CURRICULUM VITAE
Abhirup Datta

Part I

PERSONAL DATA

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health.
615 North Wolfe Street, Baltimore, MD 21205.

(410) 502 2988 | abhidatta@jhu.edu | abhidatta.com

EDUCATION, TRAINING

Degrees

- PhD, University of Minnesota, Twin Cities, MN, Biostatistics (2016).
- MStat, Indian Statistical Institute, Kolkata, India, Specialization in Math-Stat-Probability (2010).
- BStat (Hons.), Indian Statistical Institute, Kolkata, India (2008).

PROFESSIONAL EXPERIENCE

- 2016 – present: Assistant Professor, Department of Biostatistics, Johns Hopkins University.
- 2010 – 2012: Quantitative analyst, Morgan Stanley.

PROFESSIONAL ACTIVITIES

Professional Memberships

- American Statistical Association
- International Biometric Society (Eastern North American Region (ENAR))
- International Indian Statistical Association (IISA)
- The International Environmetric Society (TIES) of the International Statistical Institute (ISI)

Program Development

- Review-panel National Science Foundation (NSF) Division of Mathematical Sciences (DMS), 2020
- Scientific Program Committee for the 2020 International Indian Statistical Association conference (IISA).
- Student poster competition judge, IISA INDSTAT (2019).
- Session Chair, IISA INDSTATS (2019).
- Session Chair, Joint Statistical Meetings (2019).
- Session Organizer, Joint Statistical Meetings (2019).
- Student paper award reviewer, Section on Bayesian Statistical Science, Joint Statistical Meetings (2019).
- Session Organizer, Joint Statistical Meetings (2018).
- Student paper award reviewer, Section on Bayesian Statistical Science, Joint Statistical Meetings (2018).
- Session Chair, Joint Statistical Meetings (2014).

EDITORIAL ACTIVITIES

Editorial Board

2020 -   Journal of Computation and Graphical Statistics

Peer Review Activities *The numbers in parentheses indicate the count of manuscripts reviewed, excluding revisions.*

Advances in Statistical Climatology Meteorology and Oceanography (1), Annals of Applied Statistics (4), Bayesian Analysis (1), Biometrics (4), Biometrika (1), Brazilian Journal of Probability and Statistics (1), Computational Statistics and Data Analysis (CSDA) (3), Electronic Journal of Statistics (1), Environmetrics (2), Harvard Data Science Review (1), IEEE Transactions on Pattern Analysis and Machine Intelligence (1), Journal of Agricultural Biological and Environmental Statistics (JABES) (1), Journal of the American Statistical Association Applications and Case Studies (JASA-ACS) (2), Journal of the American Statistical Association Theory and Methods (JASA-TM) (8), Journal of Computation and Graphical Statistics (JCGS) (4), Journal of Multivariate Analysis (JMVA) (1), Journal of the Royal Statistical Society Series B (JRSSB) (1), Journal of the Royal Statistical Society Series C (JRSSC) (1), Sankhya A (2), Scientific Reports (1), Spatial Statistics (3), Statistica Sinica (3), Statistical Computing (2), Statistics in Medicine (1)

HONORS AND AWARDS

○ Honorable mention: Lindley Award, International Society for Bayesian Analysis (ISBA) (2020).
○ Honorable mention: Savage Award Applied Methodology, International Society for Bayesian Analysis (ISBA) (2018).
○ ASA Outstanding Statistical Application Award, American Statistical Association (2017).
○ ENAR Distinguished Student Paper Award, International Biometric Society, Austin, TX (2016).
○ Delta Omega Honorary Society Student Inductee (Pi Chapter), Minneapolis, MN (2016).
○ Best Student Seminar Presentation Award, Division of Biostatistics, University of Minnesota, Minneapolis, MN (2016).
○ Inter-disciplinary Doctoral Fellowship 2015-16, Division of Biostatistics, University of Minnesota Graduate School, Minneapolis, MN (2015).
○ JSM Student Paper Award, American Statistical Association, Section on Bayesian Statistical Science, Boston, MA (2014). Also selected for best paper award in Statistics and the Environment Section
○ Best Paper Award, Division of Biostatistics, University of Minnesota, Minneapolis, MN (2014).

○ **Teaching, Advising, and Mentoring:**
○ 2019-2020 AMTRA Award (Advising, Mentoring, & Teaching Recognition) by JHSPH Student Assembly (2020)
○ Excellence in Teaching, Johns Hopkins Bloomberg School of Public Health, Fourth Quarter, Probability IV (2018).
○ Outstanding Teaching Assistant Award, Division of Biostatistics, University of Minnesota, Minneapolis, MN (2014).

○ **Peer review:**
○ Top 1% of reviewers in Mathematics, Publons (2018)
○ Top 1% of reviewers in Mathematics, Publons (2017)
○ **Conference Travel Awards:**

○ ISBA World Meeting, International Society for Bayesian Analysis, Edinburgh, UK (2018).
○ G70 Conference, Duke University, Durham, NC (2015).
○ ISBA World Meeting, International Society for Bayesian Analysis, Cancun, Mexico (2014).
○ Conference on Non-parametric Statistics for Big Data and Celebration to Honor Professor Grace Wahba, University of Wisconsin, Madison, WI (2014).
○ Pan-American Advanced Study Institute on Spatio-Temporal Statistics, Travel grant from National Science Foundation, Buzios, Brazil (2014).

PUBLICATIONS *The white numbers in black boxes indicate first author (including equal contributions) or senior author manuscripts, * indicates equal contributions.*

Published Peer-Reviewed Articles

[1] **Datta, A,** Saha, A, Levy-Zamora, M, Buehler, Colby, Hao, L, Xiong, F, Gentner DR, Koehler K ▶ Statistical field calibration of a low-cost PM2.5 monitoring network in Baltimore *Atmospheric Environment (to appear)*

[2] Finley AO, **Datta A**, Banerjee S. (2019) ▶ spNNGP R package for Nearest Neighbor Gaussian Process models *Journal of Statistical Software (Accepted with minor revisions)*

[3] **Datta A**, Pita, A, Rao, A, Sithole, B, Mnisi, Z, and Baral, S. ▶ Size Estimation of Key Populations in the HIV Epidemic in eSwatini using incomplete and misaligned capture-recapture data *Annals of Applied Statistics (to appear)*

[4] **Datta A**, Fiksel J, Amouzou A, Zeger S. ▶ Regularized Bayesian transfer learning for population level etiological distributions *Biostatistics (to appear)*

[5] **Datta A**, Zou H. ▶ A note on cross-validation for Lasso under measurement errors *Technometrics (to appear)*

[6] **Datta A**, Banerjee S, Hodges JS., Gao, L. (2019) ▶ Spatial disease mapping using Directed Acyclic Graph Auto-Regressive (DAGAR) models *Bayesian Analysis 14(4), 1221–1244*

[7] Gao, L., **Datta A**, Banerjee S, ▶ Spatial Modeling for Correlated Cancers Using Bivariate Directed Graphs *Annals of Cancer Epidemiology (to appear)*

[8] Flores-Moreno H, Fazayeli F, Banerjee A, **Datta A**, Kattge J, Butler EE, Atkin O, Wythers K, Chen M, Anand M, Bahn M, Burrascano S, Byun C, Cornelissen J, Craine J, Gonzalez-Melo A, Hattingh W, Jansen S, Kraft N, Kramer K, Laughlin D, Minden V, Niinemets U, Onipchenko V, Penuelas J, Soudzilovskaia N, Reich PB. (2019) ▶ Robustness of trait connections between multiple plant organs across environmental gradients, growth forms *Global Ecology and Biogeography 28(12), 1806–1826*

[9] Taylor-Rodriguez D, Finley AO, **Datta A**, Babcock C, Andersen H, Cook BD, Morton DC, Banerjee S. (2019) ▶ Spatial Factor Models for High-Dimensional, Large Spatial Data: An Application in Forest Variable Mapping *Statistica Sinica 26(29) 1155–1180*

[10] Zhang L, **Datta A**, Banerjee S. (2019) ▶ Practical Bayesian Inference for Massive Spatial Data on Modest Computing Environments *Statistical Analysis and Data Mining: The ASA Data Science Journal 12.3:197-209.*

[11] Finley AO, **Datta A**, Cook BC, Morton DC, Andersen HE, Banerjee S. (2019) ▶ Efficient algorithms for Bayesian Nearest Neighbor Gaussian Processes *Journal of Computational and Graphical Statistics 1-14.*

[12] Heaton MJ, **Datta A**, Finley AO, Furrer R, Guhaniyogi R, Gerber F, Gramacy RB, Hammerling D, Katzfuss M, Lindgren F, Nychka DW, Sun F, Zammit-Mangion A. (2019) ▶ A Case Study Competition Among Methods for Analyzing Large Spatial Data *Journal of Agricultural, Biological and Environmental Statistics 24(3) 398–425.*

[13] **Datta A**, Lin W, Rao A, Diouf D, Kouame A, Edwards JK, Bao L, Louis TA, Baral SB (2019) ▶ Bayesian estimation of MSM population in Côte d'Ivoire *Statistics and Public Policy 6(1), 1-13.*

[14] **Datta A**, Zou H, Banerjee S. (2019) ▶ Bayesian high-dimensional regression for change point analysis *Statistics and Its Interface 12(2), 253-264.*

[15] Edwards JK, Hileman S, Donastorg Y, Sanchez R, Zadrozny S, Baral SB, Hargreaves J, Fearon E, Zhao J, **Datta A**, Weir SS. (2018) ▶ Estimating sizes of key populations at the national level: considerations for study design, analysis *Epidemiology 29(6): 795–803*

[16] Saha A, **Datta A.** (2018) ▶ BRISC: Bootstrap for rapid inference on spatial covariances *Stat e184*

[17] **Datta A**, Zou H. (2017) ▶ CoCoLasso for High-dimensional Error-in-variables Regression *Annals of Statistics 45(6): 2400-2426*

[18] *Butler EE, * **Datta A.** / ⋯ 48 authors ⋯ / Reich, PB. (2017) ▶ Mapping local and global variability in plant trait distributions *Proceedings of the National Academy of Sciences 114(51): E10937–E10946*

[19] **Datta A**, Banerjee S, Finley AO, Hamm NAS, Schaap M. (2016) ► Non-separable Dynamic Nearest Neighbor Gaussian Process Models for Large Spatio-temporal Data with Application to Particulate Matter Analysis *Annals of Applied Statistics 10(3): 1286-1316*

[20] **Datta A**, Banerjee S, Finley AO, Gelfand AE. (2016) ► On nearest-neighbor Gaussian process models for massive spatial data *Wiley Interdisciplinary Reviews: Computational Statistics 8(5) 162-171*

[21] **Datta A**, Banerjee S, Finley AO, Gelfand AE. (2016) ► Hierarchical Nearest Neighbor Gaussian Process models for Large Geostatistical Datasets *Journal of the American Statistical Association 111(514) 800-812*

Manuscripts Under Review

[22] Fiksel J, Zeger, S, **Datta A** ► A Transformation-free Linear Regression for Compositional Outcomes and Predictors

[23] Fiksel J, **Datta A**, Amouzou A, Zeger S. ► Generalized Bayesian Quantification Learning

[24] Wythers KR, Butler EE, Flores-Moreno, H, Chen M, **Datta A**, Ricciuto DE, Atkin OK, Kattge J, Thorton PM, Banerjee A, Reich PB ► Improved logic and parameterization of maintenance respiration alter seasonal and spatial carbon cycling output from a global land surface model

Software

1 **BRISC (2018) (**9305 **CRAN downloads as of July 11, 2020)**
BRISC is an R-package on CRAN for rapid estimation, prediction and inference for large spatial data in a frequentist setup. BRISC estimation and prediction relies on nearest neighbor approximations of the spatial Gaussian Process likelihood, and uses a scalable paramteric bootstrap to provide inference for all spatial parameters. To our knowledge, currently BRISC is the only R-package that provides confidence intervals in a frequentist setup for all parameters including the spatial variance and range of Gaussian Process. Inference from BRISC is highly competitive with those obtained on Bayesian approaches relying on MCMC, while being manifold times faster.

2 **calibratedVA (2018) (Github download stats not available)**
calibatedVA is an R-package on Github for local calibration of national and sub-national cause specific mortality (CSMF) estimates produced by algortihms based on verbal autopsy data. These computer coded verbal autopsy (CCVA) algorithms usually rely on non-local gold standard training data and can be inaccurate in a local context. calibratedVA uses the output of the CCVA algorithm and limited amount of local gold standard data to update the CSMF estimates using a fast Bayesian hierarchical model. calibratedVA also has an ensemble calibration option where outputs from multiple CCVA algorithms are used to produce an unified calibrated CSMF estimate. the package can also be used in other general contexts to calibrate any discrete classifier (or a set of classifiers) based on limited local labeled data.

3 **spNNGP (2017) (**10518 **CRAN downloads as of July 11, 2020)**
spNNGP is an R package on CRAN for fully Bayesian analysis of massive spatial data. Spatial analysis of point process data is usually computationally expensive requiring memory and computations that are quadratic and cubic in the number of locations where data is observed. spNNGP implements a class of scalable Nearest Neighbor Gaussian Process models that uses memory and computations that are linear in the size of the data. spNNGP enables fast fully Bayesian inference of all parameters and proper uncertainty quantified predictions at new locations. An MCMC-free hybrid Bayesian conjugate NNGP is also included which is super fast even for spatial datasets with millions of locations. The new version of spNNGP also has the option to run Bayesian spatial GLM for binary spatial data using Nearest Neighbor Gaussian Processes.

4 **codalm (2020) (**313 **CRAN downloads as of July 11, 2020)**
codalm is an R-package for linear modeling of compositional data (coda). It implements a simple transformation-free regression of a compositional outcome on a compositional prediction using an M-estimation method. Estimates of the regression-coefficient matrix, bootstrap-based confidence intervals are provided. A permutation based test of linear association is also offered.

TEACHING

Advisees

1 Saha, Arkajyoti, Doctor of Philosophy, Biostatistics (2016 – present, co-advised with Nilanjan Chatterjee).
2 Dey, Debangan, Doctor of Philosophy, Biostatistics (2017 – present, co-advised with Vadim Zipunnikov).
3 Gilbert, Brian, Doctor of Philosophy, Biostatistics (2019 – present, co-advised with Betsy Ogburn).
4 Heffernan, Claire, Doctor of Philosophy, Biostatistics (2019 – present)
5 Fiksel, Jacob, Doctor of Philosophy, Biostatistics (2015 – 2020).
6 Pita, Andrew, Master of Science, Biostatistics (2017 – 2019)

Academic Advisees

1 Kuo, Albert. Doctor of Philosophy, Biostatistics (2017 – 2019).
2 Fu, Martina. Doctor of Philosophy, Biostatistics (2019).

Thesis Committees / Thesis Reader

1 Pita, Andrew, Master of Science, Biostatistics (2019).

Preliminary Oral Participation   * committee chair

1 Wang, Guoqing. Doctor of Philosophy, Biostatistics (2020, expected).
2 Dey, Debangan. Doctor of Philosophy, Biostatistics (2020).
3 Saha, Arkajyoti. Doctor of Philosophy, Biostatistics (2019).
4 Windle, Michael. Doctor of Philosophy, Epidemiology (2019).
5 Kim, Ji Soo. Doctor of Philosophy, Biostatistics (2019).

Final Oral Participation   * committee chair

1 Kim, Ji Soo. Doctor of Philosophy, Biostatistics (2020, expected).
2 Fiksel, Jacob, Doctor of Philosophy, Biostatistics (2020).
3 Wang, Craig. Doctor of Philosophy, University of Zurich, Department of Mathematics and Department of Computational Science (2019).
4 Lee, Youjin, Doctor of Philosophy, Biostatistics (2019).
5 Colston, Josh, Doctor of Philosophy, International Health (2018).

Classroom Instruction - Principal Instructor

- 140.724 Probability Theory IV (2020).
- 140.724 Probability Theory IV (2019).
- Biostatistics PhD seminar (2019).
- 140.724 Probability Theory IV (2018).
- 140.850 Advanced spatial statistics (2018).
- 140.850 Scalable methods for large spatial data (2017).

Classroom Instruction - Invited Guest Lecturer (JHSPH)

○ 340.680.01 Environmental and Occupational Epidemiology (2018).

Classroom Instruction - Invited Lecturer (Other)

○ Full day short course on Bayesian models for high dimensional spatial data, Joint Statistical Meetings (2017).

RESEARCH GRANT PARTICIPATION *Bold titles indicates grants funded as a Principal Investigator.*

Ongoing Research Support

- *Cholera Burden and Transmission Modeling (Bill & Melinda Gates Foundation)*
  Dates: Sep 2019 to Jul 2022.
  Principal Investigator: Justin Lessler.
  Responsibility: Co-investigator.

- ***Highly multivariate geo-statistics using graphical models (NSF DMS-1915803)***
  Dates: July 2019 to June 2022.
  Principal Investigators: Abhirup Datta.
  Responsibility: Principal Investigator.

- *Comprehensive Mortality Surveillance for Action (COMSA)- Mozambique (Bill & Melinda Gates Foundation)*
  Dates: Jan 2017 to Dec 2019.
  Principal Investigator: Agbessi Amouzou.
  Responsibility: Co-investigator.

- *Individualized spatial topology in functional neuroimaging (NIBIB R01)*
  Dates: July 2018 to Mar 2021.
  Principal Investigator: Martin Lindquist.
  Responsibility: Co-investigator.

- *Improved Heritability Estimation by Spatial Mapping of Genetic Relationships (University of Minnesota (Prime: NIH R21))*
  Dates: July 2018 to Jun 2020.
  Principal Investigator: Saonli Basu.
  Responsibility: Principal investigator on sub-contract.

Completed Research Support

- ***Statistical Maps of Air Quality in Baltimore City Using Low-Cost Monitoring Data (Bloomberg American Health Initiative Spark Award)***
  Dates: July 2018 to June 2019.
  Principal Investigators: Abhirup Datta and Kirsten Koehler.
  Responsibility: Principal Investigator.

- *Project SOAR – Supporting Operational AIDS Research (USAID SH142)*
  Dates: Oct 2016 to Mar 2019.
  Principal Investigator: Deanna Kerrigan.
  Responsibility: Statistical Consultant.

ACADEMIC SERVICE

Department of Biostatistics

○ Member, Graduate students admissions committee (2019 - present)
○ Member, Curriculum committee, Biostatistics Retreat (2018)
○ Member, Faculty Recruitment Committee (2017).
○ Organizer, Biostatistics departmental seminars (2017).

PRESENTATIONS

Scientific Meetings, * invited, # posters

1. *Aug 2020 (expected) Joint Statistical Meetings, Philadelphia, PA

2. *Mar 2020 Eastern North American Region Meetings (ENAR), International Biometric Society, Nashville, TN

3. *Dec 2020 International Indian Statistical Association Conference, Mumbai, India.

4. *Aug 2019 International Statistical Institute World Congress, Kuala lampur, Malaysia.

5. *Aug 2019 Joint Statistical Meetings, Denver, CO.

6. *May 2019, LRI Causes and Etiologies Meeting, Baltimore, MD.

7. *Aug 2018 Joint Statistical Meetings, Vancouver, Canada.

8. *Jun 2018 ISBA World Meeting, Edinburgh, UK.

9. *Mar 2018 Eastern North American Region Meetings (ENAR), International Biometric Society, Atlanta, GA.

10. *Dec 2017 International Indian Statistical Association Conference, Hyderabad, India.

11. *Dec 2017 10th International Conference of the ERCIM WG on Computational and Methodological Statistics, London, UK.

12. *Nov 2017 American Public Health Association Annual Meeting, Atlanta, GA.

13. *Oct 2017 UNAIDS Reference Group Fall Meeting 16-18 October 2017, London, UK.

14. July 2017 IMS New Researcher's Conference, Baltimore, MD.

15. July 2017 Spatial Statistics Conference, Lancaster, UK.

16. *Feb 2017 CDC Consultation Conference on Key Populations, CDC, Atlanta, GA.

17. *Dec 2016 Platinum Jubilee International Conference on Applications of Statistics, Calcutta University, Kolkata, India.

18. Aug 2016 Joint Statistical Meetings, Chicago, IL.

19. Mar 2016 Eastern North American Region Meetings (ENAR), International Biometric Society, Austin, TX.

20. Dec 2015 International Indian Statistical Association Conference, Pune, India.

21. Mar 2015 Eastern North American Region Meetings (ENAR), International Biometric Society, Miami, FL.

22. Aug 2014 Joint Statistical Meetings, Boston, MA.

23. # Jul 2014 ISBA World Meeting, Cancun, Mexico.

24. # Jun 2014 Conference on Nonparametric Statistics for Big Data and Celebration to Honor Professor Grace Wahba, Madison, WI.

25. Jun 2014 Pan-American Advanced Study Institute on Spatio-Temporal Statistics, Buzios, Brazil.

26. Mar 2014 Eastern North American Region Meetings (ENAR), International Biometric Society, Baltimore, MD.

27. Aug 2010 Mahalanobis International Symposium on Statistics, Kolkata, India.

Invited Seminars

28. Sept 2019, Department of Statistics, Penn State University, State College, PA

29. Feb 2019, Department of Biostatistics, UCLA, Los Angeles, CA

30. Jan 2019 Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata, India.

31. Feb 2017 Department of Mathematics and Statistics, University of Maryland , Baltimore County, MD.

32 Feb 2016 Department of Statistical Science, Duke University, Durham, NC.

33 Feb 2016 Department of Biostatistics, University of Michigan, Ann Arbor, MI.

34 Feb 2016 Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

35 Feb 2016 Department of Statistics, University of California, Irvine, CA.

36 Feb 2016 Department of Biostatistics, University of North Carolina, Chapel Hill, NC.

Other Meetings and Events, # posters

37 # Mar 2019 SEARCH Scientific Advisory Committee meeting, Yale University, New Haven, CT.

38 Nov 2017 Small Area Estimation and Spatial Statistics (SAESS) Working Group, Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

39 Nov 2017 SAESS Working Group Open House, Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

40 Apr 2017 SAESS Working Group, Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

41 Mar 2017 Department of Biostatistics Student Recruitment Event, Johns Hopkins University, Baltimore, MD.

42 Dec 2016 Department of Biostatistics Faculty Meeting, Johns Hopkins University, Baltimore, MD.

43 Nov 2016 SAESS Working Group.

44 Sept 2016 Grand Rounds Seminar, Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

45 Nov 2016 President's Emergency Plan for Aids Relief, Washington DC.

46 Sept 2016 SAESS Working Group, Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

47 Jan 2016 Division of Biostatistics Student Seminar, University of Minnesota, Minneapolis, MN.

48 Dec 2015 Dow Sustainability Innovation Student Challenge Award (SISCA), University of Minnesota, Minneapolis, MN.

49 # Apr 2015 G70: A Celebration of Alan Gelfand's 70th Birthday, Duke University, Durham, NC.

50 # Apr 2015 University of Minnesota School of Public Health Annual Research Day, Minneapolis, MN.

51 # Apr 2014 University of Minnesota School of Public Health Annual Research Day, Minneapolis, MN.

52 # Oct 2013 University of Minnesota U-Spatial Symposium, Minneapolis, MN.