

CURRICULUM VITAE

Abhirup Datta

Address: Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health.
615 North Wolfe Street, #E3640, Baltimore, MD 21205.

Phone: (410) 502 2988

Email: abhidatta@jhu.edu

Education

Degree	Year	Institution
PhD, Biostatistics	2016	University of Minnesota, Twin Cities, MN
MStat (Specialization in Math-Stat-Probability)	2010	Indian Statistical Institute, Kolkata, India
BStat (with Honors)	2008	Indian Statistical Institute, Kolkata, India,

Professional Experience

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

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)

Editorial Board

2020 - Journal of Computation and Graphical Statistics

Peer Review Activities *The numbers of manuscripts reviewed exclude revisions.*

Advances in Statistical Climatology Meteorology and Oceanography (1), Annals of Applied Statistics (5), Bayesian Analysis (1), Biometrics (4), Biometrika (1), Computational Statistics and Data Analysis (CSDA) (3), Electronic Journal of Statistics (1), Environmental Science and Technology (ES&T) (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) (3), 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 (4), Statistica Sinica (3), Statistical Computing (2), Statistics in Medicine (2).

Honors and Awards

Research:

- Honorable mention: Lindley Prize, 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 JHU 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)

PUBLICATIONS *The white numbers in black boxes indicate first author (including equal contributions) or senior author manuscripts, * indicates a mentored student or post-doctoral fellow of Dr. Datta; [†] indicates equal contributions.*

Articles (Published, In press, Minor revisions)

- 1** *Fiksel J, Zeger S, **Datta A** (2021) ► A Transformation-free Linear Regression for Compositional Outcomes and Predictors *Biometrics (In press)*
- 2** *Fiksel J, **Datta A**, Amouzou A, Zeger S. (2021) ► Generalized Bayes Quantification Learning under Dataset Shift *Journal of the American Statistical Association Theory and Methods (In press)*
- 3** **Datta, A**, *Saha, A, Levy-Zamora, M, Buehler, Colby, Hao, L, Xiong, F, Gentner DR, Koehler K (2020) ► Statistical field calibration of a low-cost PM2.5 monitoring network in Baltimore *Atmospheric Environment* 242, 117761, ISSN 1352-2310
- 4** Finley AO, **Datta A**, Banerjee S. (2020) ► spNNGP R package for Nearest Neighbor Gaussian Process models *Journal of Statistical Software (Accepted)*
- 5** **Datta A**, *Pita, A, Rao, A, Sithole, B, Mnisi, Z, and Baral, S. (2020) ► Size Estimation of Key Populations in the HIV Epidemic in eSwatini using incomplete and misaligned capture-recapture data *Annals of Applied Statistics*, 14(3), 1207–1241

- 6 **Datta A**, *Fiksel J, Amouzou A, Zeger S. (2020) ► Regularized Bayesian transfer learning for population level etiological distributions *Biostatistics*, ISSN 1465-4644
- 7 **Datta A**, Zou H. (2019) ► A note on cross-validation for Lasso under measurement errors *Technometrics*, 62(4), 549–556
- 8 **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
- 9 Gao, L., **Datta A**, Banerjee S, (2020) ► Spatial Modeling for Correlated Cancers Using Bivariate Directed Graphs *Annals of Cancer Epidemiology* 4, ISSN 2616-4213
- 10 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
- 11 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
- 12 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.
- 13 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.
- 14 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.
 - Best Paper award for 2018-2019 in the Journal of Agricultural, Biological and Environmental Statistics by the International Biometric Society
- 15 **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.
- 16 **Datta A**, Zou H, Banerjee S. (2019) ► Bayesian high-dimensional regression for change point analysis *Statistics and Its Interface* 12(2), 253-264.
- 17 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
- 18 *Saha A, **Datta A**. (2018) ► BRISC: Bootstrap for rapid inference on spatial covariances *Stat e184*
 - American Statistical Association Section on Statistical Computing Student paper award for A. Saha at Joint Statistical Meetings, 2018.

- One of two papers selected for ‘Highlights of the Stat journal’ session at International Statistical Institute World Congress, 2019.

- 19 **Datta A**, Zou H. (2017) ► CoCoLasso for High-dimensional Error-in-variables Regression *Annals of Statistics* 45(6): 2400-2426
- 20 [†] 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
- 21 **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
 - American Statistical Association Outstanding Statistical Application award (2017).
 - Eastern North American Region (ENAR) distinguished student paper award for A. Datta (2016).
- 22 **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
- 23 **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 Theory and Methods* 111(514) 800-812
 - One of top 5 most cited papers in the Journal of the American Statistical Association between 2016-2020.
 - American Statistical Association Section on Bayesian Statistics (SBSS) student paper award (2014) for A. Datta.

Manuscripts Submitted

- 24 *Saha A, Basu S, **Datta A** ► GLS-style Random forests for spatially dependent data *Under Major Revision at the Journal of the American Statistical Association Theory and Methods*
- 25 *Dey D, **Datta A**, Banerjee S ► Graphical Gaussian Processes for highly multivariate spatial data
 - American Statistical Association Section on Bayesian Statistics (SBSS) student paper award (2020) for D. Dey.
- 26 Wang G, **Datta A**, Lindquist M ► Bayesian Functional Registration of fMRI Data
- 27 **Datta A** ► Sparse Cholesky matrices in spatial statistics
- 28 Gao, L., **Datta, A.**, Banerjee, S. ► Hierarchical Multivariate Directed Acyclic Graph Auto-Regressive (MDAGAR) models for spatial diseases mapping
- 29 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
- 30 Butler EE, Wythers KR, Flores-Moreno, H, Ricciuto DM, **Datta A**, Banerjee A, Atkin OK, Kattge J, Thorton PE, Mathur A, Burrascano S, Byun JHC, Forey E, Jansen S, Kramer K, Minden V, and Reich PB ► The influence of functional diversity on terrestrial carbon uptake

Software

- 1 BRISC (2018) (13870 CRAN downloads as of Feb, 2021)
- 2 calibratedVA (2018) (Github download stats not available)
- 3 spNNGP (2017) (14728 CRAN downloads as of Feb, 2021)
- 4 codalm (2020) (4063 CRAN downloads as of Feb, 2021)
- 5 RandomForestsGLS (2021) (911 CRAN downloads as of Feb, 2021)

PhD Advisees

- 1 Saha, Arkajyoti, Doctor of Philosophy, Biostatistics (2016 – present, co-advised with Nilanjan Chatterjee).
- 2 Dey, Debanjan, Doctor of Philosophy, Biostatistics (2017 – present, co-advised with Vadim Zipunov).
- 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).

ScM Advisees

- 1 Xiang, Chen, Master of Science, Biostatistics (2020 – present)
- 2 Pita, Andrew, Master of Science, Biostatistics (2017 – 2019)

Classroom Instruction - Principal Instructor (JHSPH)

- 140.724 Probability Theory IV (2021, 2020, 2019, 2018).
- 140.850 Biostatistics PhD Seminar on Advanced spatial statistics (2021, 2019, 2018).
- 140.850 Scalable methods for large spatial data (2017).

Short courses

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

Research Grants as Principal Investigator

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