Spin models for efficient prediction of massive spatial data

Milan Žukovič
Pavol Jozef Šafárik University in Košice, Košice, Slovakia

The steadily increasing volume of Earth observation data collected by remote sensing techniques requires the development of new methods capable of efficient (preferably real time) and automated processing. Such processing includes filling gaps with missing values, which often occur due to different reasons. Application of standard kriging methods in such situations is hindered by their high computational complexity and also necessity of certain user-specified inputs. As an alternative to the traditional methods, a computationally efficient approach inspired from statistical physics has been proposed [1]. Instead of the empirical variogram used in geostatistics it employs Boltzmann-Gibbs random fields with joint densities that model spatial correlations by means of short-range interactions. It can be applied to both gridded and scatted data, nevertheless, is still relies on the assumption that the data follow Gaussian distribution.

In this presentation, we extend the concept of deriving correlations from local interactions to non-Gaussian data by means of classical spin models [2]. Special focus is laid on the application of the planar rotator (or XY) model, modified [3] and generalized [4] to account for spatial correlations and data distributions characteristic for geophysical and environmental data sets. The proposed methods include both non-parametric and parametric approaches and employ efficient hybrid conditional Monte Carlo (MC) simulations. They turn out to be competitive with more traditional approaches in terms of the prediction performance but at the same time computationally very efficient with only roughly linear increase of the CPU time with the data size. Generalization of the methods proposed for 2D regular grids to scattered data in arbitrary dimension will be outlined.

Further, it will be demonstrated how the short-range nature of the interactions between the spin variables can be utilized to parallelize the algorithm on graphics processing units (GPUs) and thus dramatically improve its computational performance [5]. The achieved computational speedups (up to almost 500 times on large grids) compared to single-processor calculations, allow processing of massive data sets comprising millions of points automatically in less than one second on an ordinary GPU-equipped PC. Another advantage of the GPU implementation is that it facilitates a rather simple but general approach to modelling spatial heterogeneity by introducing spatial variability to model parameters. Then the predictions can be obtained based on “local-equilibrium” conditional MC simulations.

References

[1] D. T. Hristopulos, SIAM J. Sci. Comput. 24, 2125 (2003).
[2] M. Žukovič, D. T. Hristopulos, Phys. Rev. E 80, 011116 (2009).
[3] M. Žukovič, D. T. Hristopulos, Phys. Rev. E 98, 062135 (2018).
[4] M. Žukovič, D. T. Hristopulos, Physica A 612, 128509 (2023).
[5] M. Žukovič, M. Borovský, M. Lach, D.T. Hristopulos, Math. Geosci. 52, 123 (2020).