux nominaux quand ils sont calculés utilisant les quantiles asymptotiques. Dans cette présentation nous discutons une modification des statistiques de test de Duchesne et Roy (2004) utilisation la transformation introduite dans Chen et Deo (2003). L'objectif de cette présentation est de comparer par simulations les statistiques de test modifiées et les statistiques de test spectrales originales.

Session 12G Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SSC3006**Observational Studies: Methods and Applications****Méthodes et applications pour les études observationnelles****[MS-234]****Approximating Relative Risk Using Logistic Regression Revisited****Approximation du risque relatif en utilisant la régression logistique : un second aperçu**RUSSELL STEELE, *McGill University*, IAN SHRIER, *Jewish General Hospital*

We will revisit a common approximation (most commonly attributed to Zhang and Yu) for relative risk that uses the odds ratio estimates from logistic regression. Other authors have shown that this particular approximation can perform poorly although they have not shown where it will perform poorly. We will show that a related, but different form of the approximation allows for better insight into when the rule of thumb will perform well and where it will not. The alternate form of the approximation uses basic principles of causal inference and Bayesian model selection.

Nous allons revoir une approximation connue (le plus généralement attribuée à Zhang et à Yu) pour le risque relatif qui emploie l'estimation de rapport de cotes de la régression logistique. D'autres auteurs ont démontré que la performance de cette approximation particulière peut être faible, cependant ils n'ont pas spécifié où se trouve cette faiblesse. Nous montrerons qu'une forme connexe mais différente de l'approximation permet une meilleure compréhension de quand la méthode empirique performe bien ou non. La forme alternative de l'approximation utilise des principes de base d'inférence causale et de la sélection de modèles bayésiens.

[MS-235]**Bayesian Propensity Score Analysis for Observational Data****Analyse bayésienne de score de propension pour des données d'observation**LAWRENCE McCANDLESS & PAUL GUSTAFSON, *University of British Columbia*, PETER C AUSTIN, *University of Toronto*

We consider Bayesian regression adjustment for the propensity score in observational studies with a binary exposure and binary outcome. The method uses two logistic regression models with the propensity score as a latent variable. The first regression models the relationship between the outcome, exposure and propensity score, while the second regression models the relationship between the propensity score and measured confounders. Markov chain Monte Carlo is used for posterior simulation. We contrast the performance of Bayesian propensity score analysis with frequentist methods when applied to real and simulated data.

Nous considérons l'ajustement d'un modèle de régression bayésien pour les scores de propension dans les études observationnelles avec une exposition binaire et des résultats binaires. La méthode emploie deux modèles de régression logistique, le score de propension servant comme variable latente. La première régression modélise le rapport entre les résultats, l'exposition et le score de propension, alors que la deuxième modélise le rapport entre le score de propension et les facteurs parasites mesurés. La méthode de Monte Carlo par chaîne de Markov est employée pour les simulations a posteriori. Nous comparons la performance de l'analyse bayésienne de scores de propension à des méthodes fréquentistes à travers des exemples d'application basés sur des données réelles et simulées.

[MS-236]**Estimation of Parameters in Logistic Regression Models with Multiplicative Measurement Error
Estimation des paramètres d'un modèle de régression logistique en présence d'erreur de mesure multiplicative**Hyang Mi KIM, Yutaka YASUI & Igor BURSTYN, *University of Alberta*

We consider both sample- and group-based strategies in multiplicative exposure and logistic response models. With sample-based strategy, a calibration method fails to adjust measurement error when the error variance is large. The geometric mean has been used as a surrogate estimator in the group-based strategy. However, simulations show that using the geometric mean leads to either considerable over- or under- estimation of the association parameter in logistic models. We adopt Bayesian normal mixture modeling to reduce bias due to multiplicative measurement error in logistic regression models for both strategies and demonstrate its superiority.

Nous considérons des stratégies basées sur l'échantillon et le groupe pour des modèles d'exposition multiplicative et de réponse logistique. Avec une stratégie basée sur l'échantillon, les méthodes de calibration ne peuvent ajuster l'erreur de mesure lorsque la variance de l'erreur est grande. Cependant, des simulations montrent qu'utiliser la moyenne géométrique mène à une sous- ou sur-estimation considérable du paramètre d'association des modèles logistiques. Nous adoptons la modélisation bayésienne de mélange normal pour réduire le biais causé par l'erreur de mesure multiplicative dans les modèles de régression logistique sous les deux stratégies et démontrons sa supériorité.

[MS-237]**Modelling Relative Risk of Cancer Incidence in a Cohort Study
Modélisation du risque relatif de l'incidence du cancer au moyen d'une étude de cohortes**Maria LORENZI, *BC Cancer Agency*

In occupational cohort studies, it is of prime interest to determine how cancer risk changes with increasing exposure to agents in the workplace. Using data from a cohort of aluminum smelter workers, we present a series of models relating the risk of bladder cancer, lung cancer, and non-Hodgkin's lymphoma to cumulative exposure to coal tar pitch volatiles, measured as a time-dependent, continuous variable. Additive and multiplicative relative risk models are considered.

Dans les études des cohortes professionnelles, il est de grande importance de déterminer comment le risque du cancer change avec l'exposition aux divers agents dans le milieu du travail. Utilisant les données d'une cohorte d'ouvriers dans une fonderie d'aluminium, nous présentons une série de modèles qui explorent la relation entre le risque du cancer de la vessie, le cancer des poumons et du lymphome non hodgkinien au niveau d'exposition aux vapeurs et aérosols du brai du goudron de l'houille, mesuré comme une variable continue qui change en fonction du temps. Des modèles de risque relatif additifs et multiplicatifs sont considérés.

[MS-238]**A Spatiotemporal Investigation of Echinococcus Multilocularis Infections Among Red Foxes in Lower Saxony, 1991 - 2005
Une investigation spatiotemporelle des cas d'infections de Echinococcus Multilocularis chez les renards rouges de la Basse-Saxe, de 1991 à 2005**Olaf BERKE, *University of Guelph and University of Veterinary Medicine, Hannover, Germany*

Echinococcosis is a dangerous and an emerging zoonosis caused by the small fox tapeworm (*Echinococcus multilocularis*). Data about the infection status of about 8.400 foxes were sampled

L'échinococcose est une zoonose émergente dangereuse causée par le cestode du renard (*Echinococcus multilocularis*). Des données d'infection d'environ 8 400 renards ont été recueillies dans 43 régions de la province allemande de Basse-

from 43 regions in the German province of Lower Saxony over three investigation periods 1991–1994, 1994–1997 and 2003–2005. The median regional period prevalence increased from 5.8% to 11.4% and 16.9%. But what happened at a finer spatial scale? Spatial data patterns were stable over time and include a disease cluster and aspatial trend. Kriging is used in combination with cluster detection by the scan statistic to produce exploratory relative risk maps.

Saxe au cours de trois périodes : 1991-1994, 1994-1997 et 2003-2005. La prévalence régionale médiane a passé de 5,8% à 11,4% à 16,9%. Mais que s'est-il passé à une échelle spatiale plus fine ? Les données spatiales étaient stables dans le temps et incluaient une grappe de maladie et une tendance spatiale. Le krigeage est utilisé en combinaison à la détection de grappes par une statistique de balayage afin de produire des cartes exploratoires du risque relatif.

[MS-239]

Statistical Inference in Predator-prey System Modeled under Intrinsic Period Property

L'inférence statistique dans le système prédateur-proie modélisé sous la propriété intrinsèque de période

Severien NKURUNZIZA & S. Ejaz AHMED, *University of Windsor*

In this communication, we consider the inference problem for the ratio of two interaction parameters, so-called conversion efficiency of the Lotka-Volterra ordinary differential equations system (ODEs). The stochastic model under consideration views the actual population sizes as random perturbations of the solutions to this ODEs. Namely, we assume that the perturbations follow correlated Ornstein-Uhlenbeck processes. An interesting and unusual feature is that no assumption is made that the random variables are independent. In particular, the statistical testing procedure is presented. In this context, we establish the uniformly most powerful unbiased test for the conversion efficiency parameter. The asymptotic properties of the proposed test are derived. A simulation study is conducted to assess the performance of the suggested testing procedure in a practical setting.

Dans cette communication, nous considérons le problème d'inférence pour le ratio de deux paramètres d'interaction, soit la conversion d'efficacité du système d'équations différentielles ordinaires (EDO) de Lotka-Volterra. Le modèle stochastique à l'étude traite les tailles réelles de population en tant que perturbations aléatoires des solutions à ces EDOs. Nous supposons que les perturbations suivent des processus Ornstein-Uhlenbeck corrélés. Une caractéristique intéressante et peu commune est que nous ne présumons pas l'indépendance des variables aléatoires. En particulier, la procédure du test statistique est présentée. Dans ce contexte, nous établissons le test non biaisé uniformément le plus puissant pour la conversion du paramètre d'efficacité. Les propriétés asymptotiques du test proposé sont dérivées. Une étude par simulation a été faite afin d'évaluer la performance de la procédure du test suggéré.

Session 12H Mercredi 31 mai • Wednesday, May 31, 15:00–16:30 SSC3014

Modelling and Analysis in Sample Surveys

Modélisation et analyse pour l'échantillonnage

[MS-240]

Heavy-tails and the Central Limit Theorem in Business Surveys

Ailes lourdes et le théorème limite centrale dans les enquêtes auprès des entreprises

Jack LOTHIAN, *Statistics Canada*

Heavy-tailed distributions are ubiquitous in business surveys. While the functional form of these distributions vary, most of the 'best-fit' are positive-definite right-skewed heavy-tailed sub-exponential

Les distributions à aile lourde sont omniprésentes dans les enquêtes auprès des entreprises. Quoique la forme fonctionnelle de ces distributions change, la plupart des « meilleurs ajustements » sont des distributions définies-positives, asy-

distributions with finite first moments. This paper explores the empirical size distributions of Canadian corporations over a 25-year period. A log-normal distribution was an excellent fit for the size distribution, a log-Laplace for the growth rates, and a log-Subbotin for the ratio of 2 size measures. The Central Limit Theorem applied to the size measure but possibly not to the growth rates and it probably does not hold for some ratios of economic variables.

métriques à droite, à ailess lourdes, sous-exponentielles avec des premiers moments finis. Cet article explore les distributions de la taille empirique des sociétés canadiennes sur une période de 25 ans. Une distribution log-normale était un excellent ajustement pour la distribution de la taille, une log-Laplace pour les taux de croissance, et une log-Subbotin pour le ratio de 2 mesures de taille. Le théorème central limite s'applique pour la mesure de taille mais probablement pas aux taux de croissance et il n'est probablement pas valide pour quelques ratios de variables économiques.

[MS-241]

A Survey of Evangelical Mennonite Mission Conference of Churches: A Look at Factors that Impact Growth
Une étude de la conférence des églises de la mission évangélique mennonite et des facteurs en influençant la croissance

Eddy REMPEL, *McMaster University*, Lehana THABANE, *Department of Clinical Epidemiology and Biostatistics, McMaster University, Centre for Evaluation of Medicines, St Joseph's Healthcare*

The Evangelical Mennonite Mission Conference (EMMC) is a group of 40 churches (30 in Canada and a few in the United States, Mexico, and Belize). The EMMC wishes to examine their growth as measured by increases in attendance, membership, and number of churches.

La Conférence de la mission évangélique mennonite (CMEM) est un groupe de 40 églises (30 au Canada et quelques-unes aux États-Unis, au Mexique et au Belize). La CMEM souhaite examiner sa croissance telle que mesurée par l'augmentation de l'assistance, du nombre de membres et du nombre d'églises.

This study identifies factors that contribute to increased EMMC growth. It also determines whether these differ by geographic region, church size, and growth status.

Cette étude identifie des facteurs contribuant à la croissance de la CMEM. Elle détermine aussi si ces facteurs diffèrent selon la région géographique, la taille de l'église et l'état de croissance.

We combined EMMC annual church records and national census demographics of the church neighbourhood with surveys of all senior pastors and church leaders identified by the EMMC. This presentation examines factors which are associated with increased growth.

Nous avons combiné les dossiers annuels des églises de la CMEM et les données de recensement des quartiers des églises avec des sondages des prêtres senior et des dirigeants d'églises identifiés par la CMEM. Cette présentation examine les facteurs associés à une croissance accrue.

[MS-242]

Evaluating Likert Items Using Inter-Item θ
Estimation des items de Likert en utilisant l'inter-item θ

Floyd RUDMIN, *University of Tromsø, Norway*

Likert scales are evaluated by inter-item correlations, i.e., Cronbach alpha, to index how much the items elicit similar responses. Problems are to simultaneously identify deviant items and deviant respondents, to do this with small samples, using correlations that are inflated by outliers. Also, items may cluster into sub-scales, and respondents may cluster into sub-classes. The present study will articulate and demonstrate the use of polar coordi-

Les échelles de Likert sont évaluées par les corrélations entre items, i.e., alpha de Cronbach, pour répertorier combien les individus répondent d'une façon semblable. Les problèmes consistent à identifier simultanément les caractéristiques à l'écart et les répondants à l'écart, dans de petits échantillons, en utilisant les corrélations qui sont affectées par des données aberrantes. Également, les items peuvent être groupés dans des sous-échelles, et les répondants peuvent être groupés dans des sous-classes. La présente étude articulera

nates (r, θ) to compute inter-item θ for each pairing of items for each individual respondent to evaluate Likert items, in comparison to the Cronbach method, to item response theory, and to confirmatory factor analysis.

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.

et démontrera l'utilisation des coordonnées polaires (r, θ) pour calculer l'inter-item θ pour chaque paire d'items, pour chaque répondant afin d'évaluer les échelles de Likert, par rapport à la méthode de Cronbach, à la théorie de réponse d'items, et à l'analyse factorielle confirmatoire.

Author Index • Index des auteurs

- ABARIN, Taraneh 124
 ABDOUS, Belkacem 114
 ABRAHAM, Bovas 153
 ADEWALE, Adeniyi 140
 ADLER, Robert 68
 AHMED, S. Ejaz 78, 79, 126, 160, 171, 176
 AHN, Soohan 162
 ALBERS, Willem 149
 AL-KHALIDI, Abdul Sattar 98
 AL-KHASAWNEH, Mohanad 171
 ALVO, Mayer 99, 152
 AN, Lihua 160
 ANDERSON, Clive 86
 ANDREWS, David 101
 ASHLOCK, Daniel 106
 ASIMIT, Alexandru 125
 ASIMIT, Jennifer 76
 ATENAFU, Eshetu 141
 AUSTIN, Peter C 174
 AXELROD, David 133

 BADESCU, Alex 154
 BADESCU, Andrei 162
 BALAN, Raluca 157
 BANKS, David 132
 BASHIR, Shaheena 172
 BEAUDOIN, David 115, 127
 BEAUMONT, Jean-François 130
 BELLHOUSE, David R. 136
 BENGHABRIT, Youssef 155
 BERKE, Olaf 175
 BERRY, Donald 148
 BEYENE, Joseph 106
 BHAMANI, Azim 143
 BINGHAM, Derek 80, 86, 164
 BOCCI, Cynthia 130
 BOLSTAD, Ben 63
 BORDENAVE, Charles 108
 BOUZAACHANE, Khadija 155
 BOYETT, James M. 122
 BRAUN, W. John 76, 77
 BRAUN, Willard 64

 BREED, Greg 117
 BRESLOW, Norman 89
 BRETTSCHEIDER, Julia 63
 BREWSTER, John 164
 BRILLINGER, David 145
 BRIOLLAIS, Laurent 158, 159
 BRION, Vladislav 99
 BUBENIK, Peter 168
 BULL, Shelley 119
 BURKE, Murray 85
 BURNHAM, Alison 91
 BURSTYN, Igor 175

 CABILIO, Paul 122
 CAMPBELL, David 94
 CAO, Jiguo 94, 141
 CARABARIN, Alberto 99
 CARRIÈRE, Keumhee Chough 81
 CARRILLO-GARCIA, Ivan 102
 CARTER, Edward 172
 CHANG, Theodore 168
 CHAPMAN, Judy-Anne 133
 CHEN, Edward J. 143
 CHEN, Gemai 150
 CHEN, Jiahua 153, 170
 CHENG, Cheng 100
 CHENG, Ji 113
 CHENOURI, Shojaeddin 115
 CHIPMAN, Hugh 79, 96, 139
 CHOI, Yun Hee 103, 159
 CHRISTENS-BARRY, William 133
 COLLIN, François 63
 COLUBI, Ana 153
 COOK, Richard 66, 141
 CORAK, Miles 151
 CORBETT, Bradley 70
 COREY, Paul 121
 CRAIU, Radu 119
 CROY, B. Anne 143
 CSÖRGŐ, Miklós 84

 DABROWSKI, André 61

- DAVIES, Katherine 76
 DAVISON, Matt 72
 DAWSON, Donald 109
 DEAN, Charmaine 89
 DEBLOIS, Sylvie 130
 DEERING, Kathleen 82
 DE LEON, Alexander 65
 DENG, Dianliang 98
 DOLOVICH, Lisa 128
 DOWD, Michael 73, 116
 DREKIC, Steve 135
 DUCHESNE, Pierre 172, 173
 DUCHESNE, Thierry 114, 115
 DUPÉRÉ, Véronique 165
 DUPUIS, Debbie 61, 124

 EINMAHL, Uwe 92
 ESCOBAR, Michael 95
 EVANS, Mike 95
 EVRA, Rose-Carline 130

 FAN, Zhaozhi 97
 FELDMAN-STEWART, Deb 127
 FERLAND, René 156
 FERNANDO, Rajulton 69
 FLEMMING, Joanna 117
 FOGEL, Gary 105
 FRIENDLY, Michael 135
 FU, Yuejiao 133, 170

 GAGNÉ, Christian 172
 GAO, Ruochu 138
 GARRIDO, José 167
 GASTWIRTH, Joseph 63
 GEL, Yulia R. 63
 GENEST, Christian 115, 136
 GERLACH, Richard 155
 GHORI, Razi 78
 GHOSH, Debashis 133
 GHOSH, Sunita 148
 GIBBS, Alison 74
 GIBSON, Edward 128
 GIL, Maria 153
 GINÉ, Evarist 119, 120
 GLOBERSON, Amir 131
 GOMBAY, Edit 85
 GONZALEZ-RODRIGUEZ, Gil 153
 GOTTARDO, Raphael 118
 GRAMACY, Robert 71
 GREENWOOD, Celia 131

 GUPTA, Mayetri 118
 GUSTAFSON, Paul 67, 174
 GUTTORP, Peter 74

 HA, Hyung-Tae 110
 HABIBI, Reza 94
 HAIDAR, Bassam 61
 HAQ, Mahbulbul 128
 HARRINGTON, Justin 138
 HARSHMAN, Richard 62, 64
 HARTI, Mostaf 155
 HASAN, M. Tariqul 158
 HAWKINS, Douglas 149
 HE, Wenqing 160
 HE, Xin 123
 HEINEMEYER, Andreas 86
 HILL, Harvey 75
 HOOKER, Giles 94, 141
 HOQUE, Zahirul 155
 HORROCKS, Julie 143
 HOSSAIN, Shakhwat 79
 HOWARD, Allan 75
 HU, Ze-Chun 134
 HUANG, Mei Ling 92
 HUANG, Rong 166
 HUI, Wai Lok Wallace 63
 HUNSBERGER, Paul 131
 HUSSEIN, Abdulkadir 78
 HYNDMAN, Cody 156

 IBRAHIM, Joseph 118
 IVANOFF, B. Gail 99
 IVANOFF, Gail 157

 JAMES, Mike 117
 JANGA, Sarath 105
 JANKOWSKI, Hanna 157
 JENNISON, Christopher 84
 JIANG, Depeng 62
 JIANG, Tiefeng 93
 JIANG, Zhichang 65
 JIN, Zi 75
 JOCELYN, Wisner 129
 JOHNSON, Edward 88
 JONES, Bruce 125
 JONSEN, Ian 117
 JORDAN, Adriana 157
 JØRGENSEN, Bent 137

 KACZOROWSKI, Janusz 128

- KALBFLEISCH, Jack 137
 KALLENBERG, Wilbert C.M. 149
 KANG, Sohee 95
 KARUNANAYAKE, Chandima 147
 KENNEDY, Marc 86
 KERMAN, Ron 92
 KHAN, Abdul 111
 KHAN, Shahjahan 125, 170
 KHAN, Sohail 126
 KIBRIA, B. M. Golam 77
 KIM, Hyang Mi 175
 KIM, Peter 168
 KING, Maxwell 80
 KLOTZ, Laurence 112
 KORU-SENGUL, Tulay 104
 KOULIS, Theodoro 72
 KOVACEVIC, Milorad 102, 166
 KULIK, Rafal 120
 KULPERGER, Reg 77, 154
 KUSTRA, Rafal 161

 LACOURSE, Eric 165
 LAKHAL-CHAIEB, Lajmi 127
 LANDRIAULT, David 163
 LAROCQUE, Denis 109
 LARSEN, Michael 87
 LAVALLÉE, Pierre 102
 LAWLESS, Jerry 66
 LAZAR, Emese 154
 LEE, Chu-In 111
 LEE, Herbert K. H. 71
 LEITH, Chris 145
 LI, Pengfei 170
 LIANG, Yuanyuan 81
 LIN, Chunfang 80
 LIN, Chunfang (Devon) 103
 LIN, X. Sheldon 167
 LIN, Xihong 67
 LINKLETTER, Crystal 80
 LITTNAN, Charles 145
 LIU, Lin 111
 LIX, Lisa 82
 LOAIZA, Edilberto 151
 LOBLAW, Andrew 112
 LOCKHART, Richard 80
 LOEPPKY, Jason 86
 LOMAS, Mark 86
 LORENZI, Maria 175
 LOTHIAN, Jack 176

 LOUGHIN, Tom 80
 LU, Wilson 80
 LU, Xuewen 97
 LUNDY, Margaret E. 62

 MA, Zhi- Ming 134
 MACDONALD, Peter 138
 MACH, Lenka 129
 MANTEL, Harold 143
 MARCEL, Yves 61
 MARTSYNYUK, Yuliya 158
 MASON, David 120
 MASON, David M. 119
 MATTHEWS, E. David 103
 McALPINE, Rob 88
 McBRIDE, Mary 114
 McCANDLESS, Lawrence 174
 McCARTHY, Lisa 128
 McCULLOCH, Charles 90
 McDONALD, David 108
 McDUFFIE, Helen 147
 McLEOD, Ian 155, 160
 McLEOD, Rob 164
 McRAE, Kenneth 112
 MELNIKOV, Alexander 134
 MIAO, Weiwen 63
 MICHAILIDIS, George 164
 MIHRAM, Danielle 101
 MIHRAM, G. Arthur 101
 MILLER, Naomi 133
 MOLENBERGHS, Geert 148
 MOLNAR, Danielle Sirriani 128
 MORALES, Manuel 167
 MORENO-HAGELSIEB, Gabriel 105
 MOSESOVA, Sofia 96
 MOSHONOV, Hadas 95
 MUDRYK, Walter 144
 MULAYATH VARIYATH, Asokan 153
 MURDOCH, Duncan 61
 MURRAY, Scott 150
 MYERS, Ransom 117

 NAGY, Bela 86
 NATHOO, Farouk 89
 NEAL, Radford 131
 NEATON, James 83
 NEUHAUS, John 90
 NEVALAINEN, Jaakko 109
 NEWCOMBE, Pat 144
 NEWLANDS, Nathaniel 75

- NG, Peggy 74
 NKURUNZIZA, Severien 176

 O'BRIEN, Susan 128
 O'HAGAN, Anthony 86
 OJA, Hannu 109
 O'ROURKE, Keith 107
 OUALKACHA, Karim 169
 OULD HAYE, Mohamedou 154

 PAHWA, Punam 147, 148
 PARK, DO-Hwan 123
 PARKER, Peter 163
 PATERSON, Andrew 119
 PATRANGENARU, Victor 168
 PELLETIER, Éric 129
 PENG, Jianan 111
 PENG, Yingwei 98
 PEPLER, Debra 62
 PERCACCIO, Anthony 92
 PETERSON, Ivars 66
 PHIPPS, Sean 122
 PICKA, Jeffrey 169
 PLANTE, Jean-François 100
 PLATT, H.W. (Bud) 112
 PODDER, Mohua 142
 POND, Gregory 81
 PONG, Hoki 143
 POULIN, Jennifer 173
 POWLES, A.C. Peter 128
 PRITCHARD, Zachary 129
 PROUTIERE, Alexandre 108
 PROVOST, Serge 110
 PULLENAYEGUM, Eleanor 83
 PUNTANEN, Simo 136
 PURSEY, Stuart 131

 QIAN, Jin 133
 QU, Pingping 118

 RABE-HESKETH, Sophia 165
 RAI, Shesh N. 122
 RAMASWAMI, Vaidyanathan 162
 RAMSAY, James O. 94, 141
 RANJAN, Pritam 80, 164
 RAO, J.N.K. 125, 146
 REESE, Shane 86
 REESOR, Mark 72
 REMPEL, Eddy 177
 REN, Qunshu 146

 RIVEST, Louis-Paul 127, 169
 ROBERTS, Georgia 146
 ROMANIUK, Yulia 134
 ROSA, Guilherme 133
 ROSALSKY, Andrew 93
 ROWEIS, Sam 131
 ROY, Roch 173
 ROY, Valérie 114
 RUBIN-BLEUER, Susana 147
 RUDMIN, Floyd 177

 SAGR, Anwar 171
 SAHA, Krishna 80
 SAIDI, Abdelnasser 147
 SAIDI, Abdessamad 173
 SALEH, A K Md Ehsanes 125
 SALEH, A. K. Md. E. 154
 SAWKA, Anna 107
 SCHAUBEL, Douglas 137
 SCHMULAND, Byron 108
 SENDOVA, Kristina 167
 SERBAN, Daniel 85
 SHELLEY, Bull 159
 SHEN, Hui 139
 SHRIER, Ian 174
 SIMPSON, William 76
 SINGH, Gurbakhshash 114
 SITTER, Randy 103
 SIU, Lillian 81
 SKRONDAL, Anders 165
 SMIEJA, Marek 113
 SMITH, Bruce 145
 SMITH, Stephen 116
 SONG, Peter 137
 SPEED, Terence 63
 SREBOTNJAK, Tanja 73
 SRIVASTAVA, Deo Kumar 122
 STALLARD, Jim 121
 STANFORD, David 162
 STEELE, Russell 174
 STEWART, Brent 145
 STROHSCHIEIN, Lisa 70
 STUKEL, Diane 147
 STUKEL, Thérèse 90
 STYAN, George 136
 SU, Wanhua 96
 SUGATHADASA, Samanmalee 168
 SUN, Jack 113
 SUN, Jianguo (Tony) 123

- SUN, Jie 97
 SUN, Lei 119, 159
 SUN, Liuquan 123
 SUN, Shuying 131
 SUN, Wei 134
 SUN, Ye 110
 SUTCLIFFE, Paul 61
 SZYSZKOWICZ, Barbara 84

 TAKAHARA, Glen 145
 TARDIF, Charles 129
 TAYLOR, Jeremy 67, 98
 TAYLOR, Jonathan 68, 69
 TEKPETEY, Francis 143
 THABANE, Lehana 113, 128, 177
 THÉBERGE, François 152
 THOMPSON, Mary 65
 TITHECOTT, Al 88
 TOLUSSO, David 141
 TONG, Xingwei 123
 TORABI, Mahmoud 125
 TSAI, Yu-Ling 61
 TU, Dongsheng 127

 VANDAL, Nathalie 100
 VAN DEN HEUVEL, Marianne 143
 VAN HOREBEEK, Johan 139
 VICTORIA-FESER, Maria-Pia 124
 VIERTL, Reinhard 152
 VINING, Geoff 115
 VOLODIN, Andrei 93

 WANG, Miao 127
 WANG, Qiying 84
 WANG, Xikui 81
 WANG, Xu Sunny 79
 WANG, Zilin 65
 WATIER, François 156
 WELCH, William J. 79, 86, 139, 142
 WELLNER, Jon 89
 WHITEHEAD, Tyson 72
 WHITLOCK, Richard 113
 WIENS, Douglas P. 140
 WILLAN, Andrew 83, 106
 WILLIAMS, Kathryn 61
 WILLMOT, Gordon 166
 WONG, Augustine 110
 WOODWARD, Ian 86
 WOOLFORD, Douglas 77, 162
 WORSLEY, Keith 68, 69

 WU, Changbao 102
 WU, LongYang 159
 WU, Longyang 159

 XIE, Hansheng 144
 XU, Liqin 113
 XU, Xiaojian 102

 YAN, Mi 128
 YASUI, Yutaka 175
 YE, Wen 67
 YE, Yining 137
 YI, Grace 68
 YI, Y. Yi 103
 YI, Yanqing 81
 YING, Shih-Wa 114
 YOUNG, S. Stanley 142
 YU, Binbing 98
 YUAN, Xiaobin 100
 YUAN, Yan 133

 ZAGDANSKI, Adam 161
 ZAMAR, Ruben H. 142
 ZAMFIRESCU, Ingrid-Mona 157
 ZAREPOUR, Mahmoud 94
 ZEESHAN, Zeeshan 131
 ZHANG, Lijun 122
 ZHANG, Lingyun 150
 ZHANG, Liying 112
 ZHANG, Ying 155
 ZHANG, Yu 140
 ZHAO, Yang 82
 ZHENG, Zheng 75
 ZHOU, Lutong 64
 ZHOU, Qian 137, 142
 ZHU, Mu 96, 142
 ZHU, Yongtao 65
 ZIDEK, James V. 100

Author Index (per session)

Index des auteurs (par session)

ABARIN, Taraneh	07G	BOYETT, James M.	07F
ABDOUS, Belkacem	06H	BRAUN, W. John	02G
ABRAHAM, Bovas	10F	BRAUN, Willard	0
ADEWALE, Adeniyi	08F	BREED, Greg	07B
ADLER, Robert	02B	BRESLOW, Norman	04C
AHMED, S. Ejaz	02H, 07H, 10H, 12F, 12G	BRETTSCHNEIDER, Julia	0
AHN, Soohan	11B	BREWSTER, John	12A
ALBERS, Willem	10B	BRILLINGER, David	09B
AL-KHALIDI, Abdul Sattar	04G	BRIOLLAIS, Laurent	10H
AL-KHASAWNEH, Mohanad	12F	BRION, Vladislav	04H
ALVO, Mayer	04H, 10D	BROWNE, Ryan	10E
AN, Lihua	10H	BUBENIK, Peter	12D
ANDERSON, Clive	03D	BULL, Shelley	07C
ANDREWS, David	05A	BURKE, Murray	03C
ASHLOCK, Daniel	06C	BURNHAM, Alison	04D
ASIMIT, Alexandru	07G	BURSTYN, Igor	12G
ASIMIT, Jennifer	02G		
ATENAFU, Eshetu	08G	CABILIO, Paul	07E
AUSTIN, Peter C	12G	CAMPBELL, David	04F
AXELROD, David	08B	CAO, Jiguo	04F, 08G
		CARABARIN, Alberto	04H
BADESCU, Alex	10F	CARRIÈRE, Keumhee Chough	02I
BADESCU, Andrei	11B	CARRILLO-GARCIA, Ivan	06A
BALAN, Raluca	10G	CARTER, Edward	12F
BANKS, David	08A	CHANG, Theodore	12D
BASHIR, Shaheena	12F	CHAPMAN, Judy-Anne	08B
BEAUDOIN, David	06H, 07H	CHEN, Edward J.	08H
BEAUMONT, Jean-François	07I	CHEN, Gemai	10B
BELLHOUSE, David R.	08D	CHEN, Jiahua	10F, 12E
BENGHABRIT, Youssef	10F	CHENG, Cheng	04H
BERKE, Olaf	12G	CHENG, Ji	06G
BERRY, Donald	10A	CHENOURI, Shojaeddin	06H
BEYENE, Joseph	06D	CHIPMAN, Hugh	02H, 04F, 08F
BHAMANI, Azim	08G	CHOI, Yun Hee	06B, 10H
BINGHAM, Derek	02H, 03D, 12A	CHRISTENS-BARRY, William	08B
BOCCI, Cynthia	07I	CHU, Rachel	02F
BOLSTAD, Ben	0	CHUNG, Sharon	10E
BORDENAVE, Charles	06E	COLLIN, François	0
BOUZAACHANE, Khadija	10F	COLUBI, Ana	10D

- COOK, Richard 02A, 08G
 CORAK, Miles 10C
 CORBETT, Bradley 02C
 COREY, Paul 07E
 CRAIU, Radu 07C
 CROY, B. Anne 08G
 CSÖRGŐ, Miklós 03C
 CUI, Yin 10E

 DABROWSKI, André 0
 DAVIES, Katherine 02G
 DAVIS, Matthew 10I
 DAVISON, Matt 02D
 DAWSON, Donald 06E
 DEAN, Charmaine 04B
 DEBLOIS, Sylvie 07I
 DEERING, Kathleen 02I
 DE LEON, Alexander 0
 DENG, Dianliang 04G
 DENG, MingXia 10E
 DOLOVICH, Lisa 07H
 DOWD, Michael 02E, 07B
 DREKIC, Steve 08C
 DU, Meng 10E
 DUCHESNE, Pierre 12F
 DUCHESNE, Thierry 06H
 DUPÉRÉ, Véronique 12B
 DUPUIS, Debbie 0, 07G

 EINMAHL, Uwe 04E
 ESCOBAR, Michael 04F
 EVANS, Mike 04F
 EVRA, Rose-Carline 07I

 FAN, Shuyu 02F
 FAN, Zhaozhi 04G
 FANG, Lin 10E
 FAROOQI, Ahmad A. 10E
 FELDMAN-STEWART, Deb 07H
 FERLAND, René 10G
 FERNANDES, Kimberly 02F
 FERNANDO, Rajulton 02C
 FLEMMING, Joanna 07B
 FOGEL, Gary 06C
 FRIENDLY, Michael 08D
 FU, Yuejiao 08B, 12E

 GAGNÉ, Christian 12F
 GAO, Ruochu 08F
 GARRIDO, José 12C

 GASTWIRTH, Joseph 0
 GEL, Yulia R. 0
 GENEST, Christian 06H, 08D
 GERLACH, Richard 10F
 GHORI, Razi 02H
 GHOSH, Debashis 08B
 GHOSH, Sunita 09C
 GIBBS, Alison 02F
 GIBSON, Edward 07H
 GIL, Maria 10D
 GINÉ, Evarist 07D
 GLOBERSON, Amir 08A
 GOMBAY, Edit 03C
 GONZALEZ-RODRIGUEZ, Gil 10D
 GOTTARDO, Raphael 07C
 GRAMACY, Robert 02D
 GREENWOOD, Celia 08A
 GRIEVE, Charlotte 10E
 GUO, Pengfui 02F
 GUPTA, Mayetri 07C
 GUSTAFSON, Paul 02A, 12G
 GUTTORP, Peter 02E

 HA, Hyung-Tae 06F
 HABIBI, Reza 04F
 HAIDAR, Bassam 0
 HAN, David 10E
 HAQ, Mahbubul 07H
 HARRINGTON, Justin 08F
 HARSHMAN, Richard 0
 HARTI, Mostaf 10F
 HASAN, M. Tariqul 10H
 HAWKINS, Douglas 10B
 HE, Dan 10I
 HE, Lei 10I
 HE, Wenqing 10H
 HE, Xin 07F
 HEINEMEYER, Andreas 03D
 HILL, Harvey 02G
 HOOKER, Giles 04F, 08G
 HOQUE, Zahirul 10F
 HORROCKS, Julie 08G
 HOSSAIN, Shahadut 02F
 HOSSAIN, Shakhwat 02H
 HOWARD, Allan 02G
 HU, Ze-Chun 08C
 HUANG, Mei Ling 04E
 HUANG, Rong 12B
 HUI, Wai Lok Wallace 0

- HUNSBERGER, Paul07I
 HUSSEIN, Abdulkadir 02H
 HYNDMAN, Cody 10G
- IBRAHIM, Joseph07C
 IVANOFF, B. Gail 04H
 IVANOFF, Gail 10G
- JAMES, Mike 07B
 JANGA, Sarath 06C
 JANKOWSKI, Hanna 10G
 JENNISON, Christopher 03B
 JIANG, Depeng 0
 JIANG, Tiefeng 04E
 JIANG, Zhichang 0
 JIN, Zi 02G, 10E
 JOCELYN, Wisner 07I
 JOHNSON, Edward 04B
 JONES, Bruce 07G
 JONSEN, Ian 07B
 JORDAN, Adriana 10G
 JØRGENSEN, Bent 08E
- KACZOROWSKI, Janusz 07H
 KALBFLEISCH, Jack 08E
 KALLENBERG, Wilbert C.M. 10B
 KANG, Sohee 04F
 KARUNANAYAKE, Chandima 09C
 KENNEDY, Marc 03D
 KERMAN, Ron 04E
 KHAN, Abdul 06F
 KHAN, Shahedul Ahsan 02F
 KHAN, Shahjahan 07G, 12E
 KHAN, Sohail 07H
 KIBRIA, B. M. Golam 02G
 KIM, Hyang Mi 12G
 KIM, Peter 12D
 KING, Maxwell 02H
 KLOTZ, Laurence 06G
 KONG, Linglong 10E
 KORU-SENGUL, Tulay 06B
 KOULIS, Theodoro 02D
 KOVACEVIC, Milorad 06A, 12B
 KULIK, Rafal 07D
 KULPERGER, Reg 02G, 10F
 KUSTRA, Rafal 10H
- LACOURSE, Eric 12B
 LAKHAL-CHAIEB, Lajmi 07H
 LANDRIAULT, David 11B
- LAROCQUE, Denis 06F
 LARSEN, Michael 04A
 LAVALLÉE, Pierre 06A
 LAWLESS, Jerry 02A
 LAZAR, Emese 10F
 LEE, Chu-In 06F
 LEE, Herbert K. H. 02D
 LEITH, Chris 09B
 LI, Longhai 10E
 LI, Pengfei 12E
 LI, Song 10E
 LI, Wang 10E
 LIANG, Yuanyuan 02I, 10E
 LIN, Chunfang 02H
 LIN, Chunfang (Devon) 06A
 LIN, X. Sheldon 12C
 LIN, Xihong 02A
 LINKLETTER, Crystal 02H
 LITNAN, Charles 09B
 LIU, Lin 06F
 LIX, Lisa 02I
 LOAIZA, Edilberto 10C
 LOBLAW, Andrew 06G
 LOCKHART, Richard 02H
 LOEPPKY, Jason 03D
 LOMAS, Mark 03D
 LORENZI, Maria 12G
 LOTHIAN, Jack 12H
 LOUGHIN, Tom 02H
 LU, Wilson 02H
 LU, Xuewen 04G
 LUNDY, Margaret E. 0
- MA, Tianshu 02F
 MA, Zhi- Ming 08C
 MACDONALD, Peter 08F
 MACH, Lenka 07I
 MADDALENA, Kristen 10E
 MANTEL, Harold 08H
 MARCEL, Yves 0
 MARTSYNYUK, Yuliya 10G
 MASON, David 07D
 MASON, David M. 07D
 MATTHEWS, E. David 06B
 McALPINE, Rob 04B
 McBRIDE, Mary 06G
 McCANDLESS, Lawrence 12G
 McCARTHY, Lisa 07H
 McCULLOCH, Charles 04C

- McDONALD, David 06E
 McDUFFIE, Helen 09C
 McLEOD, Ian 10F, 10H
 McLEOD, Rob 12A
 McRAE, Kenneth 06G
 MELNIKOV, Alexander 08C
 MIAO, Weiwen 0
 MICHAILIDIS, George 12A
 MIHRAM, Danielle 04H
 MIHRAM, G. Arthur 04H
 MILLER, Naomi 08B
 MOLENBERGHS, Geert 09C
 MOLNAR, Danielle Sirriani 07H
 MORALES, Manuel 12C
 MORENO-HAGELSIEB, Gabriel 06C
 MOSESOVA, Sofia 04F
 MOSHONOV, Hadas 04F
 MUDRYK, Walter 08H
 MULAYATH VARIYATH, Asokan 10F
 MURDOCH, Duncan 0
 MURRAY, Scott 10C
 MYERS, Ransom 07B
- NAGY, Bela 03D
 NATHOO, Farouk 04B
 NEAL, Radford 08A
 NEATON, James 03B
 NEUHAUS, John 04C
 NEVALAINEN, Jaakko 06F
 NEWCOMBE, Pat 08H
 NEWLANDS, Nathaniel 02G
 NG, Peggy 02F
 NKURUNZIZA, Severien 12G
- O'BRIEN, Susan 07H
 O'HAGAN, Anthony 03D
 OJA, Hannu 06F
 O'ROURKE, Keith 06D
 OUALKACHA, Karim 12E
 OULD HAYE, Mohamedou 10F
- PAHWA, Punam 09C
 PARK, Do-Hwan 07F
 PARKER, Peter 12A
 PATERSON, Andrew 07C
 PATRANGENARU, Victor 12D
 PELLETIER, Éric 07I
 PENG, Jianan 06F
 PENG, Yingwei 04G
 PEPLER, Debra 0
- PERCACCIO, Anthony 04D
 PETERSON, Ivars 01B
 PHIPPS, Sean 07F
 PICKA, Jeffrey 12E
 PLANTE, Jean-François 04H
 PLATT, H.W. (Bud) 06G
 PODDER, Mohua 08G
 POND, Gregory 02I
 PONG, Hoki 08G
 POULIN, Jennifer 12F
 POWLES, A.C. Peter 07H
 PRITCHARD, Zachary 07I
 PROUTIERE, Alexandre 06E
 PROVOST, Serge 06F
 PU, Qiang 02F
 PULLENAYEGUM, Eleanor 02I
 PUNTANEN, Simo 08D
 PURSEY, Stuart 07I
- QIAN, Jin 08B
 QIAO, Wenjie 02F
 QIU, Chao 10I
 QU, Pingping 07C
- RABE-HESKETH, Sophia 12B
 RAFFA, Jesse 02F
 RAHEEM, Enayetur 02F
 RAI, Shesh N. 07F
 RAMASWAMI, Vaidyanathan 11B
 RAMSAY, James O. 04F, 08G
 RANJAN, Pritam 02H, 12A
 RAO, J.N.K. 07G, 09C
 REESE, Shane 03D
 REESOR, Mark 02D
 REMPEL, Eddy 12H
 REN, Qunshu 09C
 RIVEST, Louis-Paul 07H, 12E
 ROBERTS, Georgia 09C
 ROMANIUK, Yulia 08C
 ROSA, Guilherme 08B
 ROSALSKY, Andrew 04E
 ROWEIS, Sam 08A
 ROY, Roch 12F
 ROY, Valérie 06H
 RUBIN-BLEUER, Susana 09C
 RUDMIN, Floyd 12H
- SAGR, Anwar 12E
 SAHA, Krishna 02H
 SAIDI, Abdelnasser 09C

- SAIDI, Abdessamad 12F
SALEH, A K Md Ehsanes 07G
SALEH, A. K. Md. E. 10F
SAWKA, Anna 06D
SCHAUBEL, Douglas 08E
SCHMULAND, Byron 06E
SENDOVA, Kristina 12C
SERBAN, Daniel 03C
SHELLEY, Bull 10H
SHEN, Hui 08F
SHRIER, Ian 12G
SIMPSON, William 02G
SINGH, Gurbakhshash 06G
SITTER, Randy 06A
SIU, Lillian 02I
SKRONDAL, Anders 12B
SMIEJA, Marek 06G
SMITH, Bruce 09B
SMITH, Stephen 07B
SONG, Peter 08E
SONG, Yunna 02F
SPEED, Terence 0
SREBOTNJAK, Tanja 02E
SRIVASTAVA, Deo Kumar 07F
STALLARD, Jim 07E
STANFORD, David 11B
STEELE, Russell 12G
STEWART, Brent 09B
STROHSCHN, Lisa 02C
STUKEL, Diane 09C
STUKEL, Thérèse 04C
STYAN, George 08D
SU, Wanhua 04F
SUGATHADASA, Samanmalee 12D
SUN, Jack 06G
SUN, Jianguo (Tony) 07F
SUN, Jie 04G
SUN, Lei 07C, 10H
SUN, Liuquan 07F
SUN, Rena (Jie) 10E
SUN, Shuying 08A
SUN, Tao 02F
SUN, Wei 08C
SUN, Ye 06F
SUTCLIFFE, Paul 0
SZYSZKOWICZ, Barbara 03C
TAKAHARA, Glen 09B
TARDIF, Charles 07I
TAYLOR, Jeremy 02A, 04G
TAYLOR, Jonathan 02B
TEKPETEY, Francis 08G
THABANE, Lehana 06G, 07H, 12H
THÉBERGE, François 10D
THOMPSON, Mary 0
TITHECOTT, Al 04B
TOLUSSO, David 08G
TONG, Xingwei 07F
TORABI, Mahmoud 07G
TSAI, Yu-Ling 0
TU, Dongsheng 07H
VANDAL, Nathalie 04H
VAN DEN HEUVEL, Marianne 08G
VAN HOREBEEK, Johan 08F
VICTORIA-FESER, Maria-Pia 07G
VIERTL, Reinhard 10D
Vining, Geoff 07A
VOLODIN, Andrei 04E
WANG, Mengzhe 10E
WANG, Miao 07H
WANG, Qiyang 03C
WANG, Weiling 10I
WANG, Xikui 02I
WANG, Xu Sunny 02H
WANG, Zilin 0
WATIER, François 10G
WELCH, William J. 02H, 03D, 08F, 08G
WELLNER, Jon 04C
WHITEHEAD, Tyson 02D
WHITLOCK, Richard 06G
WIENS, Douglas P. 08F
WILLAN, Andrew 02I, 06D
WILLIAMS, Kathryn 0
WILLMOT, Gordon 12C
WONG, Augustine 06F
WOODWARD, Ian 03D
WOOLFORD, Douglas 02G, 11B
WORSLEY, Keith 02B
WU, Changbao 06A
WU, LongYang 10H
WU, Longyang 10H
XIE, Hansheng 08H
XU, Liqin 06G
XU, Xiaojian 06A
YAN, Mi 07H

YASUI, Yutaka	12G
YE, Wen	02A
YE, Yining	08E
YI, Grace	02A
YI, Y. Yi	06B
YI, Yanqing	02I
YIN, Yaling	10E
YING, Shih-Wa	06G
YOUNG, S. Stanley	08G
YOUNGSON, Erik	10E
YU, Binbing	04G
YUAN, Fei	02F
YUAN, Xiaobin	04H
YUAN, Yan	08B
ZAGDANSKI, Adam	10H
ZAIHRA, Tasneem	10E
ZAMAR, Ruben H.	08G
ZAMFIRESCU, Ingrid-Mona	10G
ZAREPOUR, Mahmoud	04F
ZEESHAN, Zeeshan	07I
ZHANG, Jin	02F
ZHANG, Lijun	07F
ZHANG, Lingyun	10B
ZHANG, Liying	06G
ZHANG, Ying	10F
ZHANG, Yu	08F
ZHAO, Yang	02I
ZHENG, Cynthia (Xin)	10E
ZHENG, Zheng	02G
ZHOU, Lutong	0
ZHOU, Michelle	02F
ZHOU, Qian	08E, 08G
ZHU, Mu	04F, 08G
ZHU, Qiaohao	10E
ZHU, Yongtao	0
ZIDEK, James V.	04H