Computers and Geosciences 114 (2018) 1–10

Contents lists available at ScienceDirect

Computers and Geosciences

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

Research paper

Multiple point statistical simulation using uncertain (soft) conditional data

Thomas Mejer Hansen [a,*], Le Thanh Vu [b], Klaus Mosegaard [a], Knud Skou Cordua [a]

[a] Niels Bohr Institute, University of Copenhagen, Denmark
[b] I-GIS, Risskov, Denmark

ARTICLE INFO

Keywords:
Multiple point statstics
Uncertain data
Data integration

ABSTRACT

Geostatistical simulation methods have been used to quantify spatial variability of reservoir models since the 80s. In the last two decades, state of the art simulation methods have changed from being based on covariance-based 2-point statistics to multiple-point statistics (MPS), that allow simulation of more realistic Earth-structures. In addition, increasing amounts of geo-information (geophysical, geological, etc.) from multiple sources are being collected. This pose the problem of integration of these different sources of information, such that decisions related to reservoir models can be taken on an as informed base as possible. In principle, though difficult in practice, this can be achieved using computationally expensive Monte Carlo methods. Here we investigate the use of sequential simulation based MPS simulation methods conditional to uncertain (soft) data, as a computational efficient alternative. First, it is demonstrated that current implementations of sequential simulation based on MPS (e.g. SNESIM, ENESIM and Direct Sampling) do not account properly for uncertain conditional information, due to a combination of using only co-located information, and a random simulation path. Then, we suggest two approaches that better account for the available uncertain information. The first make use of a preferential simulation path, where more informed model parameters are visited preferentially to less informed ones. The second approach involves using non co-located uncertain information. For different types of available data, these approaches are demonstrated to produce simulation results similar to those obtained by the general Monte Carlo based approach. These methods allow MPS simulation to condition properly to uncertain (soft) data, and hence provides a computationally attractive approach for integration of information about a reservoir model.

1. Introduction

During the last 30 years a number of probabilistic based methods and algorithms have been developed in the geostatistical community, that allow quantification and simulation of increasingly geologically complex structural variability, see e.g. Deutsch and Journel (1992); Guardiano and Srivastava (1993); Strebelle (2000); Remy et al. (2008); Mariethoz et al. (2010); Straubhaar et al. (2011); Mariethoz and Kelly (2011); Toftaker and Tjelmeland (2013); Tahmasebi et al. (2014); Mariethoz and Caers (2014).

State of the art simulation methods have changed from being based on 2-point statistics (covariance-based statistics) to multiple-point statistics (MPS), that allow simulation of more realistic Earth-structures. MPS is especially important used as a base for flow modeling, as traditional 2-point statistics cannot adequately describe for example realistic connectivity of geological structures, that may have significant effect on flow properties and transport, see e.g. Zinn and Harvey (2003); Renard et al. (2011). The information about the expected spatial variability of the properties in a reservoir model can be conveniently provided in form of a 'training image'/'sample model' when using MPS. Using such a training image, several methods exist for simulation of multiple realizations of reservoir models that are consistent with the spatial statistics of the training image, e.g. Guardiano and Srivastava (1993); Strebelle (2000); Mariethoz et al. (2010).

Additional information is often available from e.g. boreholes and geophysical surveys (seismic, electromagnetic,…). Ideally, this information should be combined with the geostatistical information in order to obtain a stochastic reservoir model, or realizations of such a model that are consistent with all available data/information.

Several methods have been proposed to deal with this problem of integration of information. Probabilistic inverse problem theory allow combining the available information by characterizing (or sampling from) a posterior probability function that combines the information form the geostatistical model that describes realistic earth models (in form of a prior probability density), with information from data (in form of a likelihood function) (Tarantola, 2005). Using Monte Carlo sampling

* Corresponding author.
E-mail address: tmeha@nbi.ku.dk (T.M. Hansen).

https://doi.org/10.1016/j.cageo.2018.01.017
Received 22 May 2017; Received in revised form 16 January 2018; Accepted 28 January 2018
Available online 1 February 2018

the posterior of any posterior probability can be sampled, as long as the prior model can be sampled, and the likelihood can be evaluated (Mosegaard and Tarantola, 1995; Hansen et al., 2008, 2012; Irving and Singha, 2010; Cordua et al., 2012; Hansen et al., 2013). While such a Monte Carlo based approach can in principle deal with a large variety of very complex systems, its practical use is hampered by its very high computational demands.

Another approach is typically used in geostatistics, where available (geophysical) data are converted into 'soft data' about each individual model parameter. Soft data is a loosely defined term that typically refer to uncertainty and inequality constraints about specific model parameters (Journel, 1986). Most all geostatistical simulation algorithms can make use of such 'soft' data (Remy et al., 2008; Mariethoz and Caers, 2014). However, challenges related to using current state of the art MPS simulation algorithms conditional to other geo-information has been considered widely in the literature with respect to ground water models He et al. (2014); Koch et al. (2014); Jørgensen et al. (2015); Biver et al. (2014); Høyer et al. (2015, 2017).

In the following the use of sequential simulation based MPS sampling methods will be considered for probabilistic data integration with independent uncertain conditional data, that may be available from other sources.

First, using notation from probabilistic data integration, we formulate precisely what is implicitly assumed about 'soft data' in most any MPS algorithm.

Through analysis of 3 reference models, with varying density of conditional/soft data, we demonstrate that a conventional implementation of sequential simulation based MPS simulation leads to simulations that fail to generate realizations (reservoir models) consistent with the available uncertain information (soft data).

Then, we suggest two novel approaches that allow considering the information in a more correct way using direct sampling (DS, Mariethoz et al., 2010), ENESIM (Guardiano and Srivastava, 1993), and SNESIM (Strebelle, 2000). The first use as preferential simulation path, where more informed model parameters are visited preferentially to less informed ones. The second approach involves using more than only co-located uncertain data, wihich is typically not done for most implementations of MPS. All examples are compared to those obtained by a general Monte Carlo based approach.

2. Data integration using conditional geostatistical simulation - theory

Consider that a model of the subsurface is parameterized into M model parameters $\mathbf{m} = [m_1, m_2, m_3, \ldots, m_M]$. Say information is available about the model parameters $\mathbf{m}$ from N independent sources $\mathbf{I} = [I_1 1, I_2, \ldots, I_N]$ through the probability densities $f_{I_1}(\mathbf{m}), f_{I_2}(\mathbf{m}), \ldots, f_{I_N}(\mathbf{m})$. Each probability distribution then represents a specific *state of information*. Tarantola and Valette (1982) and Tarantola (2005) demonstrate how these states of information can be combined through the *conjunction* of the states of information through

$$f_{\mathbf{I}}(\mathbf{m}) = f_{I_1}(\mathbf{m}) \wedge f_{I_2}(\mathbf{m}) \wedge \ldots \wedge f I_N(\mathbf{m}) = \nu \mu(\mathbf{m})^{(1-N)} \prod_i^N f_{I_i}(\mathbf{m}), \tag{1}$$

where ν represents a normalizing constant, $\mu(\mathbf{m})$ represents the homogeneous probability distribution or the 'state of total ignorance' (Jaynes, 1968), and $\wedge$ is the operator for 'conjunction'. Conjunction of information, as expressed through (1), is derived from axioms similar to the axioms of formal logic on conjunction of propositions, and the Radon-Nikodym theorem from measure theory (Tarantola and Valette, 1982).

If a Cartesian coordinate system is used to parameterize $\mathbf{m}$, then the homogeneous probability density function becomes a constant $\mu(\mathbf{m}) = k$ (Mosegaard and Tarantola, 2002), which is the case we will consider here. Then the problem of integrating information from independent sources into to one probability density $f_{\mathbf{I}}(\mathbf{m})$ is given by

$$f_{\mathbf{I}}(\mathbf{m}) \propto \prod_i^N f_{I_i}(\mathbf{m}). \tag{2}$$

In the present context $\mathbf{m}$ reflects model parameters describing a reservoir model, and I_1, I_2,. reflect different sources of information available (e.g. from expert information, well log data, training image and geophysical data).

Here, the special case is considered where all information available refers directly to the model parameters. The reason for this is two-fold: First, most (any) geostatistical simulation algorithms allow, in principle, to take such information into account as "soft" information (Mariethoz and Caers, 2014). Second, working with reservoir models, a lot of information about the model parameters of interest can be available in form of direct measurements from well logs, inverted well logs parameters, or indirectly from geophysical data inverted into information about the model parameters $\mathbf{m}$. Barfod et al. (2016) present a recent example of how to do this, by establishing an atlas (applicable in Denmark) that can be used to translate resistivity values (found through inversion of airborne EM data) into lithological/hydrological units with associated uncertainty.

Three types of information are available in a typical MPS based geostatistical data integration problem:

I_{TI} *Information from a training image*. This can be information from outcrops, previous analysis, well log analysis, expert information which is quantified through a geostatistical model describing (spatial) co-dependence between model parameters.

I_{hard} *Hard data*. Direct observation of one or more model parameters, without any associated uncertainty.

I_{soft} *Soft data*. Direct observation of one or more model parameters, with an associated uncertainty.

In case the information has been obtained independently, such a geostatistical problem is equivalent to the problem of inferring information about $f_{\mathbf{I}}(\mathbf{m})$ given by

$$f_{\mathbf{I}}(\mathbf{m}) \propto f_{I_{TI}}(\mathbf{m}) f_{I_{hard}}(\mathbf{m}) f_{I_{soft}}(\mathbf{m}). \tag{3}$$

Høyer et al. (2017) present one example of combining these three types of information into one stochastic model.

In principle there is no need to distinguish between hard and soft data, as both are simply data that provide information about the model parameters. So, a general geostatistical data integration problem can be formulated as

$$f_{\mathbf{I}}(\mathbf{m}) \propto f_{I_{TI}}(\mathbf{m}) f_{I_{data}}(\mathbf{m}). \tag{4}$$

Spatially independent 'data'. For many geostatistical data integration problems, the information about each model parameter is assumed spatially independent, such that

$$f_{I_{data}}(\mathbf{m}) = \prod_{i=1}^M f_{I_{data}}(m_i). \tag{5}$$

From hereon, the term 'soft information' about the model parameters is defined through equation (5). The general data integration problem of equation (4) is then reduced to

$$f_{\mathbf{I}}(\mathbf{m}) \propto f_{TI}(\mathbf{m}) f_{data}(\mathbf{m}) = f_{TI}(\mathbf{m}) \prod_{i=1}^M f_{data}(m_i). \tag{6}$$

Equation (6) represent the probability distribution that most sequential simulation based MPS methods suggest to sample from, by combining information from a geostatistical model with 'hard' (certain) and 'soft' (uncertain) data. From hereon different methods, existing and new, will be discussed that allow sampling from equation (6).