# CANSSI 2019 Annual Report

## Message from the Director

It has been an exciting eight months since I became Scientific Director of CANSSI. During that time, I have worked to increase my understanding of CANSSI's role in the Canadian statistical and inferential data science community. Of course, I have had to lean on Nancy Reid heavily as I feel my way into the position, and I am grateful for her expert advice and extensive knowledge. A personal milestone is presenting my first CANSSI Annual Report.

### Research

In December, the Board approved awards for CRT Projects 15 ("Modern Techniques for Survey Sampling and Complex Data" directed by Dr. David Haziza at l'Université de Montréal and Dr. Changbo Wu at the University of Waterloo) and 16 ("Addressing Spatial and Computational Issues in Integrated Analysis of Modern Ecological Data" directed by Dr. Laura Cowen at the University of Victoria, Dr. Simon Bonner at Western University, and Dr. Saman Muthukumarana at the University of Manitoba). The high quality of the submitted proposals made narrowing down the selection to two projects exceedingly difficult. Below is the complete list of CRT projects to date:

| Project | | Title |
|---|---|---|
| 16 | 2020-2023 | Addressing Spatial and Computational Issues in Integrated Analysis of Modern Ecological Data |
| 15 | 2020-2023 | Modern Techniques for Survey Sampling and Complex Data |
| 14 | 2019-2022 | Statistical methods for the analysis of genetic data with survival outcomes |
| 13 | 2019-2022 | Contingent Capital and Calibration of Capital Structure Models |
| 12 | 2018-2021 | Towards Sustainable Fisheries: State Space Assessment Models for Complex Fisheries and Biological Data |
| 11 | 2018-2021 | Spatial Modeling of Infectious Diseases: Environment and Health |
| 10 | 2018-2021 | Statistical Methods for Challenging Problems in Public Health Microbiology |
| 9 | 2017-2020 | Statistical Analysis of Large Administrative Health Databases: Emerging Challenges and Strategies |
| 8 | 2016-2019 | Joint Analysis of Neuroimaging Data: High-Dimensional Problems, Spatiotemporal Models and Computation |
| 7 | 2016-2019 | Rare DNA Variants and Human Complex Traits: Improving Analyses of Family Studies by Better Modeling the Dependence Structures |
| 6 | 2015-2018 | Evolving Marked Point Processes with Application to Wildland Fire Regime Modeling |
| 5 | 2015-2018 | Statistical Inference for Complex Surveys with Missing Observations |
| 4 | 2015-2018 | Modern Spectrum Methods in Time Series Analysis: Physical Science, Environmental Science and Computer Modeling |
| 3 | 2014-2017 | Statistical Modeling of the World: Computer and Physical Models in Earth, Ocean, and Atmospheric Sciences |
| 2 | 2014-2017 | Advancements to State-Space Models (SSMs) for Fisheries Science |
| 1 | 2014-2017 | Copula Dependence Modeling: Theory and Applications |

With the selection of projects 15 and 16, we can state in unqualified terms that the Collaborative Research Teams initiative has proved to be a world-leading model for supporting fundamental research in the context of application. The CRT program provides the

infrastructure to seed new interdisciplinary collaborations and train young researchers for future careers. Consider that project 7, over its three-year term, concluded with team members having given 48 talks, published or submitted 23 papers and mentored 15 HQP and project 8 concluded with team members having given 31 posters and talks, published or submitted 6 papers and mentored 8 HQP. Of course, the scientific impact of the research supported in these projects was felt far outside the statistical discipline.

Notable achievements among the active CRT projects include: trainee Kaiqiong Zhao receiving a Fonds de recherche du Québec-Santé (FRQS) PhD scholarship, trainee Li Xing taking a tenure track position at the University of Saskatchewan, trainee Di Shu taking a postdoctoral fellowship at Harvard Medical School, trainee Dongdong Li taking a postdoctoral fellowship at Harvard Medical School and winning a Thomas O. Pyle Fellowship, trainee Xin Liu taking an assistant professor position at Shanghai Finance and Economics University, Team 7 co-leaders co-editing a special issue of the Canadian Journal of Statistics, Team 9 publishing 14 papers in 2019, with 6 in revision while 8 are submitted or in progress, Team 8 co-Lead Celia Greenwood receiving a James McGill Professorship, Team 8 co-Lead Karim Oualkacha receiving a FRQS Research Scholar Junior 2 distinction, Team 12 co-Lead Alexandre Bouchard-Côté being awarded the PIMS-UBC Mathematical Sciences Young Investigator Award, Team 13 co-Lead Richard Cook presenting the 2019 Scandinavian Journal of Statistics Keynote Lecture at the Nordic-Baltic Meeting of the International Biometrics Society and being awarded the Faculty of Mathematics Research Chair at the University of Waterloo.

*After examination of barriers for faculty considering a CRT submission, CANSSI has made an important change to the proposal process. In 2020, the LOI is streamlined, with ingredients focusing on the scientific aspects of the proposed project. This substantially reduces the overhead involved with applying for a CRT project and, we anticipate, will encourage more faculty to submit proposals.*

## Training

The great majority of CANSSI's support goes towards training of highly qualified personnel, providing them with the skills and approaches to meet the challenges of research in the future. Young researchers taking part in CANSSI-supported activities have a strong record of success in finding careers that involve the development and/or the application of research in statistics and inferential data science. I would also like to point out that last year CANSSI supported 11 postdoctoral fellows across Canada – evidence of the growing importance of postdoctoral appointments to both the research landscape in Canada and to providing young researchers with a stepping stone that better prepares them for careers in academia, industry, and government.

*CANSSI is offering an exciting new program for graduate students, namely the CANSSI Graduate Student Exchange Scholarships. This program provides partial support for graduate students enrolled at a Canadian university to visit another Canadian university for an extended period (2 months for MSc and 4 months for PhD). The program is designed to foster acquisition of new knowledge and skill sets, to expose students to new areas of research and application, and to enable co-discovery of new statistics and co-supervision of thesis research. The program also*

*has the potential to encourage long-term research collaborations of the participating co-supervisors.*

## Equity, Diversity and Inclusion (EDI)

Putting it bluntly, the creation of an equitable, diverse and inclusive Canadian statistics and inferential data science research enterprise is necessary for CANSSI to achieve its mission and for the statistical and inferential data science community to respond effectively to local, national and global challenges. CANSSI will make pursuing equity, diversity and inclusion along with the associated issue of trainee success a cornerstone of its programs for the foreseeable future. To this end, we drafted a short CANSSI policy statement that reflects the new policies of NSERC and we intend to constitute a standing CANSSI Committee on EDI to draft a more comprehensive statement along with a plan to achieve CANSSI's goals. Stay tuned – this is a very exciting development.

## Maintaining CANSSI as a community-based operation

To understand how best to guide CANSSI's evolution, I initiated a plan to visit every mathematics and statistics department in Canada as a part of a systematic campaign of engagement with the Canadian statistics and inferential data science community. Since July, I have visited: Acadia University, Carleton University, Concordia University, Dalhousie University, HEC Montréal, McGill University, McMaster University, Western University, Wilfrid Laurier University, York University, and the Universities of British Columbia, Manitoba, Montréal, New Brunswick, Ottawa, Québec at Montréal, Toronto, Waterloo, and Winnipeg. These visits have been extremely valuable for providing a deeper understanding of the environment for research and training of HQP in statistics and inferential data science in Canada. My travels have been so rewarding and productive that I have decided that to make it a regular responsibility of the scientific directors.

## Operations

There have been some exciting developments for CANSSI operations. I am very happy to announce that Professor Nancy Heckman, from the Department of Statistics at the University of British Columbia, became Associate Scientific Director of CANSSI on September 1. She brings a wealth of experience to this role and the addition of an Associate Scientific Director greatly increases the ability of CANSSI to explore new activities and dimensions of operation. Nancy's appointment is partial fulfillment of UBC's role in hosting the national headquarters of CANSSI at Simon Fraser University.

CANSSI has been working hard towards the establishment of Regional Centres. Regional Centres will provide another way for faculty in different parts of Canada to organize their participation in the management of CANSSI operations and will allow CANSSI to more effectively deploy its resources in different parts of Canada. Formally established in 2018, CANSSI Québec held its inauguration on November 21. The event included talks by former Scientific Director Nancy Reid, Deputy Director John Braun, Professor Christian Genest, and me, along with a formal inauguration. Later we held the first Regional Board meeting, in which we discussed ways that CANSSI could support Québec. The formal agreement establishing CANSSI Ontario went into effect with Professor Lisa Strug serving as interim Director for a year. The

University of Toronto is contributing very generously to establish CANSSI Ontario and Lisa immediately announced a number of opportunities for the entire province.

### Thanks

CANSSI is a success due to the many colleagues who dedicate time and effort to building CANSSI to its full potential. The Associate Directors, Scientific Advisory Committee, and Board of Directors are active, hard-working, and sources of excellent and timely advice to CANSSI about research directions and operations.

I must mention the efforts of a few people in particular. It would be impossible to manage CANSSI without the daily (sometimes hourly!) assistance of Associate Scientific Director Nancy Heckman and Deputy Director John Braun. Of course, CANSSI is strongly supported by the Herculean efforts of CANSSI Scientific Coordinator Angela Plagemann and CANSSI Financial Manager Lori Kroeker.

I also want to note that Simon Fraser University has taken on its role as CANSSI's host institution with gusto, assisting in multiple dimensions as we prepare to compete for the next round of NSERC funding for CANSSI.

Finally, I want to thank the people in the Canadian statistical and inferential data science communities who have provided a welcoming environment for me and my family.

## CANSSI Programs

### Overview

CANSSI is Canada's catalyst for discovery and innovation in the statistical and inferential data sciences. To keep pace with the growing need to extract information from the tremendous amounts of available data, to predict the behavior of complex systems, and to quantify the uncertainty in information and predictions, the demand for statistical and inferential data science research grows daily in many fields. The mission of CANSSI is to support the development, application, and communication of cutting-edge statistical and data science research in the context of tackling engineering, health, scientific, and social challenges.

CANSSI supports a variety of programs designed to build multi-disciplinary research collaborations with a strong statistical foundation and to support new collaborations that link statistical and inferential data sciences to applications. CANSSI is also committed to training the next generation of Highly Qualified Personnel to lead the future development and application of statistical and inferential data science research. Our flagship program of Collaborative Research Teams covers a diverse range of methodology, models, and applications. We support a postdoctoral fellows program and a variety of training opportunities for graduate students. We support a wide array of workshops, conferences and thematic programs. We also support a distinguished visitor program.

Importantly, CANSSI's support is designed to address the challenges of geographic dispersion across Canada and to pursue CANSSI's commitment to equity, diversity, and inclusion.

# Collaborative Research Teams

## Joint Analysis of Neuroimaging Data: High-Dimensional Problems, Spatiotemporal Models and Computation (2016-2019)

Project leaders: Farouk Nathoo (University of Victoria) and Linglong Kong (University of Alberta)
Collaborators: 16 total from various disciplines in universities in Canada and the United States.

This team has wrapped up a successful three-year project. During that time, the team produced 23 papers that either appeared or were submitted to peer-reviewed journals and proceedings. Team members gave 48 talks at conferences and workshops in Canada and abroad. Topics included quantile estimation with incomplete data, statistical methods in imaging genetics, spatial models in imaging genetics, and empirical likelihood and robust regression in diffusion tensor imaging data analysis.

Notable achievements and activities in 2019 include:

- PhD student Yin Song successfully defended his PhD thesis in March and is now a Data Scientist at Tutella, Vancouver.
- Former postdoctoral fellow Li Xing is now a tenure-track assistant professor at the University of Saskatchewan.
- The team co-leaders are currently co-editing a special issue of the Canadian Journal of Statistics (CJS) with emphasis on neuroimaging data analysis. They have received an excellent response with a good number of submissions for the special issue.
- The team co-leaders have organized an invited session at the 2020 Joint Statistical Meetings that will highlight contributions from the CJS Special Issue.

There are plans to continue the work through additional funding via other mechanisms. To that end, Nathoo, Kong, Bei Jiang and the trainees met for two weeks in September to discuss continued collaboration beyond the life of the CRT.

Cumulative Project Presentations

| | **Presenter** & co-authors | **Type** | **Date** | **Meeting** | **Location** | **Title** |
|---|---|---|---|---|---|---|
| 1 | K. Greenlaw, E. Szefer, J. Graham, M. Lesperance, **F. Nathoo** | Talk | 2016 | ISBA 2016 World Meeting | Sardinia, Italy | A Bayesian Group Sparse Multl-Task Regression Model for Imaging Genetics |
| 2 | K. Greenlaw, E. Szefer, J. Graham, M. Lesperance, **F. Nathoo** | Talk | 2016 | Workshop on Pattern Recognition and Neuroimaging | Trento, Italy | A Bayesian Group Sparse Multi-Task Regression Model for Imaging Genetics |
| 3 | K. Greenlaw, E. Szefer, J. Graham, M. Lesperance, **F. Nathoo** | Talk | 2016 | SSC Annual Meeting | St. Catharines | A Bayesian Group Sparse Multi-Task Regression Model for Imaging Genetics |
| 4 | N. Croteau, **F. Nathoo,** J. Cao, R. Budney | Talk | 2016 | International Biometric Conference | Victoria | High-dimensional classification for brain decoding |
| 5 | Y. Song, **F. Nathoo** | Talk | 2016 | International Biometric Conference | Victoria | Spatiotemporal Mixture Potts Model for Combined EEG and MEG Data |

| | | | | | | |
|---|---|---|---|---|---|---|
| 6 | Y. Dengdeng, K. Linglong, I. Mizera. | Talk | 2016 | ICSA symposium | Atlanta, GA | An alternative formulation of functional partial quantile linear regression and its properties |
| 7 | **L. Kong,** X. Zhou, R. Karunamuni, Z. Hongtu | Talk | 2016 | Fields Institute HDDA VI | Toronto | Quantile Regression with Varying Coefficients for Functional Responses |
| 8 | **L. Kong,** X. Wang, Z. Zhang, H. Zhu | Talk | 2016 | Frontiers of Statistics and Data Sciences | Hong Kong | Optimal Estimation for Quantile Regression with Functional Response |
| 9 | **L. Kong,** X. Zhou, R. Karunamuni, H. Zhu | Talk | 2016 | Institute of Mathematical Statistics Asia Pacific Rim Meeting | Hong Kong | Quantile Regression with Varying Coefficients for Functional Responses |
| 10 | **L. Kong,** X. Wang, Z. Zhang, H. Zhu | Talk | 2016 | International Biometric Conference | Victoria | Optimal Estimation for Quantile Regression with Functional Response |
| 11 | **L. Kong**, H. Shu, C. He, G. Heo, M. Styner, J. Gilmore, H. Zhu | Talk | 2016 | Joint Statistical Meetings | Chicago, Illinois | Estimation for Bivariate Quantile Varying Coefficient Model |
| 12 | **L. Kong,** X. Zhou, R. Karunamuni, H. Zhu | Talk | 2016 | Laval University | Quebec City | Quantile Regression with Varying Coefficients for Functional Responses |
| 13 | **L. Kong,** X. Wang, Z. Zhang, X. Zhou, R. Karunamuni, H. Zhu | Talk | 2016 | University of Wisconsin at Madison | Madison, Wisconsin | Quantile Regression with Varying Coefficients and its Optimal Estimation for Functional Responses |
| 14 | **L. Kong,** X. Zhou, R. Karunamuni, H. Zhu | Talk | 2016 | ICSA International Conference | Shanghai China | Quantile Regression with Varying Coefficients for Functional Responses |
| 15 | **L. Kong**, H. Shu, C. He, G. Heo, J. Gilmore, H. Zhu | Talk | 2016 | Big Data and Medical Imaging Workshop | Sanya, China | Estimation for Bivariate Quantile Varying Coefficient Model |
| 16 | **L. Kong** | Talk | 2017 | KAUST | Saudi Arabia | Spatial quantile regression for functional responses in Neuroimaging data analysis |
| 17 | **L. Kong** | Talk | 2017 | ICSA Canada Chapter Symposium | Vancouver | Functional and Spatial Varying Coefficient Models for DTI Data Analysis, |
| 18 | **L. Kong** | Talk | 2017 | University of British Columbia | Vancouver | Estimation in Functional Linear Quantile Regression, |
| 19 | **L. Kong** | Talk | 2017 | SSC Annual Meeting | Winnipeg | Spatial Quantile Regression Models for High-Dimensional Imaging Data |
| 20 | **L. Kong** | Talk | 2017 | University of Calgary | Calgary | Estimation for Bivariate Quantile Varying Coefficient Model |
| 21 | **L. Kong** | Talk | 2017 | University of Alberta | Edmonton | Functional and Spatial Varying Coefficient Models for DTI Data Analysis |
| 22 | **F. Nathoo** | Talk | 2017 | CANSSI Workshop on Medical Physics and Statistics | Toronto | Analysis of Combined MEG and EEG Data |

| | | | | | | |
|---|---|---|---|---|---|---|
| 23 | **F. Nathoo** | Talk | 2017 | SSC 2017 Meeting | Manitoba | Analysis of Combined MEG and EEG Data |
| 24 | **F. Nathoo** | Talk | 2017 | Les Sixiemes Recontres R | Anglet, France | Analysis of Combined MEG and EEG Data |
| 25 | **F. Nathoo** | Talk | 2017 | Joint Statistical Meetings | Baltimore Maryland | Analysis of Combined MEG and EEG Data |
| 26 | **F. Nathoo** | Talk | 2017 | Western University | London Ontario | Analysis of Combined MEG and EEG Data |
| 27 | **F. Nathoo** | Talk | 2017 | University of Toronto | Toronto | Analysis of Combined MEG and EEG Data |
| 28 | **L. Kong** | Talk | 2018 | Hong Kong Baptist University | Hong Kong | A General Framework for Quantile Estimation with Incomplete Data |
| 29 | **L. Kong** | Talk | 2018 | CM Statistics | Pisa, Italy | A Review of Statistical Methods in Imaging Genetics |
| 30 | **L. Kong** | Talk | 2018 | MacEwan University | Edmonton | A General Framework for Quantile Estimation with Incomplete Data |
| 31 | **L. Kong** | Talk | 2018 | University of Manitoba | Winnipeg | A Review of Statistical Methods in Imaging Genetics |
| 32 | **L. Kong** | Talk | 2018 | ICSA China Conference | Qingdao, Shangdong, China | Empirical likelihood and robust regression in diffusion tensor imaging data analysis |
| 33 | **L. Kong** | Talk | 2018 | WNAR | Edmonton | A Review of Statistical Methods in Imaging Genetics |
| 34 | **L. Kong** | Talk | 2018 | SSC Annual Meeting | Montreal | Empirical likelihood and robust regression in diffusion tensor imaging data analysis |
| 35 | **L. Kong** | Talk | 2018 | HDDA-VIII | Marrakech, Morocco | Empirical likelihood and robust regression in diffusion tensor imaging data analysis |
| 36 | **L. Kong** | Talk | 2018 | McGill University | Montreal | Varying Coefficient Model for Functional Responses: Beyond Least Squares |
| 37 | **L. Kong** | Talk | 2018 | University of Toronto | Toronto | Varying Coefficient Model for Functional Responses: Beyond Least Squares |
| 38 | **P. Kim** | Talk | 2018 | SSC Annual Meeting | Montreal | The Gut-Brain Axis and Clostridium difficile Infection |
| 39 | **P. Kim** | Talk | 2018 | CPHAZ Symposium | Guelph | The Gut-Brain Axis and Clostridium difficile Infection |
| 40 | **F. Nathoo** | Talk | 2018 | SSC Annual Meeting | Montreal | A Spatial Model for Imaging Genetics |
| 41 | **F. Nathoo** | Talk | 2018 | WNAR | Edmonton | A Spatial Model for Imaging Genetics |
| 42 | **F. Nathoo** | Talk | 2018 | Joint Statistical Meetings | Vancouver | A Potts Model for Combined MEG and EEG Data |
| 43 | **F. Nathoo** | Talk | 2018 | 4th International Conference on Big Data and Information Analytics | Houston, TX | Feature Learning and Classification in Neuroimaging: Predicting Cognitive Impairment from Magnetic Resonance Imaging |

| | | | | | | |
|---|---|---|---|---|---|---|
| 44 | **F. Nathoo** | Talk | 2019 | University of Victoria Cognition and Brain Science Seminar | Victoria | Spectral Dynamic Causal Modelling of Resting-State fMRI: Relating Effective Brain Connectivity in the Default Mode Network to Genetics |
| 45 | **F. Nathoo** | Talk | 2019 | SSC 2019 Meeting | Calgary | Spectral Dynamic Causal Modelling of Resting-State fMRI: Relating Effective Brain Connectivity in the Default Mode Network to Genetics |
| 46 | **F. Nathoo** | Inv. Disc. | 2019 | Joint Statistical Meetings | Denver Colorado | |
| 47 | **F. Nathoo** | Talk | 2019 | ICSA | China | Spectral Dynamic Causal Modelling of Resting-State fMRI: Relating Effective Brain Connectivity in the Default Mode Network to Genetics |
| 48 | **F. Nathoo** | Talk | 2019 | CM Statistics, | London, UK | SVARO: Bayesian Analysis of fMRI data with Spatially-Varying Autoregressive Orders |

## Cumulative Project Papers

| | **References** |
|---|---|
| 1 | Nathoo, F.S., Greenlaw, K., Lesperance, M.L. (May, 2016). Regularization parameter selection for a Bayesian multi-level group lasso regression model with application to imaging genomics. *2016 International Workshop on Pattern Recognition in Neuroimaging (PRNI), IEEE*.**DOI:** 10.1109/PRNI.2016.7552328. Preprint available on arXiv: https://arxiv.org/abs/1603.08163. |
| 2 | Croteau, N., Nathoo, F.S., Cao, J., Budney, R. (August, 2016). High-dimensional classification for brain decoding. *Big and Complex Data Analysis: Statistical Methodologies and Applications. Springer, Edited Volume*, Ahmed, S. Ejaz (Ed.). Preprint available on arXiv:https://arxiv.org/abs/1504.02800. |
| 3 | Shen Z, Wu Z, Chang D, Zhang W, Tram K, Lee C, Kim P, Salena BJ, Li Y. (2016). A Catalytic DNA Activated by a Specific Strain of Bacterial Pathogen. *Angewandte Chemie*. 55(7): 2431-4. |
| 4 | Lee CH, Steiner T, Petrof EO, Smieja M, Roscoe D, Nematallah A, Weese JS, Collins S, Moayyedi P, Crowther M, Ropeleski M, Jayaratne P, Higgins D, Li Y, Rau NV, Kim PT. (2016). Frozen vs fresh fecal microbiota transplantation and clinical resolution of diarrhea in patients with recurrent *Clostridium difficile* infection. A randomized clinical trial. *JAMA: The Journal of the American Medical Association*. 315(2): 142-149. |
| 5 | E Cornea, H Zhu, P Kim, JG Ibrahim (2017). Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B*, 79(2): 463-482 |
| 6 | Greenlaw, K., Szefer, E., Graham, J., Lesperance, M.L., Nathoo, F.S. (2017). A Bayesian Group Sparse Mutli-Task Regression Model for Imaging Genetics. *Bioinformatics*, DOI: 10.1093/bioinformatics/btx215. |
| 7 | Szefer, E., Lu, D., Nathoo, F.S., M.F. Beg, Graham, J. (2017). Multivariate association between single-nucleotide polymorphisms in Alzgene linkage regions and structural changes in the brain: discovery, refinement and validation. *Statistical Applications in Genetics and Molecular Biology*, DOI: https://doi.org/10.1515/sagmb-2016-0077. |
| 8 | Teng, M., Nathoo, F.S., Johnson, T.D. (2017). Bayesian Computation for Log Gaussian Cox Processes: A Comparative Analysis of Methods. *Journal of Statistical Computation and Simulation*, DOI: 10.1080/00949655.2017.1326117. |

| | |
|---|---|
| 9 | Tang, Q. and Kong, L. (2017*).* Quantile regression in functional linear semiparametric model*, Statistics , Vol. 51, No. 6, 1342-1358.* |
| 10 | Che, M. , Kong, L. , Bell, R. and Yuan, Y. (2017). Trajectory Modeling of Gestational Weight: a Functional Principal Component Analysis Approach, P*LoS ONE* , 12, e0186761. |
| 11 | Zhang, L., Cobza, D., Wilman, A., and Kong, L. (2017). An unbiased penalty for sparse classification with application to neuroimaging data, *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2017), Lecture Notes in Computer Science,* Springer Berlin/Heidelberg, 10435: 55-63. |
| 12 | Zhang, L., Cobza, D., Wilman, A., and Kong, L. (2018). Significant Anatomy Detection through Sparse Classification: A Comparative Study, *IEEE Transition in Medical Imaging*, 37: 128-137. |
| 13 | Lee, C., Rush, S., Weese, J.S., Goldeh, P., Kim, P. (2018) Engraftment and Augmentation of Microbiome Following Fecal Microbiota Transplantation for Recurrent Clostridium difficile Infection, *Open Forum Infectious Diseases* 4 (Suppl 1): S380. |
| 14 | Martinez, L., Kim, M. (2018) Probing the geometry of data with diffusion Frechet functions. *Applied and Computational Harmonic Analysis*, published online 2018/2/13. |
| 15 | Yu, D., Zhang, L., Jiang, B., Mizera, I. and Kong, L. (2018). Sparse Wavelet Estimation in Quantile Regression with Multiple Functional Predictors, *Computational Statistics & Data Analysis*, accepted. |
| 16 | Tu, W., Kong, L.?, Karunamuni, R., Butcher, K., Zheng, L.-, and McCourt, R. (2018). Non-local Spatial Clustering in Automated Brain Hematoma and Edema Segmentation, *Applied Stochastic Models in Business and Industr*y, accepted. |
| 17 | Han, P. Kong, L, Zhao, J., and Zhou, X. (2018). A General Framework for Quantile Estimation with Incomplete Data, *Journal of the Royal Statistical Society: Series B,* accepted. |
| 18 | Karunamuni, R., Kong, L., and Tu, W. (2018). Efficient Robust Doubly Adaptive Regularized Regression, *Statistical Methods in Medical Research*, accepted. |
| 19 | Nathoo, F., Kong, L. and Zhu, H. (2018). A Review of Statistical Methods in Imaging Genetics, *Canadian Journal of Statistics*, DOI: 10.1002/cjs.11487. |
| 20 | Teng, M., Nathoo, F.S., Johnson, T.D. (2018). Bayesian Analysis of fMRI data with Spatially-Varying Autoregressive Orders. *Journal of the Royal Statistical Society: Series C,* DOI https://doi.org/10.1111/rssc.12320. Preprint available on arXiv here |
| 21 | Teng, M., Johnson, T.D., Nathoo, F.S. (2018). Time Series Analysis of fMRI Data: Spatial Modeling and Bayesian Computation. *Statistics in Medicine*, DOI: 10.1002/sim.7680. Preprint available on arXiv here. |
| 22 | Shi S. and Nathoo, F.S.. (2018) Feature Learning and Classification in Neuroimaging: Predicting Cognitive Impairment from Magnetic Resonance Imaging. *Proceedings of the 4th International Conference on Big Data and Information Analytics*. Preprint available on arXiv here. |
| 23 | Song, Y., Nathoo, F.S., Babul A. (2018). A Potts-Mixture Spatiotemporal Joint Model for Combined MEG and EEG Data. *Canadian Journal of Statistics*, DOI: 10.1002/cjs.11519. |

## Cumulative HQP

### MSc Students

| **Name** | **Supervisor** | **Research Topic** | **Completed** | **Current Position** |
|---|---|---|---|---|
| Cailin Harris | P. Kim | Bioinformatics | In progress | |
| Laila Yasmin | F. Nathoo | Relating resting-state fMRI effective connectivity in the default mode network to genetics with dynamic causal models and group-sparse multi-task regression | In progress | |
| Emma Smith | P. Kim | Quantifying Health with Missing Data: Measuring the Impact of Fecal Microbiota Transplantation on Health-Related Quality of Life | 2017 | PhD Biostatistics Student at Schulich School of Medicine & Dentistry |
| Boroslav Mavrin | L. Kong | Statistical Machine Learning | 2017 | Data Scientist, Janalta Interactive Inc. |

| | | | | |
|---|---|---|---|---|
| Nicole Croteau | F. Nathoo | Machine learning for forecasting expected hydrological runoff | 2016 | Statistical Consultant, UVic Statistical Consulting Centre |
| Dong Yang | L. Kong | Ensemble Based Ultrahigh Dimensional Variable Screening | 2018 | Manager at AID Cloud Technology |
| Xi Hu | L. Kong | Residual Weighted Learning For Quantile Optimal Treatment Regimes | 2018 | Data Scientist at Servus Credit Union |
| Fangfang Fu | L. Kong | Low Rank plus Sparse Decomposition of fMRI Data with Applications in Alzheimer's Disease | 2018 | Research Assistant at the University of Alberta |

PhD Students

| **Name** | **Supervisor** | **Research Topic** | **Completed** | **Current Position** |
|---|---|---|---|---|
| Yin Song | F. Nathoo | Statistical Methods for Neuroimaging Data Analysis and Cognitive Science | 2019 | Data Scientist, Tutela |
| Shan Shi | F. Nathoo | Machine Learning with applications to neuroimaging data; feature construction | In progress | |
| Stephen Rush | P. Kim | The Phylogenetic LASSO and the Microbiome: Metagenomic modeling in fecal microbiota transplantation | 2017 | Biostatistician at Nutrition-Gut-Brain Interactions Research Centre at Örebro University, Sweden |
| Ming Teng | F. Nathoo | Bayesian Computation for Spatial Data and Neuroimaging Data | 2017 | Statistician, Morgan Stanley, New York |
| Dengdeng Yu | L. Kong | Quantile Regression Methods in Functional Data Analysis | 2017 | Postdoctoral Fellow at the University of Toronto |
| Wei Tu | L. Kong | Machine Learning Models | 2019 | |
| Eugene Opoku | F. Nathoo | Methods for imaging Data and imaging genetics | In progress | |

Postdoctoral Fellows

| **Name** | **Supervisor** | **Research Topic** | **Completed** | **Current Position** |
|---|---|---|---|---|
| Li Xing | F. Nathoo | Regression for Longitudinal Analysis in Imaging Genetics with Bayesian Shrinkage Priors | 2019 | Assistant Professor, University of Saskatchewan |
| Peng Liu | L. Kong | Spatial Data Reconstruction | 2019 | Lecturer at the University of Kent |
| S. Selvaratnam | L. Kong | Regression Applications and Methodology | 2019 | Sessional Lecturer at the University of Alberta |

## Rare DNA variants and human complex traits: improving analyses of family studies by better modeling the dependence structures (2016-2019)

Project leaders: Alexandre Bureau (Université Laval) and Karim Oualkacha (Université du Québec à Montréal)
Collaborators: 8 from various disciplines in universities and institutes in Canada.

This team completed a very successful project. During that time, the team produced 6 papers that appeared in peer-reviewed journals and proceedings. Team members gave 31 posters and talks at conferences and workshops in Canada and abroad.

Notable achievements and activities in 2019 include:

- Kaiqiong Zhao received a $70000 Fonds de recherche du Québec – Santé (FRQS) scholarship for her doctoral thesis project entitled "Smooth modeling of covariate effects in bisulfite sequencing-derived measures of DNA methylation". Supervised by Celia Greenwood and Karim Oualkacha, Zhao is working on extensions of this CRT's third aim to deal with association testing of DNA methylation; she submitted a manuscript "Smooth modeling of covariate effects in bisulfite sequencing-derived measures of DNA methylation" to the *Biometrics Journal*. The manuscript is under revision.
- Trainee Christina Nieuwoudt submitted a manuscript to *Bioinformatics* describing her R package for Simulating Sequence Data in Families Ascertained for Multiple Disease-Affected Relatives to Bioinformatics. The R package SimRVSequences was submitted to CRAN on Apr 18 and a major revision was submitted on Nov 3.
- Patrick Fournier, who received funding to participate in the BIRS workshop since his master's thesis at UQAM was related to CRT Aim 1, is now enrolled in the mathematics PhD program (statistics) at UQAM under the supervision of CRT team member Fabrice Larribe.
- Celia Greenwood was awarded a James McGill Professorship in recognition of her outstanding and original research and international leadership in oncology. Her award will be held for a seven-year term.
- Team co-leader Karim Oualkacha received a FRQS Research Scholar-Junior 2 for his work "Nouvelles approches statistiques pour la modélisation des données génomiques de grandes dimensions à l'ère des technologies omiques" worth $286,986.
- Renaud Alie finished his master's degree at Université du Québec à Montréal and obtained an NSERC scholarship for a PhD at McGill University.
- Roland Dossa, the PhD trainee supervised by two CRT team members, Lajmi Lakhal-Chaieb and Karim Oualkacha, is investigating a new copula-based method to test for association between a set of rare variants and a dichotomous trait in presence of familial data. This project is part of the CRT's second aim. Dossa is invited to present his work in the weekly seminar of statistics of the Department of Mathematics and Statistics, Université Laval, on February 13, 2020.
- Alexandre Bureau, S. Gravel and S. Girard were awarded a CIHR grant "Investigating the role of non-coding variations in epilepsy", 2019-2022 (PI: S. Girard). Part of this grant will be devoted to the pursuit of the CRT aim to "incorporate genealogies into rare variant sharing analysis".

This collaboration will continue past the three-year timeline because Celia Greenwood, Karim Oualkacha, Aurélie Labbe, Lajmi Lakhal-Chaieb are now working on a project on modeling next-generation sequencing of DNA methylation data in presence of multilevel dependencies, e.g. time-varying measures and/or related subjects.

Cumulative Project Presentations

| | **Presenter** & co-authors | **Type** | **Date** | **Meeting** | **Location** | **Title** |
|---|---|---|---|---|---|---|
| 1 | C. Nieuwoudt, A. Brooke-Wilson, J. Graham | Cont talk | 2017 | SSC | Winnipeg | SimRVPedigree: An R package to simulate pedigrees ascertained for multiple relatives affected by a rare disease |
| 2 | A. Bureau | Invit talk | 2017 | Journées Généalogie et Génétique | Trois-Rivières | Analyse du partage de variants rares : ce que les généalogies peuvent contribuer. |
| 3 | K. Burkett | Invit talk | 2017 | SSC | Winnipeg | An ancestral tree-based approach to detect rare and common variants |
| 4 | K. Burkett | Invit talk | 2017 | CRT workshop | Quebec City | Markov Chain Monte Carlo sampling of gene genealogies conditional on genotype data from trios |
| 5 | F. Larribe | Invit talk | 2017 | SSC | Winnipeg | Mapping Complex Traits, Rare Variants and Interaction via the Coalescent Process with Recombination |
| 6 | F. Larribe | Invit talk | 2017 | CRT workshop | Quebec City | Extending the coalescent from population data to pedigree data |
| 7 | K. Oualkacha | Invit talk | 2017 | CRT workshop | Quebec City | Multivariate association test for rare variants controlling for cryptic and family relatedness |
| 8 | K. Oualkacha | Invit talk | 2018 | SSC | Montreal | A General Framework for Variable Selection in Linear Mixed Models with Applications to Genetic Studies with Structured Populations |
| 9 | A. Bureau | Invit talk | 2018 | SSC | Montreal | Statistical Analysis of Co-segregation of Rare Genetic Variants with a Disease in Families |
| 10 | A. Bureau | Invit talk | 2018 | Simon Fraser University | Burnaby | Statistical Analysis of Co-segregation of Rare Genetic Variants with a Disease in Families |
| 11 | K. Oualkacha | Invit talk | 2018 | BIRS workshop | Banff | A General Framework for Variable Selection in Linear Mixed Models with Applications to Genetic Studies with Structured Populations |
| 12 | I. Ruczinski | Invit talk | 2018 | BIRS workshop | Banff | Sharing of rare nucleotide and copy number variants in extended multiplex families |
| 13 | C. Nieuwoudt | Invit talk | 2018 | BIRS workshop | Banff | Simulating family studies with whole-exome sequencing of multiple affected relatives |
| 14 | A. Bureau | Invit talk | 2018 | BIRS workshop | Banff | Rare variant sharing methods with genealogies |
| 15 | L. Jiang | Invit talk | 2018 | BIRS workshop | Banff | Estimating the effects of copy number variants on intelligence quotient using hierarchical Bayesian models |

| | | | | | | |
|---|---|---|---|---|---|---|
| 16 | J. Graham | Invit talk | 2018 | BIRS workshop | Banff | Combining phenotypes, genotypes and gene genealogies to find trait-influencing variants |
| 17 | C. B. Karunarathna | Invit talk | 2018 | BIRS workshop | Banff | Using gene genealogies to localize trait-influencing variants in diploid populations |
| 18 | R. Alie | Invit talk | 2018 | BIRS workshop | Banff | Construction of coalescent trees on partially fixed pedigrees |
| 19 | K. Burkett | Invit talk | 2018 | BIRS workshop | Banff | Sampling gene genealogies conditional on genotype data from trios |
| 20 | L. Lakhal Chaieb | Invit talk | 2018 | BIRS workshop | Banff | Testing the heritability and parent-of-origin hypotheses for ages at onset of psoriatic arthritis under biased sampling |
| 21 | K. Zhao | Invit talk | 2018 | BIRS workshop | Banff | Smooth modeling of covariate effects in bisulfite sequencing-derived measures of DNA methylation |
| 22 | R. Dossa | Post | 2018 | BIRS workshop | Banff | Rare variant association test for binary traits in family-based designs via copulas |
| 23 | A. Bureau | Invit talk | 2018 | Statslab annual meeting, CRM | Montreal | Statistical Analysis of Co-segregation of Rare Genetic Variants with a Disease in Families |
| 24 | C. Nieuwoudt | Talk | 2018 | SSC | Montreal | Simulating Sequence Data for Pedigrees |
| 25 | C. Nieuwoudt | Post | 2018 | 7th Annual Canadian Human and Statistical Genetics Meeting | Harrison Hot Springs, BC | SimRVSequences: Simulating Sequence Data for Pedigrees |
| 26 | K. Burkett | Talk | 2019 | CHSGM | Montebello | Gene genealogies for detecting rare and common variants |
| 27 | L. Mangnier, A. Bureau | Talk | 2019 | CHSGM | Montebello | Building Gene-Enhancer clusters based on their contacts |
| 28 | K. Oualkacha | Talk | 2019 | Weekly seminar, UQAM | Montreal | Tests multidimensionnels d'association génétique pour variants rares et phénotypes non-normaux |
| 29 | K. Oualkacha | Talk | 2019 | Institut de Mathématiques et de Sciences Physiques | Dangbo, Benin | Tests multidimensionnels d'association génétique pour variants rares et phénotypes non-normaux |
| 30 | K. Zhao, K. Oualkacha, A. Labbe, L. Lakhal-Chaieb, C. Greenwood | Post | 2019 | CHSGM | Montebello | Modeling covariate effects in bisulfite sequencing-derived measures of DNA methylation, in the presence of overdispersion |
| 31 | K. Zhao, K. Oualkacha, A. Labbe, L. Lakhal-Chaieb, C. Greenwood | Post | 2019 | IGES | Houston TX, USA | Modeling covariate effects in bisulfite sequencing-derived measures of DNA methylation, in the presence of overdispersion |

Cumulative Project Papers

| | |
|---|---|
| 1. | Zhao, K., Jiang, L., Klein, K., Greenwood, CMT., Oualkacha, K. (2018). CpG-set association assessment of lipid concentration changes and DNA methylation on chromosome 11. *BMC Proceeding*, **12** (Suppl 9):30. https://doi.org/10.1186/s12919-018-0127-8 |
| 2. | Jiang, L., Zhao, K., Klein, K., Canty, AJ., Oualkacha, K., Greenwood, CMT. (2017). Investigating potential causal relationships between SNPs, DNA methylation and HDL. *BMC Proceeding*, **12** (Suppl 9):20. https://doi.org/10.1186/s12919-018-0117-x |
| 3. | Nieuwoudt, C., Jones, S., Brooks-Wilson, A., Graham, J. (2018). Simulating pedigrees ascertained for multiple disease-affected relatives. *Source Code for Biology and Medicine,* **13(**2). https://doi.org/10.1186/s13029-018-0069-6. |
| 4. | J. Sun, K. Oualkacha, C. Greenwood and L. Lakhal-Chaieb (2019). Multivariate association test for rare variants controlling for cryptic and family relatedness. *Canadian Journal of Statistics,* special issue featuring the CANSSI CRTs, 47 (1): 90-107. doi.org/10.1002/cjs.11475. |
| 5. | Zhao, Z., Oualkacha, K., Labbe, A., Lakhal-Chaieb, L., Greenwood, C. (2019). Smooth modeling of covariate effects in bisulfite sequencing-derived measures of DNA methylation. Revision requested by *Biometrics* |
| 6. | Nieuwoudt, C., Brooks-Wilson, A., Graham, J. (2019). SimRVSequences: an R package to simulate genetic sequence data for pedigrees. *Bioinformatics*, btz881, https://doi.org/10.1093/bioinformatics/btz881. |

Cumulative HQP

MSc Students

| Name | Supervisor | Research Topic | Completed | Current Position |
|---|---|---|---|---|
| Renaud Alie | F. Larribe | Processus de coalescence construit sur des pédigrés | 2019 | PhD Student at McGill |

PhD Students

| Name | Supervisor | Research Topic | Completed | Current Position |
|---|---|---|---|---|
| Saeed Sabbah | A. Bureau | Haplotyping of SNVs | N/A | |
| Kaiqiong Zhao | C. Greenwood | Modeling DNA methylation profiles from bisulfite sequencing data | In progress | |
| Christina Nieuwoudt | J. Graham | Simulating pedigrees ascertained for multiple disease-affected relatives | 2019 | Biostatistician at Emmes |
| Roland Dossa | K. Oualkacha | Statistical Methods for Family-Based Sequencing Studies | In progress | |
| Patrick Fournier | F. Larribe | Cartographie génétique par des graphes de recombinaisons ancestraux construits séquentiellement | In progress | |
| Loïc Mangnier | A. Bureau | Rare variant association tests in a schizophrenia context | In progress | |
| Charith Karunarathna | J. Graham | Statistical Genetics | In progress | |

## Statistical Analysis of Large Administrative Health Databases: Emerging Challenges and Strategies

Project leaders: Grace Y. Yi (University of Waterloo), Robert Platt (McGill University) and X. Joan Hu (Simon Fraser University)

Collaborators: 7 faculty in universities in Canada

This project is advising a large team of highly qualified personnel very successfully. The team published 14 papers in 2019 with 6 others are under revision while 8 more have been submitted or are in progress.

Notable achievements and activities in 2019 include:

- Six highly qualified personnel involved with the project have graduated and moved on to other exciting opportunities:
    - Di Shu graduated from the University of Waterloo in May and is currently a postdoctoral fellow at Harvard Medical School and Harvard Pilgrim Health Care Institute.
    - Li-Pang Chen graduated from the University of Waterloo in August and is currently a postdoctoral fellow at the University of Western Ontario under the supervision of Grace Yi.
    - Xin Liu finished his postdoctoral fellowship at the University of Waterloo February-July 2018 and at Simon Fraser University August 2018 – January 2019; he is currently an assistant professor at Shanghai Finance and Economics University, P. R. China.
    - Dongdong Li graduated from Simon Fraser University in December, and is now a postdoctoral research fellow at Harvard Medical School and a recipient the Thomas O. Pyle Fellowship 2019-2020.
    - Qian (Emma) Wen received her MSc from Simon Fraser University in September and is currently an instructor at Langara College, Vancouver.
    - Gabrielle Simoneau (PhD student, McGill, supervised by Platt) completed her degree. She is currently a Biostatistician at Biogen.
- There are 7 PhD students and 2 postdoctoral fellows that are continuing to work with this team at McGIll University, Simon Fraser University and the University of Waterloo.
- The team met several times at team-organized workshops and meetings.
    - Robert Platt, Grace Yi and Joan Hu organized the BIRS Workshop 2019 on “Statistical Challenges from Health Administrative Data: Emerging Challenges and Strategies”.
    - A trainee research meeting was held at Waterloo on April 29.
    - Grace Yi and Wenqing He (a project collaborator) co-organized a Data Science Research Group which will contain all the students they are supervising at the University of Waterloo and the University of Western Ontario. The group will have weekly meetings with one participant presenting relevant papers on an interesting topic. These meetings will generate useful discussions which may lead to new projects related to the projects funded by this CANSSI CRT.
    - Grace Yi also organized an invited session at the SSC 2019, with the speakers Wenqing He, and the PhD students of Joan Hu and Robert Platt and another invited session at the JSM 2019, with the speakers Wenqing He, Joan Hu, Mireille Schnitzer (a collaborator), and a PhD student of Robert Platt.

- Grace Yi gave a short course, entitled “Statistical Analysis with Noisy Data - Dealing with Measurement Error and Missing Observations”, at the Joint Statistical Meetings, Denver, 2019. Team members also gave 7 invited talks in 2019 in Canada, the United States, China and Taiwan.

## Towards Sustainable Fisheries: State Space Assessment Models (SSAMs) for Complex Fisheries and Biological Data

Project leader: Joanna Mills Flemming (Dalhousie University)
Collaborators: William Aeberhard, Chris Field, Marie Auger-Méthé, Eva Cantoni, Anders Nielsen, Louis-Paul Rivest, Huges Benoit, Daniel Duplisea, Noel Cadigan, Andrew Edwards, David Keith, Aaron MacNeil, Boris Worm, Cóilín Minto

2019 got off to a productive start with the recruitment of key highly qualified personnel arriving at the universities and several training and workshop events.

Notable achievements and activities in 2019 include:

- Yuan Yan arrived at Dalhousie University on January 2 to commence her postdoctoral fellowship under the supervision of Joanna Mills Flemming and Chris Field.
- In January, collaborator Anders Nielsen returned to Halifax to offer a week-long course titled ‘Template Model Building (TMB) for Advanced Fish Stock Assessment’. This training course was funded by the International Council for the Exploration of the Seas (ICES) and was a natural follow-up to the introductory course offered by Nielsen (and William Aeberhard) in 2018. PhD student trainee Ethan Lawler served as the Teaching Assistant and the course was at capacity with over 30 participants. All Dalhousie-based student trainees involved with this CRT had the opportunity to attend. TMB is amongst the newest modern tools for efficiently fitting highly parameterized stock assessment and other models with random effects. The course included a section on model selection, showing users how to evaluate the evidence for models that make different model structure assumptions. Strategies for obtaining projections were also reviewed and discussed.
- In early February, Mills Flemming, Field and Eva Cantoni began discussions with Research Scientist Margaret Treble, Division of Arctic Research of Fisheries and Oceans Canada (DFO) regarding a potential research project for postdoctoral fellow Yuan Yan. Together they laid out a framework that would see the development of spatiotemporal models for bycatch (of Greenland sharks initially) in Canadian Arctic fisheries.
- Mills Flemming offered a short course in Applied Fishery Data Analytics at BIO on February 20 and 21. The objective was to provide scientists within DFO an opportunity to learn about the latest statistical advances in fisheries science, understand the basic theory underpinning these techniques, and apply these methods using DFO data/case studies. Student trainees had the opportunity to showcase their research on the second day. Mills Flemming received wonderful feedback from all attendees.
- PhD student trainee Jonathan Babyn remained at the Northwest Atlantic Fisheries Centre (DFO) in St. John's to assist with the framework review for the cod stock off the south coast of Newfoundland (3Ps cod). This stock had been assessed with a survey

based (SURBA) model, and Jonathan was tasked with helping to develop a SSAM to deal with catch uncertainty and integrate more of the available data.

- Due in part to the generous support of ICES, this team's Year 2 Workshop was held at their headquarters in Copenhagen, Denmark, April 14-16. It was very well-attended with 9 collaborators and 4 student trainees. Both Mills Flemming and Aeberhard have since been invited back to ICES Headquarters to participate in various workshops.
- PhD student trainee Ethan Lawler successfully defended his thesis proposal titled 'Contemporary Spatiotemporal Methods for Fisheries Science' on April 26th. His proposal involves incorporating state-of-the-art methods in spatiotemporal statistical modelling into models for fisheries science. Due to the complex nature of fisheries data, this effort involves: state-space models for discrete-time components, dimension reduction for efficiently approximating continuous random fields, and spatial averaging. Lawler has since been preparing two manuscripts, this first of which centers on the development of an R package for modelling point-referenced spatiotemporal data. This work was motivated by collaborator Boris Worm's interest in looking at bycatch risk in Atlantic Canadian fisheries.
- PhD student trainee Jonathan Babyn gave a presentation describing his research on developing spatial age-length keys at the DFO data review meeting in support of the ongoing framework project for 3Ps cod, May 2-3. Mills Flemming attended this meeting through Webex. Babyn later presented this research at The Center for the Advancement of Population Assessment Methodology (CAPAM) workshop on the creation of frameworks for the next generation general stock assessment models. This workshop was held in collaboration with the National Institute of Water and Atmospheric Research in Wellington, New Zealand November 4-8. Babyn's abstract was selected for inclusion in a special issue proposal to the *Journal of Fisheries Research*.
- Research assistant Yihao Yin was offered full time employment at BIO and commenced his position on October 1. He will be working closely with collaborators David Keith and Jessica Sameoto to continue to develop SSAMs for select Atlantic Canadian fish stocks.
- Mills Flemming visited PhD student trainee Jonathan Babyn in St. Johns in October and attended the DFO Regional Peer Review Meeting to review the Assessment Framework for 3Ps Cod. This was an enlightening experience and also gave opportunity to discuss DFO's involvement in Babyn's thesis proposal (after he returns to Dalhousie).
- In late October postdoctoral fellow Yuan Yan submitted a manuscript titled 'Spatial Analysis of Bycatch in a Canadian Arctic Fishery' to the *Journal of Applied Ecology* (coauthored with Cantoni, Field, Treble and Mills Flemming). Excess bycatch of endangered marine species during commercial fishing trips is a challenging problem in fishery management worldwide. In order to understand spatial patterns and driving factors for both bycatch occurrence and magnitude, the manuscript proposes a spatially explicit two-part model to analyze bycatch of Greenland shark observed in the Canadian *Arctic* Greenland halibut fishery. Model selection and comparison procedures are developed and it is shown that such models can be used to identify spatiotemporal hotspots, in order to instruct proper conservation strategies and management decisions, such as seasonal closures, limiting access to spatial hotspots, and gear restrictions. Yan

presented her work in late November as an e-poster at the Statistics and Data Science Workshop held at King Abdullah University of Science and Technology in Saudi Arabia. In early December she received additional data from Margaret Treble to further explore assessment issues for select Arctic fisheries.

- Rivest hired three senior statistics undergraduate student trainees to undertake empirical descriptive analyses of the data appearing in his presentation for the Year 2 Workshop at ICES.

## Spatial Modeling of Infectious Diseases: Environment and Health

Project leader: Mahmoud Torabi (University of Manitoba)
Collaborators: Charmaine Dean (University of Waterloo), Mike Pickles (University of Manitoba), Rob Deardon (University of Calgary), Rhonda Rosychuk (University of Calgary), Cindy Feng (University of Saskatchewan), Subhash Lele (University of Alberta), and Erin Rees (Université de Montréal)

This project began in 2018. The team so far has recruited four highly qualified personnel and plans to recruit more postdoctoral fellows and graduate students. They have given five talks and published two peer-reviewed papers.

Notable achievements and activities in 2019 include:

- Leila Amiri is continuing her postdoctoral fellowship at the University of Manitoba
- Georges Bucyiburata began his postdoctoral fellowship at the University of Waterloo.
- Md Mahsin continues his PhD student at the University of Calgary
- Tingxuan Wu began work towards a PhD at the University of Saskatchewan.
- The team met in 2019 for a workshop at the University of Calgary during the SSC Meetings and they are planning a 5-day workshop at the Fields Institute (FI) in Toronto from June 8-12, 2020. Their workshop application to the FI was successful and the FI will provide $21,500 for the workshop costs in addition to the venue. They are planning to invite 34 world-renowned speakers for this event.
- Notable talks given by team members include:
  - *Advancing Knowledge about Spatial Modeling, Infectious Diseases, Environment and Health* by Torabi in March in Winnipeg
  - *A New Class of Spatiotemporal Individual-Level Models for Infectious Diseases Transmission* by Mahsin and Deardon in May in Calgary
  - *Geographically-Dependent Individual-Level Models for Infectious Diseases Transmission* by Mahsin and Deardon in August, September and December in Glasgow (Scotland), Calgary, and Dhaka (Bangladesh)

## Statistical Methods for Challenging Problems in Public Health Microbiology

Project leaders: Leonid Chindelevitch (Simon Fraser University), Alexandre Bouchard-Côté (University of British Columbia), and Liangliang Wang (Simon Fraser University)
Collaborators: Joseph Bielawski, David Buckeridge, Kelly Burkett, Cedric Chauve, Caroline Colijn, Maha Farhat, Hong Gu, William Hsiao, Shamil Sunyaev, Linda Wahl, Jianhong Wu

*This project has not proceeded as anticipated due to the personal situations of the three leads. Project lead Chindelevitch is on an extended sabbatical lead that has diminished his engagement with the project. Project co-leads Bouchard-*Côté *and Wang have been on parental leave during the past year. This has prevented the project from becoming fully operational, especially with regards to recruiting the complete team of HQP proposed for the project. Unfortunately, with the NSERC funding for institutes during the current cycle ending, we could not be flexible about extension of expenditure deadlines. Therefore, we have implemented the CANSSI policy about expenditures and re-captured some of the project funding for use on other activities. We have requested a detailed expenditure plan for the third year of the project. We believe that the project will see substantial activity in its last year.*

So far, the project team members have given 14 presentations at professional meetings related to the CRT project and have published three papers in peer-reviewed journals and proceedings. The team has involved eight highly qualified personnel in research related to the project.

Notable achievements and activities in 2019 include:

- PIMS-UBC Mathematical Sciences Young Faculty Award (Bouchard-Côté)
- Early tenure awarded, with promotion to Associate Professor (Leonid Chindelevitch)
- Newly established collaboration with Dr. Mark Brockman, joint CIHR grant (Leonid Chindelevitch): Harnessing cross-reactive T cells to eliminate latent HIV reservoirs.
- Participation in the CANSSI CRT Workshop. March 30-March 31, 2019. SFU
- HQP recruited and involved with the project include
    - Kevin Chern (MSc),
    - Miguel Biron (PhD),
    - Vittorio Romaniello (PhD),
    - Doig (MSc),
    - Jingxue Feng (PhD),
    - Jiarui Zhang (PhD),
    - Mohammad Hossein Rezaie (MSc),
    - Amir-Hosein Safari (MSc)

## Statistical Methods for the Analysis of Genetic Data with Survival Outcomes

Project Leaders: Lajmi Lakhal-Chaieb (Université Laval), Richard Cook (University of Waterloo), and Laurent Briollais (Lunenfeld Tanenbaum Research Institute of Mount Sinai Hospital, Toronto)
Collaborators: Shelley Bull (Lunenfeld Tanenbaum Research Institute of Mount Sinai Hospital, Toronto), Yun-Hee Choi (Western University), and Yildiz Yilmaz (Memorial University of Newfoundland)

This team began in 2019. The team's research brings together individuals who published seven papers on research related to this project and are now collaborating on joint papers which will be submitted in 2020. They have collectively made six presentations on research related to the project.

Notable achievements and activities in 2019 include:

- Richard Cook presenting the 2019 Scandinavian Journal of Statistics Keynote Lecture at the Nordic-Baltic Meeting of the International Biometrics Society.
- Richard Cook was awarded the Faculty of Mathematics Research Chair for his Statistical Methods for Health Research, July 1, 2019 – June 31, 2024 in the Statistics and Actuarial Science Department at the University of Waterloo.
- Notable talks/presentations made by team members in 2019:
    - *Joint Modeling of Multivariate Longitudinal and Time-to-Event Phenotypes in Genetic Association Studies of Complex Traits* by Myriam Brossard, AD Paterson, O Espin-Garcia, Radu Craiu, and Shelley Bull in May and June in Calgary and Montebello (Canada)
    - *Improving Mixture Cure Modelling of Molecular Genetic Biomarkers in Cancer Prognosis* by Penalized Maximum Likelihood by C Xu and Shelley Bull in May in Calgary
    - *Revealing the Complex Genetic Architecture of Type 1 Diabetes Complications: Joint Modeling of Multivariate Longitudinal and Time-to-Event Traits* by Myriam Brossard, AD Paterson, O Espin-Garcia, Radu Craiu, and Shelley Bull in October in Houston, TX

## Contingent Capital and Calibration of Capital Structure Models

Project Leaders: Mark Reesor (Wilfrid Laurier University), Adam Metzler (Wilfrid Laurier University) and Hatem Ben-Ameur (HEC Montréal)
Collaborators: Joe Campolieti (Wilfrid Laurier University), Rim Chérif (HEC Montréal) and Malek Ben-Abdellatif (HEC Montréal)

This team began their collaboration in 2019. Collaborators include HQP in this team consists of the following individuals:

- Fortuné Edong: MSc student, HEC Montréal, registered since spring 2019, co-directed by Hatem Ben-Ameur and Rim Chérif
- Di Meng: PhD student in the Department of Mathematics at Wilfrid Laurier University, began in Fall 2018
- Bolade Ewarawon: PhD student in the Department of Mathematics at Wilfrid Laurier University, began in Winter 2020
- Hanchao Liu: MSc student in the Department of Mathematics at Wilfrid Laurier University, began in Fall 2019

So far, the team has made five presentations on research related to the project and authored and submitted 7 papers in peer-reviewed venues.

Notable activities on the project include:

- Ben-Ameur, Ben-Abdellatif, and Chérif are working to extend the models of Ayadi, Ben-Ameur, and Fakhfakh (2016) and Ben-Ameur, Fakhfakh, and Roch (2018) in modeling the call and the conversion options embedded in corporate bonds. These authors propose an extended structural setting that accommodates an arbitrary debt structure, several seniority classes and multiple intangible assets, including illiquidity costs. This part of their project is ongoing and still at the theoretical level.

- Ben-Ameur, Ben-Abdellatif, Chérif, and Fakhfakh have been working on extending the maximum-likelihood (ML) procedure of Duan (1994), which uses the closed-form solution of Merton (1974). We propose a quasi-maximum likelihood (QML) procedure, which builds on approximate value functions of corporate securities for the model estimation step. Both ML and QML make use of the stock-price time series of the underlying public company. This part of the team's project is already very advanced.
- Ben-Ameur, Chérif, and Edong are collaborating on an ML procedure that is perfectly aligned with the second part of the project except that the QML procedure capitalizes on CDS premiums written on the underlying public company. This part is ongoing and is at the algorithm-design level.

Once those three parts of the team's research are completed, the team will start on the financial industry and the potential added value of the conversion option as a risk management tool. This portion of the research agenda is expected to happen during the third year of their research project.

# CANSSI Postdoctoral Fellows

## Started in 2018

**Bouchra Nasri**
Supervisor name: Christian Genest
Title of Project: Dependence Models for High Dimensional Time Series
Location/University of project: McGill University
Current: Nasri is now the Assistant professor of the Department of Social and Preventive Medicine at the Université de Montréal (Canada).

Summary: The main objective of this research project was to model the dependence between a large number of time series variables appearing in finance and hydrology in order to estimate the risk of occurrence of extreme events. For financial data, it is essential to model the dependence between financial assets in order to predict returns or measure risk. In addition, for hydrological data, modeling the serial dependence between the variables of inflows and between the observations of these variables makes it possible to quantify the risk of certain extreme events such as the persistence of low flows or floods. Indeed, considering the dependence between these variables can increase the precision of the estimation, and from a financial point of view, a correct forecast of these events can help save a lot of money for companies like Hydro-Québec and Manitoba Hydro. With the grant, Nasri was able to complete five articles that have all been accepted and/or published in peer-reviewed journals, a significant achievement.

**Nanwei Wang**
Supervisor name: Laurent Briollas
Title of project: Bayesian mixed graphical model learning with application to genomic data integration
Location/University of project: Mount Sinai Hospital
Current: Postdoctoral Fellowship at Pompeu Fabra University in Barcelona.

Summary: This project was designed to address the problems surrounding the integration of complex genomic data to make proper influences. An example of such a scenario would be the identification and isolation of specific genes and gene interactions that are correlated to some cancer diseases. This research could provide new insights into the real world applications of complex data, with a great potential to incite meaningful change in the health sciences. Through the studying and application of Bayesian model selection algorithms, and the usage of mixed graphical models to model the mixed types of variables in genomic data, the goals set out for the project were achieved. Beyond this, the new mixed graphical models were fitted on TCGA cancer data, and reasonable genomic networks were found. As a direct result of this academic inquiry, a scientific paper is currently being submitted to the Journal of Applied Statistics, and Wang's career is furthered.

**Linda Mhalla**
Supervisor name: Debbie Dupuis
Title of project: Causal discovery of extreme events
Location/University of project: HEC Montréal
Current: Postdoctoral Fellowship advised by Debbie Dupuis at HEC Montréal

Summary: This project studied the causal discovery of extreme events, in particular events that are unlikely to occur but would have disastrous consequences. The project began with two goals. One goal was to adapt certain causal inference methods based on conditional quantiles at the tails of bivariate distributions. This part of the research was successful. The second objective was to develop a new causal inference method for extreme values in multivariate data where the co-occurrence of these events has a common cause or an underlying cause. This objective was not completed. This project helped to increase our understanding of the consequences of extreme events, whether they be financial or climate-related. In addition, the research helped identify the most important interactions governing the co-occurrence of these types of events.

**Luc Villandré**
Supervisor name: Jean-François Plante
Title of project: A Bayesian estimation algorithm for spatiotemporal big data
Location/University of project: HEC Montréal
Current: Postdoctoral Fellowship is continuing for a second year

Summary: Satellite imagery has greatly increased our ability to create maps containing detailed geophysical characteristics of the Earth. Ideally, we'd have a detailed portrait for any grid on the map. Unfortunately, heavy cloud cover can cause missing data. Statistical methods to impute the missing data for this volume of spatiotemporal data still need to be developed. Villandré's research is developing an algorithm for the statistical estimation of spatiotemporal big data. The algorithm has been named INLA-MRA, which combines the names of the two methods used: INLA for integrated nested Laplace approximation and MRA for multi-resolution approximation. Villandré's research has shown that this algorithm produces estimates that are very close to the actual values. The algorithm produces a quality prediction that is surprisingly robust. A prototype of this algorithm is available on Github and can be used on any spatiotemporal data set.

**Whitney Huang**
Supervisors: Francis Zwiers and Adam Monahan
Title of Project: Estimating Concurrent Climate Extremes Using the CanRCM4 Large Ensemble
Location/University: University of Victoria, Pacific Climate Impacts Consortium
Current: tenure track professor at Clemson University

Summary: Concurrent extreme events are the simultaneous occurrence of extreme values for multiple variables. Examples include coastal storm surge and heavy precipitation, compound drought and heat, and concurrent wind and precipitation extremes. These events can produce some of the largest impacts on both human society and environmental systems and therefore it is crucial to properly assess the risk of these events. During this CANSSI postdoctoral fellowship, Whitney Huang worked on developing a statistical modeling framework for estimating concurrent extremes. Specifically, a conditional approach (by conditioning the timing of the concurrent extremes of interest) that combines advanced statistical techniques from multivariate extreme value modeling and non-parametric quantile regression to tackle this problem was proposed. Huang also made use of the CanRCM4 large ensemble not only to better estimate the concurrent extremes and their changes but also to assess the statistical performance of the proposed methods. This work was recognized by being invited to several conference presentations.

## Started in 2019

**Dengdeng Yu**
Supervisor name: Dehan Kong
Title of project: Integrating different modalities of baseline imaging data to predict the risk of Alzheimer's disease progression
Location/University: University of Toronto

**Haolun Shi**
Supervisor name: Jiguo Cao
Title of project: High Dimensional Functional Data Analysis
Location/University: Simon Fraser University

**Mamadou Yauck**
Supervisor name: Erica Moodie
Title of project: Causal inference for network data
Location/University: McGill University

**Eric Rose**
Supervisor name: Erica Moodie
Title of project: Design of data analyses of optimal dynamic treatment regimes from observational data, and the development of sensitivity analyses to investigate violations of the typical assumptions required for such analyses.
Location/University: McGill University

# Conferences and Workshops

**Analysis of Life History Data with Multistate Models, April 8, Calgary, AB**

This workshop was hosted at the University of Calgary, with the presenters being Richard Cook and Jerry Lawless from the University of Waterloo. There were 130 participants with 65 who attended in person and 65 who joined online. The majority were from across North America, with additional participants joining from Europe and Asia. The objective of this conference was to not only impart knowledge about the benefits of multistate research models, but also to help proliferate the usage of this analysis framework. Multistate models are especially applicable to the medical field, where continuous records over time will show a patient's transitions through states, effectively recording long-term illnesses or chronic health issues. Due to the nature of the conference, over 35% of participants were involved with the health science field, one of the most applicable fields for the information given in the seminar.

**Statistics Graduate Student Research Day, April 25, Toronto, ON**

Hosted by the Department of Statistical Sciences and Statistics Graduate Student Union at the University of Toronto, this conference encouraged thought and discussion on the impacts of new research methods on the current state of statistical analysis. There were 74 attendees. Talks were presented by were spoken to by four invited lecturers:

- Richard Samsworth, from the University of Cambridge.
- Bhramar Mukherjee, from the University of Michigan.
- Edward George, from the University of Pennsylvania.
- Elizabeth Schifano, from the University of Connecticut.

In addition, there were talks by three graduate students and one postdoctoral fellow from UofT.

**2019 Atlantic Causal Inference Conference, May 22-24, Montreal, QC**

The Atlantic Causal Inference Conference (ACIC) is a yearly conference which has been held in the US until this year. This year's edition, held at McGill University and hosted by the departments of Epidemiology, Biostatistics, and Occupational Health, and Mathematics and Statistics, was a landmark success. This three-day conference kicked off with a full day of workshops, ranging from an introductory lecture on instrumental variables, to much more complex topics, such as Graphical Model Identification Theory for Causal Inference. The 290 conference participants were given seminars and lectures by 53 speakers. To add to the impact, the conference organizers have observed that local research has improved due to the exposure to recent technological and methodological advancements that were shown during the conference.

**Industry-Academic Big Data Analytics, May 3, Kelowna, BC**

The Big Data Analytics Workshop, held at University of British Columbia – Okanagan, was designed to bring members of industry, academia, students, and practitioners in the area of big data together. With over 170 participants from varied backgrounds, a deep and diverse group of attendees participated in a short, speed networking session that was extraordinarily effective, furthering the goal of fostering new collaborations and partnerships between industry and academia. Students have already been interviewed for internship/full-time

positions as a direct result of the exposure from the conference, as well as faculty members being involved with corporate entities to arrange partnership grants.

**R à Québec 2019, May 13-16, Québec City, QC**

R in Quebec (RAQ) 2019 is the second hosting of this conference, which is the largest North American French conference of its kind. This year, it was held at Université Laval. RAQ is an interdisciplinary conference aimed at bringing together members of all areas of employment to learn together and foster growth. RaQ is a conference entirely designed on providing workshops and training in R, one of the fastest growing languages in statistical fields.

Nearly 300 individuals registered for the conference, who were given an inspiring lecture about the evolution of data analysis and the appropriate languages to assess them. Keynote speaker Professor Bates is particularly well known for his work on linear mixed models, and as a member of R core. With 19 workshops, 23 conferences and 2 master classes, this two day conference was crammed full of important information relevant to anyone in the field, and most importantly was available in French, something that is highly underrepresented in the North American statistics society. Giving French-speaking individuals an opportunity to network and form new academic relationships is a unique benefit of this conference, and one that is of high importance.

**Canadian Statistics Student Conference 2019, May 25, Calgary, AB**

The Canadian Statistics Student Conference (CSSC) was initiated seven years ago to raise student and graduate involvement in the field of statistics, as well as increasing participation in the annual SSC meeting among students. The conference at the University of Calgary consisted of multiple undergraduate and graduate speakers, a workshop on computational statistics and many presentations addressing skills that are imperative to students' futures, such as networking, professional opportunities, and career development. The student talks allowed current students to showcase their work to their peers and receive invaluable feedback that will help them develop in the future. This impact was further enhanced by the low stress environment which heavily increased student participation and involvement, one of the main intentions of the CSSC. The conference was closed out by the keynote speaker Charmaine Dean, the first female keynote speaker to present at the CSSC. Dean is the Vice-President for Research at the University of Waterloo and spoke on her experience with interdisciplinary work and how she became a successful leader in her field.

**Data Science Boot Camp, June 10-21, Saskatoon, SK**

The Data Science Boot Camp, hosted at the University of Saskatchewan, is a Summer School education seminar, in which 50 students and professionals were exposed to some core areas of data science. The three main courses offered each offered 11 hours of lectures, on the topics of Machine Learning, Analysis of High Dimensional Data, and Data Visualization, along with four additional case studies which added 20 hours of instruction in total. Participants are expected to have gained a working knowledge of data science, including basic concepts and ideas, problem solving skills, computer programming, and in general, tools for extracting and communicating meaningful information from complex data. The participants worked and studied together, while being actively involved in social events, which may lead to networking

for their future careers. The presenters were sourced from many leading universities across Canada, as well as Arizona and Texas.

**The 2019 Classification Society Annual Meeting, June 19-22, Edmonton, AB**

The 2019 Classification Society annual meeting took place at MacEwan University, featuring talks from North American and European-based researchers specializing in the areas of classification. The program featured not only established researchers with academic rank but also students from both the undergraduate and graduate levels in both oral and poster settings. The range of topics was diverse, including presentations on mixture modelling methodology, implementation, and applications, on record linkage problems using Irish census records, on Bayesian clustering approaches for discrete data, and on applications of clustering methodology in studies of sleep health in older adults. All of the speakers presented their work in lecture style, with 50 minute keynote speeches of and invited 25 minute lectures.

The goal of the Classification Society has always been to provide North American researchers an opportunity to share their research and work with others. Following this aim, the 2019 Annual Meeting was the first to be held in Western Canada, reaching new demographics and forming new connections. With over half the Canadian attendants being from western provinces, this goal was achieved, and benefitted the attendants immensely.

**2019 Digital Humanities Summer Institute**

CANSSI supported Professor David Campbell (Carleton University) who presented a week-long short course at the DHSI on statistics and data science topics.

**2019 Diversity in Mathematics, July 21-August 2, Vancouver/Burnaby BC**

This two-week camp was held at two major British Columbia institutions, the University of British Columbia and Simon Fraser University. Prioritization of Diversity is something very important to the statistical field, and this year's camp was titled “Undergraduate Women’s Summer School”, showing a focus on an important, yet underrepresented group in all STEM fields. With 23 undergraduate and high school participants attending the camp this year, who all experienced multiple lectures by experienced professionals in the field, as well as day trips to the corporate headquarters of adMare BioInnovations and DWave Systems, both of which are cutting edge technology companies who use higher level statistics in their operations. adMare is a pan-Canadian enterprise with global reach, which source promising research to create new companies of scale and encouraging growth of already existing companies into Canadian anchors. DWave Systems is the leader in the development and delivery of quantum computing systems, software, and services.

On these field trips students were exposed to the real-world applications of their field of interest, potentially inspiring those same students to pursue careers in statistical fields. 10 lecturers, both from the forefront of academic society and from private industry, joined with representatives from the likes of Amazon to highlight the importance of statistical methods within successful businesses. Participant responses to a questionnaire about the nature of the camp were overwhelmingly positive, with many individuals taking the time to include small changes that could make the camp even better, both in attracting participants and providing a better experience to all further participants.

**NISS Writing Workshop for Junior Researchers, July 28-30, Denver, CO**

The National Institute of Statistical Sciences (NISS) Writing Workshop for Junior Researchers is a yearly workshop in which recent PhD graduates are given necessary instruction on journal writing and the submission of grant proposals. Participants were asked to provide a recent sample of their writing, which was reviewed and critiqued, in order to provide applicable and individualized assistance to every participant. Beyond the written assistance provided, there were also two tutorials that covered scientific writing formats, as well as proper paper organization. Other topics included ethical issues, issues of journal choice, and how to properly respond to reviewer's comments.

The second day was more centred on specific issues covering the writing process and details of grammar, sentence structure, and word choice, which is especially helpful for any participants whose native language is not English. This workshop in particular will have a direct, meaningful impact on every one of the attendants, who will now be better equipped to write persuasive applications for grants, while also having their research papers structured in a way that may lead to a higher chance of publication.

**4th ICSA - Canada Chapter Symposium, August 9-11, Kingston, ON**

The 4th ICSA-Canada Chapter symposium was hosted at Queen's University and consisted of two plenary talks, 26 invited sessions, 1 contributed session, and a one-day workshop with two short courses on modern high-dimensional and statistical learning methods. The scientific sessions covered a wide range of statistical topics including Bayesian statistical methods, functional data analysis, statistical methods in biostatistics, and high dimensional data. The Annual General Meeting of ICSA-Canada Chapter took place on the first day. The members had a long discussion on the current state of the organization and future plans.

The intended outcome of this event is to develop and grow the Canadian chapter, strengthen the connections between chapter members and the wider statistical community, especially fast-growing regions such as Asia, as well as the communities from other disciplines. The event promoted the practical and theoretical development of statistical methodology, and the training of highly qualified personnel.

**2019 Variational Analysis and Optimization Workshop, August 12-13, Kelowna, BC**

The 2019 Variational Analysis and Optimization Workshop was hosted at the Kelowna campus of the University of British Columbia, in which nearly 40 attendees were given lectures by the 20 invited speakers, coming from Australia, Canada, China, New Zealand, and USA. Each day had 10 invited talks. On Monday, the two plenary speakers were Stephen Simons from the University of California, and B. Mordukhovich from Wayne State University. On Tuesday, the two plenary speakers were Regina Burachik from the University of South Australia, and Michel Thera from the University of Limoges, France. Eighteen of the talks are focused on Convex Analysis, Variational Analysis, and Optimization. The remaining two talks on Statistics are given by John Braun from the UBC Okanagan and Julie Zhou from the University of Victoria.

The talks covered recent progress on monotone operator theory, splitting algorithms, nonconvex optimization, optimal control, R-optimal designs, and data sharpening in Statistics. Participants discuss research problems and exchange ideas during coffee breaks, lunch times, and other discussion periods throughout.

**Nexus 2019 - International Data Science Conference, November 14, Winnipeg, MB**
The International Data Science Conference took place at the University of Manitoba, which featured three keynote speakers, who are all prominent members of industries that rely on statistics and data sciences greatly. Joel Semeniuk, the opening keynote, is the CEO & Founder of HORIZONthree, who spoke on Innovation, and how the Fourth Industrial Revolution is necessitating a complete refresh of how we innovate in the modern environment. The luncheon keynote speech was delivered by Grant Barkman, the President and CEO of Decision Works, whose speech was titled "Vision into Action - Art of Now". Barkman spoke on the impacts of technology on our everyday lives, and how the monitoring of individuals through their devices can be utilised by data scientists to find deep patterns and predict human action. He also spoke on the many ethical issues and complications that come from privacy concerns, and promoted growth in a positive direction. The closing keynote was given by Tyler Baadley, the Vice President of Westcap Management, a venture capital & private equity fund manager. He spoke on the many options for the commercialization and industrialization of AI and machine learning. Between each keynote speech, there were panel discussions and concurrent sessions, which gave all attendees a chance to table their ideas, and participate in meaningful exchanges between peers in their fields.

**Ottawa Hockey Analytics Conference 2019, November 15-16, Ottawa, ON**
The 5th annual Hockey Analytics conference was hosted once again at Carleton University. This year the attendants of the conference participated in a workshop, followed by scientific talks, panels, and poster competitions for undergraduate and graduate students. The winning students had the opportunity to present at the conference, as well as submitting their expanded work for publication. Continuing this beneficial opportunity, there has been a history of presenting students acquiring funding from many sources due to their presentation topic. The conference's continued success has garnered quite a following and social media presence, which only continues to benefit all individuals involved.

**Vancouver Machine Learning: Genomics (VanML 2019), December 16, Vancouver, BC**
The inaugural Vancouver Machine Learning: Genomics (VanML 2019) conference was held on December 16 at the Harbour Centre. The conference involved invited talks from six researchers from Canada, USA, UK and Norway. At the conclusion of the invited talks, Caroline Colijn , Canada 150 Research Chair in Mathematics for Infection Evolution and Public Health, led a discussion period in which the audience explored topics and questions raised by the speakers.

Lecture topics covered by the speakers included population genetics (including ancient human genomes), sequential Monte Carlo methods and probabilistic programming for phylogenetics, machine learning methods for predictive genome wide association studies, pathogen studies (including visualizations of and geographic aspects of epidemics), and proteomics. The main thread connecting these topics was machine learning methods for genomics and bioinformatics, and Caroline Colijn's discussion focused on issues of model uncertainty in genomics. This conference achieved its two goals, being to engage researchers from SFU with the international community and to provide a diverse and accessible environment for student attendees.

## Student Support for SAMSI Workshops

CANSSI supported 4 students who attended Undergraduate Workshops at SAMSI. Two attended in February and two in May. The Undergraduate Workshop in February began with tutorials on R and MATLAB, then touched on different topics in statistics research. It also included a panel on career opportunities. The second day of the workshop had the students working on team projects and presenting their work at the end of the day. At the May Undergraduate Modeling, students worked in small groups on specific problems drawn from several fields. Problem topics included predictive modeling for tropical cyclones with historical storm data, analysis of the heart disease data, imaging genetics studies on Alzheimer's disease, differential equation models which can be used for modeling biological population growth, stock market fluctuations, and heat transfer, and forecast the impact of the implementation of the Patient Protection and Affordable Care Act and the Health Care and Education Reconciliation Act. The groups used either MATLAB or R to do the computing for their analyses.

## Undergraduate Datathons

**University of Toronto Sports Analytics Hackathon, January 19, Toronto, ON**
The second annual University of Toronto Sports Analytics Hackathon was a great success. The hackathon brought together students from various academic backgrounds interested in data science, statistics, and machine learning and gave them the opportunity to work on a real-world dataset under the mentorship of experts in these fields. The hackathon was structured so that every team presented to the judges individually and then the judges deliberated and selected the finalists who gave longer presentations in front of everyone. The initial presentations gave participants, mentors and the organizing committee the opportunity to talk with teams not currently presenting to judges about their work. The biggest highlight at this stage of presentations was a surprise appearance from Glen Grunwald, president of Canada Basketball. Many teams got to present their work to him one on one.

**Data Miners 2019, March 23-24, Montreal, QC**
This first Data Miners datathon, organized by the Data Science for Business (DSB) committee was very successful. This event was an attempt to bridge the divide between universities and industry, strategy and data science, as well as organizations and promising new talent. This event brought together students, faculty and industry representatives in data science from all over Montreal. 77.5% of the participants were undergraduate students (contrary to what the organizers expected). 32.5% of the participants were women. Students came primarily to learn more about data science tools. Participants were able to submit their predictions to obtain real-time feedback on their model's performance. The teams all used different programming tools to work on the case study, such as Excel, R, SAS and Python. The DSB committee hopes to organize a second Data Miners datathon, with improvements based on feedback from this year and even more participation. The DSB was able to recruit majority of the participants to take part in their other activities later in the year, including R and Python programming workshops, visits to businesses and the HEC Forecast Conference.

**ASA DataFest, March 29-31, Edmonton, AB**

Hosted at MacEwan University, this event is a collaboration with the University of Alberta. In addition to CANSSI, event sponsors included Alberta Innovates, PIMS, The Canada Revenue Agency, Pearson Publisher, University of Alberta (U of A), and MacEwan University. 42 students organized in 9 teams, 5 from U of A and 4 from MacEwan, analyzed data on Women's Rugby, presented their findings, and received awards during this intense 48-hour event. Students were supported by volunteers including graduate students from the U of A, faculty members from U of A and MacEwan, and 4 members from the CRA. The panel of judges were Yan Yuan (Public Health, U of A), Michael Buro (Computing Science, U of A), Rui Hu (Math & Stats, MacEwan), Indratmo (Computer Science, MacEwan), Eric Jinsan Yu, and Bronwen Russell from the CRA.

**ASA DataFest, May 3-5, Toronto, ON**
Held at the University of Toronto, this fourth ASA DataFest was very successful. DataFest is like a hackathon, for undergraduate students, except the problem is a data analysis problem, rather than a programming problem. Teams of students get a dataset on Friday afternoon and work on the problem until Sunday afternoon when they present their results. After two days of intense data wrangling, analysis, and presentation design, each team is allowed a few minutes and no more than two slides to impress a panel of judges. Prizes are given for Best in Show, Best Visualization, and Best Use of External Data.

A key feature of ASA DataFest is that it brings together the data science community. Undergraduate students do the work, but they are assisted by roving consultants who are graduate students, faculty, and industry professionals.

**CANSSI National Case Study Competition, September 3 – November 2, Across Canada**
This year's CANSSI National Case Study Competition (NCSC) challenge was to create a better customer experience for BC Ferries with sailing analysis and delay prediction in BC Ferries sailings around Vancouver harbours. The dataset consisted of 61,880 sailings occurring between August 2016 and March 2018. The task was to predict delay of sailings. In the regional competitions and national poster championship, students were judged based on the accuracy of their delay predictions (percent correct), and also a report in which they discussed their methods and results and additional insight about the data provided by their analysis.

The CANSSI NCSC was open to all students enrolled in undergraduate or graduate programs at Canadian universities. The data for this competition was made available on September 3, and students were able to submit their solutions online until October 2 for feedback on their prediction accuracy. Carleton University, Concordia University, MacEwan University, Simon Fraser University and the University of New Brunswick hosted regional. Winners of the regional competitions were invited to compete in the final national poster championship at Simon Fraser University at the CANSSI Headquarters on November 2.

# Administration and Governance

**Directorate**
Don Estep, Simon Fraser University, Scientific Director as of July 2019
Nancy Heckman, University of British Columbia, Associate Scientific Director as of July 2019
W. John Braun, University of British Columbia – Okanagan, Deputy Director
Joanna Mills Flemming, Dalhousie University, Associate Director, Atlantic

Yogendra Chaubey, Concordia University, Regional Director, CANSSI Québec as of July 2019
Lisa Strug, University of Toronto, Regional Director, CANSSI Ontario as of August 2019
Mohammad Jafari Jozani, University of Manitoba. Associate Director, Prairies
Karen Buro, MacEwan University, Associate Director, Alberta-British Columbia

*Nancy Reid, University of Toronto, Scientific Director until June 2019*
*Erica Moodie, McGill University, Associate Director, Québec until June 2019*
*Paul McNicholas, McMaster University, Associate Director, Ontario until June 2019*

**Staff**
Lori Kroeker, Financial Administrator
Angela Plagemann, Scientific Coordinator

**Board of Directors**
Wesley Yung, Statistics Canada (Chair)
Paul Kovacs, Western University (Vice-Chair)
Michael Boehnke, University of Michigan
Don Estep, Simon Fraser University
Mark Girolami, Imperial College London
Joan Hu, Simon Fraser University
Adrian Levy, Dalhousie University
Lisa Lix, University of Manitoba
Farouk Nathoo, University of Victoria
Nancy Reid, University of Toronto
Laxmi Sushama, McGill University

**Scientific Advisory Committee**
Amy Braverman, Jet Propulsion Laboratory
Raymond Carroll, Texas A&M University
Don Estep, Simon Fraser University
Yulia Gel, University of Texas at Dallas
Xuming He, University of Michigan
Sallie Keller, Institute of Virginia Tech
Rogemar Mamon, Western University
Doug Nychka, Colorado School of Mines
David Stephens, McGill University
Marina Vannucci, Rice University

## CANSSI Member Institutes

Acadia University
Brock University
Carleton University
Concordia University
Dalhousie University
HEC Montréal
MacEwan University
McGill University
McMaster University
Memorial University
Queen's University
Simon Fraser University
Université de Montréal
Université de Sherbrooke
Université du Québec à Montréal
Université Laval
University of Alberta
University of British Columbia
University of British Columbia Okanagan
University of Calgary
University of Guelph
University of Manitoba
University of New Brunswick
University of Ottawa
University of Regina
University of Saskatchewan
University of Toronto
University of Victoria
University of Waterloo
University of Windsor
University of Winnipeg
Western University
Wilfrid Laurier University
York University

# Annual General Meeting Highlights

The CANSSI Annual General Meeting took place at McGill University on May 25, 2019. A quorum of the members of the corporation were present. Members passed several motions including an amendment to the By-Law, the election of 6 members of Board of Directors, and the approval of the financial statement.

This was Nancy Reid's last meeting as Scientific Director of CANSSI. After she highlighted many of CANSSI's achievements during her 5-year term, the members of the corporation put forward the following motion which was enthusiastically and unanimously carried: It was resolved that the Members express their sincere thanks to Nancy Reid for her leadership setting CANSSI on a solid path to continued growth, during her term as the Director.

# CANSSI Media

CANSSI's main method of communication is via websites: in English, www.canssi.ca, and in French, www.incass.ca. Information about programs, deadlines and news items can be found there. The website also includes advice for new researchers as well as lists of employment ads from member institutes and postdoctoral opportunities across the country. We send periodic emails to member institutes as well as those who have opted into the CANSSI mailing list. We also use the SSC's mailing list to send out information about the opportunities. We have a Facebook page, a Twitter feed and a YouTube channel. We post deadline reminders and program information on Facebook and Twitter and the YouTube channel features videos from Distinguished Visitor Lectures.

# CANSSI Financial Information

## Research Grants

The fiscal year ending March 31, 2019 was to be the last of a five-year funding arrangement in which CANSSI directs the use of a portion of research funds provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Ontario Ministry of Research and Innovation (MRIS) to the following Canadian math institutes: Pacific Institute for the Mathematical Sciences (PIMS), Fields Institute for Research in Mathematical Sciences (Fields) and Centre de Recherches Mathématiques (CRM). The funding arrangement has now been extended to the end of March 2021.

**Expenditures for the year ended March 31, 2019**

| | PIMS | CRM | FIELDS | Total |
|---|---|---|---|---|
| **Research** | | | | |
| CRTs | 149,805 | 36,401 | 240,223 | 426,429 |
| Postdoctoral fellowships | 52,500 | 28,568 | 22,500 | 103,568 |
| Short programs and workshops | 39,673 | | 34,622 | 74,295 |
| | 241,979 | 64,969 | 297,345 | 604,292 |
| Scientific Coordinator | 67,142 | | | 67,142 |
| **Total Expenditures** | **309,120** | **64,969** | **297,345** | **671,434** |

### Summary of Funding and Expenditures for the 5 years ended March 31, 2019

| | PIMS | CRM | FIELDS |
|---|---|---|---|
| Five Year Funding Commitment (Ended March 31, 2019) | **1,000,000** | **525,000** | **1,000,000** |
| Cumulative Expenditures to March 31, 2018 | 677,704 | 470,839 | 666,606 |
| Expenditures for Year Ended March 31, 2019 | 309,120 | 64,969 | 297,345 |
| | 986,824 | 535,808 | 963,951 |
| **Remaining Balance at March 31, 2019** | **13,176** | **(10,808)** | **36,049** |
| Retroactive Increase from NSERC: 2018-2019 | 19,900 | 10,450 | 9,800 |
| 2019-2020 Funding Commitment | 219,900 | 115,450 | 209,800 |
| **Total Funds Available to March 31, 2020** | **252,976** | **115,092** | **255,649** |

## CANSSI Inc.

### Income Statement for the Year Ended March 31, 2019

| | |
|---|---|
| **REVENUE** | |
| Institutional Memberships | 120,000 |
| Simon Fraser University | 8,467 |
| University of Toronto Department of Statistical Sciences | 5,000 |
| **Total Other Revenue** | 13,467 |
| **TOTAL REVENUE** | 133,467 |
| **EXPENSE** | |
| **Scientific** | |
| Health Sciences Collaborating Centres | 25,000 |
| Short programs, workshops and conferences | 33,155 |
| **Total Scientific Expenses** | 58,155 |
| **General & Administrative Expenses** | |
| Board Expenses | 12,480 |
| Communications & Promotion | 6,617 |
| Deputy Director | 10,000 |
| Director's Expenses | 5,369 |
| Operating | 6,108 |
| Salaries | 27,416 |
| **Total General & Admin. Expenses** | 67,990 |
| **TOTAL EXPENSE** | 126,145 |
| **NET INCOME** | 7,322 |

## Relationship with Simon Fraser University

The University is now the host site for CANSSI and provides access to certain facilities as described in the collaboration agreement between SFU and CANSSI, dated June 15, 2018. In addition, the University, through the Big Data Hub, provides administrative support for communications, event coordination, and business development. The university also established an internal research account, in the amount of $98,900 per fiscal year. The term of the collaboration agreement is 5 years.

# Thank You from CANSSI

The Natural Sciences and Engineering Research Council of Canada (NSERC)
Centre de recherches mathématiques (CRM)
The Fields Institute for Research in Mathematical Sciences
The Pacific Institute for the Mathematical Sciences (PIMS)
The Statistical Society of Canada (SSC)
and our institutional members:
Acadia University
Brock University
Carleton University
Concordia University
Dalhousie University
HEC Montréal
MacEwan University
McGill University
McMaster University
Memorial University
Queen's University
Simon Fraser University
Université de Montréal
Université de Sherbrooke
Université du Québec à Montréal
Université Laval
University of Alberta
University of British Columbia
University of British Columbia Okanagan
University of Calgary
University of Guelph
University of Manitoba
University of New Brunswick
University of Ottawa
University of Regina
University of Saskatchewan
University of Toronto
University of Victoria
University of Waterloo
University of Windsor
University of Winnipeg
Western University
Wilfrid Laurier University
York University