samsi

NSF·Duke·NCSU·UNC

Climate Program Remote Sensing Workshop
February 12-14, 2018

SPEAKER TITLES/ABSTRACTS

Rajarshi Guhaniyogi
University of California, Santa Cruz

"DISK: a divide and conquer Bayesian approach to large scale kriging"

Flexible hierarchical Bayesian modeling of massive data is challenging due to poorly scaling computations in large sample size settings. This talk is motivated by spatial process models for analyzing geostatistical data, which typically entail computations that become prohibitive as the number of spatial locations becomes large. We propose a three-step divide-and-conquer strategy within the Bayesian paradigm to achieve massive scalability for any spatial process model. We partition the data into a large number of subsets, apply a readily available Bayesian spatial process model on every subset in parallel, and optimally combine the posterior distributions estimated across all the subsets into a pseudo-posterior distribution that conditions on the entire data. The combined pseudo posterior distribution is used for predicting the responses at arbitrary locations and for performing posterior inference on the model parameters and the residual spatial surface. We call this approach "Distributed Kriging" (DISK). It offers significant advantages in applications where the entire data are or can be stored on multiple machines. Under the standard theoretical setup, we show that if the number of subsets is not too large, then the Bayes risk of estimating the true residual spatial surface using the DISK posterior distribution decays to zero at a nearly optimal rate. While DISK is a general approach to distributed nonparametric regression, we focus on its applications in spatial statistics and demonstrate its empirical performance using a stationary full-rank and a nonstationary low-rank model based on Gaussian process (GP) prior. A variety of simulations and a geostatistical analysis of the Pacific Ocean sea surface temperature data validate our theoretical results.