CURRICULUM VITAE Abhirup Datta

PROFESSIONAL DATA

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 Webpage: abhidatta.com Github: https://github.com/abhirupdatta Twitter: @dattascience

EDUCATION AND TRAINING

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

PROFESSIONAL EXPERIENCE

Johns Hopkins University

- 2022 present: Associate Professor, Department of Biostatistics, Johns Hopkins University.
- 2016 2021: Assistant Professor, Department of Biostatistics, Johns Hopkins University.
- 2017 present: Affiliate Faculty, The Spatial Science for Public Health Center, Johns Hopkins University.

Other Non-JHU Professional Experience

• 2010 – 2012: Quantitative analyst, Morgan Stanley, New York.

HONORS AND AWARDS

Early Career Awards:

- (2024) Committee of Presidents of Statistical Societies (COPSS) Emerging Leader Award. This award recognizes statistical scientists within 10 years of terminal-degree who show evidence of and potential for leadership and who will help shape and strengthen the field
- (2023) Abdel El-Shaarawi Early Investigator's Award, The International Environmetrics Society (TIES). This award is given annually to an researcher within 12 years of completion of their PhD for their statistical contributions to Environmental Sciences.
- (2021) Early Career Award in Statistics and Data Sciences (ECADS), International Indian Statistical Association (IISA). This annual award recognizes most outstanding contributions in Applications and Statistical Practice by a person under 41 years of age.
- (2021) Early Investigator Award (EIA), American Statistical Association Section on Statistics and the Environment (ENVR). This annual award recognizes outstanding contributions to environmental statistics in the first 12 years after obtaining terminal degree.
- (2018) Honorable mention (Finalist): Savage Award (Applied Methodology), International Society for Bayesian Analysis (ISBA).

Other Research Awards:

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

Teaching, Advising, and Mentoring awards:

- (2020) JHU AMTRA Award (Advising, Mentoring, & Teaching Recognition) by JHSPH Student Assembly for 2019-2020.
- Five Excellence in Teaching awards, Johns Hopkins Bloomberg School of Public Health,
 - (1) (2023, Term 4) Probability Theory IV
 - (2) (2023, Term 3) Probability Theory III
 - (3) (2022) Probability Theory IV
 - (4) (2021) Probability Theory IV
 - (5) (2018) Probability Theory IV

 (2014) Outstanding Teaching Assistant Award, Division of Biostatistics, University of Minnesota, Minneapolis, MN.

Peer review awards:

- (2018) Top 1% of reviewers in Mathematics, Publons.
- (2017) Top 1% of reviewers in Mathematics, Publons.

RESEARCH GRANT AWARDS AS PRINCIPAL INVESTIGATOR

 Geospatial AI driven air-quality dashboard for Baltimore Data Science and AI Institute, Johns Hopkins University Dates: July 2024 to June 2026. Principal Investigators: Abhirup Datta. Award amount: \$200,00.00 Responsibility: Principal Investigator.

NIEHS R01: Statistical methods for air-pollution studies using low-cost monitors The National Institute of Environmental Health Sciences (NIEHS) R01ES033739 Dates: February 2022 to November 2026. Principal Investigators: Abhirup Datta. Award amount: \$1,326,376.00 Responsibility: Principal Investigator.

 NIBIB R01: Individualized spatial topology in functional neuroimaging National Institute Of Biomedical Imaging And Bioengineering (NIBIB) R01EB026549
 Dates: April 2024 to March 2028.
 Multiple Principal Investigators (M-PI): Abhirup Datta, Martin Lindquist (PI), Tor Wager Award amount: \$ 2,196,809.00
 Responsibility: Multiple Principal Investigator (M-PI).

Broadening the applicability of minimally-invasive-tissue-sampling (MITS)-based verbal autopsy (VA) calibration to improve global mortality estimates

Bill & Melinda Gates Foundation Dates: December 2021 to April 2024. Principal Investigators: Abhirup Datta. **Award amount: \$1,099,940.00 Responsibility: Principal Investigator.**

- NSF-DMS award: Highly multivariate geo-statistics using graphical models
 National Science Foundation Division of Mathematical Sciences (NSF-DMS) award DMS-1915803
 Dates: July 2019 to June 2023.
 Principal Investigators: Abhirup Datta.
 Award amount: \$180,000
 Responsibility: Principal Investigator.
- Air Pollution and COPD Hospitalizations in Baltimore in the Context of COVID-19 Alliance for a Healthier World COVID-19 Launchpad Grant

Dates: June 2020 to May 2021. Principal Investigators: Kirsten Koehler and Abhirup Datta. **Responsibility: Co-Principal Investigator.**

 NIDA R21: Improved Heritability Estimation by Spatial Mapping of Genetic Relationships National Institute on Drug Abuse (NIDA) R21DA046188
 Dates: July 2018 to Jun 2020.
 Principal Investigator: Saonli Basu.
 Responsibility: Principal investigator on sub-contract.

 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.
 Award amount: \$71,000
 Responsibility: Principal Investigator.

PUBLICATIONS

- * before a number indicates manuscripts where Dr. Datta is the first author or senior/corresponding author;
- ** indicates a PhD, ScM, or post-doctoral advisee of Dr. Datta;

⁺ indicates equal contributions.

Statistical Methodology

- ***67. ****Zhan W, **Datta A** (2024) Neural networks for geospatial data *Journal of the American Statistical Association Theory and Methods (In Press)* **DOI**
- ***66. ****Gilbert B, **Datta A** (2024) Visibility graph-based covariance functions for scalable spatial analysis in non-convex domains *Biometrics (In Press)* **DOI**
- ***65.** [†]Tang B, ^{†**}Pramanik S, Zhao Y, Caffo B, **Datta A** (2024) Direct Bayesian Regression for Distribution-valued Covariates *Electronic Journal of Statistics (In press)* ArXiV
- 64. Bonas M, **Datta** A, Wikle CK, Boone EL, Alamri FS, Hari BV, Kavila I, Simmons SJ, Jarvis SM, Burr WS, Pagendam D, Chang W, Castruccio S (2024) Assessing Predictability of Environmental Processes with Statistical and Machine Learning Models *Environmetrics*
 - Selected as a discussion paper for Environmetrics.
- ***63. ****Heffernan C, Koehler K, Levy-Zamora M, Buehler C, Gentner D, Peng RD, **Datta**, **A** (2024) A machine learning based interrupted time series framework for studying causal changes in air pollution due to policy interventions: A case study in COVID-19 lockdowns *American Journal of Epidemiology (In Press)*
 - Student paper award for C. Heffernan from the American Statistical Association Section on Statistics and the Environment at the Joint Statistical Meetings (2023)
- 62. Wang G, **Datta A**, Lindquist M (2024) Improved fMRI-based Pain Prediction using Bayesian Group-wise Functional Registration *Biostatistics* 25(3), 885-903
- ***61. ****Heffernan C, Peng RD, Gentner D, Koehler K, **Datta, A** (2023) A dynamic spatial filtering approach to mitigate underestimation bias in field calibrated low-cost sensor air-pollution data *Annals of Applied Statistics vol.* 17, no. 4, pp. 3056–3087, 2023.
 - Student paper award for C. Heffernan at the EnviBayes 2023 conference of the International Society for Bayesian Analysis (ISBA).
- ***60. ****Saha A, Basu S, **Datta A** (2023) Random forests for spatially dependent data *Journal of the American Statistical Association Theory and Methods* 118.541 665-683.
- ***59.** Datta A. (2023) Invited Discussion of "Saving Storage in Climate Ensembles: A Model-Based Stochastic Approach" *Journal of Agricultural Biological and Environmental Statistics 28.2* (2023): 352-357.
- 58. **Dey D, **Datta A**, Banerjee S (2023) Modeling Multivariate Spatial Dependencies Using Graphical Models *New England Journal of Statistics and Data Science (Accepted)*
- ***57. ****Dey D, **Datta A**, Banerjee S (2022) Graphical Gaussian Processes for highly multivariate spatial data *Biometrika*, 109(4), 993-1014.

- Student paper award for D. Dey from the American Statistical Association Section on Bayesian Statistics (SBSS) at the Joint Statistical Meetings (2021).
- ***56. ****Fiksel J, **Datta A**, Amouzou A, Zeger S. (2022) Generalized Bayes Quantification Learning under Dataset Shift *Journal of the American Statistical Association Theory and Methods*, 117(540), 2163-2181.
- 55. Wikle CK, **Datta A**, Hari BV, Boone EL, Sahoo I, Kavila I, Castruccio S, Simmons SJ, Burr WS, Chang W (2022) An Illustration of Model Agnostic Explainability Methods Applied to Environmental Data *Environmetrics*, *34*(1), *e*2772.
- 54. Finley AO, **Datta A**, Banerjee S. (2022) spNNGP R package for Nearest Neighbor Gaussian Process models *Journal of Statistical Software*, 103(1), 1–40.
- 53. **Saha A, **Datta A**, Banerjee S (2022) Scalable Predictions for Spatial Probit Linear Mixed Models Using Nearest Neighbor Gaussian Processes *Journal of Data Science*, 20(4), 533-544,
- ***52. ****Saha A, Basu S, and **Datta A.**, (2022). RandomForestsGLS: An R package for Random Forests for dependent data *Journal of Open Source Software*, 7(70), 3780
- 51. Gao L., **Datta A**, Banerjee S (2022) Hierarchical Multivariate Directed Acyclic Graph Auto-Regressive (MDAGAR) Models for Spatial Diseases Mapping *Statistics in Medicine*, 41(16): 3057–3075
- ***50. ****Fiksel J, Zeger S, **Datta A** (2021) A Transformation-free Linear Regression for Compositional Outcomes and Predictors *Biometrics*, *78*, *974–987*.
- 49. Wang G, Datta A, Lindquist M (2021) Bayesian Functional Registration of fMRI Data *Annals of Applied Statistics*, *16*(3), *1676-1699*.
- ***48.** Datta A (2021) Sparse nearest-neighbor Cholesky matrices in spatial statistics *Wiley Interdisciplinary Reviews: Computational Statistics*, e1574.
- ***47. Datta A**, ******Fiksel J, Amouzou A, Zeger S. (2020) Regularized Bayesian transfer learning for population level etiological distributions *Biostatistics*, *ISSN* 1465–4644
- 46. Gao, L., **Datta A**, Banerjee S, (2020) Spatial Modeling for Correlated Cancers Using Bivariate Directed Graphs *Annals of Cancer Epidemiology 4*, *ISSN 2616-4213*
- ***45.** 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*
- ***44. Datta A**, Zou H. (2019) A note on cross-validation for Lasso under measurement errors *Technometrics*, 62(4), 549–556
- ***43. 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
- 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

- 41. 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.
- 40. 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* 28.2 (2019): 401-414.
- 39. 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
- ***38.** 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.
- ***37.** Datta A, Zou H, Banerjee S. (2019) Bayesian high-dimensional regression for change point analysis *Statistics and Its Interface* 12(2), 253-264.
- ***36. ****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.
- ***35.** Datta A, Zou H. (2017) CoCoLasso for High-dimensional Error-in-variables Regression *Annals of Statistics* 45(6): 2400-2426
- 34. **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
- ***33. 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).
- ***32.** 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 for A. Datta at the Joint Statistical Meetings (2014).

Scientific Applications

- Rosenblum, M, Chin, ET, Ogburn, EL, Nishimura, A, Westreich, D, Datta, A, Vanderplas, S, Cuellar, M and Thompson, WC (2023) Incorrect Statistical Reasoning in Guyll et al. Leads to Biased Claims about Strength of Forensic Evidence *Proceedings of the National Academy of Sciences (Accepted)*
- ***30. ****Fiksel J, ******Gilbert B, Wilson E, Kalter H, Kante A, Akum A, Blau D, Bassat Q, Macicame I, Gudo E, Black R, Zeger S, Amouzou A, **Datta A** (2023) Correcting for verbal autopsy misclassification bias in cause-specific mortality estimates *American Journal of Tropical Medicine and Hygiene 108.5 Suppl: 66.*
- 29. Levy-Zamora M, Buehler C, **Datta, A**, Gentner D, Koehler K (2023) Identifying optimal colocation calibration periods for low-cost sensors *Atmospheric Measurement Techniques* 16.1: 169-179.
- 28. Wilks M, Green T, Rule AM, Zamora ML, Buehler C, **Datta A**, Gentner DR, Putcha N, Hansel NN, Kirk GD, Raju S, McCormack M, Koehler K (2023) Evaluation of Calibration Approaches for Indoor Deployments of PurpleAir Monitors *Atmospheric Environment* 119944.
- 27. Rosenblum, M, Chin, ET, Ogburn, EL, Nishimura, A, Westreich, D, **Datta, A**, Vanderplas, S, Cuellar, M and Thompson, WC (2023) Misuse of Statistical Method Results in Highly Biased Interpretation of Forensic Evidence in Guyll et al. *Law.*, *Probability, and Risk (Accepted)*
- ***26. ****Gilbert B, ******Fiksel J, Wilson E, Kalter H, Kante A, Akum A, Blau D, Bassat Q, Macicame I, Gudo E, Black R, Zeger S, Amouzou A, **Datta A** (2023) Multi-cause calibration of verbal autopsy-based cause-specific mortality estimates of children and neonates in Mozambique *American Journal of Tropical Medicine and Hygiene 108.5 Suppl: 78.*
- 25. Weber L, **Saha A, **Datta A**, Hansen K, Hicks S (2023) nnSVG: scalable identification of spatially variable genes using nearest-neighbor Gaussian processes *Nature Communications* 14.1: 4059.
- 24. Lin J, Buehler C, **Datta A**, Gentner D, Koehler K, Levy-Zamora M (2023) Laboratory and Field Evaluation of a Low-cost Methane Sensor and Key Environmental Factors for Sensor Calibration *Environmental Science: Atmospheres* 3.4: 683-694.
- 23. Ivalda M, Almamy K, Wilson E, **Gilbert B, Nhachungue S, Monjane C, Adriano A, Chicumbe S, Jani, I, Kalter H, **Datta A**, Zeger S, Black R, Samo G, Amouzou A (2023) Countrywide Mortality Surveillance for Action COMSA in Mozambique: Results from a national sample vital registration system for mortality and cause of death *American Journal of Tropical Medicine and Hygiene 108.5 Suppl: 5.*
- 22. Seal S, Datta A, Basu S (2022) Efficient Estimation of SNP Heritability using Gaussian Predictive Process in Large scale Cohort Studies *PLOS Genetics*, *18*(4): *e*1010151.
- 21. Patton AN, **Datta A**, Levy-Zamora M, Buehler C, Xiong F, Gentner D, Koehler K (2022) Non-linear probabilistic calibration of low-cost environmental air pollution sensor networks for neighborhood level spatiotemporal exposure assessment *Journal of Exposure Science and Environmental Epidemiology* 32.6: 908-916.

- 20. Zamora ML, Buehler C, Lei H, **Datta A**, Xiong F, Gentner D, Koehler, K (2022) Evaluating the performance of using low-cost sensors to calibrate for cross-sensitivities in a multipollutant network *Environmental Science and Technology Engineering* 2.5: 780-793.
- 19. Butler EE, Wythers KR, Flores-Moreno, H, Ricciuto DM, **Datta A**, Banerjee A, Atkin OK, Kattge J, Thorton PE, Madhur A, Burrascano S, Byun C, Cornelissen JHC, Forey E, Jansen S, Kramer K, Minden V, and Reich PB (2022) Increasing functional diversity in a global land surface model illustrates uncertainties related to parameter simplification *Journal of Geophysical Research Biogeosciences* 127.3 (2022): e2021JG006606.
- 18. Butler EE, Wythers KR, Flores-Moreno, H, Chen M, **Datta A**, Ricciuto DE, Atkin OK, Kattge J, Thorton PM, Banerjee A, Reich PB (2021) Updated respiration routines alter spatio-temporal patterns of carbon cycling in a global land surface model *Environmental Research Letters* 16(10) *p.* 104015.
- *17. 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
- 16. 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
- 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
- *14. ⁺ 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

Under Revision

- ***13. ****Gilbert B, Ogburn EL, **Datta A** Consistency of common spatial estimators under spatial confounding (Major Revision, Biometrika) ArXiV
- ***12.** Pramanik S, Zeger S, Blau D, **Datta A** Modeling Structure and Country-specific Heterogeneity in Misclassification Matrices of Verbal Autopsy-based Cause of Death Classifiers (Major Revision, Annals of Applied Statistics) ArXiV
- *11. **Dey D, Banerjee S, Lindquist M, Datta A Graph-constrained Analysis for Multivariate Functional Data using Graphical Gaussian Processes (Major revision, Journal of multivariate analysis) ArXiV
- ***10. ****Gilbert B, **Datta A**, Casey JA, Ogburn EL A causal inference framework for spatial confounding (Revision, Statistical Science) ArXiV

*9. **Saha A, Datta A Random forests for binary geospatial data (Invited for resubmission, Biometrika) ArXiV

Submitted

- 8. Zhao Y, Datta A, Tang B, Zipunnikov V, Caffo B Density-on-Density Regression ArXiV
- 7. Perez-Saez J, Zheng Q, / ... 20 authors ... / Lessler J, **Datta A**, Azman AS, Lee EC A decade of shifting cholera burden in Africa and its implications for control: a statistical mapping analysis

Published Open-access Software

6. *Zhan W, **Datta A** (2024) **GeospaNN** – A Python Package for Geospatial Neural Networks https://pypi.org/project/geospaNN/

Description: GeospaNN is a Python package for using neural networks for geospatial data. GeospaNN implements the NNGLS method of Zhan and Datta (2024, JASA). NNGLS is a graph neural network (GNN) for fast estimation and spatial prediction (kriging) from large geospatial datasets, by combining multi-layer perceptrons, Gaussian processes, and the generalized least squares (GLS) loss.

5. *Saha A, **Datta A** (2022) **BRISC** – R package for Fast Inference for Large Spatial Datasets https://cran.r-project.org/web/packages/BRISC/

Downloads: 38,000+ CRAN downloads as of Sep, 2024

Description: BRISC is an R-package on CRAN for rapid estimation, prediction, bootstrap inference for and simulation of large spatial data in a frequentist setup using Nearest Neighbor Gaussian Processes (Datta et al, 2016, JASA). To our knowledge, BRISC is the only R-package that provides frequentist confidence intervals 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.

4. Finley A, **Datta A**, Banerjee S (2022) **spNNGP** – Spatial Regression Models for Large Datasets using Nearest Neighbor Gaussian Processes https://cran.r-project.org/web/packages/spNNGP/index.html

Downloads: 29,000+ CRAN downloads as of Sep 2024

Description: spNNGP is an R package on CRAN for scalable fully Bayesian analysis of massive spatial data using Nearest Neighbor Gaussian Processes (Datta et al, 2016, JASA). 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 fast Bayesian spatial GLM for binary spatial data.

*3. **Fiksel J, Datta A (2021) https://cran.rstudio.com/web/packages/codalm/index.html

Downloads: 23000+ CRAN downloads as of Jan 2024)

Description: codalm is an R-package on CRAN 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.

2. *Saha A, **Datta A** (2022) **RandomForestsGLS** – Random Forests for Dependent Data https://cran.r-project.org/web/packages/RandomForestsGLS/index.html

Downloads: 2000+ CRAN downloads as of Jan 2024)

Description: RandomForestsGLS is an R-package for fitting non-linear regression models on dependent data (spatial and temporal) with Generalised Least Square (GLS) based Random Forests (RF-GLS) detailed in Saha, Basu and Datta (2020). For spatial data, 'RandomForests-GLS' combines the strengths of Random Forest and Gaussian Process to estimate and predict non-linear functions using nearest neighbor Gaussian Process. For time-series data, 'RandomForestsGLS' uses the AR (auto-regressive) process covariance structure with Random Forests for estimation.

1. **Pramanik S, Wilson E, **GIlbert B, **Fiksel, J, **Datta A** (2024) **calibratedVA** – Biascorrection of cause-specific mortality estimates from verbal autopsy data https://github.com/abhirupdatta/CalibratedVA

Description: calibatedVA is an R-package on Github for calibration of national and subnational cause specific mortality (CSMF) estimates produced by algorithms based on verbal autopsy data. calibratedVA also has an ensemble calibration option where outputs from multiple VA 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.

MENTORING

PhD Advisees

Current:

- 20. Zhang, Wentao, Doctor of Philosophy, Biostatistics (2020 present, co-advised with Hongkai Ji)
- 19. Song, Jiafang, Doctor of Philosophy, Biostatistics (2021 present)
- 18. Anik Burman, Doctor of Philosophy, Biostatistics (2022 present)

Graduated:

- 17. Heffernan, Claire, Doctor of Philosophy, Biostatistics (2019 2024)
 - Next position Merck.
- 16. Gilbert, Brian, Doctor of Philosophy, Biostatistics (2019 2023, co-advised with Betsy Ogburn).
 - Currently Postdoctoral fellow at the New York University (NYU)
- 15. Dey, Debangan, Doctor of Philosophy, Biostatistics (2017 2022, co-advised with Vadim Zipunnikov)
 - Currently Postdoctoral fellow at the National Institute of Mental Health (NIMH)
- 14. Saha, Arkajyoti, Doctor of Philosophy, Biostatistics (2016 2021, co-advised with Nilanjan Chatterjee).
 - Incoming **Tenure-track Assistant Professor** for 2024 at Department of Statistics, University of California Irvine
 - Currently Postdoctoral fellow at the Department of Statistics at University of Washington
- 13. Fiksel, Jacob, Doctor of Philosophy, Biostatistics (2015 2020).
 - Currently at Vertex Pharmaceuticals, Boston, MA

Post-doctoral fellows

- 12. Dr. Sandipan Pramanik (2022)
- 11. Bora Jin (August, 2023)

ScM Advisees

- 10. Lin, Yi-Ting, Master of Science, Biostatistics (2022 2023)
 - Currently PhD student at Department of Biostatistics, University of Michigan
- 9. Xiang, Chen, Master of Science, Biostatistics (2020 2021)
 - Currently PhD student at Department of Biostatistics, University of California Los Angeles
- 8. Pita, Andrew, Master of Science, Biostatistics (2017 2019)

JHU Diversity Summer Internship Program (DSIP) Mentorship

Mentor for summer internship projects for undergraduate students from under-represented groups.

- 7. Hall, Byron, Intern (Summer 2023)
- 6. Thomas, Bella, Intern (Summer 2023)
- 5. Griffin, Karen, Intern (Summer 2022)

Other mentored students

- 4. Jungin Choi (PhD student, 2023 , Biostatistics, co-advisor: Martin Lindquist)
- 3. Dr. Bohao Tang (PhD, 2023, Biostatistics, advisor: Brian Caffo)
- 2. Dr. Guoqing Wang (PhD, 2022, Biostatistics, advisor: Martin Lindquist)
- 1. Wenyi Lin (ScM, 2017, Biostatistics, advisor: Scott Zeger)

TEACHING

Classroom Instruction - Principal Instructor (JHSPH)

- $\circ~2018\text{-}2023~140.724$ Probability Theory IV (4th term, 6 years).
- 2023, 2025 140.724 Probability Theory III (3^{*rd*} term).
- 2021, 140.850 Advanced Spatial Statistics (3rd term).
- 2019, Biostatistics PhD seminar (3rd term).
- 2018, 140.850 Advanced spatial statistics (3rd term).
- $\circ~2017, 140.850$ Scalable methods for large spatial data (4th term).

Short Courses

- Invited short course on Statistical and machine learning for spatial data, International Biometric Society Meeting (2024, scheduled).
- Full day invited short course on Bayesian models for high dimensional spatial data, Joint Statistical Meetings (2017).

ACADEMIC SERVICE

Scientific Advisory Committees

2023 - Pathology-informed Reference Death Archive for the World Health Organization

Program Leadership

Program Committee Joint Statistical Meetings 2025 Program Chair 2024-2025 for the American Statistical Association Section on the Statistics and the Environment

Editorial Boards

- 2024 Journal of the American Statistical Association
- 2023 Journal of the Royal Statistical Society Series B
- 2023 Journal of the Royal Statistical Society Series A
- 2022 Biometrics
- 2022 Sankhya (Series B)
- 2020 Journal of Computation and Graphical Statistics

Peer Review Activities

- 1. Advances in Statistical Climatology Meteorology and Oceanography
- 2. American Journal of Epidemiology (AJE)
- 3. Artificial Intelligence and Statistics Conference 2021 (AISTATS 2021)
- 4. Annals of Applied Statistics
- 5. Annals of Statistics
- 6. Bayesian Analysis
- 7. Biometrics
- 8. Biometrika
- 9. Biostatistics
- 10. Brazilian Journal of Probability and Statistics
- 11. Canadian Journal of Statistics
- 12. Computational Statistics and Data Analysis (CSDA)
- 13. Electronic Journal of Statistics

- 14. Environmental Health Policy (EHP)
- 15. Environmental Science and Technology (ES&T)
- 16. Environmetrics
- 17. Harvard Data Science Review
- 18. IEEE Transactions on Pattern Analysis and Machine Intelligence
- 19. Journal of Agricultural Biological and Environmental Statistics (JABES)
- 20. Journal of the American Statistical Association Applications and Case Studies (JASA-ACS)
- 21. Journal of the American Statistical Association Theory and Methods (JASA-TM)
- 22. Journal of Computation and Graphical Statistics (JCGS)
- 23. Journal of Multivariate Analysis (JMVA)
- 24. Journal of the Royal Statistical Society Series B (JRSSB)
- 25. Journal of the Royal Statistical Society Series C (JRSSC)
- 26. New England Journal of Statistics and Data Science (NEJSDS)
- 27. Sankhya A
- 28. Scientific Reports
- 29. Spatial Statistics
- 30. Statistica Sinica
- 31. Statistics and Computing
- 32. Statistics in Medicine

Johns Hopkins Bloomberg School of Public Health

• Elected member of the Faculty Senate (2021 – 2023)

Department of Biostatistics, Johns Hopkins University

- Honors and Awards committee (2021 present)
- Co-leader of the Bayesian Learning and Spatio-temporal (BLAST) modeling working group, (2020 present)
- Graduate students admissions committee (2019 present)
- Curriculum committee, Biostatistics Retreat (2018)
- Co-leader of the Spatial Statistics and Small area estimation (SAESS) working group, (2016 2018)
- Faculty Recruitment Committee (2017)
- Organizer, Biostatistics departmental seminars (2017)

External Letters of support for Promotion or Tenure

George Mason University

Grant Review Panels

- o (2023) Faculty Innovation Fund Panel, Johns Hopkins Bloomberg School of Public Health
- (2020) National Science Foundation (NSF) Division of Mathematical Sciences (DMS)
- (2023) National Science Foundation (NSF) Methodology, Measurement, and Statistics (MMS) Program

Program Development

- 2023 Eastern North American Region (ENAR) conference Program Committee
- Ad-hoc committee of the International Society for Bayesian Analysis (ISBA) for early career awards (2021)
- Scientific Program Committee for the International Indian Statistical Association conference (IISA) (2021).

Professional Memberships

- American Statistical Association
- International Biometric Society (Eastern North American Region (ENAR))
- International Indian Statistical Association (IISA)
- International Society for Bayesian Analysis (ISBA)

Scientific Communication

PLENARY AND KEYNOTE LECTURES

- ¹ July 2024 (scheduled) Keynote Lecture, 10th International Workshop on Spatio-Temporal Modeling (METMA XI), Lancaster, UK
- 2 July 2023 Plenary Lecture, The International Environmetrics Society Meeting, Peterborough, Canada

INVITED PANELS

- ³ July, 2024 Panelist at Grants360 Discussion of funding from the National Science Foundation (NSF), Johns Hopkins Bloomberg School of Public Health
- ⁴ May, 2024 Panelist at Conference on Evaluating the Science of Geospatial AI, Harvard University
- ⁵ Oct, 2023 Panelist in Bridging the Gap: Exploring the Role of AI in the Climate and Health Intersection, Johns Hopkins India Institute, Washington DC
- ⁶ Aug, 2023 Panelist in American Statistical Association Graduate Student Mentoring Session at the Joint Statistical Meetings 2023, Toronto, CA
- 7 2020 Panelist in Faculty workshop on Mentoring PhD students, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

INVITED PRESENTATIONS

- 8 Aug 2024 Joint Statistical Meetings, Oregon, Portland
- 9 Nov 2023 The International Environmetrics Society Webinar Series on Interdisciplinary applications of statistics and Machine Learning for environmental applications
- Nov 2023 Department of Statistics, Federal University of Minas Gerais (UFMG), Sao Paulo, Brazil (virtual)
- 11 Oct 2023 Child and Adolescent Cause of Death Estimation Meeting, World Health Organization, Geneva, Switzerland
- 12 Oct 2023 Department of Biostatistics, Yale University, New Haven, CT
- 13 Oct 2023 Department of Statistics, Cornell University, Ithaca, NY
- 14 Oct 2023 MITS Surveillance Alliance Annual Meeting, Nairobi, Kenya
- 15 Sep 2023 EnviBayes workshop, Fort Collins CO
- ¹⁶ Sep 2023 Department of Biostatistics, University of Utah, Salt Lake City, UT
- 17 Sep 2023 Department of Statistics, Brigham Young University, Provo, UT
- 18 Sep 2023 (scheduled) Department of Statistics, Virginia Tech University, Blacksburg, VA

- 19 Jul 2023 EcoSTAT Conference, Tokyo, Japan (virtual)
- 20 Jun 2023, International Chinese Statistical Association Conference, Ann Arbor, MI
- 21 Jun 2023 SIAM Conference on Optimization, Seattle, WA
- 22 Jun 2023 The Bill and Melinda Gates Foundation, Seattle, WA
- 23 May 2023 Center for Statistics and the Social Sciences, University of Washington, Seattle, WA
- 24 May 2023, Statistics, Computer Science, and Mathematics Department, The Public University of Navarre, Pamplona, Spain
- 25 May 2023, BIRS workshop on Machine Learning and Spatial Extremes, Granada, Spain
- ²⁶ Dec 2022 International Indian Statistical Association Conference, Bengaluru, India.
- 27 Dec 2022, CMStat conference, Kings College, London, UK
- 28 Oct 2022, Data Science Conference, Texas A&M University, College Station, TX
- ²⁹ Oct 2022, Southern Regional Council on Statistics (SRCOS) Conference, Jekyll Island, GA
- ³⁰ Sep 2022, Department of Population, Family and Reproductive Health, Johns Hopkins University, Baltimore, MD
- 31 Aug 2022, Statistical Data Science Workshop, University of Bologna, Italy
- 32 Aug 2022 Joint Statistical Meetings, Washington DC.
- 33 Jul 2022, International Biometrics Conference (IBC2022)
- 34 May 2022 North-Eastern Statistics Symposium (NESS) (virtual)
- ³⁵ Mar 2022 Eastern North American Region Meetings (ENAR), International Biometric Society, Houston, TX
- 36 Feb 2022, Department of Biostatistics, McGill University, Montreal, CA
- 37 Nov 2021 East Asia Chapter of the International Society for Bayesian Analysis (EAC ISBA) (virtual)
- 38 Oct 2021, Department of Statistics, Purdue University, West Lafayette, IA
- 39 Sep 2021, Department of Biostatistics, University of Minnesota, Twin Cities
- 40 Aug 2021 Joint Statistical Meetings, Seattle, WA
- ⁴¹ May 2021 Center for Disease Control (CDC) Division of Global HIV and TB's Key Population Surveillance Team (virtual)
- 42 Apr 2021 IHME-CHAMPS Convening (virtual)
- 43 Apr 2021 Department of Biostatistics, NYU School of Global Public Health, New York City, NY
- 44 Apr 2021 MCEE Cause of Death (COD) Misclassification Methods Meeting (virtual).
- ⁴⁵ Mar 2021 SIAM Conference on Computational Science and Engineering (virtual)
- ⁴⁶ Oct 2020, Department of Statistics, Iowa State University, Ames, IA

- 47 Oct 2020, Department of Biostatistics, Virginia Commonwealth University, Richmond, VA
- 48 Sep 2020, RTI International, Raleigh, NC
- 49 Sep 2020 MITS Surveillance Alliance Meeting (virtual)
- 50 Aug 2020 Joint Statistical Meetings (virtual)
- 51 Mar 2020 Eastern North American Region Meetings (ENAR), International Biometric Society (virtual)
- 52 Dec 2020 International Indian Statistical Association Conference, Mumbai, India.
- 53 Sept 2019, Department of Statistics, Penn State University, State College, PA
- ⁵⁴ Aug 2019 International Statistical Institute World Congress, Kuala lampur, Malaysia.
- 55 Aug 2019 Joint Statistical Meetings, Denver, CO.
- 56 May 2019, LRI Causes and Etiologies Meeting, Baltimore, MD.
- ⁵⁷ [#] Mar 2019 SEARCH Scientific Advisory Committee meeting, Yale University, New Haven, CT.
- ⁵⁸ Mar 2019, Child Health and Mortality Prevention Surveillance (CHAMPS) program, Emory University, Atlanta, GA.
- 59 Feb 2019, Department of Biostatistics, UCLA, Los Angeles, CA
- ⁶⁰ Jan 2019 Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata, India.
- 61 Aug 2018 Joint Statistical Meetings, Vancouver, Canada.
- 62 Jun 2018 ISBA World Meeting, Edinburgh, UK.
- 63 Jun 2018, MITS Surveillance Alliance Inaugural Meeting, Barcelona, Spain
- ⁶⁴ Mar 2018 Eastern North American Region Meetings (ENAR), International Biometric Society, Atlanta, GA.
- 65 Dec 2017 International Indian Statistical Association Conference, Hyderabad, India.
- ⁶⁶ Dec 2017 10th International Conference of the ERCIM WG on Computational and Methodological Statistics, London, UK.
- ⁶⁷ Nov 2017 American Public Health Association Annual Meeting, Atlanta, GA.
- 68 Oct 2017 UNAIDS Reference Group Fall Meeting 16-18 October 2017, London, UK.
- 69 Feb 2017 CDC Consultation Conference on Key Populations, CDC, Atlanta, GA.
- 70 Feb 2017 Department of Mathematics and Statistics, University of Maryland , Baltimore County, MD.
- 71 Nov 2016 President's Emergency Plan for Aids Relief, Washington DC.
- 72 Feb 2016 Department of Statistical Science, Duke University, Durham, NC.
- 73 Feb 2016 Department of Biostatistics, University of Michigan, Ann Arbor, MI.

- 74 Feb 2016 Department of Biostatistics, Johns Hopkins University, Baltimore, MD.
- 75 Feb 2016 Department of Statistics, University of California, Irvine, CA.
- ⁷⁶ Feb 2016 Department of Biostatistics, University of North Carolina, Chapel Hill, NC.
- 77 Dec 2016 Platinum Jubilee International Conference on Applications of Statistics, Calcutta University, Kolkata, India.