Computers & Geosciences 89 (2016) 57–70

Contents lists available at ScienceDirect

Computers & Geosciences

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

Case study

GOSIM: A multi-scale iterative multiple-point statistics algorithm with global optimization

Liang Yang [a,b], Weisheng Hou [a,b,*], Chanjie Cui [a,b], Jie Cui [c]

[a] School of Earth Science and Geological Engineering, Sun Yat-sen University, Guangzhou, 510275, China
[b] Guangdong Provincial Key Laboratory of Mineral Resource Exploration & Geological Processes, Guangzhou, 510275, China
[c] Guangdong Techno-economy Research and Development Center, Guangzhou, 510070, China

ARTICLE INFO

Article history:
Received 10 July 2015
Received in revised form
19 December 2015
Accepted 29 December 2015
Available online 7 January 2016

Keywords:
MPS algorithm
Global optimization
Iterative algorithm
Geostatistical simulation

ABSTRACT

Most current multiple-point statistics (MPS) algorithms are based on a sequential simulation procedure, during which grid values are updated according to the local data events. Because the realization is updated only once during the sequential process, errors that occur while updating data events cannot be corrected. Error accumulation during simulations decreases the realization quality. Aimed at improving simulation quality, this study presents an MPS algorithm based on global optimization, called GOSIM. An objective function is defined for representing the dissimilarity between a realization and the TI in GOSIM, which is minimized by a multi-scale EM-like iterative method that contains an E-step and M-step in each iteration. The E-step searches for TI patterns that are most similar to the realization and match the conditioning data. A modified PatchMatch algorithm is used to accelerate the search process in E-step. M-step updates the realization based on the most similar patterns found in E-step and matches the global statistics of TI. During categorical data simulation, *k*-means clustering is used for transforming the obtained continuous realization into a categorical realization. The qualitative and quantitative comparison results of GOSIM, MS-CCSIM and SNESIM suggest that GOSIM has a better pattern reproduction ability for both unconditional and conditional simulations. A sensitivity analysis illustrates that pattern size significantly impacts the time costs and simulation quality. In conditional simulations, the weights of conditioning data should be as small as possible to maintain a good simulation quality. The study shows that big iteration numbers at coarser scales increase simulation quality and small iteration numbers at finer scales significantly save simulation time.

1. Introduction

As a developing technique originated about a decade ago, multiple-point statistics (MPS) has been applied in many fields, such as reservoir forecasting (Pyrcz and Deutsch, 2014), mineral resources (Jones et al. 2013; Boucher et al., 2014) and climate modeling (Jha et al., 2013). The main purpose of MPS algorithms is to simulate multiple models via capturing the spatial structure in training images (TIs), while simultaneously following the constraints of conditioning data (Journel and Zhang, 2006; Caers, 2011). The similarity between realizations and TI, also called pattern reproduction, is one of the major factors in evaluating MPS algorithms (Tan et al., 2014; Mariethoz and Caers, 2015). Although MPS algorithms have been studied for some time, further pattern reproduction quality improvements need to be achieved.

The first MPS algorithm presented by Guardiano and Srivastava (1993) is rarely used in practical applications due to its high CPU demand. Since then, various MPS algorithms have been proposed. Some algorithms that are called pixel-based algorithms simulate one pixel at a time (Strebelle, 2002; Mariethoz et al., 2010; Straubhaar et al., 2011). Some other algorithms are pattern-based algorithms because the entire pattern is pasted at a time during simulation (Zhang et al., 2006; Arpat and Caers, 2007; Honarkhah and Caers, 2010; Tahmasebi et al., 2012, 2014; Mahmud et al., 2014). Different from these TI-based algorithms, HOSIM uses spatial cumulants to represent high-order information that can be directly calculated from abundant data (Dimitrakopoulos et al., 2010; Mustapha and Dimitrakopoulos, 2011). HOSIM is promising in the mineral resources field, where numerous borehole data are available.

Both existing pixel-based and pattern-based MPS algorithms follow a sequential simulation procedure. In these MPS algorithms, the update of data event is confined by a local region and errors might occur. Because the grid is traversed only once, the error cannot be corrected and errors accumulate due to the sequential

* Corresponding author at: School of Earth Science and Geological Engineering, Sun Yat-sen University, B410 Dihuan Building, No. 135 Xingangxi Road, Guangzhou, Guangdong Province 510275, China.
E-mail address: houwsh@mail.sysu.edu.cn (W. Hou).

http://dx.doi.org/10.1016/j.cageo.2015.12.020
0098-3004/

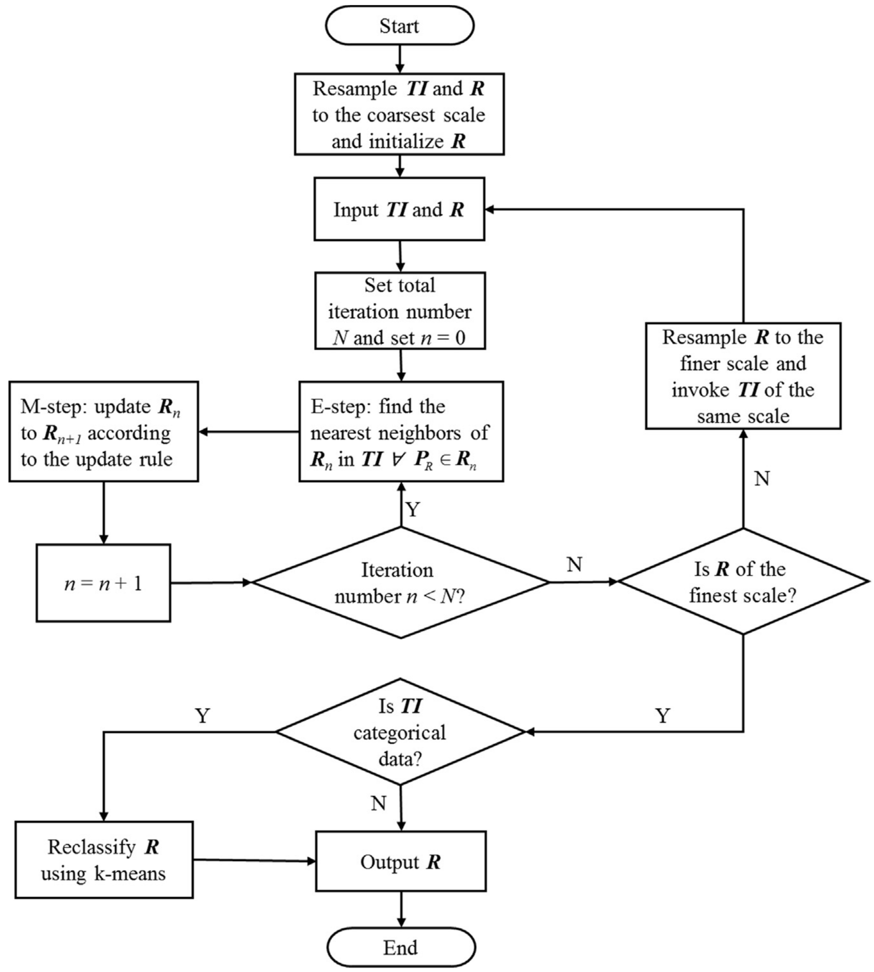

Fig. 1. The GOSIM algorithm flowchart.

process. After a sequence of error accumulations, the final realization may deviate from the TI. Although it has been noted that, a realization obtained with a raster path is less sensitive to the pixels or patterns simulated at the beginning compared to a random path (Parra and Ortiz, 2011; Tahmasebi et al., 2012), the issue of error accumulation remains a hindrance of MPS.

In the field of computer graphics, example-based texture synthesis has a goal that is similar to that of MPS (Wei et al., 2009; Mariethoz and Lefebvre, 2014). Current example-based texture synthesis algorithms can be classified as pixel-based methods (Efros and Leung, 1999; Wei and Levoy, 2000), patch-based methods (Efros and Freeman, 2001; Kwatra et al., 2003) and optimization-based methods (Kwatra et al., 2005; Kopf et al., 2007). The first two algorithm types are similar to the pixel-based and pattern-based methods of MPS, while the optimization-based methods have no analogous MPS algorithm. The optimization-based methods refine realizations using an iterative process, thus they avoid the issue of error accumulation. However, the original global optimization algorithm proposed by Kwatra et al. (2005) is relatively inefficient, and cannot be directly applied to categorical data simulations, 3D simulations and conditional simulations. Although Kopf et al. (2007) extends the optimization-based method to 3D cases and allows the addition of soft constraints to control the synthesis outcome, the algorithm cannot be directly used in 3D simulation cases with 3D TI because it utilizes 2D exemplar to synthesize 3D results, and is not suitable for hard point data conditioning. A thorough discussion of the similarities and differences between texture synthesis and MPS can be found in Mariethoz and Lefebvre (2014).

In this study, we present an MPS algorithm based on the global optimization scheme, hereafter GOSIM. GOSIM is adapted from the global optimization algorithm proposed by Kwatra et al. (2005). Different from the algorithm presented by Kwatra et al. (2005), GOSIM can be applied to categorical data simulations, 3D simulations and conditional simulations. In addition, GOSIM uses a modified version of PatchMatch (Barnes et al., 2009) to accelerate the pattern search process. This paper is organized as follows. Section 2 provides important background and terminology. The details of how GOSIM is implemented in different situations are explained in Section 3. Section 4 compares several simulation tests between GOSIM, MS-CCSIM and SNESIM, by means of human vision, ensemble averages and analysis of distance (ANODI). In addition, a sensitivity analysis of GOSIM is presented in Section 4.

2. Background and terminology

GOSIM is based on a simple assumption. If a realization ($\boldsymbol{R}$) visually resembles a training image ($\boldsymbol{TI}$), then as many patterns as possible in $\boldsymbol{R}$ come from $\boldsymbol{TI}$. Let $\boldsymbol{P}_{TI}$ denote a pattern in $\boldsymbol{TI}$ and $\boldsymbol{P}_R$ denote a pattern in $\boldsymbol{R}$, the distance (or dissimilarity) between $\boldsymbol{R}$ and $\boldsymbol{TI}$ is:

$$d(\boldsymbol{R}, \boldsymbol{TI}) = \sum_{\boldsymbol{P}_R \in \boldsymbol{R}} \min_{\boldsymbol{P}_{TI} \in \boldsymbol{TI}} D(\boldsymbol{P}_R, \boldsymbol{P}_{TI}) \quad (1)$$

The implication of Eq. (1) is that for each $\boldsymbol{P}_R \in \boldsymbol{R}$, its most