With scientific data in the natural sciences and in public health available at geocoded locations, investigators are increasingly turning to spatial process models for carrying out statistical inference. Over the last decade, spatial process models have become especially popular for spatial modeling, given their flexibility and power to fit models that would be infeasible with classical methods as well their avoidance of possibly inappropriate asymptotic assumptions. However, fitting hierarchical spatial models often involves expensive matrix decompositions whose computational complexity increases in cubic order with the number of spatial locations. This renders such models infeasible for large spatial data sets. In this talk, I will propose a new class of well-defined spatial stochastic processes called Nearest-Neighbor Gaussian Process models that can be exploited as a dimension-reducing prior embedded within a rich and flexible hierarchical modeling framework to deliver exact Bayesian inference. We develop a computationally efficient Markov chain Monte Carlo algorithm, where the number of floating point operations (flops) is linear in the number of spatial locations, thereby delivering massive scalability. We compare this with our earlier work on low-rank predictive process models and demonstrate its use in inferring on the spatial distribution of forest biomass from the US Forest Inventory database spanning the continental US.

Joint work with Abhirup Datta and Andrew O. Finley.