

On Dynamic Nearest-Neighbor Gaussian Process Models for High-Dimensional Spatiotemporal Datasets.

With the growing capabilities of Geographic Information Systems (GIS) and user-friendly software, statisticians today routinely encounter geographically referenced data containing observations from a large number of spatial locations and time points. Over the last decade, hierarchical spatial-temporal process models have become widely deployed statistical tools for researchers to better understanding the complex nature of spatial and temporal variability. However, fitting hierarchical spatial-temporal models often involves expensive matrix computations with complexity increasing in cubic order for the number of spatial locations and temporal points. This renders such models unfeasible for large data sets. In this talk, I will present two approaches for constructing well-defined spatial-temporal stochastic processes that accrue substantial computational savings. Both these processes can be used as "priors" for spatial-temporal random fields. The first approach constructs a low-rank process operating on a lower-dimensional subspace. The second approach constructs a Nearest-Neighbor Gaussian Process (NNGP) that can be exploited as a dimension-reducing prior embedded within a rich and flexible hierarchical modeling framework to deliver exact Bayesian inference. Both these approaches lead to Markov chain Monte Carlo algorithms with floating point operations (flops) that are linear in the number of spatial locations (per iteration). We compare these methods and demonstrate its use in inferring on the spatial-temporal distribution of ambient air pollution in continental Europe using spatial-temporal regression models with chemistry transport models.

Joint work with Abhirup Datta and Andrew O. Finley.