

# CURRICULUM VITAE

Hong Gu  
Professor, Department of Mathematics and Statistics  
Dalhousie University  
6316 Coburg Rd. PO BOX 15000  
Halifax, NS, B3H 4R2, Canada

E-mail : hgu@mathstat.dal.ca  
Telephone Number: 1-902-4947161

## EDUCATION:

- 1985-1989      B.Sc. Peking Normal University, Department of Mathematics, majoring in Mathematics.
- 1991-1994      M.Sc. Peking University, Department of Probability and Statistics. Thesis topic: Mechanism of Stochastic Resonance in Bistable System.
- 1995-1999      Ph.D. University of Hong Kong, Department of Statistics and Actuarial Science. Thesis topic: Influence Diagnostics in Principal Components and Canonical Analyses.

## EMPLOYMENT HISTORY

- 1989, Aug.-1991, July      Teacher in Datong No. 3 Middle School, Datong City, Shanxi Province, China.
- 1994, Aug.-1995, Nov.      Lecturer in the Department of Mathematics & Mechanics, Beijing University of Science and Technology, Beijing.  
Courses taught:  
Advanced Mathematics, Mathematical Analysis
- 1995, Dec.-1998, Nov.      Tutor in the Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong.  
Courses tutored:  
Introductory Statistics, Elementary Statistical Method  
Applied Mathematics-Statistics, Probability and Statistics,  
Statistical Methods in the Physical Sciences.

- 1998, Dec.-1999, July      Research Assistant. Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong.
- 1999, Aug.-2001, July      Post-doc fellow and Part-time lecturer. Dept. of Statistics and Actuarial Science, University of Waterloo.  
Courses taught:  
Stat 330: Statistical Theory and Methods  
Stat 202: Statistics for the Life Sciences
- 2001, July- 2002, Mar.      Post-doc fellow and Part-time lecturer. Dept. of Math. & Stat., Dalhousie University
- 2002, Apr.- 2007, June      Assistant Professor. Dept. of Math. & Stat., Dalhousie University.
- 2007, July- 2016, June      Associate Professor. Dept. of Math. & Stat., Dalhousie University.
- 2016, July-                      Professor. Dept. of Math. & Stat., Dalhousie University.

### **Research Grants and Awards:**

- 2020, 9-2020, 12, Mitacs Globalink Research Award, \$6000 in total (Yun Cai on dentistry project).
- 2020, 8-2021, 2, Mitacs Globalink Research Award, \$6000 in total.
- 2019, 9-2023, 8, Mitacs accelerate-PVSC, \$106,666.67 in total.
- 2019, 6-2020, 2, Mitacs -EhEye Inc. (Patriot One Technologies), \$15K in total.
- 2018, 10-2019/3, NSERC Engage, \$25K in total.
- 2017-2022, NSERC Discovery, \$120K in total.
- 2016, 5-2016, 10 NSERC Engage, \$25K in total.
- 2012-2017 NSERC Discovery, \$60K in total.
- 2006-2012 NSERC Discovery, \$90K in total.
- 2002-2006 NSERC Discovery, \$62K in total.
- 2002-2007 Startup funding, Faculty of Science, \$20K.
- 2003-2008 CFI New Opportunities fund, \$1.9million, with 6 others in Faculty of computing science, (PI:Theodore Chiasson).

- 1998: Wang Guangwu Scholarship, The University of Hong Kong.
- 1995-1998: Fully funded Ph.D. scholarship from the Hong Kong government.
- 1997: Conference grant award from the University of Hong Kong for presenting paper at *the Joint Statistical Meetings*, held in Anaheim, Orange County, USA.
- 1994: Award for Excellent M.Sc. Dissertation, Peking University.
- 1993: Guang Hua Award for Outstanding Postgraduate Student, Peking University.

### **Student supervision and co-supervision**

#### **Current HQP:**

Shanglun Li, PhD candidate. (2021/1-). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie). Microbial dynamics modelling.

Paul Bjorndahl, PhD candidate. (2020/1-) Co-supervised with Joseph Bielawski. (Biology Dept. in Dalhousie). Modelling the dynamics of subcommunities of ocean microbiome.

Xinyue Zhang, PhD candidate. (2019/1-). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie).

Chaoyue Liu, PhD candidate. (2016/5-). Co-supervised with Robert Beiko (Computer science). Pattern recognition in microbial ecology.

Lihui Liu. PhD candidate. (2014/1-). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie) and Johan Van Limbergen (Division of Pediatric Gastroenterology & Nutrition, IWK). Joint statistical modelling of metagenomic-genetics data and inference of their associations with inflammatory bowel disease.

Ling Shen. PhD candidate. (2014/9-). Co-supervised with Chris Field. Statistical data mining methods for multi-class problems, with applications to aid in diagnosis of Emergency Department Patients.

Yun Cai, PhD candidate.(2014/9- ). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie). Non-negative matrix factorization, with application in analysis of microbiome data.

#### **Post-doctorate supervised:**

Wei Zhou, Postdoc fellow. (2018/8-2021/9). Co-supervised with Bielawski (Biology Dept. in Dalhousie). Prediction for the upcoming cyanobacteria bloom in lakes based on the lake microbial community analysis.

Mahdi Shafiei, Postdoc fellow. (2011/8-2012/8). Co-supervised with Bielawski (Biology Dept. in Dalhousie). Network modelling of complex microbial communities.

#### **Graduated PhD Students:**

Melanie Abeysundera. PhD. (2007-2011). Co-supervised with Chris Field. *Phylogenetic analysis of multiple genes based on spectral methods*. Employee of Statistics Canada.

Xiaofei Shi, Ph.D. (2001-2006). Co-supervised with Chris Field. *Phylogenetic inferences and pattern classification*. Statistical Analysis Manager, ebays, China.

#### **Graduated MSc Students:**

Mia Parenteau, MSc candidate. (2018/9-2020/12). Violence detection in crowd footage: statistical features using transformed optical flow.

Wanru Jia, MSc candidate. (2018/9-2020/12). Co-supervised with Toby Kenney. Edge Detection Operators for X-ray Images Based on Hessian Matrices.

Mary Molly Hayes, MSc (2017/9-2020/5) Co-supervised with Morgan Langille (Pharmacology, Dalhousie University). Cross-Study Analyses of Microbial Abundance Using Generalized Common Factor Methods.

Paul Bjorndahl, MSc (2017/9-2019/12) Co-supervised with Joseph Bielawski. (Biology Dept. in Dalhousie). Statistical Approaches for Spatial-Temporal Dynamics of Microbial Communities in the Red Sea.

Junqiu Gao, MSc (2017/9-2019/8) Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie). Ornstein-Uhlenbeck Process and Optimal Sampling for Analysis of Microbiome Data.

Tianshu Huang, MSc. (2015/9-2017/9). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie). Semi-parametric Principal Component Analysis for Poisson Count Data with application to microbiome data analysis.

Rana Bashwih, MSc. (Bioinformatics) (2013/9-2016/8). Co-supervised with Joseph Bielawski (Biology). Inference and investigation of marine microbial community structures in the global oceans.

Chang Chen, MSc. (2014/9-2016/8). Co-supervised with Robert Beiko (Computer science). Negative binomial modelling and applications for microbial count data.

Hao He, MSc. (2014/9-2016/4). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie). Robust Ranking and selection with heavy-tail priors and its applications in market basket analysis.

Chongci Tang, MSc. (2014/9-2016/4). Co-supervised with Joseph Bielawski (Biology). Statistical approaches for matching the components of complex microbial communities.

Chaoyue Liu, MSc. (2014/9-2016/4). Co-supervised with Robert Beiko (Computer science). Gene clustering based on co-occurrence with correction for common evolutionary history.

Li Li, MSc (2011/9-2016/4). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie). A likelihood based clustering method for detection of recombination for DNA and Amino Acid sequences. Statistical Analyst, Canadian Center for Vaccinology.

Yun Cai, MSc. (2013/9-2014/8). Co-supervised with Kenney (Math. and Stat. Dept., Dalhousie). Supervised non-negative matrix factorization for analysis of microbiome data.

Michael Butler, MSc. (2013/9-2014/8). The use of Machine Learning algorithms to aid in diagnosis of adult emergency department patients.

Wei Chen, MSc. (2010/9-2012/12). Co-supervised with Bielawski. The exploration of effect of model misspecification and development of an adequacy-test for substitution model in phylogenetics. Research and Statistics officer, NS Dept. of Health & Wellness.

Wei Dai, MSc. (2010/9-2013/8). Co-supervised with Kenney. *A new test to build confidence region using balanced minimum evolution method*. HR predictive analyst in Honeywell inc..

Jiachi Zhu, MSc. (2010/9-2012/8). Co-supervised with Xu (Dept. of Economics, Dalhousie). Impact of hedging on the firm value and stock return. IAM Technical Assistant/Analyst, Feedlot Health Management.

Caroline Urquhart, MSc (2006-2011). Co-supervised with Bielawski. *Analysis of Prokaryotic Metabolic Networks*. Business Analyst at Johnson Inc.. (Caroline took sick leave from 2007–2010 and came back in the fall term 2010 to finish her thesis.)

Mei Chen, MSc (2006-2007). Co-supervised with Chris Field. *Model comparisons between general time reversible model and Barry Hartigan model*. Senior analyst for Canadian Institute for Health Informatics.

Melanie Abeysundera, (2005-2006). Co-supervised with Chris Field. *Predicting protein structure and functions using spectral techniques*.

Li Li, MSc, (2005-2006). Co-supervised with Joe Bielawski. *Development of spherical data analysis techniques for phylogenomics*. PhD Student at Univ. of Toronto.

Le Bao, MSc, (2004-2005). Co-supervised with Joe Bielawski. *Generalized fixed effect models and likelihood based clustering in codon substitution models*. PhD at Univ. of Washington, now a faculty member in Department of Statistics, Pennsylvania State University.

Chang Dan, MSc, (Dept. of Economics , Essay, 2004-2005). Co-supervised with Xu. *Does Hedging Add Firm Value? Empirical Evidence From Canadian Oil and Gas Companies*. Data analyst in a financial institute in China.

Shuyan Li, MSc. (2004-2005). Co-supervised with Bruce Smith. *Histogram-Based Mutual Information Estimation*.

Krista Collins, MSc. (2003-2005). Co-supervised with Chris Field. *Examining the periodicity in DNA and Amino Acid sequences using the spectral envelope*. Employee of Statistics Canada.

David Campbell, M.Sc. (2002 -2003). *Application of Wavelets in Clustering Time Series*. (Faculty member in Simon Fraser University now).

### **Honours' theses:**

Wang, Jiachen (2020/9-2021/5) Gene networks related to nine potato traits by NMF. (a collaborative work).

Wang, Yuexu (2020/9-2021/5) Gene networks for potato plants by DBSCAN clustering method.

Ma, Jiajun (2020/9-2021/5) Gene networks related to nine potato traits by NMF. (a collaborative work).

Zhang, Ziqi (2020/9-2021/5) Prediction of human body temperature and presence of fever based on thermal camera images.

Renny Doig (2018/9-2019/4). Prediction of multiple traits in potato plants using gene expression profiles.

Mia Parenteau (2018/1-2018/4). Using Gene Expression as a Means to Predict Response in Potato Data.

Amir Farrag (2017/9-2018/4). Honours in Mathematics. Data Mining for Detection of Acetabular Cartilage Delamination in Femoroacetabular Impingement Patients.

Ying Chi (2017/4-2017/8). Principal Component Analysis of Fatty Acids in Atlantic Salmon.

Xinyue Zhang (2017/1-2017/8) GAM and ARIMA models for prediction for the patient emergency visits on different hospital sites.

Molly Hayes (2016/9-2017/4) Common principal component analysis by removing unwanted variance.

Ziwei Jin (2016/5-2017/4) Common linear discriminant analysis (CLDA) as a prediction method by removing unwanted sources of metagenomics variance.

Yike Hou (2016/5-2016/8) Clustering analysis based on the MLE estimates of the phase and periodicity for the time series of ocean microbial marker-gene data.

Qi Li (2016/5-2016/12). Time series modelling and prediction for the patient emergency visits on different hospital sites.

Jing Zhang (2013/9 - 2014/4). Analyzing the Impact of Ocean Acidification on Bacterial Community Structure Using Clustering analysis. Risk analyst, GV Funds.

Li Li (2009-2010). Testing the periodicity of annually reoccurring bacterial communities from ocean conditions. Statistical Analyst, Canadian Center for Vaccinology.

Cuiwei Luo (2009-2010). Predicting the stocks from the market indices.

Sylvia de la Ronde (2007-2008). Co-supervised with Beiko (Computer science, Dalhousie). *Identifying environmental factors in lake bacteria compositions*. Research and Statistical Officer, Government of Nova Scotia, Department of Labour and Advanced Education, Universities and Colleges Division.

Shuai Gao (2007-2008). Linear discriminant analysis in high dimensional space: simulation studies.

Scott Wile (2005-2006). Asymmetric adjustments in Canadian retail gasoline market. Employee of Statistics Canada.

### **Others:**

Junqiu Gao (2019/9-2020/1). Co-supervised with Ying Zhang (Acadia Univ.)  
Simulation and sampling methods on multivariate time series data.  
Xueli Xu (2019/8-2019/10) Visiting PhD student from Nankai Univ..

Libai Xu (2019/8-2019/9) Visiting PhD student from Nankai Univ..

Hao He (2016/5-2016/8). Co-supervised with Chris Field. Model based clustering analysis of the microbial time sequences.

Yanxia Sun, Mitacs Globalink summer intern. (2015/6-2015/8). Metagenomic time series analysis. MSc candidate, Peking Union Medical College in Basic Medicine.

Mengying Wu, Mitacs Globalink Program Student Placement (2014/7-2014/9).  
Project: "Optimization of cost efficiency in multi-stage tests,application in drug-discovery process". Masters of Finance, MIT Sloan School of Management.

Nader Fallah (Visiting PhD Student, 2007-2008). *Nonlinear Poisson regression using neural networks*. Post-doc fellow at UBC.

Emad Bahrami Samani (2006-2007, Honours' thesis supervised distantly, Amirkabir Univ. of Technology, Tehran, Iran). *A Novel Hybrid GMM/SVM Architecture for Protein Secondary Structure Prediction*. PhD Student in University of Southern California.

Le Bao (2005-2006, Research Associate). Co-supervised with Joe Bielawski.  
*Improving the estimation of parameters in codon models*.

### **Papers submitted or under revision:**

Xinyue Zhang, Hong Gu and Toby Kenney. Feature and Structure Selection of Neural Networks.

T. Kenney, H. He, H. Gu. Prior Distributions for Ranking Problems (Under Revision, 42 pages, arXiv: <https://arxiv.org/abs/1610.08779>)

T. Kenney, H. Gu. The Adequate Bootstrap. (Under revision, ArXiv: <https://arxiv.org/abs/1608.05913>)

L. Xu, X. Xu, D. Kong, H. Gu, T. Kenney. Stochastic Generalized Lotka-Volterra Model with An Application to Learning Microbial Community Structures (Under revision for Journal of the American Statistical Association, 52 pages)

### **Publications:**

(note that underlined names correspond to authors who were students at the time)

L. Liu, H. Gu, J. Van Limbergen, T. Kenney. (2020). SuRF: a New Method for Sparse Variable Selection, with Application in Microbiome Data Analysis. *Statistics in Medicine*. 2020;1–23. <https://doi.org/10.1002/sim.8809>

Mia T. Parenteau, Hong Gu, Bernie J. Zebarth, Athyna N. Cambouris, Jean Lafond, Alison Nelson, Judith Nyiraneza, Charlotte Davidson, Martin Lagüe, José Héctor Galvez, Martina V. Strömvik and Helen H. Tai (2020). Data Mining Nitrogen-Responsive Gene Expression for Source–Sink Relations and Indicators of N Status in Potato. *Agronomy* **2020**, *10*, 1617; doi:10.3390/agronomy10101617.

T. Kenney, J. Gao, H. Gu. (2020) Application of OU processes to modelling temporal dynamics of the human microbiome, and calculating optimal sampling schemes. *BMC Bioinformatics*, 21:450 <https://doi.org/10.1186/s12859-020-03747-4>.  
<https://rdcu.be/b8qva>

Kenney T, Gu H, Huang T. (2020) Poisson PCA: Poisson measurement error corrected PCA, with application to microbiome data. *Biometrics*. 2020;1–16. <https://doi.org/10.1111/biom.13384>

Wei Chen, Toby Kenney, Joseph Bielawski and Hong Gu (2019). Testing adequacy for DNA substitution Models. *BMC Bioinformatics* 2019 **20**:349. <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-2905-3>.

Katherine A. Dunn, Toby Kenney, Hong Gu and Joseph P. Bielawski (2019). Improved inference of site-specific positive selection under a generalized parametric codon model when there are multinucleotide mutations and multiple nonsynonymous rates. *BMC Evolutionary Biology*, 2019:22 <https://doi.org/10.1186/s12862-018-1326-7>

Chaoyue Liu\*, Benjamin Wright\*, Emma Allen-Vercoe, Hong Gu and Robert Beiko (2018). Phylogenetic clustering of genes reveals shared evolutionary trajectories and putative gene functions. *Genome Biology and Evolution*, Volume 10, Issue 9, 1 September 2018, Pages 2255–2265, <https://doi.org/10.1093/gbe/evy178>

Moamen Bydoun, Andra Sterea, Henry Liptay, Andrea Uzans, Weei-Yuan Huang, Gloria J. Rodrigues, Ian Weaver, Hong Gu, David M Waisman (2018) S100A10, a



Novel Biomarker in Pancreatic Ductal Adenocarcinoma. *Molecular Oncology*.12(11):1895-1916. doi: 10.1002/1878-0261.

Nidhin Nandhakumar, Ehsan Sherkat, Evangelos E.Milios, Hong Gu and Michael Butler (2017). Clinically Significant Information Extraction from Radiology Report. *Proceedings of the 2017 ACM Symposium on Document Engineering (DocEng 2017)*, p153-162. (refereed conference in CS).

Yun Cai, Hong Gu and Toby Kenney (2017) Learning Microbial Community Structures with Supervised and Unsupervised Non-negative Matrix Factorization. *Microbiome*, 5:110. <https://doi.org/10.1186/s40168-017-0323-1>.

Mahdi Shafiei, Katherine Dunn, Eva Boon, Shelley MacDonald, David Walsh, Hong Gu and Joseph P Bielawski (2015). BioMiCo: a supervised Bayesian model for inference of microbial community structure. *Microbiome*, 3(8).

Mahdi Shafiei, Katherine A. Dunn, Hugh Chipman, Hong Gu, Joseph P. Bielawski (2014). BiomeNet: A Bayesian model for inference of metabolic divergence among microbial communities. *PLoS Comput Biol* 10(11).

Melanie Abeysondera, Toby Kenney, Chris Field, Hong Gu (2014). Combining Distance Matrices on Identical Taxon Sets for Multi-gene Analysis with Singular Value Decomposition. *PLOS ONE*, 9(4).

Toby Kenney and Hong Gu (2012). Hessian calculation for phylogenetic likelihood based on the pruning algorithm and its applications, *Statistical Applications in Genetics and Molecular Biology*. 11(4).

Melanie Abeysondera, Chris Field and Hong Gu (2012). Phylogenetic analysis based on spectral methods. *Molecular Biology and Evolution*. 29(2):579-97.

Hong Gu, Katherine A. Dunn and Joseph P. Bielawski (2012) Likelihood Based Clustering (LiBaC) for Codon Models in "Codon Evolution: mechanisms and models", edited by GM Cannarozzi and A Schneider. Oxford University Press.

Gu, H., Kenney, T. and Zhu, M. (2010). Partial Generalized Additive Models: An Information-theoretic Approach for Dealing with Concurvity and Selecting Variables. *Journal of Computational and Graphical Statistics*. September 1, 2010, 19(3): 531-551.

Katherine A. Dunn, Joseph P. Bielawski, Todd J. Ward, Caroline Urquhart and Hong Gu (2009). Reconciling ecological and genomic divergence among lineages of *Listeria* under an "extended mosaic genome concept". *Mol Biol Evol*. vol 26 (11): 2605-2615.

Fallah N., Gu H., Mohammad K., Seyyedsalehi S.A., Nourijelyani K and Eshraghian M. (2009). Nonlinear Poisson regression using neural networks: a simulation study. *Neural Computing & Applications*. Vol 18(8), 939-943.

Morine MJ, Gu H, Myers RA, Bielawski JP. (2009). Trade-Offs Between Efficiency and Robustness in Bacterial Metabolic Networks Are Associated with Niche Breadth. *J Mol Evol.* Vol 68 (5): 506-515.

Chakraborty, H. and Gu, H. (2009). A Mixed Model Approach for Intent-to-Treat Analysis in Longitudinal Clinical Trials with Missing Values. RTI Press publication No. MR-0009-0903.

Xiaofei Shi, Hong Gu and Chris Field (2008). Pattern classification of phylogeny signals. *Statistical Applications in Genetics and Molecular Biology*: Vol. 7: Iss. 1, Article 30.

Le Bao, Hong Gu, Katherine A. Dunn and Joe Bielawski (2008). Likelihood Based Clustering (LiBaC) for Codon Models, a method for grouping sites according to similarities in the underlying process of evolution. *Mol Biol Evol.* Vol. 25(9), pp. 1995-2007.

Emad Bahrami Samani, M. Mehdi Homayounpour and Hong Gu (2007). A Novel Hybrid GMM/SVM Architecture for Protein Secondary Structure Prediction. *Applications of Fuzzy Sets Theory*. Springer Berlin / Heidelberg, Volume 4578/2007.

Le Bao, Hong Gu, Katherine Dunn and Joe Bielawski (2007). Methods for selecting fixed-effect models for heterogeneous codon evolution, with comments on their application to gene and genome data. *BMC Evolutionary Biology* 2007, 7(suppl 1):S5.

Wing K. Fung, Hong Gu, Liming Xiang and Kelvin K. W. Yau (2006). Assessing local influence in principal component analysis with application to hematology study data. *Statistics in Medicine*, Vol. 26, Issue 13, Pages 2730 – 2744.

Xiaofei Shi, Hong Gu and Chris Field (2006). Testing a clade in Phylogenetic trees. *Mol Biol Evol.* Vol. 23, Number 10, pp. 1976-1983(8).

Krista Collins, Hong Gu, and Chris Field (2006) Examining Protein Structure and Similarities By Spectral Analysis Technique. *Statistical Applications in Genetics and Molecular Biology*: Vol. 5: No. 1, Article 23.

Lorenzo Vega-Montoto, Hong Gu and Peter D. Wentzell. (2005). Mathematical Improvements to Maximum Likelihood Parallel Factor Analysis: Theory and Simulations. *Journal of Chemometrics*. 19: 216-235.

Xiaofei Shi, Hong Gu, Edward Susko and Chris Field (2005). The Comparison of the Confidence Regions in Phylogeny. *Mol Biol Evol.* 22: 2285-2296.

Hugh A. Chipman and Hong Gu (2005). Interpretable Dimension Reduction. *Journal of Applied Statistics*, 32, 969-987.

Hugh A. Chipman and Hong Gu (2001). "Discussion of 'Flexible regression modeling with adaptive logistic basis functions' by P. M. Hooper". *Canadian Journal of Statistics*, 29,370--374.

- Hong Gu and Wing K. Fung (2001). Influence diagnostics in common principal components analysis. *Journal of Multivariate Analysis*, 79, 275-294.

- Hong Gu and Wing K. Fung (2001). Local influence for the restricted likelihood

with applications. *Sankhya*, 63, 250-259.

- Hong Gu and Wing K. Fung (2000). Influence diagnostics in common canonical variates analysis. *Annals of the Institute of Statistical Mathematics (AISM)*, 52, 753-766.

- Hong Gu and Wing K. Fung (1998). Assessing local influence in canonical correlation analysis. *Annals of the Institute of Statistical Mathematics (AISM)*, 50, PP.755-772.

- Wing K. Fung and Hong Gu (1998). The second order approximation to sample influence curve in canonical correlation analysis. *Psychometrika*, 63, PP. 263-269.

- Wing K. Fung and Hong Gu (1998). Discussion on the paper by Hodges (Some algebra and geometry for hierarchical models, applied to diagnostics). *J. R. Statist. Soc. B*, 60, PP.531-532.

### **Book Chapters:**

-A Novel Hybrid GMM/SVM Architecture for Protein Secondary Structure Prediction (2007). [\*Applications of Fuzzy Sets Theory\*](#). Springer Berlin / Heidelberg, Volume 4578/2007.

-Hong Gu, Katherine A. Dunn and Joseph P. Bielawski (2012). Likelihood Based Clustering (LiBaC) for Codon Models. Book Chapter in "Codon Evolution: Mechanisms and Models", edited by GM Cannarozzi and A Schneider. Oxford University Press.

### **Invited lectures and Conference Presentations:**

-Principal Component Analysis for microbiome data by correcting the measurement errors and sequencing depths, BIRS Workshop 19w5221 Emerging Statistical Challenges & Methods for Analysis of Human Microbiome Data, Banff, Canada, Sep.16, 2019.

-Poisson Measurement Error Corrected PCA, with Application to Microbiome Data, ICSA Canada Chapter, invited session, Kingston, Canada, Aug. 10, 2019.

-Statistical learning methods in human microbiome analysis and emergency diagnoses. Invited talk in Formulating Success Research Connector, in AARMS IPSW workshop, July 14-19, 2019.

- Semiparametric Transformed PCA, removing Poisson measurement error for dimension reduction. International workshop on perspectives on high-dimensional data analysis. 2019, June 24-27, Uppsala University, Sweden.

- Poisson Measurement Error corrected PCA for Microbiome Data. AARMS CRG workshop on Statistical analysis and machine learning with application in medicine, biology, environmental sciences. 2019, May 9th-10th. Dalhousie Univ..

- Poisson Measurement Error corrected PCA for Microbiome Data. CANSSI Mathematical and Statistical methods for public health epidemiology workshop, 2019, Mar. 30-31. SFU, Vancouver, BC. Canada.
  
- Statistical learning methods in emergency diagnoses and human microbiome analysis. The Science Atlantic AARMS workshop 2017, Oct. 14-15. Fredericton, NB, Canada.
  
- Data exploratory methods for microbiome data analysis. 2017 ICASA Applied Statistics Symposium. June 25-28. Chicago, IL, USA.
  
- Prior distributions for ranking problems. The 1st International Conference on Econometrics and Statistics (EcoSta 2017). June 15-17, Hong Kong University of Science and Technology, Hong Kong.
  
- Learning Microbial Community Structures with Supervised and Unsupervised methods. 13th Annual Guelph Biomathematics and Biostatistics Symposium, May 17th 2017, Waterloo, ON, Canada.
  
- Robust Ranking Using Heavy-tailed Prior Distributions. The International Conference on Robust Statistics (ICORS) 2016, July 4-8. Geneva, Switzerland.
  
- Machine learning diagnoses on patients presenting abdominal pain. The International Federation of Classification Societies (IFCS) 2015 conference, Bologna, Italy.
  
- Machine learning diagnoses on patients presenting abdominal pain. Invited talk in Statistical society of Canada annual meeting 2015, Halifax, Canada.
  
- Non-negative Matrix Factorisation for Analysis of Metagenomic Data. Invited talk, The 2nd International Society for NonParametric Statistics (ISNPS) conference 2014, Cadiz, Spain.
  
- Development of an Adequacy-Test for the Substitution Models in Phylogenetic Analysis. Invited talk in IMS-China International Conference on Statistics and Probability, June 30 - July 4, 2013, Chengdu, China.
  
- Estimation of Darwinian positive selection using generalized codon-based Models. SSC meeting, Wolfville, NS. June 2011.
  
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Matej Bel University. June, 2009.
  
- Spectral Analysis Techniques and Their Applications in Bioinformatics. Invited colloquium talk in Matej Bel University. May, 5<sup>th</sup>, 2009.
  
- Interpretable Dimension Reduction. Invited lecture in Matej Bel University. May, 13<sup>th</sup>, 2009.

- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Charles University. April, 2009.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Vienna University of Science and Technology. Apr., 2009.
- Spectral Analysis Techniques and Their Applications in Bioinformatics. Invited lecture in Center for Integrative Bioinformatics Vienna (CIBIV). Apr., 2009.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in University of Vienna. March, 2009.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Australian National University. Feb. 2009.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Peking University. Jan. 16th, 2009.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Yunnan University. Dec. 2nd, 2008.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Zhejiang University. Nov. 20<sup>th</sup>, 2008.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in Anhui University of Technology and Science. Nov. 16<sup>th</sup>, 2008.
- Spectral Analysis Techniques and Their Applications in Bioinformatics. . Invited lecture in university of Science and Technology of China. Nov. 18<sup>th</sup>, 2008.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in university of Science and Technology of China. Nov. 10<sup>th</sup>, 2008.
- Partial Generalized Additive Models: An Information-theoretic Approach for Selecting Variables and Dealing with Concurvity. Invited lecture in university of Melbourne. Oct. 2008.
- Partial generalized additive models: An information-theoretic approach for avoiding concurvity. Invited lecture in university of Hong Kong, May 27 to 29, 2007.
- Applications of information theory in Generalized Additive Models. The 15<sup>th</sup> International Conference forum on interdisciplinary Mathematical and Statistical Techniques. Shanghai, P.R. China, May 20 to 23, 2007.

- Predicting Protein Structure and Examining Similarities of Protein Structure by Spectral Analysis Techniques. Australia, Canberra. ANU BioinfoSummer2006, Dec.4-8, 2006.
- Avoiding the concurvity in generalized additive models-PGAM. 17<sup>th</sup> COMPSTAT symposium of the IASC. Rome, Italy. Aug. 28 to Sep. 1, 2006.
- Partial generalized additive models: An information-theoretic approach for avoiding concurvity. 69<sup>th</sup> Annual meeting of the institute of mathematical statistics. July 30-Aug. 4, 2006. -- Also chair the session of “Semiparametric and Econometric Models”.
- Phylogenomic Analysis of *Listeria* Reveals Divergent Evolution among Lineages with Different Abilities to Respond to Environmental Stress. Poster presentation at The 1<sup>st</sup> International Conference in Phylogenomics, University of Montreal, Canada, March, 2006
- Likelihood Based Clustering (LiBaC) for Codon Models, a new method for inferring amino acid sites subject to positive and negative selection pressure. Poster presentation at The 1<sup>st</sup> International Conference in Phylogenomics, University of Montreal, Canada, March, 2006.
- Interpretable dimension reduction. Invited talk for MALNIS seminar in Computer Science. Dec., 2004.
- Interpretable dimension reduction in drug discovery. Invited lecture in 26th Annual Midwest Biopharmaceutical Statistics Workshop. May, 2003.
- Statistical analysis of visual field progression. Presented in IPS (International Perimetric society meeting), June 2002.
- Detecting change in visual field data. ENAR meeting held at Arlington VA, Mar. 2002.
- Invited Discussion of Peter Hooper’s paper Flexible regression modelling with adaptive logistic basis functions, Presented at SSC Meeting, held in SFU, June, 2001.
- Interpretable Dimension Reduction. Presented at SSC Meeting, held in SFU, June, 2001.
- Interpretable Dimension Reduction. Presented at MITACS Annual General Meeting, held in Toronto, June, 2000.
- On the sample influence curve in canonical correlation analysis. Presented at the Joint Statistical Meetings, held in Anaheim, Orange County, USA, August, 1997. --- Also published in the American Statistical Association 1997 Proceedings of the Statistical Computing Section, pp. 66-69.
- An approximation to the sample influence curve in canonical correlation analysis. Presented at the International Symposium on Contemporary Multivariate Analysis and its Applications, held in Hong Kong, May 1997.

## **Professional Services:**

**Referee for journals:**

*Statistics in Medicine*  
*The Computational statistics and data analysis*  
*The institute of statistical mathematics*  
*Canadian Journal of Statistics*  
*Statistical Applications in Genetics and Molecular Biology*  
*BMC Bioinformatics*  
*Bioinformatics*  
*Statistics and computing*  
*Plos one*  
*Plos computational Biology*  
*Biometrics*  
*Molecular Biology and Evolution*  
*Environmetrics*  
*Microbiome*  
*iScience*  
*Mathematics*  
*Computational and Structural Biotechnology Journal*  
*Agronomy*

**Reviewer for others:**

NSERC Discovery Grants  
Scholarly review of tenure and promotion application, Bowdoin College  
Review for a full professor application, UNB.  
Scholarly review of tenure and promotion application, Univ. of Waterloo

**University Administrative contributions**

University committees: Senate Discipline Committee, 2003-2004.  
Faculty committees: The Faculty Honorary Degrees Committee, 2003-2006.  
Undergraduate adviser (July 2007- July 2008).  
Statistical colloquium chair (since Sep. 2004-July 2007).

## Departmental committees:

Appointment committee for the position in Environmental Statistics, UFA  
committees, ad hoc curriculum committee, etc..

Statistical colloquium chair (July 2009-Dec. 2009).  
Director of Statistics (July 2009-Jan. 2010).  
Director of Statistics (July 2010-Dec. 2011).  
Director of Statistics (Jan. 2013-June 2016).  
Undergraduate advisor (July 2013-Dec. 2013).  
Graduate coordinator in statistics (Jan. 2014-July, 2014).

Faculty committees: Chair advisory committee for Biology Department Chair. (Oct. 2013-Dec. 2013).

Department committees: Instructor positions committee (For both Math. & Stat.) (2013-2015);  
Math Director Search Committee.

Director of Statistics (Jan. 2020-Dec. 2020)  
Instructor position search committee for learning center director (Mar. 2020)  
Search committee for LTA in statistics (May, 2020)  
Tenure track position on data science search committee (2020)  
Hiring committee for spousal tenure-track appointment for Theo Johanson-Freyd (2020)  
Search Committee for the Part-Time Academic (PHAR 3011 and PHAR 2020).

### **External services**

AARMS summer school co-director (2013-2014). (co-applicant for funding from CANSSI and Field's Institute).

AARMS Industrial Collaboration Committee (2017).

AARMS Industrial Problem Solving Workshop 2018 (2018/07/3-2018/07/06), local organizer. (Applicant for NSERC for this event)

AARMS CRG workshop on Statistical analysis and machine learning with application in medicine, biology, environmental sciences. 2019, May 9th-10th. Dalhousie Univ..Organizer.

BIRS workshop organization committee for workshop titled "Emerging Statistical Challenges and Methods for Analysis of Human Microbiome Data", 2019/09/15-2019/09/20.

Maritime Statistical and Health Sciences Collaborating Centre (MSHSCC) (2018/04-2021/04), Co-investigator.

### **Other graduate student supervision at Dalhousie and other universities:**

#### **Committee member/thesis reader for M.Sc theses of**

T.Ikeda (Statistics, Dalhousie)

Liwen Zhou (Statistics, Dalhousie)

Caren Rose (Statistics, Dalhousie)

Melissa Morine (Biology, Dalhousie) "Network analysis of cellular metabolism and genome evolution in obligate intracellular bacteria". (Admission to Candidacy Examination (ATC) defense)

Melissa Morine (Biology Dept.), "Functional topology and evolution in prokaryotic metabolic networks". (External examiner).



Mitra Jazayeri (La Trobe University, 2008), “Approximating Cross-validation results for SIR and SAVE preceded by Principal Component Dimension Reduction”. (External examiner).

He Gao (Statistics, Dalhousie), 2009/8.

Sylvia de la Ronde(Statistics, Dalhousie). 2009/4.

Lihui Liu (Statistics, Dalhousie), 2010/8.

Khatoon Alobaidan (Statistics, Acadia), “Generalizing mutual clusters: a measure of cluster compactness’, 2014/9.

Jie Ning (CS, Dalhousie), 2015/08.

Laura (Dong) Lin (Statistics, Dalhousie), 2016/04.

Hongyue Wang (Statistics, Dalhousie), 2016/04.

Michael Hall (CS, Dalhousie), 2016/07.

Thomas Bennett (Acadia Univ.) 2016/9. Nonparametric and semiparametric outlier detection and quality control in seasonal time series data.

Nidhin NANDHAKUMAR (CS, Dalhousie), “Clinical Significant Information and Phrase Extraction from Radiology Reports”. 2017/7.

Mingzhu Wang (Statistics, Dalhousie), “The Influence of Utility Functions on Insurance Choices”. 2018/4.

Jonathan Bradet-Legris (Statistics, Dalhousie), “Distribution and Elemental Composition of Picoplankton in the North Pacific Ocean”. 2020/11.

Etai Markowski (Statistics, Dalhousie), “A comparison of methods for constructing confidence sets of phylogenetic trees using maximum likelihood”. 2021/03.

**PhD committee//thesis reader and Proposal or theses defense:**

Xiaomeng Wan (CS, Dalhousie), “Statistical analysis of dynamic graphs”. (external examiner of aptitude defense). 2007.

Mahdi Shafiei (CS, Dalhousie), “Probabilistic Models for Document and Term Clustering”. (external examiner of aptitude defense). 2007.

Mahdi Shafiei (CS, Dalhousie), PhD thesis defense. 2009.

Siyuan Hou (Chemistry, Dalhousie), PhD thesis defense. 2012.

Philippe Fullsack (Statistics, Dalhousie), PhD proposal defense. 2013/12.

Ling Ding (Statistics, University of Melbourne), “Regression Clustering Using Gibbs Sampler and Optimal Cluster Number Estimation”, External examiner for PhD thesis, 2016/5.

Ahmad Pesaranghader (CS, Dalhousie), “simDEF: Definition-based Semantic Similarity Measure of Gene Ontology Terms for Functional Similarity Analysis of Genes”. External examiner for PhD RAD defense. 2016/07.

Ahmad Pesaranghader (CS, Dalhousie), “deepSimDEF: Deep Definition-based Semantic Similarity of Gene Ontology-Term and Gene-Product Embeddings for Functional Analysis of Genes”. External examiner for PhD proposal defense. 2017/06.

Tang, Yaohua (Department of Statistics & Actuarial Science, University of Hong Kong), “Some topics on deep learning and text analytics”. External examiner for PhD defense. 2017/08.

Ahmad Pesaranghader (CS, Dalhousie), “Concept Embedding for Deep neural Functional Analysis of Genes and Deep Neural Word Sense Disambiguation of Biomedical Text”. PhD defense. 2019/06.

You, Jia (Department of Statistics & Actuarial Science, University of Hong Kong), Fast automated artificial intelligence evaluation system for acute ischemic stroke. External examiner for PhD defense. 2020/08.

### **Courses taught (in Dalhousie University only)**

2001 fall: Stat.4350/5350 Applied Multivariate Analysis

2002 winter: Stat. 2060

2002 Fall: Stat. 4350/5350, Stat. 1060, Stat. 2060

2003 Winter: Stat. 2060, Stat.1060

2003 fall: Stat.4350/5350, Stat.1060

2004 winter: Stat.5500B, Data Mining and Statistical learning; Stat. 1060

2004 fall: Stat.4620/5620, Data Analysis, Stat.1060

2005 winter: Stat.5500B, Data Mining and Statistical learning; Stat. 1060, Stat. 2060

2005 fall: Stat.4620/5620, Data Analysis, Stat.1060

2006 winter: Stat.5500B, Data Mining and Statistical learning; Stat. 1060

2006 fall: Stat.4620/5620, Data Analysis, Stat.1060

2007 winter: Stat. 2060, Introduction to Probability and Statistics. Stat. 1060.

2007 fall: Stat. 4620/5620, Data Analysis. DISP (Dalhousie Integrated Science Program).

2008 winter: Stat. 2060

2009/10 fall: DISP (Dalhousie Integrated Science Program). Winter: maternity leave.  
2010/11 fall: DISP (Dalhousie Integrated Science Program).  
2011/12 fall: DISP (Dalhousie Integrated Science Program). Winter: maternity leave.  
2012/13 fall: sabbatical leave, winter Stat. 5750.  
2013/14 fall: Stat. 2080. Winter Stat. 5750.  
2014/15: winter: Stat 3460. Stat. 5750.  
2015/16: fall: Stat. 3360. Winter: Stat 3460.  
2016/17: fall: Stat 3360. Winter: Stat 3460, Stat. 5750.  
2017/18: fall: Stat 3360. Stat 5620 (reading course). Winter: Stat 3460, Stat. 5750.  
2018/19: fall: Stat 3360. Winter: Stat 3460, Stat. 5750.  
2019/20 fall: Stat 3360.  
2020 sabbatical leave  
2020/21 Winter Stat 3460 online  
2021/22 Fall Stat 3360, Stat 4350/5350