

Geostatistical Inference Under Preferential Sampling

Peter J Diggle

(School of Health and Medicine, Lancaster University
and
Johns Hopkins University School of Public Health)

February 19, 2009

Abstract

Geostatistics involves the fitting of spatially continuous models to spatially discrete data (Chilès and Delfiner, 1999). Preferential sampling arises when the process that determines the data-locations and the process being modelled are stochastically dependent. Conventional geostatistical methods assume, if only implicitly, that sampling is non-preferential. However, these methods are often used in situations where sampling is likely to be preferential. For example, in mineral exploration samples may be concentrated in areas thought likely to yield high-grade ore.

We give a general expression for the likelihood function of preferentially sampled geostatistical data and describe how this can be evaluated approximately using Monte Carlo methods. We present a model for preferential sampling, and demonstrate through simulated examples that ignoring preferential sampling can lead to seriously misleading inferences. We describe an application of the model to a set of bio-monitoring data from Galicia, northern Spain, in which making allowance for preferential sampling materially changes the results of the analysis

The talk is based on joint work with Raquel Menezes and Ting-Li Su (Diggle, Menezes and Su, 2009).

Chilès, J-P and Delfiner, P. (1999). *Geostatistics*. New York : Wiley.

Diggle, P.J. Menezes, R. and Su, T-L. (2009). Geostatistical inference under preferential sampling (with Discussion). *Applied Statistics* (to appear)