Computers & Geosciences 104 (2017) 35–53

Contents lists available at ScienceDirect

Computers & Geosciences

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

Research paper

A training image evaluation and selection method based on minimum data event distance for multiple-point geostatistics

Wenjie Feng[a], Shenghe Wu[a,*], Yanshu Yin[b], Jiajia Zhang[a], Ke Zhang[a]

[a] State Key Laboratory of Petroleum Resource and Prospecting, College of Geosciences, China University of Petroleum (Beijing), Beijing, China
[b] School of Geosciences, Yangtze University, Wuhan, China

ARTICLE INFO

Keywords:
Multiple-point geostatistics
Candidate training images
Evaluation and optimal selection
Minimum data event distance

ABSTRACT

A training image (TI) can be regarded as a database of spatial structures and their low to higher order statistics used in multiple-point geostatistics (MPS) simulation. Presently, there are a number of methods to construct a series of candidate TIs (CTIs) for MPS simulation based on a modeler's subjective criteria. The spatial structures of TIs are often various, meaning that the compatibilities of different CTIs with the conditioning data are different. Therefore, evaluation and optimal selection of CTIs before MPS simulation is essential. This paper proposes a CTI evaluation and optimal selection method based on minimum data event distance (MDevD). In the proposed method, a set of MDevD properties are established through calculation of the MDevD of conditioning data events in each CTI. Then, CTIs are evaluated and ranked according to the mean value and variance of the MDevD properties. The smaller the mean value and variance of an MDevD property are, the more compatible the corresponding CTI is with the conditioning data. In addition, data events with low compatibility in the conditioning data grid can be located to help modelers select a set of complementary CTIs for MPS simulation. The MDevD property can also help to narrow the range of the distance threshold for MPS simulation. The proposed method was evaluated using three examples: a 2D categorical example, a 2D continuous example, and an actual 3D oil reservoir case study. To illustrate the method, a C++ implementation of the method is attached to the paper.

1. Introduction

The goal of multiple-point geostatistics (MPS) is to reproduce the geological patterns contained in training images (TIs) into realizations. Therefore, TIs can be regarded as one of the key factors for determining the quality of realizations (Arpat and Caers, 2007; Boisvert et al., 2007; Hu and Chugunova, 2008; Mariethoz and Caers, 2014; Pérez et al., 2014; Strebelle and Journel, 2001; Wu and Zhang, 2008; Zhang et al., 2006). In order to obtain sufficient TIs for MPS simulation, various methods have been proposed, such as object-based methods (Boucher et al., 2010; Pyrcz et al., 2008), process-based methods (Pyrcz et al., 2009), and process-mimicking methods (Lopez et al., 2009). Based on these methods, a number of programs have been developed, such as Fluvsim (Deutsch and Tran, 2002), TiGenerator (Maharaja, 2008), Alluvsim (Pyrcz et al., 2009), Flumy (Deviese, 2010), Tetris (Boucher et al., 2010), and TiConverter (Fadlelmula et al., 2016). Additionally, modelers may also construct TIs using deterministic methods, and TI databases have been proposed and developed (Colombera et al., 2012; Pyrcz et al., 2008).

TIs offer a framework that adds subjectivity to earth science models, although subjectivity is a double-edged sword as modeling assumptions are explicitly laid out in the TI (Mariethoz and Caers, 2014). Various types of TIs can be obtained by different methods and tools, however, there is a very limited number of methods that can evaluate the compatibility of candidate training images (CTIs) with the conditioning data. As a consequence, selecting an appropriate TI is one of the main issues when using MPS simulation.

Variograms can be used to quantify the consistency of a TI based on two-point spatial statistics derived from both TIs and the conditioning data, but the higher order TI consistency will not be considered (Pérez et al., 2014). Ortiz and Deutsch (2004) came up with a method for TI selection through comparing the cumulative distribution of runs (Mood, 1940) of the TI with the cumulative distribution of runs observed in 1D wells. Boisvert et al. (2007) proposed another method based on the comparison of multiple point histograms for vertical one-dimensional patterns. Case studies show that the above methods can narrow the list of possible TIs. However, these methods do not consider the compatibilities of the complex spatial structures constituted by the

* Correspondence to: State Key Laboratory of Petroleum Resource and Prospecting, College of Geosciences, No.18 Fuxue Road, Changping District, Beijing, China.
E-mail address: reser@cup.edu.cn (S. Wu).

http://dx.doi.org/10.1016/j.cageo.2017.04.004
Received 17 August 2016; Accepted 7 April 2017
Available online 11 April 2017

Table 1
An example parameter set and its associated descriptions.

Parameter value	Description	Line number
0.5	Fraction of grid nodes to be considered during calculation(F_n)	1
0	CTI Type (0: categorical, 1: continuous)	2
6	Number of categories (if CTI type=1, this parameter will be ignored)	3
1	Number of templates	4
0	Template type (0: simple rectangular templates; 1: user defined templates)	5
15 15 3	Size of template 1 in the x, y, and z directions	6
Template1.dat	File name of template file 1	7
1	MDevD type (1: for categorical CTIs, see Eq. (2); 2-for continuous CTIs, see Eq. (3))	8
Conditioning data.dat	File name of conditioning data	9
128 183 25	Size of conditioning data grid in the x, y, and z directions	10
2	Number of CTIs	11
TI_A.dat	File name of CTI_A	12
128 183 25	Size of CTI A in the x, y, and z directions	13
TI_B.dat	File name of CTI_B	14
128 183 25	Size of CTI B in the x, y, and z directions	15
0	Threshold of MDevD	16
5	The least amount of conditioning data nodes necessary for CDE construction	17
0.5	The least fraction of nodes in TDE necessary for data event distance calculation, (0, 1)	18
69071	Random number seed	19
MDevD.out	File path of MDevD properties	20

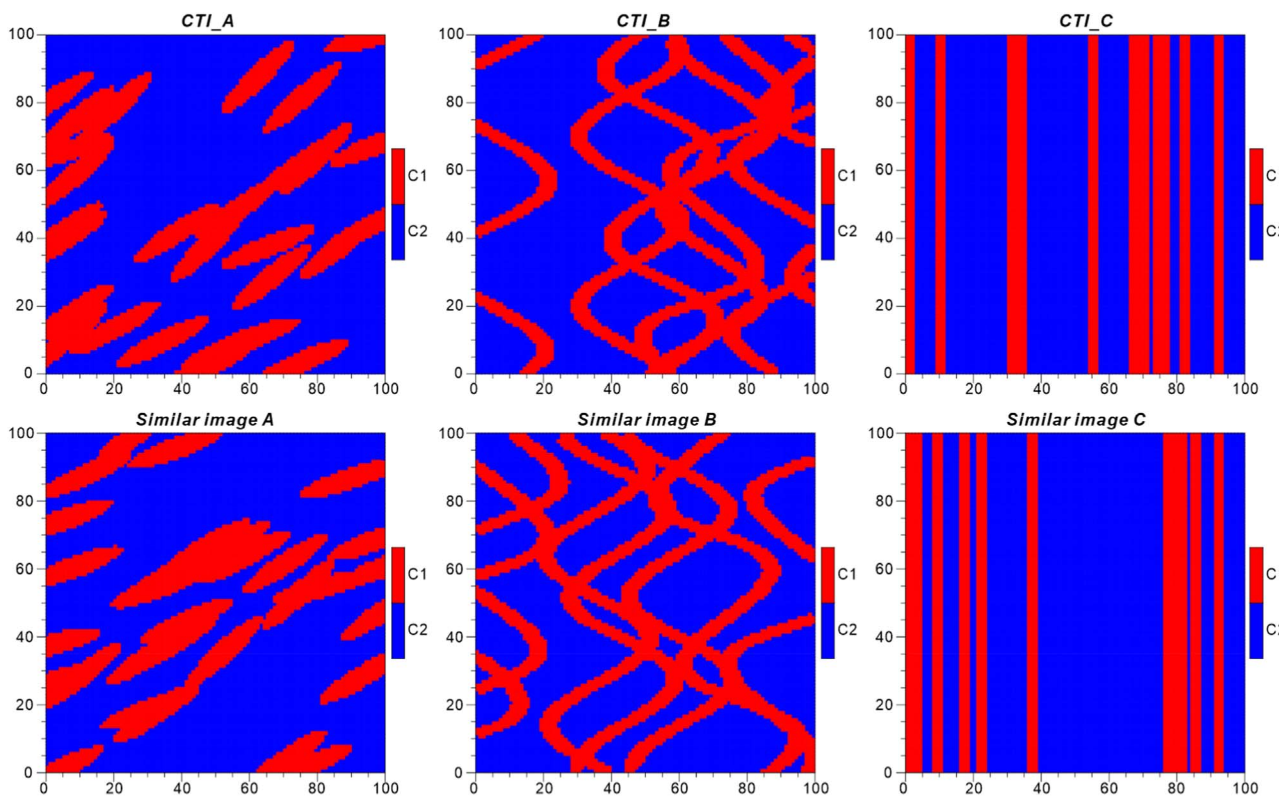

Fig. 1. Binary images used as CTIs (upper row) and for random conditioning data extraction (lower row) (Pérez et al., 2014).

grid nodes from multiple wells because the calculation path is 1D along the well bore.

Eskandaridalvand (2008) proposed a spiral searching based MPS method that enables acquisition of a distribution of compatible training nodes for each conditioning node and a unique distribution of maximum compatible nodes. These distributions can be used to derive a measure of TIs' consistency with the conditioning data. Inspired by this method, a practical tool was developed to rank TIs according to their relative compatibility with the conditioning data, and to obtain an absolute measurement quantifying the consistency between TIs and conditioning data in terms of spatial structure (Pérez et al., 2014). The indicator variable for compatibilities used in this method is similar to the data event distance described by Mariethoz et al. (2010). However, the indicator gives the same weight to all the nodes of the data event regardless of their location relative to the central node.

As an alternative to the methods mentioned above, we note that data event distance, which gives distance-based weights to nodes in a data event, can precisely and quantitatively indicate the similarity error between two data events (Arpat, 2005; Mariethoz et al., 2010). As stated by Arpat (2005), modelers can decide ahead of time whether the actual data conflicts with the TI by calculating the similarity of hard/soft data images to hard/soft TIs. The minimum data event distance