Conditional Simulation with Patterns

G. Burc Arpat and Jef Caers

An entirely new approach to stochastic simulation is proposed through the direct simulation of patterns. Unlike pixel-based (single grid cells) or object-based stochastic simulation, pattern-based simulation simulates by pasting patterns directly onto the simulation grid. A pattern is a multi-pixel configuration identifying a meaningful entity (a puzzle piece) of the underlying spatial continuity. The methodology relies on the use of a training image from which the pattern set (database) is extracted. The use of training images is not new. The concept of a training image is extensively used in simulating Markov random fields or for sequentially simulating structures using multiple-point statistics. Both these approaches rely on extracting statistics from the training image, then reproducing these statistics in multiple stochastic realizations, at the same time conditioning to any available data. The proposed approach does not rely, explicitly, on either a statistical or probabilistic methodology. Instead, a sequential simulation method is proposed that borrows heavily from the pattern recognition literature and simulates by pasting at each visited location along a random path a pattern that is compatible with the available local data and any previously simulated patterns. This paper discusses the various implementation details to accomplish this idea. Several 2D illustrative as well as realistic and complex 3D examples are presented to showcase the versatility of the proposed algorithm.

KEY WORDS: Geostatistics, pattern similarity, multiple-point statistics.

INTRODUCTION

Stochastic conditional simulation is an important tool for modeling spatial phenomena. The goal of stochastic simulation is to create multiple, realistic representations, termed realizations, of a studied spatial phenomenon, that are constrained (conditioned) to any available data. Each realization should reflect the spatial continuity believed to exist. In most applications, spatial continuity is quantified through a spatial covariance or variogram. This variogram is modeled from experimental data, then reproduced in each realization. Several publications have pointed out (Journel, 1993; Caers and Journel, 1998; Strebelle, 2000, 2002; Caers and Zhang, 2004) the limitation of the variogram model in capturing spatial...
continuity of actual phenomena. Being a two-point statistic, describing correlation between any two points in space only, the variogram cannot model strongly connected and/or curvilinear geometries of the variable considered. Recent work has suggested an alternative approach to reproducing realistic spatial continuity, based on the concept of “training images.” The training image is a conceptual, but explicit, 2D or 3D, depiction of the spatial continuity studied. Training images may come from actual data such as from outcrops or other exhaustive datasets deemed representative for the area being modeled. Alternatively, a training image could be a single but large, unconditional realization of a stochastic simulation method, e.g. using a Boolean or object-based model. Either way, a training image need not be locally constrained to any (hard and soft) data, need not have the same dimensions as the study area, but should reflect a style of spatial continuity interpreted as similar to the actual phenomenon.

Several approaches have been developed for generating conditional realizations reproducing spatial continuity similar as the training image. Consider the simulation of a random function $Z(u)$ over a regularly gridded domain $u = (x, y, z) \in A$.

When implementing the approach of simulating Markov random fields (MRF) with higher order interaction using McMC (Tjelmeland, 1996), one models the conditional probability of $Z(u)$ given a template of neighboring values around $u$ through an exponential-type parametric model (known up to a normalization constant) involving higher order (more than two) interactions (termed cliques). Under this form the joint distribution of all $Z(u)$ can be written analytically. The unknown normalization constant requires that simulation is done iteratively, in particular, using the Metropolis-Hasting sampler. While theoretically elegant, the application of the method to practical cases runs into several problems. The parameter estimation of the exponential model using maximum likelihood is tedious because of the unknown normalization constant of the conditional pdfs. Moreover, the model parameters are difficult to interpret in terms of actual patterns or statistics in the training image. Secondly, simulation is slow due to McMC; issues of convergence arise. The methodology is essentially borrowed from Bayesian image analysis and works well for image restoration problems. Few large 3D application in the earth sciences, typified by sparse data and complex priors, has been reported.

Strebelle (2002) proposes a more ad-hoc approach based on an original idea of Guardiano and Srivastava (1993). Strebelle uses a sequential simulation framework, hence is non-iterative, to simulate realizations. Sequential simulation requires one to calculate, at each visited location, the conditional distribution of $Z(u)$ given any direct observations on $Z(u)$ (termed hard data) and previously simulated nodes. Instead of modeling this conditional probability using parametric model as done in MRF, the conditional distribution is estimated directly from the training image as follows. At each nodal location $u$ along the random path, the data values and configuration of any hard data and previously simulated values.