Spatial Statistics 16 (2016) 53–71

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Conditioning multiple-point statistics simulations to block data

Julien Straubhaar[a,*], Philippe Renard[a], Grégoire Mariethoz[b]

[a] The Centre for Hydrogeology and Geothermics (CHYN), University of Neuchâtel, rue Emile-Argand 11, CH-2000 Neuchâtel, Switzerland

[b] Institute of Earth Surface Dynamics (IDYST), University of Lausanne, UNIL-Mouline, Geopolis, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 17 August 2015
Accepted 6 February 2016
Available online 17 February 2016

Keywords:
Geostatistical simulation
Multiple-point statistics
Block data
Downscaling

ABSTRACT

Multiple-points statistics (MPS) allows to generate random fields reproducing spatial statistics derived from a training image. MPS methods consist in borrowing patterns from the training set. Therefore, the simulation domain is assumed to be at the same resolution as the conceptual model, although geometrical deformations can be handled by such techniques. Whereas punctual conditioning data corresponding to the scale of the grid node can be easily integrated, accounting for data available at larger scales is challenging. In this paper, we propose an extension of MPS able to deal with block data, *i.e.* target mean values over subsets of the simulation domain. Our extension is based on the direct sampling algorithm and consists to add a criterion for the acceptance of the candidate node scanned in the training image to constrain the simulation to block data. Likelihood ratios are used to compare the averages of the simulated variable taken on the informed nodes in the blocks and the target mean values. Moreover, the block data may overlap and their support can be of any shape and size. Illustrative examples show the potential of the presented algorithm for practical applications.

* Corresponding author.
E-mail addresses: julien.straubhaar@unine.ch (J. Straubhaar), philippe.renard@unine.ch (P. Renard), gregoire.mariethoz@unil.ch (G. Mariethoz).

http://dx.doi.org/10.1016/j.spasta.2016.02.005

1. Introduction

The multiple-point statistics (MPS) methods have become very popular in earth sciences, because they allows to generate highly heterogeneous random fields reproducing the spatial statistics of a conceptual geological model, the training image, given by the user. These methods overcome some limitations of classical geostatistical simulation techniques based on two-point statistics: variogram-based methods such as sequential Gaussian simulation, sequential indicator simulation (*sisim*) (Deutsch and Journel, 1998), transition probability based approaches such as *TProgs* (Carle, 1996), or Markovian-type categorical prediction (MCP) based on a maximum entropy principle (Allard et al., 2011). Among the existing MPS simulation algorithms, *snesim* (Strebelle, 2002) and *impala* (Straubhaar et al., 2011, 2013) successively populate each node of the simulation grid by randomly drawing a facies category according to a probability distribution conditioned to the data event centered at the simulated node, computed from a catalog of patterns found in the training image. A storage based on a tree structure is used in *snesim* ensuring computational time efficiency, and a list-based catalog employed in *impala* guarantees low memory requirements. Using a catalog implies to consider only categorical variables and patterns of fixed geometry. A multiple grid approach (Tran, 1994) is employed in theses algorithms to capture large scale structures while keeping data events of reduced size. On the other hand, the direct sampling algorithm (Mariethoz et al., 2010) is a distance-based MPS algorithm. To simulate a node, the method consists in randomly scanning the training image until the pattern in the training image is compatible with the pattern retrieved from the simulation grid and centered at the simulated node. Then, the central node value is copied and pasted from the training image to the simulation grid. The compatibility between two patterns is related to a distance. This basic simulation principle leads to a very flexible method. Indeed, not using any catalog of patterns, categorical as well as continuous variables can be considered by defining an appropriate distance between data events, and the geometry of the patterns can vary during the simulation allowing to reproduce large scale structures without using a multiple grid approach. In particular, punctual conditioning data can be simply assigned in the simulation grid at the beginning of the simulation, whereas methods based on a multiple grid approach implies some precautions to properly address punctual data (Straubhaar and Malinverni, 2014). Distance-based MPS algorithms include also techniques consisting in pasting patches of the training image in the simulation grid at a time instead of only one pixel value, such as *filtersim* (Zhang and Journel, 2006) or *simpat* (Arpat and Caers, 2007). These latter methods use patterns database built from the training image and are also based on multiple grid approaches. Other patch-based MPS algorithms not using multiple grids nor databases consist in pasting overlapping boxes of pixels along a raster path, by minimizing a cross-correlation function over the overlapping region in the algorithm *ccsim* (Tahmasebi et al., 2012), or by minimizing an error between the common area followed by an optimal cut through this area in the algorithm *conditional image quilting (CIQ)* (Mahmud et al., 2014). Theses methods allow to better model the connectivity of the structures, but make the conditioning difficult. Therefore, the direct sampling method is appealing due to its simplicity and its flexibility. In particular it can easily be extended to the simulation of multivariate fields (Mariethoz et al., 2010, 2012), providing an intuitive tool to manage various types of nonstationarities.

No matter what MPS technique is considered, the simulation domain is filled by borrowing patterns from the training image, which is assumed to have the same resolution as the simulation grid. Hence, whereas punctual conditioning data (corresponding to the scale of the simulation domain) can be straightforwardly handled, conditioning a simulation at local scale with data defined at a larger scale is quite challenging. Classical parametric methods based on covariance models can be used to integrate data with different support sizes (Liu and Journel, 2009; Journel, 1999). Essentially based on cokriging theory, such techniques nevertheless require point-to-point, point-to-block and block-to-block (cross-) covariance models and imply Gaussian assumptions.

In this paper, we propose an extension of the direct sampling algorithm able to deal with block data, *i.e.* target values for the average of the simulated variable on subsets of the simulation domain. The principle is to use the block data as a criterion for accepting a candidate location in addition to the comparison of the patterns. The current averages accounting for the already simulated pixels in the blocks plus the candidate value are compared to the target mean values and then a related mismatch for each block data is computed. Indeed, we follow a similar strategy as for multivariate simulations