Geophysical Research Abstracts
Vol. 17, EGU2015-2179, 2015
EGU General Assembly 2015

Local Geostatistical Models and Big Data in Hydrological and Ecological Applications

Dionissios Hristopulos
Technical University of Crete, Mineral Resources Engineering, Chania, Greece (dionisi@mred.tuc.gr)

The advent of the big data era creates new opportunities for environmental and ecological modelling but also presents significant challenges. The availability of remote sensing images and low-cost wireless sensor networks implies that spatiotemporal environmental data to cover larger spatial domains at higher spatial and temporal resolution for longer time windows. Handling such voluminous data presents several technical and scientific challenges. In particular, the geostatistical methods used to process spatiotemporal data need to overcome the dimensionality curse associated with the need to store and invert large covariance matrices. There are various mathematical approaches for addressing the dimensionality problem, including change of basis, dimensionality reduction, hierarchical schemes, and local approximations.

We present a Stochastic Local Interaction (SLI) model that can be used to model local correlations in spatial data. SLI is a random field model suitable for data on discrete supports (i.e. regular lattices or irregular sampling grids). The degree of localization is determined by means of kernel functions and appropriate bandwidths. The strength of the correlations is determined by means of coefficients. In the “plain vanilla” version the parameter set involves scale and rigidity coefficients as well as a characteristic length. The latter determines in connection with the rigidity coefficient the correlation length of the random field. The SLI model is based on statistical field theory and extends previous research on Spartan spatial random fields [2,3] from continuum spaces to explicitly discrete supports. The SLI kernel functions employ adaptive bandwidths learned from the sampling spatial distribution [1]. The SLI precision matrix is expressed explicitly in terms of the model parameter and the kernel function. Hence, covariance matrix inversion is not necessary for parameter inference that is based on leave-one-out cross validation. This property helps to overcome a significant computational bottleneck of geostatistical models due to the poor scaling of the matrix inversion [4,5]. We present applications to real and simulated data sets, including the Walker lake data, and we investigate the SLI performance using various statistical cross validation measures.

References
[1] T. Hofmann, B. Schlkopf, A.J. Smola, Annals of Statistics, 36, 1171-1220 (2008).
[2] D. T. Hristopulos, SIAM Journal on Scientific Computing, 24(6): 2125–2162 (2003).
[3] D. T. Hristopulos and S. N. Elogne, IEEE Transactions on Signal Processing, 57(9): 3475–3487 (2009)
[4] G. Jona Lasinio, G. Mastrantonio, and A. Pollice, Statistical Methods and Applications, 22(1):97–112 (2013)
[5] Sun, Y., B. Li, and M. G. Genton (2012). Geostatistics for large datasets. In: Advances and Challenges in Space-time Modelling of Natural Events, Lecture Notes in Statistics, pp. 55-77. Springer, Berlin-Heidelberg.