

Giles Hooker

Associate Professor
Department of Statistics and Data Science
Department of Computational Biology
Cornell University

<http://www.bscb.cornell.edu/~hooker>
<http://blogs.cornell.edu/modelmeanings/>



Contact

Department of Bio. Stat. and Comp. Bio.,
1186 Comstock Hall
Cornell University
Ithaca, NY, 14853

E-mail gjh27@cornell.edu
Phone: (1 607) 255 1638
Fax: (1 607) 255 4698

Education

PhD Statistics, Stanford University 2004.

MSc Statistics, Stanford University, 2002.

BSc Mathematics Honours 1, Australian National University, 1999.

BA Political Science, Australian National University, 1998.

Honours and Awards

2011 NSF CAREER Award

2004 TA of the Year, Department of Statistics, Stanford University.

2004 Laha Travel Award, Institute of Mathematical Statistics

2003 TA Award - Department of Statistics, Stanford University.

2000 Fulbright Tim Matthews Memorial Award

2000 Shandong Province Award for Excellence in Teaching

1999 University Medal in Mathematics, Australian National University

1999 Hana Neumann Prize, Australian National University

1999 Australian National University Honours Scholarship

1997 University of New South Wales Summer Research Scholarship

Professional Experience

- March 2019** - Associate Professor, Department of Statistics and Data Science and Department of Computational Biology, Cornell University.
- July 2016 - December 2019** Director of Undergraduate Studies, Biometry and Statistics Major, Cornell University
- May 2012 - March 2019** Associate Professor, Department of Biological Statistics and Computational Biology and Department of Statistical Science, Cornell University.
- June 2012 - June 2016** Director of Graduate Studies, Graduate Field of Statistics, Cornell University.
- June 2012 - September 2013** Director of Graduate Studies, Graduate Field of Biometry, Cornell University.
- August 2006 - May 2012** Assistant Professor, Department of Biological Statistics and Computational Biology and Department of Statistical Science, Cornell University.
- October 2004- August 2006** Post-Doctoral Fellow, McGill University.
- June - September 2003** Intern, Robert Bosch Corp.
- June 2002 - August 2002** Intern, AT&T Research.
- Feb-July 2000** English Teacher, Chiang Chen Industrial Institute.
- Dec 1999 - Feb 2000** Research Assistant, Research School of Information Sciences and Engineering.
- Dec 1998 - Feb 1999** Intern, Commonwealth Bank of Australia.
- Dec 1997 - Feb 1998** Research Scholar, University of New South Wales.

Visiting and Adjunct Positions

- July 2013** - Associate Professor, Department of Healthcare Policy and Research, Weill Cornell Medical College.
- May 2014 - January 2021** Visiting Fellow, Research School of Finance, Actuarial Studies and Applied Statistics, Australian National University
- April - June 2014** Health Care Analytics Working Group, IBM Research.
- Feb - April 2014** Department of Biostatistics, Johns Hopkins University.
- Sept - Dec 2013** Department of Mathematics and Statistics, University of Melbourne.
- Sept - Dec 2010** Program on the Analysis of Object Data, Statistics and Applied Mathematical Sciences Institute.

Grants Funded

- NSF DEB-1933497, 2020: “Collaborative Research: A general approach to partitioning contributions from multiple drivers affecting individuals, populations, and communities”, \$750,000, co-PI with S. Ellner, P. Adler and R. Snyder.
- NSF TRIPODS 1740882, 2017: “Data Science for the Social Good”, \$1,497,238, co-PI with K. Weinberger, J. Kleinberg, S. Strogatz and D. Schmoys.
- NSF DMS-1712554, 2017: “Statistical Inference Using Random Forests and Related Methods”, \$335,078, PI with L. Mentch.
- NVIDIA Hardware Grant, 2014: “Linking Convolutional Neural Networks, Random Forests and Statistical Inference”, PI.
- NIH R03DA036683, 2014: “Shortening the SOAPP-R: Computer-based opioid risk assessment”, \$160,579, co-PI with M. Finkelman.
- NSF DEB-1353039, 2014: “Integral Projection Models for Populations in Varying Environments: Construction and Analysis”, \$652,847, co-PI with S. Ellner, P. Adler and R. Snyder.
- NSF DEB-125619, 2013: “Effects of Rapid Consumer Evolution on Community Dynamics: predictions and Tests in a (Nearly) Natural Food Web”, \$200,000, co-PI with S. Ellner and N. Hairstone.
- CTSC Community Engagement Award, 2011: “HCV Testing in NYC Commercial Sex Venues”, \$20,000, co-I with K. Marks and D. Daskalakis.
- NSF CDI Type II 1125098, 2011: “Bircast: Novel Machine Learning Methods for Understanding Continent-Scale Migration”, \$1,217,895. co-I with S. Kelling, and T. Diettrich.
- NSF DMS-1053252 CAREER, 2010: “Diagnostics and Experimental Design for Nonlinear Dynamics”, \$400,000. PI.
- NSF CMG-0934735, 2009: “Functional modeling of climate-ecosystem dynamics”, \$350,000, co-PI with S. Ray and M. Friedl.
- NSF DEB-0813743, 2008: “Rapid Evolution and the Dynamics of Complex Ecological Communities”, \$539,957, co-PI with S. Ellner, L. Jones and N. Hairstone.
- Hatch NYC-150446, 2007: “Experimental Design for Nonlinear Processes in Agriculture”, \$30,000, PI.

Books

- Ramsay, James O. and **Giles Hooker**, 2017, “Dynamic Data Analysis: Modeling Data with Differential Equations”, Springer.
- Ramsay, James O., **Giles Hooker** and Spencer Graves, 2009, “Functional Data Analysis in R and Matlab”, Springer.

Journal Papers and Book Chapters

1. Warmenhoven, John, Norma Bargary, Dominik Liebl, Andrew Harrison, Mark Robinson, Edward Gunning and **Giles Hooker**, “PCA of Waveforms and Functional PCA: A Primer for Biomechanics”, *Journal of Biomechanics*, in press.
2. Zhou, Zhengze, and **Giles Hooker**, 2020, “Unbiased Measurement of Feature Importance in Tree-Based Methods”, *Transactions in Knowledge Discovery and Data Mining*, in press.
3. Snyder, Robin, Stephen P. Ellner and **Giles Hooker**, “Time and chance: using age partitioning to understand how luck drives variation in reproductive success”, *The American Naturalist*, in press.
4. Ghosal, Indrayudh and **Giles Hooker**, 2020, “Boosting Random Forests to Reduce Bias; One-Step Boosted Forest and its Variance Estimate”, *Journal of Computational and Graphical Statistics*, in press.
5. Ye, Zi and **Giles Hooker**, 2020, “Local Quadratic Estimation of the Curvature in a Functional Single Index Model”, *Scandinavian Journal of Statistics*, 47(4)1307-1338.
6. Ye, Zi, **Giles Hooker** and Stephen P. Ellner, 2020, “The Jensen Effect and Functional Single Index Models: Estimating the Ecological Implications of Nonlinear Reaction Norms”, *Annals of Applied Statistics*, 14(3):1236-12341.
7. Coleman, Tim, Lucas Mentch, Daniel Fink, Frank La Sort, **Giles Hooker**, Wesley Hochachka and David Winkler, 2020, “Statistical Inference on Tree Swallow Migrations”, *Journal of the Royal Statistical Society, Series C*, in press.
8. **Giles Hooker**, Sophia Brumer, Chrisopher Zappa and Edward Monahan “Inferences to be Drawn from a Consideration of Power-Law Descriptions of Multiple Data Sets Each Comprised of Whitecap Coverage, WB, and 10-m Elevation Wind Speed Measurements”, in *Recent Advances in the Study of Oceanic Whitecaps*, P. Vlahos and E. C. Monahan (Eds).
9. Kilian, Nicole, Yongden Zhang, Lauren LoMonica, **Giles Hooker**, Derek Toomre, Choukri Ben Mamoun and Andreas. M. Ernst, 2020, “Trafficking and Localization of S-Palmitoylated Proteins in Plasmodium falciparum-Infected Erythrocytes”, *BioEssays*, 1900145.
10. Javeed, Aurya, and **Giles Hooker**, 2020, “Timing Observations of Diffusions”, *Statistics and Computing*, 30:405-417.
11. J. Wen, P. Köhler, G. Duveiller, N.C. Parazoo, T.S. Magney, **G. Hooker**, L. Yu, C. Y. Chang, and Y. Sun, 2020 “Generating a Long-Term Record of High-Resolution Global Solar-Induced Chlorophyll Fluorescence (SIF) by Harmonizing Multiple Satellite Instruments: A Case Study for Fusing GOME-2 and SCIAMACHY”, *Remote Sensing of Environment*, in press.
12. Sinclair, David G., and **Giles Hooker**, 2019, “Sparse Inverse Covariance Estimation for High-throughput microRNA Sequencing Data in the Poisson Log-Normal Graphical Model”, *Journal of Statistical Computation and Simulation*, in press.

13. Ellner, Stephen P., Snyder, Robin E., Adler, Peter B. and **Giles Hooker**, 2019, “An Expanded Modern Coexistence Theory for Empirical Applications”, *Ecology Letters*, 22(1):3-18.
14. Goryaynov, Alexander, Nicole Kilian, Mark Lessard, Derek Toomre, James Rothman, **Giles Hooker** and Jörg Bewersdorf, 2018 “Assessing photodamage in live-cell STED microscopy”, *Nature Methods*, 15:755-756.
15. Wu, Yuefeng and **Giles Hooker**, 2018, “Asymptotic Properties for Methods Combining Minimum Hellinger Distance Estimates and Bayesian Nonparametric Density Estimates”, *Entropy*, 20(12):955.
16. Tredennick, Andrew T., Brittany J. Teller, Peter B. Adler, 2018, **Giles Hooker** and Stephen P. Ellner, “Size-by-environment interactions: a neglected dimension of species’ responses to environmental variation”, *Ecology Letters*, 21(12):1757-1770.
17. Adler, Peter B., Andrew Kleinhesselink, **Giles Hooker**, Brittany Teller and Stephen P. Ellner, 2018, “Weak interspecific interactions in a sagebrush steppe: evidence from observations, models, and experiments”, *Ecology*, 99(7):1621-1632.
18. **Hooker, Giles** and Cliff Hooker, 2018, “Machine Learning and the Future of Realism”, *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 9(1):174-182.
19. Thorbergsson, Leifur and **Giles Hooker**, 2018, “Experimental Design for Partially Observed Markov Decision Processes”, *Journal of Uncertainty Quantification*, 6(2):549-567.
20. **Hooker, Giles** and Lucas Mentch, 2018, “Bootstrap Bias Corrections for Ensemble Methods” *Statistics and Computing*, 28(1):77-86.
21. Mentch, Lucas and **Giles Hooker**, 2017, “Formal Hypothesis Tests for Additive Structure in Random Forests”, *Journal of Computational and Graphical Statistics*, 26(3):589-597.
22. Earls, Cecilia and **Giles Hooker**, 2017, “Combining Functional Data Registration and Factor Analysis”, *Journal of Computational and Graphical Statistics*, 26(2):296-305.
23. Earls, Cecilia and **Giles Hooker**, 2017, “Adapted Variational Bayes for Functional Data Registration, Smoothing, and Prediction”, *Bayesian Analysis*, 12(2):557-582.
24. Liu, Chong, Surajit Ray and **Giles Hooker**, 2017, “Functional Principal Components Analysis of Spatially Correlated Data”, *Journal of Computational and Graphical Statistics*, 27(6):1639-1654.
25. Mentch, Lucas and **Giles Hooker**, 2016, “Quantifying Uncertainty in Random Forests via Confidence Intervals and Hypothesis Tests”, *Journal of Machine Learning Research*, 17(3):1-41
26. **Hooker, Giles**, James O. Ramsay and Luo Xiao, 2016, “CollocInfer: Collocation Inference in Differential Equation Models”, *Journal of Statistical Software*, 75(2)
27. **Hooker, Giles** and Steven Roberts, 2016, “Maximal Autocorrelation Functions in Functional Data Analysis”, *Statistics and Computing*, 26(5):945-950.

28. Hall, Peter and **Giles Hooker**, 2016, “Truncated Linear Models for Functional Data”. *Journal of the Royal Statistical Society, Series B*, 78(3):637-653
29. **Hooker, Giles**, 2016, “Consistency, Efficiency and Robustness of Conditional Disparity Methods”, *Bernoulli*, 22(2):857-900
30. Teller, Brittany J., Peter B. Adler, Collin B. Edwards, **Giles Hooker**, Robin E. Snyder and Stephen P. Ellner, 2016, “Linking demography with drivers: climate and competition”, *Methods in Ecology and Evolution*, 7(2):171-183.
31. Grinspan, Zachary, M., JS Shapiro, Erika L. Abramson, **Giles Hooker**, Rainu Kaushal and Lisa M. Kern, 2015, “Predicting Frequent ED Use By People with Epilepsy with Health Information Exchange Data”, *Neurology*, 85(12):1031-1038.
32. **Hooker, Giles** and Stephen P. Ellner, 2015, “Goodness of Fit Diagnostics in Nonlinear Dynamics: Mis-specified Rates or Mis-specified States?”. *Annals of Applied Statistics*, 9(2):754-776.
33. **Hooker, Giles**, Kevin K. Lin and Bruce Rogers, 2015, “Control Theory and Experimental Design in Diffusion Processes”, *Journal of Uncertainty Quantification*, 3(1):234-264.
34. McLean, Matthew W., **Giles Hooker** and David Ruppert, 2014, “Restricted Likelihood Ratio Tests for Linearity in Scalar-on-Function Regression”, *Statistics and Computing*, 25(5):997-1008.
35. Hiltunen, Teppo, Stephen P. Ellner, **Giles Hooker**, Laura E. Jones, Nelson G. Hairston, 2014, “Eco-evolutionary Dynamics in a Three-Species Food Web with Intraguild Predation: Intriguingly Complex” in *Advances in Ecological Research, Vol. 50 – Eco-Evolutionary Dynamics*, Jordi Moya-Laraño, Jennifer Rowntree and Guy Woodard, Editors.
36. Hiltunen, Teppo, Nelson G. Hairstone, **Giles Hooker**, Laura E. Jones and Stephen P. Ellner, 2014, “A newly discovered role of evolution in previously published consumer-resource dynamics”, *Ecology Letters*, 17(8):915-923.
37. **Hooker, Giles** and Anand N. Vidyashankar, 2014, “Bayesian Model Robustness via Disparities”, *TEST*, 23(3):556-584.
38. McLean, Matthew W., **Giles Hooker**, Ana-Maria Staicu, Fabian Schiepl and David Ruppert, 2014, “Functional Generalized Additive Models”, *Journal of Computational and Graphical Statistics*, 23(1):249-269.
39. Earls, Cecilia, and **Giles Hooker**, 2014, “Bayesian Covariance Estimation and Inference in latent Gaussian Process Models”, *Statistical Methodology*, 18:79-100
40. Asencio, Maria, **Giles Hooker** and H. Oliver Gao, 2014, “Functional Convolution Models”, *Statistical Modeling*, 14(4):1-21.
41. Gibbons, Robert D., **Giles Hooker**, Matthew D. Finkelman, David J. Weiss, Paul A. Pilkonis, Ellen Frank, Tara Moore and David J. Kupfer, 2013, “Computerized Adaptive Diagnosis of Depression Using the CAD-MDD”, *Journal of Clinical Psychiatry*, 74(7): 669-674.

42. Jesty, S.A., S.W. Jung, J.M. Cordeiro, T.M. Gunn, J.M. Di Diego, S. Hemsley, B.G. Kornreich, **G. Hooker**, C. Antzelevitch, N.S. Moise, 2013, “Cardiomyocyte calcium cycling in a naturally occurring German shepherd dog model of inherited ventricular arrhythmia and sudden cardiac death”, *Journal of Veterinary Cardiology* 15(1): 5-14.
43. **Hooker, Giles** and James Ramsay, 2012, “Learned-Loss Boosting”, *Computational Statistics and Data Analysis*, 56:3935-3944.
44. Campbell, David, **Giles Hooker** and Kim McAuley, 2012, “Parameter Estimation in Differential Equation Models with Constrained States”, *Journal of Chemometrics*, 56:322-332.
45. **Hooker, Giles** and Saharon Rosset, 2012, “Prediction-Focussed Regularization Using Data-Augmented Regression”, *Statistics and Computing*, 1:237-248.
46. Liu, Chong, Surajit Ray, **Giles Hooker** and Mark Friedl, 2012, “Functional Factor Analysis for Periodic Remote Sensing Data”, *Annals of Applied Statistics*, 6:601-624.
47. Zeremski, Marija, **Giles Hooker**, Marla A. Shu, Emily Winkelstein, Queenie Brown, Don C. Des Jarlais, Leslie H. Tobler, Barbara Rehmann, Michael P. Busch, Brian R. Edlin, and Andrew H. Talal, 2011, “Induction of CXCR3- and CCR5-associated Chemokines during Acute Hepatitis C Virus Infection.”, *Journal of Hepatology*, 55:545-553.
48. **Hooker, Giles**, Stephen P. Ellner, Laura de Vargas Roditi and David J. D. Earn, 2011, “Parameterizing State-space Models for Infectious Disease Dynamics by Generalized Profiling: Measles in Ontario”, *J. Royal Society Interface*, 8:961-975.
49. Finkelman, Matthew, **Giles Hooker** and Zhen Wang, 2010, “Prevalence and Magnitude of Paradoxical Results in Multidimensional Item Response Theory”, *J. Educational and Behavioral Statistics*, 35:744-761.
50. **Hooker, Giles**, 2010, “On Separable Tests, Correlated Priors and Paradoxical Results in Multidimensional Item Response Theory”, *Psychometrika*, 75:694-707.
51. Fink, Daniel, Wesley M. Hochachka, Benjamin Zuckerberg, David W. Winkler, Ben Shaby, M. Arthur Munson, **Giles Hooker**, Mirek Riedewald, Daniel Sheldon and Steve Kelling, 2010, “Spatiotemoral Exploratory Models for Broad-scale Survey Data”, *Ecological Applications*, 20:2121-22147.
52. Atam, Ercan and **Giles Hooker**, 2010, “An Identification-based State Estimation Method for a Class of Nonlinear Systems”. *J. Systems and Control Engineering*, 224:349-359.
53. **Hooker, Giles** and Matthew Finkelman, 2010, “Paradoxical Results and Item Bundles”. *Psychometrika*, 75:249-271.
54. Kelling, Steve, Wesley M. Hochachka, Daniel Fink, Mirek Riedewald, Rich Caruana, Grant Ballard and **Giles Hooker**, 2009, “Data Intensive Science: A New Paradigm for Diversity Studies”. *Biosciences*, 59:613-620.
55. **Hooker, Giles**, Matthew Finkelman and Armin Schwartzman, 2009, “Paradoxical Results in Multidimensional Item Response Theory”. *Psychometrika*, 74:419-442.

56. **Hooker, Giles**, 2009, “Forcing Function Diagnostics for Nonlinear Dynamics”. *Biometrics*, 65:928-936.
57. Gelzer, Anna, Marcus L. Koller, Niels F. Otani, Jeffrey J. Fox, M. W. Enyeart, **Giles Hooker**, Mark L. Riccio, Carlo R. Bartoli and Robert F. Gilmour, 2008, “Dynamic Mechanisms for Initiation of Ventricular Fibrillation in vivo”, *Circulation*, 118:1123-1129.
58. Ramsay, James O., **Giles Hooker**, David Campbell and Jiguo Cao, 2007. “Parameter Estimation for Nonlinear Differential Equations: A Smoothing-Spline Approach”. *Journal of the Royal Statistical Society, Series B*, with discussion 69:741-796.
59. **Hooker, Giles**, 2007. “Generalized Functional ANOVA Diagnostics for High Dimensional Functions of Dependent Variables”. *Journal of Computational and Graphical Statistics*. 16:709-732.
60. Norris, Robert, Jessica Ngo, Karen Nolan and **Giles Hooker**, 2005. “Volunteers are Unable to Properly Apply Pressure Immobilization in a Simulated Snakebite Scenario”. *Journal of Wilderness and Environmental Medicine*, 16:16-21.
61. Shirts, Michael, Eric Bair, **Giles Hooker** and Vijay Pande, 2003. “Equilibrium Free Energies from Non-equilibrium Estimates Using Maximum Likelihood Methods”. *Physical Letters Review*. 91(14):140601.

Refereed Conference Proceedings

62. Tan, Sarah, Matvey Soloviev, **Giles Hooker**, and Martin T. Wells, 2020, “Tree Space Prototypes: Another Look at Making Tree Ensembles Interpretable” *Foundations of Data Science*.
63. Lengerich, Benjamin, Sarah Tan, Chun-Hao Chang, **Giles Hooker** and Rich Caruana, 2020, “Purifying Interaction Effects with the Functional ANOVA: An Efficient Algorithm for Recovering Identifiable Additive Models”, *AISTATS*.
64. Tan, Sarah, Rich Caruana, Giles Hooker and Yin Lou, 2018, “Distill-and-Compare: Auditing Black-Box Models Using Transparent Model Distillation”, *AAAI/ACM Artificial Intelligence, Ethics, and Society 2018*.
65. Kang, Keegan and **Giles Hooker**, 2017, “Random Projections with Control Variates”, *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods*.
66. Kang, Keegan and **Giles Hooker**, 2016, “Improving the Recovery Of Principal Components with Semi Deterministic Random Projections”, *Proceedings of the 50th Annual Conference on Information Science and Systems*.
67. Kang, Keegan and **Giles Hooker**, 2016, “Block Correlated Deterministic Random Projections”, *Proceedings of the 6th Conference on Computational Mathematics, Computational Geometry and Statistics*.

68. Lou, Yin, Rich Caruana, Johannes Gehrke and **Giles Hooker**, 2013, “Accurate Intelligible Models with Pairwise Interactions”, *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
69. **Hooker, Giles** and Matthew Finkelman, 2004. “Sequential Analysis for Learning Modes of Browsing”. *WEBKDD 2004: Proceedings of the Sixth International Workshop on Knowledge Discovery from the Web*.
70. **Hooker, Giles**, 2004. “Diagnosing Extrapolation: Tree-Based Density Estimation”. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
71. **Hooker, Giles**, 2004. “Discovering Additive Structure in Black Box Functions”. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
72. Hegland, Markus, **Giles Hooker** and Stephen Roberts. 1999. “Finite Element Thin Plate Splines in Density Estimation”. In *Computational Techniques and Applications, Proceedings of the Ninth Biennial Conference: CTAC99*. Journal of the Australian Mathematics Society.

Technical Reports and Commentary

1. Warmenhoven, John, Andrew Harrison, Daniel Quintana, **Giles Hooker**, Edward Gunning and Norma Bargary, 2020, “Unlocking Sports Medicine Research Data while Maintaining Participant Privacy via Synthetic Datasets”, *SportRxiv Preprints*
2. Ellner, Stephen P., Robin E. Snyder, Peter B. Adler, **Giles Hooker** and Sebastian J. Shreiber, 2020, “Comment on Pande et al. (2020): Why invasion analysis is important for understanding coexistence” *Ecology Letters*, in press.
3. **Giles Hooker** and Lucas Mentch, 2019, “Please Stop Permuting Features: An Explanation and Alternatives”, ArXiv preprint.
4. Zhou, Yichen and **Giles Hooker**, 2016, “Interpreting Models via Single Tree Approximation”. ArXiv preprint.
5. **Hooker, Giles** and Lucas Mentch, 2016, “Comments On: A Random Forest Guided Tour”, *TEST* 25(2):254-260.
6. Wu, Yuefeng and **Giles Hooker**, 2013, “Hellinger Distance and Bayesian Non-Parametrics: Hierarchical Models for Robust and Efficient Bayesian Inference”.
7. **Hooker, Giles**, 2013, “On the Identifiability of the Functional Convolution Model”, Technical Report BU-1681-M, Department of Biological Statistics and Computational Biology, Cornell University.
8. **Hooker, Giles**, 2013, “A review of Boosting: Foundations and Algorithms by Schapire and Freund”, *Journal of the American Statistical Association*, 108(502):750-754..

9. **Hooker, Giles** and Anand Vidyashankar, 2011, “Consistency and Efficiency of Conditional Disparity Methods”, Technical Report BU-1670-M, Department of Biological Statistics and Computational Biology, Cornell University.
10. **Hooker, Giles** and Stephen P. Ellner, 2010, “On Forwards Prediction Error”, Technical Report BU-1679-M, Dept. Bio. Stat. and Comp. Bio., Cornell
11. **Hooker, Giles**, 2010, “Comments on: Dynamic Relations for Sparsely Sampled Gaussian Processes”, *TEST*, 19:50-53.
12. Finkelman, Matthew, **Giles Hooker** and Zhen Wang, 2009, “Unidentifiability and Lack of Monotonicity in the Multidimensional Three-Parameter Logistic Model”. Technical Report BU-1678-M, Dept Bio. Stat. and Comp. Bio., Cornell
13. **Hooker, Giles** and Lorenz T. Biegler, 2007. “ILOPT and Neural Dynamics: Tips, Tricks and Diagnostics”, Technical Report BU-1676-M, Dept Bio. Stat. and Comp. Bio., Cornell
14. **Hooker, Giles**, 2007. “Theorems and Calculations for Smoothing-Based Profiled Estimation of Differential Equations”. Technical Report BU-1671-M, Dept Bio. Stat. and Comp. Bio., Cornell
15. Schwartzman, Armin, Matthew Finkelman and **Giles Hooker**, 2004. “The Stanford Statistics Songbook: A Musical Tribute”. Technical Report, Department of Statistics, Stanford
16. **Hooker, Giles** and Fuliang Weng, 2004. “Subset Selection in Large, Sparse Systems: An application of the Forward Stagewise approach to Natural Language Processing”. Technical Report, Robert Bosch Corporation.
17. **Hooker, Giles**, 1999. “Developing a Spline-Smoothed Density”. Technical Report, Research School of Information Sciences and Engineering, Australian National University.

Published Abstracts

Seto, Skyler, Sarah Tan, **Giles Hooker** and Martin T. Wells, “A Double Parametric Bootstrap Test for Topic Models”, *NIPS 2017 Symposium on Interpretable Machine Learning*.

Tan, Sarah, Rich Caruana, **Giles Hooker** and Yin Lou, 2017, “Detecting Bias in Black-Box Models Using Transparent Model Distillation”, *NIPS 2017 Symposium on Interpretable Machine Learning*.

Tan, Sarah, **Giles Hooker** and Martin T. Wells, 2016, “Tree Space Prototypes: Another Look at Making Tree Ensembles Interpretable”, *NIPS 2016 Workshop on Interpretable Machine Learning in Complex Systems*.

Tan, Sarah, **Giles Hooker** and Martin T. Wells, 2016, “ Probabilistic Matching: Incorporating Uncertainty to Correct for Selection Bias”, *NIPS 2016 Causal Inference Workshop*.

Monahan, Edward C, **Giles Hooker** and Christopher J. Zappa, 2015, “The Latitudinal Variation in the Wind-Speed Parameterization of Oceanic Whitecap Coverage: Implications for

Global Modelling of Air-Sea Gas Flux and Sea Surface Aerosol Generation”, *American Meteorological Society*.

Software Development

Author: “AdaLoss: Adaptive-Loss Boosting”, Matlab routines.

Author: “CollocInfer: Collocation Inference in Nonlinear Stochastic Dynamics”, R functions and Manual.

Developer: `fda` library in R and Matlab for Functional Data Analysis.

Author: “ProfileDE: Smoothing Methods for Nonlinear Dynamics”. Manual and Matlab Software.

Manuscripts Under Review or Revision

Hooker, Giles and Hanlin Shang, “Selecting the Derivative of a Functional Covariate in Scalar-on-Function Regression”.

Zhou, Zhengze, Lucas Mentch and **Giles Hooker**, “V-Statistics and Variance Estimation”.

Tredennick, Andrew T., **Giles Hooker**, Stephen P. Ellner and Peter B. Adler, 2020, “A practical guide to selecting models for exploration, inference, and prediction in ecology”.

Sharmistha Sikdar and **Giles Hooker**, “Price Dynamics on Amazon Marketplace: A Machine Learning Approach”.

Zhou, Yichen and **Giles Hooker**, “Tree Boosted Varying Coefficient Models”.

Zhou, Yichen, Zhengze Zhou and **Giles Hooker**, “Approximation Trees: Statistical Stability in Model Distillation”.

Zhou, Yichen and **Giles Hooker** “Boulevard: Regularized Stochastic Gradient Boosted Trees and Their Limiting Distribution”.

Sarah Tah, Rich Caruana, **Giles Hooker** and Yin Lou, “Auditing Black-Box Models Using Transparent Model Distillation With Side Information”

Sharmistha Sikdar and **Giles Hooker**, “A Hidden Semi-Markov Model of Customers’ Multi-channel Engagement Dynamics”

Sinclair, David G. and **Giles Hooker**, “An Expectation Maximization Algorithm for High-Dimensional Model Selection for the Ising Model with Misclassified States”

Editorial Service

- Program Committee Member, NeurIPS 2020 workshop on “Machine Learning for the Developing World (ML4D): Improving Resilience”
- Editorial Board Member, Journal of Machine Learning Research, 2020 -
- Program Committee member, ICLR Workshop on Debugging Machine Learning Models, 2019
- Program Committee member, ACM FAT* 2019
- Editorial Board Member, STATS.org, 2015 -
- Associate Editor, Journal of the American Statistical Association, 2014 - 2021.
- Associate Editor, Journal of the Royal Statistical Society, 2013 - 2015.
- Associate Editor, Electronic Journal of Statistics, 2010 - 2013.
- Program Committee member, International Workshop on Statistical Modeling, July 2009.

Professional Service

- Organizer, Workshop on the Interface of Machine Learning and Statistical Inference, Banff International Research Station, January 2018.
- Co-organizer, Workshop on Distributed Data for Dynamics and Manifolds, Banff International Research Station, September 2017.
- Co-organizer, Invited Session on “Machine Learning and Statistical Inference: Building Breiman’s Bridge”, JSM 2017.
- Co-organizer, Cornell Day of Statistics, September 2015, 2016.
- Advisory Board, Pediatric Epilepsy Research Foundation grant for “Care Management for Pediatric Epilepsy: Predictive Modeling and Evaluation”.
- Organizer, Invited Session on “Statistical Inference with Random Forests and Related Ensemble Learners”, ENAR 2015.
- Organizer, Workshop on Statistics and Nonlinear Dynamics in Biology and Medicine, Banff International Research Station, July 2014.
- Organizer, Topic Contributed Session on “Disparities as Nonparametric Interfaces to Parametric Models”, Joint Statistical Meetings, August 2013.
- Organizer, Workshop on Recent Advances in Statistical Inference for Mathematical Biology, Mathematical Biology Institute, 2012.
- Working group leader, “Dynamics and Inference”, SAMSI Program of the Analysis of Object Data, 2010-2011.
- Organizer, Topic Contributed Session on “New Advances in Disparity and Divergence Methods for Statistical Inference”, Joint Statistical Meetings, August 2010.

- Organizer, Workshop on Statistical Methods for Dynamic Systems Models, Vancouver, June 2009.
- Organizer, Topic Contributed Session on “Statistical Methods for Differential Equation Models with Scientific Applications”, Joint Statistical Meetings, August 2008.
- Organizer, Invited Session on “Advances in Functional Data Analysis”, Joint Statistical Meetings, August 2008.
- Organizer, Topic Contributed Session on “Statistics for Nonlinear Dynamics”, Joint Statistical Meetings, August 2007.
- Organizer, Workshop on Statistical Methods for Modeling Dynamical Systems, Centre des Recherches Mathematiques, Montreal. July 2007.

Public Engagement

Blog: Of Models and Meanings

<http://blogs.cornell.edu/modelmeanings/>

Australian Fulbrighter, 03/2020, “A Matter of Chance”,

<https://www.fulbright.org.au/news-and-events/2020/03/a-matter-of-chance/>

PBS NOVA Scientific Advisor, “Prediction by the Numbers”, 02/28/2018

<http://www.pbs.org/wgbh/nova/physics/prediction-numbers.html>

STATS.org 03/30/2017: Machine Learning: What journalists need to know

<http://www.senseaboutscienceusa.org/machine-learning-journalists-need-know/>

STATS.org 11/15/2016: What the 2016 Presidential Election Taught us about Polling, Prediction

<http://www.stats.org/2016-presidential-election-taught-us-polling-predictions/>

geneticexperts.org: 12/16/2015 New analysis finds environmental factors, not random mutations, contribute to majority of cancer risk

<http://geneticexperts.org/environmental-factors-not-random-mutations-contribute-to-majority-of-cancer-risk/>

STATS.org 11/24/2015: Racism, redcards and rabbit holes

<http://www.stats.org/racism-redcards-and-rabbit-holes/>

STATS.org 08/11/2015: Online dating and the statistical dark arts

<http://www.stats.org/online-dating-and-the-statistical-dark-arts/>

Short Courses

Functional Data Analysis: Methods and Computing, *Department of Ecology, Yale University*, April 20, 2017

Functional Data Analysis: Methods and Computing, *Joint Statistical Meetings*, Seattle, August 8, 2015

Functional Data Analysis, *International Workshop on Statistical Modeling*, Glasgow, July 3, 2010

Cornell Statistical Consulting Unit Workshop: Introduction to Functional Data Analysis, Spring 2008, Spring 2009

Teaching

STSCI 6940: Multivariate and Functional Data, Fall 2020.

STAT 8024: Special Topics in Applied Statistics: Machine Learning and Statistics, Australian National University, Spring 2020.

ILRST/ORIE/STSCI 7170: Linear Models Fall 2018

BTRY/STSCI 4030, STSCI 5030: Linear Models with Matrices Fall 2016-2018

BTRY 4090: Theory of Statistics Fall 2015

BTRY 6020: Statistical Methods II Spring 2009, 2010, 2011, 2015

BTRY 7950: Graduate Student Consulting Unit Spring/Fall 2006-2008, 2011, 2014

BTRY 6520: Computationally Intensive Methods in Statistics Fall 2014

STSCI 7740: Predictive Competitions Spring 2013

BTRY/STSCI 4520/3520: Statistical Computing¹ Spring 2012, 2013, 2017-2019

BTRY 6940: Inference in Nonlinear Dynamics Fall 2012

BTRY 7180: Generalized Linear Models Fall 2011

BTRY 6150: Applied Functional Data Analysis² Fall 2008

BTRY 694: Theory of Multivariate Statistics Spring 2008, Fall 2009

BTRY 694: Statistical Learning Theory Fall 2007

BTRY 694: Functional Data Analysis Spring 2007

Math 204: Principles of Statistics II. Department of Mathematics and Statistics Winter 2005 (McGill University)

2001 - 2004 Teaching Assistant, Statistics Dept, Stanford University.

Tutor, Math 100, Mathematics Dept, Australian National University, 1999

Mathematics and Philosophy Tutor, Burgmann College, Australian National University, 1997

¹Course materials have since been used at University of Pittsburgh and Washington State University

²Course materials have since been used at Simon Fraser University, Colorado State University, Glasgow University, University of Newcastle, Johns Hopkins University, University of Manitoba, Chungnam National University, Singapore Institute of Technology

Students

Zhengze Zhou	PhD		Current
Indrayudh Ghosal	PhD		Current
Hui Fen Tan	PhD (with M. Wells)	2019	Facebook
<i>Interpretable Approaches to Opening Up Black Box Models</i>			
Zi Ye	PhD	2019	Microsoft
<i>Functional Single Index Models and Jensen Effect</i>			
Yichen Zhou	PhD	2019	Google Data Science
<i>Asymptotics and Interpretability of Decision Trees and Decision Tree Ensembles</i>			
David Sinclair	PhD	2017	Google: Android Analytics
<i>Model Selection Results for High Dimensional Graphical Models on Binary and Count Data with Applications to FMRI and Genomics</i>			
Keegan Kang	PhD	2017	Lectuer, Singapore U. Tech. & Design
<i>Data Dependent Random Projections</i>			
Lucas Mentch	PhD	2015	Asst. Prof. University of Pittsburgh
<i>Statistical Inference with Ensemble Methods in Machine Learning</i>			
Leifur Thorbergson	PhD	2014	Memorial Sloan Kettering Cancer Center
<i>Experimental Design for Partially Observed Markov Decision Processes</i>			
Cecilia Earls	PhD	2014	Lecturer, Cornell University
<i>Bayesian Hierarchical Gaussian Process Models for Functional Data Analysis</i>			
Chong Liu	PhD (BU with S. Ray)	2013	State Street Global Advisors
<i>Functional Principal Component and Factor Analysis of Spatially Correlated Data</i>			
Mathew McLean	PhD	2013	Postdoc with Matt Wand
	(with D. Ruppert)		
<i>On Generalized Additive Models for Regression with Functional Data</i>			

Yuefeng Wu Postdoc 2009-2011 Asst. Prof. Missouri State, St Louis
Bayesian Nonparametrics in Minimum Hellinger Distance Estimates

Maria Asencio MSc 2010 Mosaic ATM
Functional Convolution Models

Graduate Student Committees and External Examiner Service

** indicates substantial supervision and co-authorship.

Juan Beltran	PhD	2019	Computational Biology
Sharmistha Sikdar**	PhD	2019	Marketing
Aurya Javeed**	PhD	2018	Applied Mathematics
Guo Yu	PhD	2018	Statistics
Claudiu Dinicu	MSc	2018	Statistics
Marie-Hélène Descaray	PhD	2017	Statistics: External Examiner École Polytechnique Fédérale de Lausanne
Mara Bernardi	PhD	2016	Statistics: External Examiner Politecnico di Milano
Kelly Kirtland	PhD	2016	Statistics
Lei Huang	PhD	2016	Computational Biology
Galina Nogin	MSc	2016	Statistics
Tiffany Ho	MSc	2016	Computational Biology
Yanning Liu	PhD	2016	Statistics
Jon Steingrimsson	PhD	2015	Statistics
Muting Wan	PhD	2014	Statistics

Andre Martins	PhD	2014	Computational Biology
Yin Lou**	PhD	2014	Computer Science
Ben Dalziel	PhD	2014	Ecology and Evolutionary Biology
Xun Wang	PhD	2013	Civil and Environmental Engineering
Laura de Vargas Roditi**	PhD	2013	Computational Biology
Daniel Ly	PhD	2013	Mechanical Engineering
Scott Asay	PhD	2013	Accounting
Caitlin Cunningham	PhD	2012	Statistics
Luo Xiao	PhD	2012	Statistics
Nikolaus Karampatziakis	PhD	2012	Computer Science
Michael Adams	MSc	2011	Food Sciences
Dennis Oransky	MSc	2011	Statistics
Yingxing Li	PhD	2010	Statistics
David Clement	PhD	2010	Statistics
Nam Nguyen	PhD	2010	Computer Science
Darrell Sonntag	PhD	2009	Civil and Environmental Engineering
Kathryn Barger	PhD	2008	Statistics
Daria Sorokina	PhD	2007	Computer Science
Emmanuel Asaba	MSc	2007	Civil and Environmental Engineering

Department and University Service

- CALS Statistics and Data Science Curricula Task Force, 2019
- Provosts Task Force on Data Science, 2016-2017
- Faculty/Lecturer Search Committee Chair
 - Biological Statistics and Computational Biology: 2015, 2018
 - Statistical Science/Industrial and Labor Relations: 2017, 2018
- Faculty Search Committee Membership
 - Biological Statistics and Computational Biology: 2008, 2009, 2011, 2012
 - Industrial and Labor Relations: 2011
 - Statistical Science: 2014, 2015, 2016
- *Ad hoc* tenure committees: 2014, 2015, 2016
- Peer teaching evaluations, 2012-2017.
- Website re-design, Department of Biological Statistics and Computational Biology, 2009
- Curriculum Committee, College of Agriculture and Life Sciences, 2009-2012

Invited Talks (last 10 years)

2020

“Optimal Adaptive Design of Experiments for Stochastic Dynamic Systems”, Statistics Seminar, University of Pennsylvania, November 2020.

“Ensembles of Trees and CLT’s: Inference and Machine Learning”, Statistics Seminar, Johns Hopkins University, September 2020.

“Statistics and Dynamical Systems”, Introductory Overview Lecture with Sy-Miin Chow, Joint Statistical Meetings, August 2020.

“Ensembles of Trees and CLT’s: Inference and Machine Learning”, Statistics Seminar, Australian National University, May 2020.

“Ensembles of Trees and CLT’s: Inference and Machine Learning”, Statistics Seminar, Monash University, March 2020.

“Ensembles of Trees and CLT’s: Inference and Machine Learning”, Statistics Seminar, University of Melbourne, March 2020.

“Optimal Adaptive Design of Experiments for Stochastic Dynamic Systems”, Neyman Seminar, University of California at Berkeley, January 2020.

2019

“Ensembles of Trees and Inference Boosting, U and V Statistics”, Statistics Seminar, Temple University, November 2019.

“Ensembles of Trees and CLT’s: Inference and Machine Learning”, Neyman Seminar, University of California at Berkeley, November 2019.

“Inference and Gradient Boosting”, Conference of the Israel Statistics Association, Tel Aviv, June 2019.

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, Tel Aviv University, June 2019.

“Optimal Adaptive Design of Experiments for Stochastic Dynamic Systems”, Statistics Seminar, Ben Gurion University, June 2019.

“Inference and Gradient Boosting”, Workshop in Celebration of Jerry Friedman’s 80th birthday, May, 2019.

“Boosting and Bagging and Uncertainty Quantification”, Statistics Seminar, University of Pennsylvania, April, 2019.

2018

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, Iowa State University, October 2018.

“Explanations, Interpretation, Uncertainty Quantification”, Machine Learning in Medicine Symposium, Cornell University, September 2018.

“An ODE to Statistics: Inference about Nonlinear Dynamics”, PIMS-IAM Distinguished Colloquium, University of British Columbia, January 2018.

2017

“Testing Curvature in Functional Single Index Models: Inference for Ecological Reaction Norms”, Joint Statistical Meetings, August 2017.

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, Imperial College London, May 2017.

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, Nottingham University, May 2017.

“Decision Trees and CLT’s: Machine Learning and Ecological Inference”, Ecology Seminar, Yale University, April 2017.

“Tests for Lack of Fit and Missing State Variables in Ordinary Differential Equation Models”, Statistics Seminar, University of Cyprus, April 2017.

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, University of Cyprus, February 2017.

2016

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, Simon Fraser University, November 2016.

“Truncated Linear Models for Functional Data”, Royal Statistical Society Meetings, JRSSB Editors Invited Session, September 2016.

“On the Range of Integration of a Functional Linear Model”, CRoNoS Workshop on Functional Data Analysis, August 2016.

“Testing High Dimensional Interactions with Random Forests”, COMPSTAT, August 2016.

“Testing High Dimensional Interactions with Random Forests”, International Chinese Statistics Symposium, July 2016.

“Three Unidentifiable Problems in Functional Data Analysis”, Workshop on Functional Data Analysis, Les Diablerets, May 2016.

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, Temple University, February 2016.

2015

“Tests for Lack of Fit and Missing State Variables in Ordinary Differential Equation Models”, York University, October 2015.

“Tests for Lack of Fit and Missing State Variables in Ordinary Differential Equation Models”, University of Michigan, September 2015.

“Modeling Covariance in Functional Data Analysis”, Joint Statistical Meetings, August 2015.

“An ODE to Statistics: Inference for Nonlinear Dynamics”, Statistics Seminar, Syracuse University, January 2015.

“Subsample Trees and CLTs: Inference for Machine Learning”, Biostatistics Seminar, Weil Cornell Medical College, January 2015.

2014

“Subsample Trees and CLTs: Inference for Machine Learning”, Statistics Seminar, North Carolina State University, October 2014.

“Subsample Trees and CLTs: Inference for Machine Learning”, Artificial Intelligence Seminar, Cornell University, October 2014.

“Robustness, Inference and Gradient Matching”, BIRS Workshop on Statistics and Nonlinear

Dynamics in Biology and Medicine, July 2014.

“Domain Selection and Truncated Functional Linear Models”, IMS Meetings, July 2014.

“Experimental Design in Stochastic Dynamic Systems”, Mathematics and Business Analytics Seminar, IBM Research, June 2014.

“Ensemble Trees and CLTs: Inference for Machine Learning”, Health Care Analytics Seminar, IBM Research, April 2014.

“Experimental Design in Stochastic Dynamic Systems”, Statistics Seminar, University of British Columbia, January 2014.

“Tests for Lack of Fit and Missing State Variables in Ordinary Differential Equation Models”, Statistics Seminar, Simon Fraser University, January 2014.

2013

“Robust and Efficient Analysis of Conditionally-Specified Models via Disparities”, Summer Camp, Australian National University, December 2013.

“Paradoxical Results in Multidimensional Item Response Theory”, Applied Statistics Seminar, Australian National University, November 2013.

“Robust and Efficient Analysis of Conditionally-Specified Models via Disparities”, Statistics Seminar, University of Melbourne, October 2013.

“Control Theory and the Design of Dynamic Experiments”, Statistics Seminar, University of California at Davis, 2013.

“Measures of Robustness in Regularized Estimates”, Joint Statistical Meetings, August 2013.

“Measures of Robustness in Regularized Estimates”, International Conference on Robust Statistics, July 2013.

“Functional Convolution Models: Design and Domain Selection”, Statistical Society of Canada Meetings, June 2013.

“Control Theory and the Design of Dynamic Experiments”, Mathematics Seminar, De Paul University, April 2013.

“Robust and Efficient Analysis of Conditionally-Specified Models via Disparities”, Statistics Seminar, University of Wisconsin, Madison, April 2013.

“Control Theory and the Design of Dynamic Experiments”, Mathematics Seminar, Karlsruhe Institute für Technologie, January 2013.

2012

“Control Theory and the Design of Dynamic Experiments”, Greek Stochastics Meetings, Au-

gust, 2012.

“Control Theory and the Design of Dynamic Experiments”, Statistical Society of Canada Meetings, June 2012.

“Detecting Evolution in Experimental Ecology: Diagnostics for Missing State Variables”, Statistics Seminar, Cornell University, March 2012.

“Robust and Efficient Analysis of Conditionally-Specified Models via Disparities”, Statistics Seminar, University of Waterloo, January 2012.

“Robust and Efficient Analysis of Conditionally-Specified Models via Disparities”, Statistics Seminar, University of Chicago, January 2012.

“Detecting Evolution in Experimental Ecology: Diagnostics for Missing State Variables”, Ecology and Evolution Seminar, University of Chicago, January 2012.

“Detecting Evolution in Experimental Ecology: Diagnostics for Missing State Variables”, Biostatistics Seminar, Johns Hopkins University, January 2012.

2011

“Bayesian Robustness via Disparities”, International Conference on Advances in Probability and Statistics, Hong Kong, December 2011.

“Robustness, Efficiency and Regularization in Model Selection: LASSO with Disparities”, Conference on Statistical Concepts and Methods for the Modern World, Colombo, Sri Lanka, December 2011.

“Robustness and Efficiency in Conditionally Specified Models via Disparities”, 2011 International Forum on Modern Statistics and Econometrics, Wang Yanan Institute for Studies in Economics, Xiamen, China, December 2011.

“Detecting Evolution in Experimental Ecology: Diagnostics for Missing State Variables”, Applied Mathematics Seminar, University of Colorado, December 2011.

“Detecting Evolution in Experimental Ecology: Diagnostics for Missing State Variables”, Colloque de Statistique de Montréal, Centre de Recherches Mathématiques, Montreal, December 2011.

“Robust and Efficient Analysis of Conditionally-Specified Models via Disparities”, University of Michigan, September 2011.

“Functional Convolution Models, Design and Domain Selection”, The University of Melbourne, June 2011.

“Robust and Efficient Analysis of Non-Stationary Data via Disparities”, Macquarie University, June 2011.

“Functional Convolution Models”, ENAR, March 2011.