Le Bao

Email: lebao@psu.edu, Office: Wartik 514-C, Phone: 206-295-9044

Research interests:

Bayesian methods, machine learning, mixture models, stochastic modeling, computational methods, infectious diseases dynamics, and bioinformatics.

Professional position:

2011 – present Assistant Professor, Department of Statistics, The Pennsylvania State University

2014 – present Associate director of Center for Advanced Data Assimilation and Predictability Techniques (ADAPT) , The Pennsylvania State University

Education:

2006 – 2011 Ph.D. in Statistics, University of Washington, Seattle, WA

2004 – 2005 M.S. in Statistics, Dalhousie University, Canada

2000 – 2004 B.S. in Applied Mathematics, Peking University, China

Awards:

2010 Chinese Government Award for Outstanding Students Abroad

2009 Z.W. Birnbaum Award for Outstanding General Exam University of Washington

Refereed publications:

1. Malhotra, R., Elleder, D., **Bao, L.**, Hunter, D. R., Poss, M., Acharya, R. (2016). A pipeline for identifying integration sites of mobile elements in the genome using next-generation sequencing. *Bioinformatics and Computational Biology.* [In Press].
2. Li R., Dudek S.M., Kim D., Hall M.A., Bradford Y., Peissig P.L., Brilliant M.H., Linneman J.G., McCarty C.A., **Bao L.**, and Ritchie M.D. (2016) Identification of genetic interaction networks via an evolutionary algorithm evolved Bayesian Network. *Bio Data Mining.* [In Press].
3. Thomas J. and **Bao L.** (2016). Modeling the dynamics of an HIV epidemic. *Dynamic Demographic Analysis.* [In Press].

4. **Bao L.**, Raftery A.E., Reddy A. (2015) Estimating the sizes of populations at risk of HIV infection from multiple data sources using a Bayesian hierarchical model. *Statistics and Its inference*. 8(2): 125–136.
5. **Bao L.**, (2015). Comment on Quasi-Monte Carlo sampling by Gerber and Chopin. *Journal of the Royal Statistical Society: Series B*, 77: 560.
6. **Bao L.**, Elleder D., Malhotra R., DeGiorgio M., Maravegias T., Horvath L., Carrel L., Gillin C., Hron T., Fábryová H., Hunter D. and Poss M. (2014) Computational and statistical analyses of insertional polymorphic endogenous retroviruses in a non-model organism. *Computation*. 2: 221-245.
7. **Bao L**., Ye J., Hallett T.B. (2014) Incorporating incidence information within the UNAIDS Estimation and Projection Package framework: a study based on simulated incidence assay data. *AIDS* 28: S515-S522.
8. Brown T., **Bao L.**, Eaton J.W., Hogan D.R., Mahy M., March K., Mathers B.M., Puckett R. (2014) Improvements in prevalence trend fitting and incidence estimation in EPP 2013. *AIDS* 28: S415-S425.
9. Kamath P.., Elleder D., **Bao L.**, Cross P., Poss M. (2013) The population history of endogenous retroviral elements in mule deer (Odocoileus hemionus). *Journal of Heredity*, 105: 173-187.
10. **Bao L.** (2012). A new infectious disease model for estimating and projecting HIV/AIDS epidemics. *Sexually Transmitted Infections*, 88: i58-i65.
11. **Bao L.**, Salomon J.A., Brown T., Raftery A.E., and Hogan D.R. (2012). Modelling national HIV/AIDS epidemics: revised approach in the UNAIDS Estimation and Projection Package 2011. *Sexually Transmitted Infections*, 88: i3-i10.
12. Clark S.J., Thomas J., and **Bao L.** (2012). Estimates of Age-Specific Reductions in HIV Prevalence in Uganda: Bayesian Melding Estimation and Probabilistic Population Forecast with an HIV-enabled Cohort Component Projection Model. *Demographic Research* 27: 743-774.
13. **Bao L.**, Fricks J. and Haran M. (2012). Comment on the Mechanistic Modeling and Inference for Cell Motility by Manolopoulou et al. *Journal of the American Statistical Association*, 107: 869−871.
14. **Bao, L.,** Hogan D., Raftery A.E., and Salomon J.A. (2011). A flexible model for estimating and projecting trends in HIV/AIDS epidemics. *Technical report, UNAIDS, Geneva*.

15. Meila M. and **Bao L.** (2010). An exponential model for infinite rankings. *Journal of Machine Learning Research*, 11: 3481−3518.
16. Raftery A.E. and **Bao L.** (2010). Estimating and projecting trends in HIV/AIDS generalized epidemics using incremental mixture importance sampling. *Biometrics,* 66: 1162-1173.
17. **Bao L.**, Raftery A.E. (2010). A stochastic infection rate model for estimating and projecting national HIV prevalence rates. *Sexually Transmitted Infections,* 86: ii93-ii99.
18. Brown T., **Bao L.**, Raftery A.E., Salomon J.A., Baggaley R.F., Stover J., and Gerland P. (2010). EPP 2009: bringing the UNAIDS Estimation and Projection Package into the ART era. *Sexually Transmitted Infections,* 86: ii3-ii10.
19. **Bao L.**, Gneiting T., Grimit E., Guttrop P. and Raftery A.E. (2010). Bias Correction and Bayesian Model Averaging for Ensemble Forecasts of Surface Wind Direction. *Monthly Weather Review*, 138:1811-1821.
20. **Bao L.,** Zhu, Z. and Ye, J. (2009). Modeling oncology gene pathways network with multiple genotypes and phenotypes via a copula method. *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 237- 246.
21. Meila M. and **Bao L.** (2008). Estimation and clustering with infinite rankings. *Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence*, 24:393-402.
22. **Bao L.**, Gu H., Dunn, K.A. and Bielawski J. (2008). Likelihood Based Clustering (LiBaC) for Codon Models, a method for grouping sites according to similarities in the underlying process of evolution. *Molecular Biology and Evolution.* 25:1995-2007.
23. **Bao L.**, Gu H., Dunn K.A. and Bielawski J. (2007). Methods for selecting fixed-effect models for heterogeneous codon evolution, with comments on their application to gene and genome data. *BMC Evolutionary Biology*. 7 Suppl 1:S5.
24. Mitnitski A, **Bao L.** and Rockwood K. (2007). A cross-national study of transitions in deficit counts in two birth cohorts: Implications for modeling ageing. *Experimental Gerontology.* 42:241-246.
25. Mitnitski A, **Bao L.** and Rockwood K. (2006). Going from bad to worse: a stochastic model of transitions in deficit accumulation, in relation to mortality. *Mechanisms of Ageing and Development.* 127: 490-493.

Papers under review:

26. **Bao L.,** Ben S., Niu X., Tang Y., Brown T., Ghys P. and Eaton J. Incorporating Hierarchical Structure Into Dynamic Systems: An Application Of Estimating HIV Epidemics At Sub-National And Sub-Population Level. http://arxiv.org/abs/1602.05665
27. **Bao L.,** Niu X. Mahy M., and Ghys P. Estimating HIV Epidemics for Sub-National Areas. http://arxiv.org/abs/1508.06618
28. **Bao L.**, Hunter D.R., and Poss M. Assigning polymorphic endogenous retrovirus integration sites using a mixture model. http://arxiv.org/abs/1510.00028
29. **Bao L.,** Niu X. and Yeung K.Y. Inference of Gene Associations using Model-based Clustering and Permutation Test in Time-course Gene Expression Data.
30. Liu L., Ou L., **Bao L.**, and Hogan D. Age-specific distributions for cause of death among children aged 1-59 months.
31. Cheng W.F., Gao X., Bao L., Mitchell D.C., Wood C., Sliwinski M.J., Smiciklas-Wright H., Still C., Rolston D., Jensen G.L. Obesity as a Risk Factor for Developing Functional Limitation Among Older Adults: Results from Conditional Inference Tree Analysis.

Reports:

32. UNAIDS (2015). Guidelines on monitoring the impact of the HIV epidemic using population-based surveys. http://www.unaids.org/en/resources/documents/2015/population_based_surveys

Invited talks:

1. 2016 JSM, Chicago, 08/03/2016, *Estimating HIV Epidemics at Fine Scales*

2. The Third Taihu International Statistics Forum, Shanghai, China, 07/10/2016, *Statistical models for Estimating and Predicting HIV Epidemics*

3. The School of Mathematics, Peking University, 06/27/2016, *Statistical models for Estimating and Predicting HIV Epidemics*

4. The School of Economics, Xiamen University, 06/22/2016, *Statistical models for Estimating and Predicting HIV Epidemics*

5. National Center for AIDS/STD Control and Prevention, China CDC, 06/16/2016, *Estimating HIV Epidemics at Fine Scales*

6. Hershey/Institute for Personalized Medicine, Penn State University, 04/29/2016, *Network-based discriminant analysis*

7. CTSI BERD seminar, Penn State University, 03/29/2016, *Leaving No One Behind -- Estimating HIV Epidemics At Fine Scales*

8. Department of Global Health, University of Washington, 02/29/2016, *Estimating HIV Epidemics At Sub-National And Sub-Population Level*

9. Department of Statistics, Temple University, 11/20/2015, *An Efficient Way of Estimating HIV Epidemics in Sub-National Areas and Sub-Populations*

10. UNAIDS Reference Group on Estimates, Models and Projections, London, 10/26/2015, *Incorporating PMTCT data in EPP: Investigation of level of continuity required, data quality requirements; Further testing of the hierarchical model.*

11. Department of Statistics, Columbia University, 09/28/2015, *Estimating HIV Epidemics for Sub-National Areas*

12. Big Data Social Science IGERT program, Penn State University, 09/09/2015, *Estimating HIV Epidemics for Sub-National Areas*

13. Bioinformatics and Genomics Retreat, Penn State University, 08/29/2015, *Assigning viruses from sequence count data via a mixture model*

14. UNAIDS Reference Group on Estimates, Models and Projections, Boston, 06/04/2015, *Hierarchical model - revisions, accuracy, efficiency, use with key populations*

15. UNAIDS Reference Group on Estimates, Models and Projections, Boston, 06/03/2015, *Incorporating PMTCT data into EPP fitting*

16. ENAR, Miami, 03/17/2015, *Compression of Complex Data with an Example of Time Series Gene Expression and Biomarker*

17. Department of Genomic, Penn State University, 02/27/2015, *Compression of Complex Data with an Example of Time Series Gene Expression*

18. Bill & Melinda Gates Foundation, Seattle, 02/17/2015, *Towards a Target Product Profile for HIV Incidence Assay Development*

19. Information Communication Technology for Development (ICT4D), Penn State University, 12/17/2014, *HIV Epidemics Study*

20. The Methodology Center, Penn State University, 11/13/2014, *A Hierarchical Model for Estimating HIV/AIDS Epidemics*

21. UNAIDS Reference Group on Estimates, Models and Projections, UNAIDS, Geneva, 10/29/2014, *Considerations for incorporating PMTCT as an additional data source with own calibrating parameter*

22. UNAIDS Reference Group on Estimates, Models and Projections, UNAIDS, Geneva, 10/28/2014, *Hierarchical approach for generating sub-national estimates within the EPP framework*

23. UNAIDS Reference Group on Estimates, Models and Projections, Seattle, 04/25/2014, *Bayesian hierarchal model for sharing information across areas and countries*

24. Institute for Health Metrics and Evaluation, University of Washington, Seattle, 04/23/2014, *Statistical Models for Estimating and Predicting HIV/AIDS Epidemics*

25. National Institutes of Health (NIH), NICHD, Bethesda, Maryland, 02/20/2014, *Compression of Complex Data with an Example of Time Series Gene Expression*

26. UNAIDS Reference Group on Estimates, Models and Projections, Spain, 08/12/2013, *Use of incidence assays within the EPP framework*

27. The 2nd Taihu International Statistics Forum, Soo Chow University, China, 07/07/2013, *Inference of Gene Associations using Model-based Clustering and Adjusted Rand Index in Time-course Gene Expression Data*

28. IMS-China, International Conference on Statistics and Probability, Southwestern University of Finance and Economics, China, 07/02/2013, *Inference of Gene Associations using Model-based Clustering and Adjusted Rand Index in Time-course Gene Expression Data*

29. ENAR, Orlando, 03/11/2013, *A new infectious disease model for estimating and projecting HIV/AIDS epidemics*

30. Working group on Stochastic Modeling and Computational Statistics, State College, PA, 11/15/2012, *Incremental Mixture Importance Sampling*

31. UNAIDS Reference Group on Estimates, Models and Projections, London, 09/24/2012, *A new infectious disease model for estimating and projecting HIV/AIDS epidemics*

32. Second Biostatistics Symposium, Renmin University, Beijing, 07/09/2012, *New Methods for Estimating and Projecting National HIV/AIDS Prevalence Rates*

33. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 06/15/2012, *Incremental Mixture Importance Sampling for Estimating and Predicting HIV/AIDS Epidemics*

34. Department of Mathematics and Statistics, University of Minnesota, Duluth, 05/04/2012, *Statistical Models for Estimating and Predicting HIV/AIDS Epidemics*

35. UNAIDS Reference Group on Estimates, Models and Projections, Boston, 04/02/2012, *New Improvements on R-Flexible Model*

36. Student Advisory Committee (SAC) Seminar, Department of Statistics, The Pennsylvania State University, State College, PA, 03/21/2011, *New Methods for Estimating and Projecting National HIV/AIDS Prevalence Rates*

37. Department of Statistics, University of Washington, Seattle, WA, 05/17/2011, *Statistical Models for Estimating and Projecting HIV/AIDS Epidemics*

38. Center of Statistical Science, Peking University, Beijing, China, 04/12/2011, *Statistical Models for Estimating HIV Prevalence and the At-risk Population Size*

39. Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, 03/31/2011, *Statistical Models for Estimating HIV Prevalence and the At-risk Population Size*

40. Department of Statistics, The Pennsylvania State University, State College, PA, 02/10/2011, *Statistical Models for Estimating HIV Prevalence and the At-risk Population Size*

41. Department of Statistics, National University of Singapore, Singapore, 01/21/2011, *Estimating the Size of Populations at High Risk of HIV Using a Bayesian Hierarchical Model*

42. UNAIDS Reference Group on Estimates, Models and Projections, Boston, 10/26/2010, *A Flexible Model for Estimating and Projecting HIV Prevalence Rates*

43. Joint Statistical Meetings, Vancouver, Canada, 08/01/2010, *The Bayesian Hierarchical Model for Estimating the Size of HIV At-risk Populations in Bangladesh*

44. UNAIDS Reference Group on Estimates, Models and Projections, Glastonbury, 05/21/2010, *A Stochastic Infection Rate Model for Estimating and Projecting National HIV Prevalence Rates*

45. UNAIDS Reference Group on Estimates, Models and Projections, London, 10/14/2009, *The Bayesian Model for Estimating the size of HIV at-risk Population*

46. Working group on Model-Based Clustering Summer Session, Seattle, WA, 07/10/2008, *Incremental Mixture Importance Sampling*

47. Pacific Northwest Weather Workshop, Seattle, WA, 02/29/2008, *Probabilistic Wind Direction Forecasting Using Bayesian Model Averaging*

Grant:

2014-2015 PI, Diagnostics Modelling Consortium -- Evaluating Impacts of Incidence Assays, funded by Imperial College of Science, Technology and Medicine

2014-2015 PI, Development of hierarchical models for estimating health indicators across countries, funded by World Health Organization.

2016-2018 PI, Development of methods to produce spatial estimates of HIV epidemics, funded by UNAIDS.

2016-2017 Co-PI, HWRF Initialization by Comparing with and Adoption of the PSU WRFENKF Method, funded by National Oceanic and Atmospheric Administration.

Teaching:

1. STAT/IST 557, Data Mining, Fall 2011/Fall 2012/Spring2014/Spring2015/Fall 2016
2. STAT554, Categorical Data Analysis, Fall 2014/Fall 2015/Fall 2016
3. STAT/MATH415, Introduction to Mathematical Statistics, Fall 2013/Spring 2016
4. STAT897D, Applied Data Mining, Fall 2012

Scientific Software:

1. IMIS: R-package for Incremental Mixture Importance Sampling.
2. Estimation and Projection Package (EPP): it is used to estimate and project adult HIV prevalence and incidence from surveillance data.
3. Codeml_FE: Comprehensive set of fixed-effect models of codon evolution.
4. LiBaC: Clustering analysis under a probabilistic model of codon evolution.

Professional Memberships:

Institute of Mathematical Statistics, American Statistical Association, the International Society for Bayesian Analysis

Service to the Profession:

2008—present Member of the UNAIDS Reference Group on HIV Estimates, Modelling and Projections who advises on the techniques to be used in EPP and Spectrum. Participated EPP/Spectrum Technical Review Meetings and UNAIDS Reference Group Meetings.

2013—present Statistical Adviser for PLOS One

2015 Consultant, Meeting for HIV/AIDS Response of China.

2014 Consultant, Meeting for HIV/AIDS Response of China.

2010 Consultant, Regional Training on Methods for Size Estimation of Most-at-Risk Populations in the Asia-Pacific Region.

2009 Consultant, Training of Trainers Workshop on HIV/AIDS Estimates and Projections.

Reviewer for the following journals:

Annals of Applied Statistics, Biostatistics, Computational Statistics and Data Analysis, Journal of the American Statistical Association, Pattern Recognition, Statistics and Its Interface, Statistics in Medicine, Technometrics, (in Statistics)

BMC Genetics, BMC Bioinformatics, BioData Mining, Bioinformatics, PLOS one, (in Bioinformatics)

Epidemiology, International Journal of Infectious Diseases, Journal of Acquired Immune Deficiency Syndromes, Journal of AIDS and Clinical Research, Journal of the International AIDS Society, Sexually Transmitted Infections, (in Epidemiology)

Journal of Agricultural, Biological, and Environmental Statistics, The Journal of Nutrition, Health and Aging, Population Health Metrics, The South Pacific Journal of Natural and Applied Sciences, (in Global health)