

# TIES 2023 PLENARY PRESENTATION

Dr. Abhirup Datta  
*Department of Biostatistics*  
*Bloomberg School of Public Health*  
*John Hopkins University*  
[abhidatta.com](http://abhidatta.com)

## Machine learning in model-based geostatistics

### Abstract

Spatial generalized linear mixed-models, consisting of a linear covariate effect and a Gaussian Process (GP) distributed spatial random effect, are widely used for analyses of geospatial data. We consider the setting where the covariate effect is non-linear and propose modeling it using a flexible machine learning algorithm like random forests or deep neural networks. We propose well-principled extensions of these methods, for estimating non-linear covariate effects in spatial mixed models where the spatial correlation is still modeled using GP. The basic principle is guided by how ordinary least squares extends to generalized least squares for linear models to account for dependence. We demonstrate how the same extension can be done for these machine learning approaches like random forests and neural networks. We provide extensive theoretical and empirical support for the methods and show how they fare better than naïve or brute-force approaches to use machine learning algorithms for spatially correlated data.

### Biography



Dr. Datta is an Associate Professor in the Department of Biostatistics at Johns Hopkins University. He completed his PhD in Biostatistics from University of Minnesota. Dr. Datta's research focuses on developing spatial models for geographically indexed data. His work on Nearest Neighbor Gaussian Processes (NNGP) has become one of the most widely used methods for scalable analysis of massive geospatial data. His recent work focuses on developing theory and methodology for combining machine learning algorithms with traditional spatial modeling, and application of the methodology to air pollution and infectious disease modeling. He also works on developing Bayesian hierarchical models for multi-source data with applications in global health. His research as Principal Investigator is funded by grants from the National Science Foundation (NSF), National Institute of Environmental Health Sciences (NIEHS) and the Bill and Melinda Gates Foundation. He has received the Early Career Investigator award from the American Statistical Association Section of Environmental Health, the Young Statistical Scientist Award (YSSA) by the International Indian Statistical Association (IISA), and the Abdel El-Shaarawi Early Investigator's Award from The International Environmetrics Society (TIES).