Abstract

Geostatistical data are measurements taken at fixed locations in a spatial domain. Generally the latter are spatially continuous, as is typically the case in mining engineering, geology, soil science, and hydrology. Geostatistical models are based on the concept of spatial or spatio-temporal processes and aim to describe the underlying dependence structure. Spatial variability is modeled as a function of the distance between sampling sites. Called the 'variogram' or 'covariogram', this function is used to apply statistical methods such as estimation and/or prediction, referred to as 'kriging' in the geostatistical context. To quantify the spatio-temporal dependence, estimation techniques relying on certain hypotheses of stationarity (seldom met in reality) are applied.

Nonstationarity and covariance estimation are the underlying topics of the present thesis, which consists of four chapters.

The first chapter gives a concise overview of geostatistical definitions and notation used throughout the thesis. Prior to generalizing the concepts to multivariate and spatio-temporal processes, they are explained on spatial processes.

There exist many different forms of nonstationarity. Two of them are discussed in the second chapter. First, the case where the mean of the process depends on the location is studied. The identification of a trend is a nontrivial problem and we emphasize that there exists no trend estimation procedure for spatial processes with unknown dependence structure. Exploratory tools for the empirical variogram or for the observed process, as well as a commonly used parametric and nonparametric method for trend estimation are illustrated. A simple method that evolved out of the consequences of visual data analysis is developed, namely variogram estimation based on 'local trend estimation'. The latter separates the domain in several subdomains, or patches, on which an individual trend is estimated, and the residuals are combined throughout the entire domain to allow global estimation and/or inference. Simulations show that a simple and almost arbitrary subdivision is already sufficient to improve the results of variogram estimation. Moreover, the method does not break down when the (heuristic) decomposition does not coincide with the (true) separation of the populations. Even if the true trend is not linear the method performs better than other well-known parametric or nonparametric trend estimation techniques. To underline these statements the method is applied to real data. A second form of nonstationarity is the dependence of the covariance structure on the location. Under this circumstance classical covariogram estimation techniques are not applicable. For example, in atmospheric science one can easily imagine situations where the spatial dependence changes with time or where the maximum magnitude of variability may alter in time. For such phenomena new models are needed. Hence the remaining part of the second chapter discusses a new method of valid covariogram construction for nonstationary spatio-temporal processes. These new covariogram models are illustrated with simulations and an application to a dataset is given.

Several statistical tools are based on the covariance matrix of the underlying process. An example of such a method is (functional) principal components analysis, which aims to represent a set of possibly correlated variables into uncorrelated orthogonal components. These uncorrelated components can be constructed successively, each one extracting a maximal amount of the remaining variance. This often leads to an appreciable reduction in dimensionality, replacing the original variables by a few components. To calculate the orthogonal components the covariance matrix of a multivariate or spatio-temporal process is required. The latter is rarely known and therefore has to be estimated. As mentioned, an important aspect of geostatistical data is dependence over space as well as over time. This has to be taken into account when estimating the covariance matrix and the natural estimator of the covariance matrix is introduced in the third chapter. It is shown that it is biased under spatio-temporal dependence. This bias is studied under two different asymptotic models, namely increasing the number of observations in the domain and increasing the domain by increasing the number of locations. Using the first asymptotic model we derive a fast and accurate bias correction, whereas the second asymptotic model serves to quantify the speed of convergence of the bias and the covariance of the components of the estimated covariance matrix. As shown, under mild hypotheses the asymptotic normality of the estimated covariance matrix holds and can be used to test whether the eigenvectors of the estimated and the true covariance matrices are significantly different. This is revealed by examples, emphasizing the need for a bias correction. Furthermore, the theoretical results are illustrated with Monte Carlo simulation studies and again with an application to real data.

The most commonly used decomposition to extract stationary parts of a process is based on the separation into different scales: (deterministic) large-scale variation, smooth small-scale variation, micro-scale variation and a measurement error. Although such additive partitioning is of considerable utility it also has several drawbacks, so an alternative analysis based on state-space decompositions is presented in the fourth chapter. The space equation is a process governed by the state equation and an additional observational error, where the state at the point is a weighted mean of its neighborhood states described by a kernel function plus a spatial process. The new model takes account of diverse shapes of trends and one does not have to decide whether the process is stationary or not. As other existing decompositions can be reconstructed by the new representation, it can be seen as a generalization of existing ones. The decomposition results in a Fredholm integral equation of the second kind. By imposing separable kernels this integral equation has an explicit solution, and the model is defined by the parametrized covariogram of the spatial process and the parameters defining the kernel. In our distribution-free model we will explore different methods based on minimal distances and moment equations for the parameter estimation, and, by generalizing the concept of M-estimators to the dependent setting consistency for these new estimators is proven. The efficiency of the proposed method is discussed and the results are compared to other commonly used models by means of extensive Monte Carlo simulations and applications to real datasets. Despite its complexity the new model furnishes an efficient and competitive approach throughout the simulations, which show that for most parameters this new estimator is more precise than the ordinary least squares estimator.