STATISTICS SEMINAR

Rigele Te

Preliminary PhD Exam

Friday, November 19th, 2021
Via Zoom, 12:00 pm (Central Time)

Multivariate Spatio-Temporal Modeling and Simulation

In view of multivariate nature of general spatio-temporal data sets observed in various disciplines such as meteorology, engineering and medical science, we are interested in gaining scientific insight in the underlying processes, which incorporate the complex dependence among different variables at different locations over time. To this objective, this dissertation studies the methods and applications of the multivariate modelling, simulation and missing value imputation for some data sets in space and time.

Chapter I, Space-time data sets are often multivariate and collected at monitored discrete time lags, which are usually viewed as a component of time series in environmental science and other areas. Valid and practical covariance models are needed to characterize geostatistical formulations of these types of data sets in a wide range of applications. We propose several classes of multivariate spatio-temporal functions to model underlying stochastic processes whose discrete temporal margins are some celebrated autoregressive and moving average (ARMA) models, and obtain sufficient and/or necessary conditions for them to be valid covariance matrix functions. The possibility of taking advantage of well-established time series and spatial statistics tools makes it relatively easy to identify and fit the proposed models in practice. Finally, applications of the proposed multivariate covariance matrix functions are illustrated on Kansas weather data in terms of co-kriging, compared with some traditional space-time models for prediction.

Chapter II, We propose an efficient method for simulating multivariate spatio-temporal data on a compact two-point homogeneous space with sphere as a special case. These large scale global data sets are obtained based on truncating the series expansion of multivariate spatio-temporal random fields on this space. The algorithm can be boiled down to simply simulate a uniformly distributed random vector on a sphere, on which the great circle distance is defined. Multiple covariance models are compared to fit the simulated multivariate space-time data including the model proposed in the Chapter I. The simulation results suggest some guideline for choosing appropriate models and parametrizations for different multivariate data in space and time.

Chapter III, Motivated by dealing with incompleteness of energy data to study network situational awareness, we propose a multivariate spatio-temporal variational autoencoders (SP-VAEs) to impute the corrupted or missing information on a lattice based on Gaussian processes and Bayesian deep learning, which allows for uncertainty estimation. The missing data is learned by projecting the data space into a lower dimensional latent space without missingness, where the low dimensional dynamics is modeled with vector autoregressive Gaussian process combined with Moran's I basis functions that efficiently capture the space-time dependence structure. Model comparison will be made with several classical and deep learning-based data imputation methods on a simulated energy data set from smart meters, solar inverters, grid automation/SCADA sensors and micro PMU on a geographical lattice over time.