

Scalable Gaussian process models for analyzing large spatial and spatio-temporal datasets

Alan E. Gelfand, Duke University

Spatial process models for analyzing geostatistical data entail computations that become prohibitive as the number of spatial locations become large. This talk considers computationally feasible Gaussian process-based approaches to address this problem. We consider two approaches to approximate an intended Gaussian process; both are Gaussian processes in their own right. One, the Predictive Gaussian Process (PGP), is based upon the idea of dimension reduction. The other, the Nearest Neighbor Gaussian Process (NNGP), is based upon sparsity ideas.

The predictive process is simple to understand, routine to implement, with straightforward bias correction. It enjoys several attractive properties within the class of dimension reduction approaches and works well for datasets of order 10^3 or 10^4 . It suffers several limitations including spanning only a finite dimensional subspace, over-smoothing, and underestimation of uncertainty.

So, we focus primarily on the more recent nearest neighbor Gaussian process which draws upon earlier ideas of Vecchia and of Stein. It is a bit more complicated to grasp and implement but it is highly scalable, having been applied to datasets as large as 10^6 . It is a well-defined spatial process providing legitimate finite dimensional Gaussian densities with sparse precision matrices. Scalability is achieved by using local information from few nearest neighbors, i.e., by using the neighbor sets in a conditional specification of the model. This is equivalent to sparse modeling of Cholesky factors of large covariance matrices.

For either specification, we embed the PGP as a dimension reduction prior and the NNGP as a sparsity-inducing prior within a rich hierarchical modeling framework and outline how computationally efficient Markov chain Monte Carlo (MCMC) algorithms can be executed. We present illustrative examples for both specifications. We anticipate that the future likely lies with the NNGP since it can accommodate spatial scales that preclude dimension-reducing methods.