Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	Modeling and Simulating Depositional Sequences Using Latent Gaussian Random Fields	
Article Sub-Title		
Article CopyRight	International Association for Mathematical Geosciences (This will be the copyright line in the final PDF)	
Journal Name	Mathematical Geosciences	
Corresponding Author	Family Name	**Allard**
	Particle	
	Given Name	**Denis**
	Suffix	
	Division	
	Organization	INRAE, BioSP
	Address	84914, Avignon, France
	Phone	
	Fax	
	Email	denis.allard@inrae.fr
	URL	
	ORCID	http://orcid.org/0000-0001-7944-1906
Author	Family Name	**Fabbri**
	Particle	
	Given Name	**Paolo**
	Suffix	
	Division	
	Organization	University of Padua
	Address	Padua, Italy
	Phone	
	Fax	
	Email	paolo.fabbri@unipd.it
	URL	
	ORCID	
Author	Family Name	**Gaetan**
	Particle	
	Given Name	**Carlo**
	Suffix	
	Division	
	Organization	Ca’ Foscari University of Venice
	Address	Venice, Italy
	Phone	
	Fax	
	Email	gaetan@unive.it
	URL	

ORCID

Schedule	Received	17 February 2020
	Revised	
	Accepted	21 May 2020

Abstract	Simulating a depositional (or stratigraphic) sequence conditionally on borehole data is a long-standing problem in hydrogeology and in petroleum geostatistics. This paper presents a new rule-based approach for simulating depositional sequences of surfaces conditionally on lithofacies thickness data. The thickness of each layer is modeled by a transformed latent Gaussian random field allowing for null thickness thanks to a truncation process. Layers are sequentially stacked above each other following the regional stratigraphic sequence. By choosing adequately the variograms of these random fields, the simulated surfaces separating two layers can be continuous and smooth. Borehole information is often incomplete in the sense that it does not provide direct information about the exact layer that some observed thickness belongs to. The latent Gaussian model proposed in this paper offers a natural solution to this problem by means of a Bayesian setting with a Markov chain Monte Carlo (MCMC) algorithm that can explore all possible configurations that are compatible with the data. The model and the associated MCMC algorithm are validated on synthetic data and then applied to a subsoil in the Venetian Plain with a moderately dense network of cored boreholes.

Keywords (separated by '-')	Subsoil modeling - Stratigraphic sequence - PC prior - Stochastic 3D model - Data augmentation - Conditional simulation
Footnote Information	

Math Geosci
https://doi.org/10.1007/s11004-020-09875-0

Modeling and Simulating Depositional Sequences Using Latent Gaussian Random Fields

Denis Allard[1] · **Paolo Fabbri**[2] · **Carlo Gaetan**[3]

Received: 17 February 2020 / Accepted: 21 May 2020

Abstract Simulating a depositional (or stratigraphic) sequence conditionally on borehole data is a long-standing problem in hydrogeology and in petroleum geostatistics. This paper presents a new rule-based approach for simulating depositional sequences of surfaces conditionally on lithofacies thickness data. The thickness of each layer is 1 modeled by a transformed latent Gaussian random field allowing for null thickness thanks to a truncation process. Layers are sequentially stacked above each other following the regional stratigraphic sequence. By choosing adequately the variograms of these random fields, the simulated surfaces separating two layers can be continuous and smooth. Borehole information is often incomplete in the sense that it does not 2 provide direct information about the exact layer that some observed thickness belongs to. The latent Gaussian model proposed in this paper offers a natural solution to this problem by means of a Bayesian setting with a Markov chain Monte Carlo (MCMC) algorithm that can explore all possible configurations that are compatible with the data. The model and the associated MCMC algorithm are validated on synthetic data and then applied to a subsoil in the Venetian Plain with a moderately dense network of cored boreholes.

✉ Denis Allard
denis.allard@inrae.fr

Paolo Fabbri
paolo.fabbri@unipd.it

Carlo Gaetan
gaetan@unive.it

1 INRAE, BioSP, 84914 Avignon, France

2 University of Padua, Padua, Italy

3 Ca' Foscari University of Venice, Venice, Italy

Keywords Subsoil modeling · Stratigraphic sequence · PC prior · Stochastic 3D model · Data augmentation · Conditional simulation

1 Introduction

The case study motivating this work is a subsoil in the Venetian Plain with a moderately dense network of cored boreholes. Geologists and hydrogeologists managing this subsoil are in need of stochastic three-dimensional models of the stratigraphic sequence. The model should of course be conditioned to borehole data. The sequence of layers must correspond to the known regional stratigraphic sequence and, in addition, to the surfaces separating the layers are required to be smooth and continuous.

Simulating a depositional (or stratigraphic) sequence conditionally on boreholes data has been and still is a long-standing problem in hydrogeology and in petroleum geostatistics. In the context of reservoir modeling, Pyrcz et al. (2015) offers a comprehensive overview of the literature and a convincing conceptual framework in which methods are represented along a complexity gradient with one extreme corresponding to pixel based models with statistics and conditioning derived from the data and the other extreme representing geological concepts unconditional to local observations. As models tend to move away from the less complex extreme to the more complex one, they are less versatile and more difficult to condition (Pyrcz et al. 2015). Easy-to-condition pixel based methods thus tend to be favored when data are dense, whereas rule-based or process-based models are preferred when conditioning data is sparse.

Pixel based approaches, whether based on variograms (Matheron et al. 1987), truncated Gaussian random fields and plurigaussian random fields (Beucher et al. 1993; Galli et al. 1994; Armstrong et al. 2011; Le Blévec et al. 2017; Le Blévec et al. 2018), transiograms (Carle and Fogg 1996), or MCP (Allard et al. 2011; Sartore et al. 2016; Benoit et al. 2018b), are well known and relatively easy to handle. For these approaches, variogram and transiogram fitting is well understood and conditioning to well data is efficient, even for truncated Gaussian models (Marcotte and Allard 2018). However, one source of difficulty in the fitting procedure is the fact that the processes and the amount of information are often anisotropic. Typically, for borehole data, there is much more information along the depth than along horizontal directions.

Multiple point statistics (MPS) approaches (Strebelle 2002; Mariethoz and Caers 2014) require a training image when simulations are performed in two dimensions. Three-dimensional simulations are much more difficult to perform, since training cubes are rarely available at kilometer scales. Methods for combining images in three-dimensional simulations have been proposed (Comunian et al. 2012, 2014). But since a high degree of continuity is required for layers in this work, pixel based methods, including MPS, are not deemed appropriate.

Object models, such as Boolean models, are more difficult to fit and to condition, in particular when accounting for non-stationarity and erosion rules, see for example Syversveen and Omre (1997) and Allard et al. (2006). In addition, object models are not geologically appropriate for simulating sequences of layers.

Rule-based and process-based models incorporate some amount of understanding of the geological processes. They use rules to control the temporal sequence and spatial

Journal: **11004** Article No.: **9875** TYPESET DISK LE CP Disp.:**2020/6/8** Pages: **29** Layout: **Small-X**

position of geological objects so as to mimic geological processes. Among others, they have been applied to fluvial systems, deepwater channel systems and turbiditic lobes systems. Particular cases of interest to this work are surface-based models. For simulating lobes in a turbidite reservoir, Bertoncello et al. (2013) proposed a rule-based stacking of lobe-shape events with quite complicated sequential placement rules that depend partly on the already simulated events. The conditioning to well-log data and seismic data is achieved through sequential optimization. One of the limitations of this approach is that the variability between the conditional simulations is low, owing to the optimization approach. A second limitation recognized by the authors is that their method works best with a limited amount of data.

This paper presents a new rule-based approach for simulating depositional sequences of surfaces conditionally to lithofacies thickness data. It is a stochastic model that belongs to the *Markov rules* sub-class of rule-based methods, see Pyrcz et al. (2015) and appropriate references therein. The thickness of each layer is modeled by a transformed latent Gaussian random field allowing for null thickness. The random fields are *latent* because they can be unobserved on some parts of the domain under study, thanks to a truncation process. Layers are sequentially stacked above each other following the regional stratigraphic sequence. By choosing adequately the variograms of these random fields, the simulated surfaces separating two layers can be continuous and smooth. Conditioning to the observed borehole data is made possible thanks to constrained Gaussian conditioning, as will be explained later on.

A problem that has been barely addressed in geostatistical models for depositional sequences is the fact that borehole information is often incomplete in the sense that it does not provide direct information regarding the exact layers that have been observed. For example, let us consider that the stratigraphic sequence of the study domain contains several repetitions of a given lithofacies, say Clay. Consider also that the recorded data at one given borehole measures one single thickness for Clay. A first possibility is that there is actually only one Clay layer at this location, but it could be any of the several Clay layers of the regional stratigraphic sequence. Simulations should therefore account for this uncertainty. A second possibility is that the measurement actually corresponds to two (or more) Clay layers, one on top of the other, with missing intermediate layers at this location. In this case, the measured thickness should be shared between two layers. The latent Gaussian model proposed in this paper offers a natural solution to this problem by means of a Bayesian setting with a Markov Chain Monte Carlo (MCMC) algorithm that can explore all possible configurations compatible with the data. Notice that the approach proposed in Bertoncello et al. (2013) does not address this problem at all.

The rest of this paper is organized as follows. Section 2 is devoted to the conceptual model. In particular, the difference between the (unique) regional stratigraphic sequence, referred to as the parent sequence, and the observed sequences is detailed. Section 3 presents the stochastic model. In Sect. 4 all details for Bayesian inference with an MCMC algorithm are given. It is then validated on a synthetic data set in Sect. 5. Finally, it is successfully applied to the Venetian Plain that motivated this work in Sect. 6. Some concluding remarks are then given in Sect. 7.

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

2 The Conceptual Model

2.1 Notations

Let us consider a spatial domain $\mathcal{S} \in \mathbb{R}^2$ and an interval $\mathcal{T} \subset \mathbb{R}^+$, which will correspond to "depth". Note that depth can be converted into time through depositional processes, which is the reason why $t \in \mathcal{T}$ is used to denote depth. Let us also consider a family of K lithofacies, $\mathcal{C} = \{C_1, \ldots, C_K\}$. The aim of this work is to build a process $X = \{X(s,t)\}$, defined at any point $(s,t) \in \mathcal{S} \times \mathcal{T}$ and taking values in $\mathcal{C}$. In other words, at each location is associated one and only one lithofacies. The process must be continuous almost everywhere and the discontinuity surfaces should be smooth and have a general horizontal orientation. The process X is observed along depth at a finite number of locations $s_1, \ldots, s_n$ and each observation corresponds to a drilled core, referred to as boreholes in the rest of this work.

Let $X_i = \{X(s_i, t), t \in \mathcal{T}\}$ be one of these observations at site s_i, $i = 1, \ldots, n$, where n is the number of sites. The observation X_i is piece-wise constant, with M_i discontinuities at different depths each time a new layer is encountered. The resulting information is a sequence of facies and depths, referred to as the *observed sequence*, $(\mathbf{C}_i^o, \mathbf{T}_i^o)$, where $\mathbf{C}_i^o = (C_{1,i}^o, \ldots, C_{M_i,i}^o)$ with $C_{j,i}^o \in \mathcal{C}$ for $j = 1, \ldots, M_i$, and $\mathbf{T}_i^o = (T_{1,i}^o, \ldots, T_{M_i,i}^o)$ with $T_{j,i}^o \in \mathcal{T}$ and $T_{1,i}^o < \cdots < T_{M_i,i}^o$. The depths are measured with respect to a ground-level $T_{0,i}$. The thicknesses of each observed layer $\mathbf{Z}_i^o = (Z_{1,i}^o, \ldots, Z_{M_i,i}^o)$ can be derived from the depths, with $Z_{j,i}^o = T_{j,i}^o - T_{j-1,i}^o$, $j = 1, \ldots, M_i$. Finally, the last layer is assumed to be completely observed, that is the depth $Z_{M_i,i}^o$ is assumed to be not censored.

2.2 Parent Sequence

The working hypothesis is that there exists a common lithological sequence of facies, hereafter referred to as the "parent sequence," which is compatible with all observed sequences in the area of study in the sense that each observed sequence can be obtained from the parent sequence by deleting some layers of the parent sequence.

This sequence can result from the prior knowledge of the scientists. Alternatively, it can be derived from the observed data. From a mathematical viewpoint, there always exists a parent sequence. For example, it can easily be obtained by simply stacking all observed sequences into a single sequence. Then, each observed sequence of layers is simply obtained by "deleting" all other observed sequences. Obviously, this parent sequence is of no modeling interest, but it is mathematically important since it provides a proof of the existence of this concept. In general, very long parent sequences are uninteresting from a modeling point of view. In accordance with a parsimony principle, one should seek the shortest possible parent sequences. Clearly, there is only a finite number of parent sequences of minimal length. Such parent sequences could be built using discrete optimization algorithms, or they could be provided by scientists, based on prior geological knowledge. Either way, how minimal parent sequences are obtained is a subject out of the scope of the present research, and this route is not pursued any longer.

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

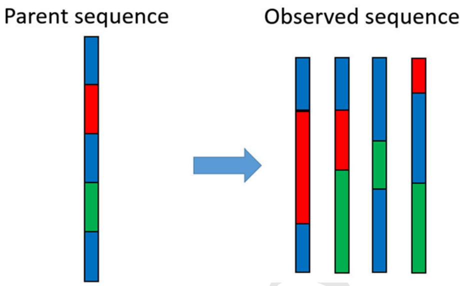


Fig. 1 Parent sequence and four possible incomplete observed sequences. Since the parent sequence is conceptual, thicknesses are only meaningful in the observed sequences

From now on it will be considered that the parent sequence is known, and that it is one of the minimal length parent sequences. The parent sequence of length M will be denoted $\mathbf{C} = (C_1, \ldots, C_M)$, $C_i \in \mathcal{C}$, with $M \geq \max\{K, M_1, \ldots, M_n\}$.

2.3 From the Parent Sequence to the Observed Sequences

When analyzing sequences of lithofacies, it is quite common that some facies are unobserved at one or several boreholes. In order to allow for this, each observed sequence at each site s_i is therefore a subset of a complete sequence $(\mathbf{C}, \mathbf{T}_i)$ corresponding to the parent sequence. The corresponding vector of complete thickness is $\mathbf{Z}_i$, and, in contrast to the observed ones, some thickness $Z_{j,i} = T_{j,i} - T_{j-1,i}$, $j = 1, \ldots, M$ can be equal to zero. In this case, the corresponding layer is unobserved at location s_i. When $M_i < M$, the sequence at s_i is an *incomplete sequence*, and $\mathbf{C}_i^o$ is a subsequence of $\mathbf{C}$. The complete data will be denoted $\mathbf{X} = \{(\mathbf{C}, \mathbf{Z}_i),\ i = 1, \ldots, n\}$ and $\mathbf{X}^o = \{(\mathbf{C}_i^o, \mathbf{Z}_i^o),\ i = 1, \ldots, n\}$ will denote the observed data. In the following, $O(\cdot)$ will denote the mapping such that $\mathbf{X}^o = O(\mathbf{X})$. Figure 1 illustrates a parent sequence and four different possible observed sequences.

3 Statistical Setting

3.1 Stochastic Model

The stochastic model requires a univariate model for the marginal distribution of the thicknesses and a spatial model to account for the lateral continuity of the layers. Thicknesses are modeled using positive zero inflated random variables in order to account for the many 0s resulting from incomplete observed sequences. Among many possible models, latent truncated Gaussian models (Allcroft and Glasbey 2003; Baxevani and Lennartsson 2015; Benoit et al. 2018a), also referred to as Tobit models (Liu et al. 2019) in econometrics, are flexible models that easily allow geostatistical modeling. Spatial dependence among the thicknesses belonging to a same layer is introduced by means of a truncated Gaussian random field. More precisely, for $j = 1, \ldots, M$, let $W_j(s)$, $s \in \mathcal{S}$ be a standardized Gaussian random field that, for simplicity, will be supposed stationary with covariance function $\text{cov}[W_j(s), W_j(s')] = \rho_j(s-s'; \xi_j)$, where ρ_j is a parametric correlation function and ξ_j the vector of associated parameters. The thickness field $\{Z_j(s), s \in \mathcal{S}\}$ is defined as

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

$$Z_j(s) = \varphi_j \left(W_j(s) - \tau_j\right) \quad \text{if} \quad W_j(s) > \tau_j, \tag{1}$$

and $Z_j(s) = 0$ otherwise, where τ_j is a threshold and $\varphi_j(\cdot)$ is a continuous one-to-one mapping from $\mathbb{R}_+$ to $\mathbb{R}_+$. The probability of positive thickness $\Pr(Z_j(s) > 0)$ will be denoted by p_j. With this construction, null thickness has a positive probability, since $\Pr(Z_j(s) = 0) = 1 - p_j = \Phi(\tau_j) > 0$, where $\Phi(\cdot)$ is the cumulative probability function of the standard Gaussian random variable. Parameters of the stochastic model can be expressed equivalently in terms of τ_j or p_j and in the sequel the second setting is chosen. One particular case that will be used later is to set $\varphi_j(x) = \mu_j x^{\beta_j}, x > 0$ with $\beta_j, \mu_j > 0$. When $\beta_j = 1$, one gets

$$E\left[Z_j(s)\right] = \mu_j \left(\frac{\phi(\tau_j)}{1 - \Phi(\tau_j)} - \tau_j\right), \tag{2}$$

and

$$\mathrm{Var}[Z_j(s)] = \mu_j^2 \left[1 + \frac{\phi(\tau_j)}{1 - \Phi(\tau_j)} \left(\tau_j - \frac{\phi(\tau_j)}{1 - \Phi(\tau_j)}\right)\right], \tag{3}$$

where $\phi(\cdot)$ is the density function of the standard Gaussian random variable. When β_j is not an integer, the moments of $Z_j(s)$ involve hypergeometric functions and are not reported here. From Eqs. (2) and (3), it is clear that the expectation and standard deviation of the thickness of layer j are both proportional to the parameter μ_j. The covariance function ρ_j must be smooth enough in order to generate regular thicknesses. For example, choosing that ρ_j is twice differentiable at the origin leads to a mean-squared differentiable random field W_j and, as a consequence, to a mean-squared differentiable random field for the thicknesses since φ_j is continuous and locally finite boundaries of the non null thickness sets. The depth surfaces $\{T_j(s), s \in \mathcal{S}\}$ are then obtained by adding up the thickness fields. Starting from a fixed and known ground-floor $T_0 = \{T_0(s), s \in \mathcal{S}\}$ one sets

$$T_j(s) = T_{j-1}(s) + Z_j(s) = T_0(s) + \sum_{i=1}^{j} Z_i(s), \qquad j = 1, \ldots, M.$$

Finally, the random fields W_j are assumed to be independent, since they relate to independent depositional processes.

3.2 Complete Likelihood

Since layers are assumed to be independent, the complete likelihood factorizes into a product of M likelihoods

$$L(\theta; \mathbf{X}) = \prod_{j=1}^{M} L_j(\theta_j; Z_{j,1}, \ldots, Z_{j,n}), \tag{4}$$

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

where $\theta_j = (p_j, \mu_j, \beta_j, \xi_j)$, $j = 1, \ldots, M$ and $\theta = (\theta_1, \ldots, \theta_M)$.

In the sequel $\phi_k(\cdot, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_k(\cdot, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the density and the cumulative distribution function of a k-multivariate Gaussian random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Let us consider now a layer $j \in \{1, \ldots, M\}$. For convenience, thicknesses and the corresponding locations are reordered such that the first n_j thicknesses $Z_{j,1}, \ldots, Z_{j,n_j}$ correspond to the positive values and the remaining $\ell_j = n - n_j$ ones are 0. The complete-data likelihood of the single layer j is

$$L_j(\theta_j; Z_{j,1}, \ldots, Z_{j,n}) = f_j(Z_{j,1}, \ldots, Z_{j,n_j}; \theta_j) F_j(0, \ldots, 0, |Z_{j,1}, \ldots, Z_{j,n_j}; \theta_j). \tag{5}$$

The density $f_j(Z_{j,1}, \ldots, Z_{j,n_j}; \theta_j)$ is given by

$$f(Z_{j,1}, \ldots, Z_{j,n_j}; \theta) = \phi_{n_j}(W_{j,1}, \ldots, W_{j,n_j}; \mathbf{0}, \Sigma_j) \prod_{i=1}^{n_j} J_{\varphi_j^{-1}}(Z_{j,i}), \tag{6}$$

where $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}_{n_j,n_j} = [\rho(s_i - s_k; \xi_j)]_{i,k=1,\ldots,n_j}$, $W_{j,i} = \varphi_j^{-1}\left(Z_{j,i}\right) + \tau_j$, $i = 1, \ldots, n_j$, and $J_{\varphi_j^{-1}}(Z_{j,i})$ is the Jacobian of φ_j^{-1} computed at $Z_{j,i}$. The conditional probability $F_j(0, \ldots, 0|Z_{j,1}, \ldots, Z_{j,n_j}; \theta)$ is given by

$$F_j(0, \ldots, 0|Z_{j,1}, \ldots, Z_{j,n_j}; \theta) = \Phi_{l_j}(\tau_j, \ldots, \tau_j; \mathbf{m}_j, \mathbf{V}_j), \tag{7}$$

where the mean vector $\mathbf{m}_j$ and covariance matrix $\mathbf{V}_j$ can be easily derived using the Kriging equations (Cressie 1993; Chilès and Delfiner 2012)

$$\mathbf{m}_j = \Sigma_{\ell_j,n_j} \Sigma_{n_j,n_j}^{-1} \mathbf{W}_{n_j}; \quad \mathbf{V}_j = \Sigma_{\ell_j,\ell_j} - \Sigma_{\ell_j,n_j} \Sigma_{n_j,n_j}^{-1} \Sigma_{n_j,\ell_j}, \tag{8}$$

with $\mathbf{W}_{n_j} = (W_{j,1}, \ldots, W_{j,n_j})'$ and the matrices Σ_{ℓ_j,n_j} and Σ_{ℓ_j,ℓ_j} being defined in similar ways as Σ_{n_j,n_j}. To summarize, the complete data likelihood in (4) becomes

$$\begin{aligned} L(\theta; \mathbf{X}) &= \prod_{j=1}^{M} L_j(\theta; Z_{j,1}, \ldots, Z_{j,n}) \\ &= \prod_{j=1}^{M} \phi_{n_j}(W_{j,1}, \ldots, W_{j,n}; \mathbf{0}, \Sigma_j) \\ &\quad \times \prod_{i=1}^{n_j} J_{\varphi_j^{-1}}(Z_{j,i}) \Phi_{l_j}(\tau_j, \ldots, \tau_j; \mathbf{m}_j, \mathbf{V_j}). \end{aligned} \tag{9}$$

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

In the particular case $\varphi_j(x) = \mu_j x^{\beta_j}$ that will be considered below, the Jacobian simplifies to

$$J_{\varphi_j^{-1}}(Z_{j,i}) = \frac{1}{\mu_j \beta_j} \left(\frac{Z_{j,i}}{\mu_j} \right)^{1-1/\beta_j}. \quad (10)$$

3.3 Observed Likelihood

In principle, the observed likelihood is related to the complete likelihood through

$$L(\theta; \mathbf{X}^o) = \int_{\{\mathbf{X}:\mathbf{X}^o = O(\mathbf{X})\}} L_X(\theta; \mathbf{X}) d\mathbf{X}. \quad (11)$$

However, even for moderately long parent sequence and number of 0 thicknesses, the space $\{\mathbf{X} : \mathbf{X}^o = O(\mathbf{X})\}$ is difficult to explore and the integral (11) becomes intractable. These difficulties are illustrated with two examples. At some site, let us consider an observed sequence $(\mathbf{C}^o, \mathbf{T}^o)$ and the corresponding thicknesses $\mathbf{Z}^o$. Here, the reference to the site is dropped for the sake of clearer notations. Recall that since the sequence $\mathbf{C}^o$ must be compatible with the parent sequence $\mathbf{C}$, $\mathbf{C}^o$ is obtained by deleting some layers of $\mathbf{C}$.

Table 1 shows an example of a parent sequence $\mathbf{C}$ with three categories: `Blue`, `Red` and `Green`. The observed sequence $\mathbf{C}^o$ is incomplete. Several augmented sequences $\mathbf{C}^a$ with corresponding depths $\mathbf{T}^a$ are possible. Since in the observed series the first `Blue` is followed by `Red`, the sub-sequence [`Blue-Red`] must correspond to the beginning of the parent sequence. Regarding the second occurrence of `Blue`, three cases can be distinguished: (i) it corresponds only to the third layer of $\mathbf{C}$ with 4th and 5th layers having null thickness; (ii) it corresponds only to the fifth layer, in which case the 3rd and 4th layers have null thickness; (iii) it corresponds partly to the 3rd and partly to the 5th layers. Then, only the 4th layer has 0 thickness. In this last case, an intermediate, latent, transition at depth $\tilde{T}$ with $T_2^o \leq \tilde{T} \leq T_3^o$ must be introduced. These augmented series are all possible, but some will be more likely than others, depending on the parameters of the model. In "Appendix A" an even more complex example is provided. Only some of the possible configurations are shown. They are too numerous and complex to be completely listed, even for short parent sequences.

In order to estimate the parameters of the model, a data augmentation algorithm (Tanner 1996, Ch. 5) can be exploited where the complete sequences that are compatible with the observed ones are explored. A Bayesian approach will be adopted for the inference of the parameters and a Markov Chain Monte Carlo (MCMC) algorithm will be designed in Sect. 4. But first, a simulation in which all parameters are known and all sequences are complete is shown.

3.4 Simulation

Unconditional simulation is straightforward when the transformation φ_j and the parameters $\theta_j, j = 1, \ldots, M$, are known. All that is required is to simulate M

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

Table 1 Example of a parent sequence $\mathbf{C}$ with an observed sequence $\mathbf{C}^o$ and several possible augmented sequences with corresponding transition depths and thicknesses

Parent	Observed		Possible augmented sequences								
$\mathbf{C}$	$\mathbf{C}^o$	$\mathbf{T}^o$	$\mathbf{C}^a$	$\mathbf{T}^a$	$\mathbf{Z}^a$	$\mathbf{C}^a$	$\mathbf{T}^a$	$\mathbf{Z}^a$	$\mathbf{C}^a$	$\mathbf{T}^a$	$\mathbf{Z}^a$
Blue	Blue	T_1^o	Blue	T_1^o	T_1^o	Blue	T_1^o	T_1^o	Blue	T_1^o	T_1^o
Red	Red	T_2^o	Red	T_2^o	$T_2^o - T_1^o$	Red	T_2^o	$T_2^o - T_1^o$	Red	T_2^o	$T_2^o - T_1^o$
Blue	Blue	T_3^o	Blue	T_3^o	$T_3^o - T_2^o$		T_2^o	0	Red	$\tilde{T}$	$\tilde{T} - T_2^o$
Green	–	–		T_3^o	0		T_2^o	0		$\tilde{T}$	0
Blue	–	–		T_3^o	0	Blue	T_3^o	$T_3^o - T_2^o$	Blue	T_3^o	$T_3^o - \tilde{T}$

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

Table 2 Parameters for the simulation example

	`Black`	`Red`	`Blue`	`Green`
μ, β	1	1	1	1
p	0.3	0.8	0.3	0.8
α	20	20	10	10

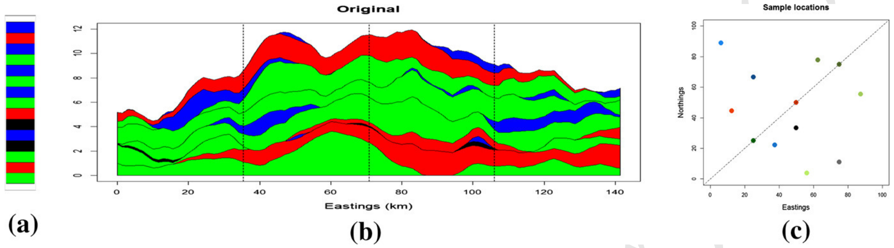


Fig. 2 Simulation experiment: **a** parent sequence of length 15, with 4 lithofacies {`Black-Red-Blue-Green`}; **b** cross-section of a two-dimensional simulation along the diagonal of $\mathcal{S} = [0, 100] \times [0, 100]$. See Table 2 for the parameters; **c** locations of the twelve boreholes

random fields W_j, $j = 1\ldots, M$ and then to apply (1) in order to transform the Gaussian process into a thickness surface. Figure 2 illustrates a cross-section of a two-dimensional simulation over $\mathcal{S} = [0, 100] \times [0, 100]$ with four lithofacies {`Black-Red-Blue-Green`} and $\varphi_j(x) = \mu_j x$, that is $\beta_j = 1$ for all categories. The parent sequence has 15 layers (see Fig. 2a) and stochastic models for layers with the same lithofacies have identical parameters. Thicknesses have been simulated using Gaussian random fields with a Matérn covariance function

$$\rho(h; \nu, \alpha, \sigma^2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{||h||}{\alpha}\right)^{\nu} K_\nu \left(\frac{||h||}{\alpha}\right), \quad h \in \mathbb{R}^2, \tag{12}$$

where $\nu > 0$ is a smoothness parameter, $\alpha > 0$ a range parameter and σ^2 the sill. Γ is the gamma function and K_ν is the modified Bessel function of the second kind of order ν. Here, the smoothness parameter has been set to $\nu = 3/2$ and $\sigma^2 = 1$, which leads to the simplified expression $\rho_j(h; \alpha_j) = (1 + ||h||/\alpha_j)\exp(-||h||/\alpha_j)$, where α_j is a range parameter. The set of the parameters in the simulation experiment is shown in Table 2.

welve synthetic boreholes have been located in $\mathcal{S}$. Three of them are placed along the diagonal at coordinates (25, 25), (50,50) and (75, 75). Nine others are randomly located (see Fig. 2c). For each category, the observed frequencies along these twelve boreholes are (0.58, 0.83, 0.28, 0.80). Notice that `Black` is highly over-represented. The average thicknesses computed along the boreholes are (0.31, 1.31, 0.62, 1.25) for each of the four categories, whilst the theoretical expectations of each category computed as per (2) are respectively (1.8, 1.1, 1.8, 1.1). Note here that `Black` and `Blue` are very unlikely to be directly stacked above each other, while it is often the case for `Red` and `Green`.

Journal: **11004** Article No.: **9875** TYPESET DISK LE CP Disp.:**2020/6/8** Pages: **29** Layout: **Small-X**

Conditional simulation is relatively easy to implement when the parameters are known and when complete sequences of thicknesses are available, including all null thicknesses. Care must be taken when simulating values from the Gaussian distribution that are below the thresholds τ_j, but otherwise the algorithm, shown in Algorithm 1, is rather straightforward. Simulations of the truncated Gaussian values are done by calling the function `rmvnorm` of the `R` package `mvtnorm` (Genz et al. 2019). The reader is referred to Chilès and Delfiner (2012) for a general exposition on unconditional simulations and conditional simulations using Kriging techniques.

Algorithm 1 Conditional simulation when all sequences and all parameters are known

Require: Data with complete sequences; transform functions φ_j, $j = 1, \ldots, M$
Require: All parameters
1: **for** $j = 1$ to M **do**
2: Compute the vector $\mathbf{W}_{n_j} = (W_{j,1}, \ldots, W_{j,n_j})$ where $W_{j,k} = \varphi_j^{-1}(Z_{j,k})$ corresponding to $Z_{j,k} > 0, k = 1, \ldots, n_j$
3: Compute $\mathbf{m}_j$ and $\mathbf{V}_j$ according to (8)
4: Draw a vector of length ℓ_j from a truncated multivariate Gaussian distribution, $\mathbf{W}_{l_j} \sim \mathcal{TN}_{\ell_j}(\mathbf{m}_j, \mathbf{V}_j; -\infty, \tau_j)$, for which each component must be below τ_j.
5: Set $\mathbf{W}_j = (\mathbf{W}_{n_j}, \mathbf{W}_{\ell_j})$
6: Simulate a Gaussian random field F_j conditionally on $\mathbf{W}_j$
7: Transform the field F_j into the thicknesses according to (1)
8: **end for**

4 Bayesian Inference with a Markov Chain Monte Carlo Algorithm

4.1 Sampling All Possible Configurations

In order to sample within all possible configurations of the augmented sequence at a given site s_i that are compatible with the parent sequence, the Markov Chain Monte Carlo (MCMC) algorithm must be able to delete a layer, to add a new layer or to displace the limit between two layers of the same category. Recall that the limit between two different categories are hard conditioning data that cannot be changed. These elementary moves, illustrated in Fig. 3, are now detailed.

Split: A state is split into two successive states of the same category. A split is only possible if it is compatible within the parent sequence. For example, in Fig. 3, the `Blue` layer at the bottom can be split into two layers since the parent sequence contains a second `Blue` layer. In Table 5 the situation in panel 4 can be obtained by splitting the state `Red`, either in panel 2 or in panel 3. When a state is split, a new transition depth, denoted t_i in Table 5, must be introduced. The thickness is split in two thicknesses accordingly.

Merge: This move is the opposite move of *Split*. Two successive states in the same category are merged together. The corresponding depth is removed and the resulting thickness is the sum of the two merged thicknesses.

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

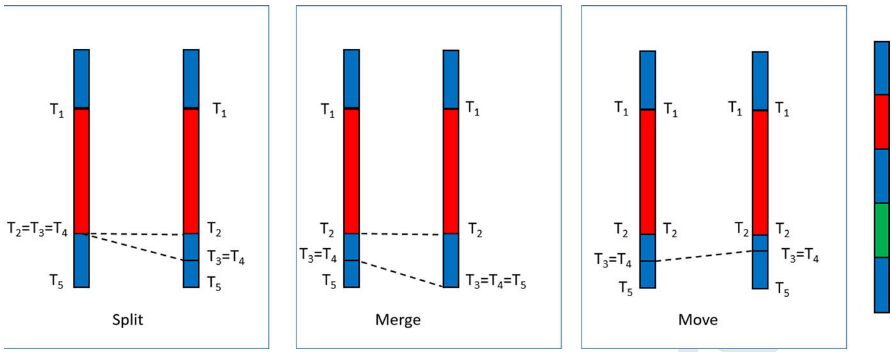


Fig. 3 Elementary moves in an incomplete observed sequence. Note that the layer `Green` is unobserved. From left to right: *Split*, *Merge* and *Displace*

Displace: Here, the augmented sequence is not changed, but the intermediate value between two successive states of the same category is changed. The corresponding thicknesses are then updated.

It is easy to verify that, starting from any initial configuration that is compatible with the parent sequence, any other configuration can be reached by combining finite numbers of *Split*, *Merge* and *Displace*. Hence, if these moves are used as building blocks of a MCMC algorithm, the resulting Markov Chain will be ergodic. At each borehole, one of the three moves is proposed with probabilities (p_S, p_M, p_D) with $p_S + p_M + p_D = 1$. If the move is possible, it is accepted according to Metropolis-Hasting acceptance ratio described in Sect. 4.3.

4.2 Choosing the Priors

Priors must be defined for all parameters of the model. For the parameters of the transform functions φ_j, $1 - p_j = \Phi(\tau_j)$ and β_j, uninformative flat priors have been chosen on the intervals (0, 1) and (0.25, 4), respectively. Regarding the covariance function, the Matérn covariance function in (12) has been chosen for its great flexibility thanks to three parameters: $\xi = (\nu, \alpha, \sigma)$, for smoothness, range and sill, respectively.

However, it is known that the joint estimation of these parameters is difficult in a Bayesian context, in particular if the number of data is small. Zhang (2004) showed that for a Matérn covariance function the only quantity that can be estimated consistently under in-fill asymptotics is $\sigma^2\alpha^{-2\nu}$. As a consequence, since the parameter μ^2 behaves as the marginal variance of the random field, using uninformative flat priors for (μ, α, ν) is expected to provide poor posterior distributions for these parameters. This was indeed confirmed on preliminary MCMC runs (results not reported here). It was thus decided to fix the smoothness parameter ν among the values (1/2, 3/2, 5/2) that would provide the highest likelihood. The above values correspond to covariance functions being the product of an exponential and a polynomial of order p with $p = 0, 1, 2$ respectively, namely $\rho(r; 1/2, \alpha, 1) = \exp(-r/\alpha)$, $\rho(r; 3/2, \alpha, 1) = (1 + r/\alpha)\exp(-r/\alpha)$ and $\rho(r; 5/2, \alpha, 1) = (1 + r/\alpha + r^2/(3\alpha^2))\exp(-r/\alpha)$.

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

Simpson et al. (2017) proposed an approach for building priors that are based on penalizing the complexity to a base model. For example, a random effect with positive variance is an extension (a more complex version) of random effect with null variance. Similarly, a random field with a finite range is an extension (a more complex version) of a random field with an infinite range. Indeed, if the range is infinite, the random field is perfectly correlated and its spatial variance is null. Penalized Complexity (PC) priors are then defined as the only priors that: (i) use the Kullback–Leibler divergence as a measure between the extended and the base models; (ii) have a penalization that increases with the distance at a constant rate.

Fuglstad et al. (2019) derived the PC priors for a Matérn covariance with parameters σ, α and ν, when ν is fixed. They showed that the joint PC prior corresponding to a base model with infinite range and zero variance when $d = 2$ is

$$\pi(\sigma, \alpha) = \lambda_\alpha \alpha^{-2} \exp\left(-\lambda_\alpha/\alpha\right) \lambda_\sigma \exp\left(-\lambda_\sigma \sigma\right), \tag{13}$$

where $\lambda_\alpha = -\ln(\epsilon_\alpha)\alpha_0$ and $\lambda_\sigma = -\ln(\epsilon_\sigma)/\sigma_0$, and the values of λ_α and λ_σ are such that $P(\alpha < \alpha_0) = \epsilon_\alpha$ and $P(\sigma > \sigma_0) = \epsilon_\sigma$. By choosing small probabilities ϵ_α and ϵ_σ, the range is lower-bounded above α_0 and the standard deviation is upper bounded at σ_0 with probability $1 - \epsilon_\alpha$ and $1 - \epsilon_\sigma$, respectively. PC priors described in (13) will be used throughout, where μ plays the role of the standard deviation as shown in Eq. (3) in Sect. 3.1.

4.3 General Description of the Algorithm

Each parameter in each category is updated iteratively in a Metropolis-within-Gibbs algorithm (Gelfand 2000). A new value is proposed according to symmetric transition kernels, for which it is equally likely to move from a current value y^c to a new value y^n than the opposite. Let θ^c and θ^n be the current and the proposed vector of parameters θ, respectively. Let further $\pi(\cdot)$ be the prior density of θ. The acceptance ratio is then

$$A(\theta^c, \theta^n) = \frac{L(\theta^n; \mathbf{X})\pi(\theta^n)}{L(\theta^c; \mathbf{X})\pi(\theta^c)}. \tag{14}$$

When sampling the configurations thanks to one of the possible moves *Split*, *Merge* and *Displace*, a new configuration $\mathbf{X}^n$ is proposed, $\mathbf{X}^c$ being the current one. In this case, the acceptance ratio is

$$A(\mathbf{X}^c, \mathbf{X}^n) = \frac{L(\theta; \mathbf{X}^n)}{L(\theta; \mathbf{X}^c)}. \tag{15}$$

The proposals are accepted if the acceptance ratios $A(\cdot, \cdot)$ are larger than one. Otherwise, they are accepted with a probability equal to the ratio. The proposal in the Metropolis-Hasting step are random walk proposals aiming at an acceptance rate above 0.5. For sampling new configurations at each borehole in turn, a possible move is drawn according to the probabilities $p_S = p_M = p_D = 1/3$. Then, it is checked

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

whether such a move is feasible within this borehole. If several moves are possible, one is selected uniformly among all possible moves in that borehole, and a new configuration is proposed. The whole procedure is summarized in Algorithm 2.

Algorithm 2 MCMC procedure

Require: Data; parent sequence; transform functions φ_j, $j = 1, \ldots, M$
Require: Initial values and priors for all parameters
Require: Number of iterations, N

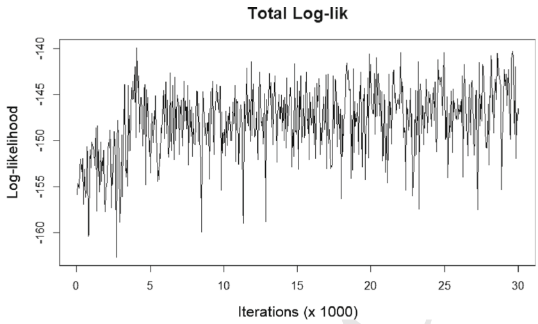
```
for i = 1 to N do
  for each parameter η ∈ {p, μ, β, α} do
    for j = 1 to M do
      Propose new η_j according to transition kernel
      Compute acceptance ratio, A using (14)
      Generate U ~ U[0, 1]; accept new η_j if (U ≤ A)
    end for
  end for
  for Borehole k = 1 to n do
    Draw a move ∈ {Split, Merge, Displace} according to the probabilities (p_S, p_M, p_D)
    Check for feasibility within borehole k
    if (move is feasible) then
      Draw uniformly one among all possible moves
      Compute acceptance ratio, A using (15)
      Generate U ~ U[0, 1]; accept the move if (U ≤ A)
    end if
  end for
end for
```

5 A Synthetic Data Example

The MCMC algorithm described above is first validated on the synthetic data-set described in Sect. 3.4 and illustrated in Fig. 2. It was coded in `R` using standard functions and our own code for the *Split*, *Merge* and *Displace* movements. Most of the running time is spent in computing the simultaneous probabilities of being below 0 in (7). This is done by calling the function `pmvnorm` of the `R` package *mvtnorm* (Genz et al. 2019; Genz and Bretz 2009). Uniform priors are used for the parameters p_j and β_j, respectively on (0, 1) and (0.25, 4), while PC priors are used for the parameters μ_j and α_j, as described in details in Sect. 4.2. Here, the setting was $\epsilon_\alpha = \epsilon_\mu = 0.01$, with $\alpha_0 = 3$ and $\mu_0 = 10$. Algorithm 2 is run for 30,000 iterations, after a burn-in period of 2,500 iterations. Values of parameters are then sampled every 50 iterations. The proposals in the Metropolis-Hasting steps follow a uniform random walk with increments in $[-0.4, 0.4]$ for μ_j and β_j, in $[-0.15, 0.15]$ for p_j and in $[-3, 3]$ for the range α_j. With these choices, the observed acceptance ratio lies between 0.43 and 0.57, depending on the parameters. This dataset being quite constrained, the acceptation ratio for exploring new configurations is only $6.78\,10^{-5}$.

Journal: **11004** Article No.: **9875** ☐ TYPESET ☐ DISK ☐ LE ☐ CP Disp.:**2020/6/8** Pages: **29** Layout: **Small-X**

Fig. 4 Complete log-likelihood as a function of iterations. The log-likelihood values are depicted every 50 iterations



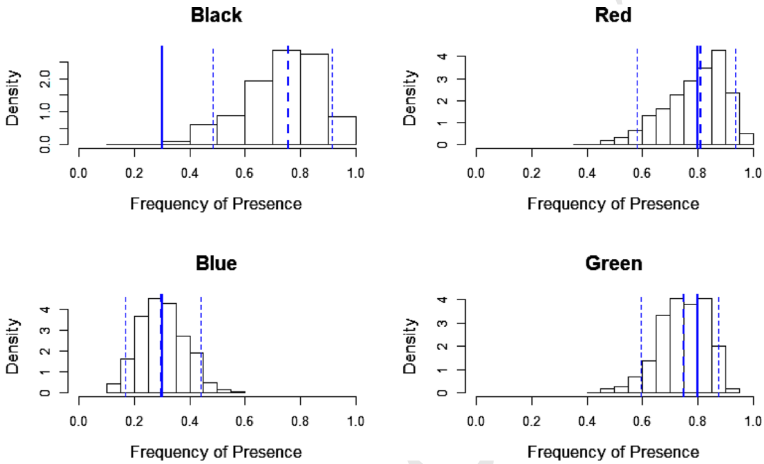


Fig. 5 Posterior histograms of the frequencies p_j. Thick continuous line: true value of the parameter. Dashed thick line: posterior median. Dashed thin lines: posterior 0.05 and 0.95 quantiles

5.1 Estimation of the Parameters

Figure 4 shows the complete log-likelihood as a function of the iterations. The mixing of the Markov chain is satisfactory and MCMC achieves convergence quite quickly. Figure 5 shows the posterior distribution of the frequency of each category. With the exception of the `Black` category, which was over-represented as already mentioned, the parameters p_j are very well estimated. Figure 6 shows the posterior cross-plot of the parameters β_j (resp. α_j) versus μ_j. One can see that there is some amount of negative correlation between β_j and μ_j, while there is some positive correlation between α_j and μ_j. These findings are quite consistent with the parametric form of the function $\varphi(x) = \mu x^\beta$ on the one hand, and with the result obtained in Zhang (2004) regarding the simultaneous estimation of the range and variance of a Matérn random field on the other hand. One can observe that the posterior median is quite close to the true value and always within the 90% posterior credibility interval, at the exception of the range parameter for the `Black` category. For this category, it should be remembered that the observed frequency was over-represented (0.58, as compared to 0.3) and that

Journal: **11004** Article No.: **9875** TYPESET DISK LE CP Disp.:**2020/6/8** Pages: **29** Layout: **Small-X**

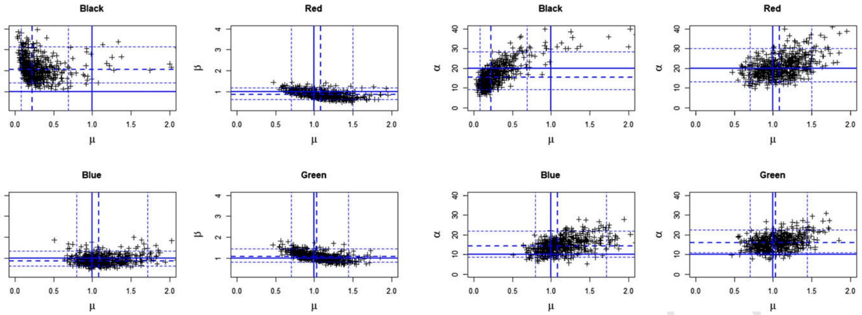


Fig. 6 Left: posterior cross-plot of β_j versus μ_j. Right: posterior cross-plot of α_j versus μ_j. Thick continuous line: true value of the parameter; dashed thick line: posterior medians; dashed thin lines: posterior 0.05 and 0.95 quantiles

the average thickness was 0.31 as compared to the theoretical expectation equal to 1.8. The maximum likelihood for the parameters (p_j, μ_j, β_j) is thus completely off the real values (0.3, 1, 1) as can also be seen on Fig. 6, where μ_j is under-estimated and β_j is over-estimated (Fig. 6). Nonetheless, given the good performances in the other categories, these results are quite promising considering that there are only 2 to 5 layers per category and that there are only 12 synthetic boreholes.

5.2 Reconstruction of the Sequences

The observed sequence is not complete on most boreholes. Augmented sequences are created during the MCMC iterations. Since they can change along the iterations, the MCMC algorithm allows us to explore different consistent reconstructions. Figure 7 shows the thickness of the 15 layers as a function of iterations for the first six synthetic boreholes. Each layer is color-coded according to its category. Similar plots were obtained for the other boreholes, but they are not shown here for the sake of concision. Firstly, it should be noted that the thicknesses do not vary often and that the variability of the thicknesses is quite different among the layers and among the boreholes. `Red` layers show constant thickness because, in the parent sequence, `Red` layers are separated by 4, respectively 6 layers (see Fig. 2). As a consequence, the conditioning makes it impossible to *Merge* or *Split* any `Red` layers. The relative low number of moves is due to the lateral correlations implied by the smoothness parameter being equal to 3/2 and the range parameter being approximately equal to 1/3 of the size of the domain. On boreholes $\sharp 1$ and $\sharp 6$, there is no `Black` layer at all. The variations are not numerous and they concern mostly the 6-layer sequence [`Green-Blue-Green-Blue-Green-Blue`] that allows some exchanges of depth through successive moves. In particular, in boreholes $\sharp 1$ and $\sharp 3$ the actual sequence is [`Green-Blue-Green-Blue`], so that some of the `Green` thickness can be exchanged between layers. Note that the total amount of `Green` thickness remains always constant. On boreholes $\sharp 2$ to $\sharp 5$, some `Black` layers are visible. The parent sequence is [`Black-Blue-Black`], but on borehole $\sharp 4$ one of the observed thickness of `Blue` is 0. As a consequence, the observed `Black` thick-

Journal: **11004** Article No.: **9875** TYPESET DISK LE CP Disp.:**2020/6/8** Pages: **29** Layout: **Small-X**

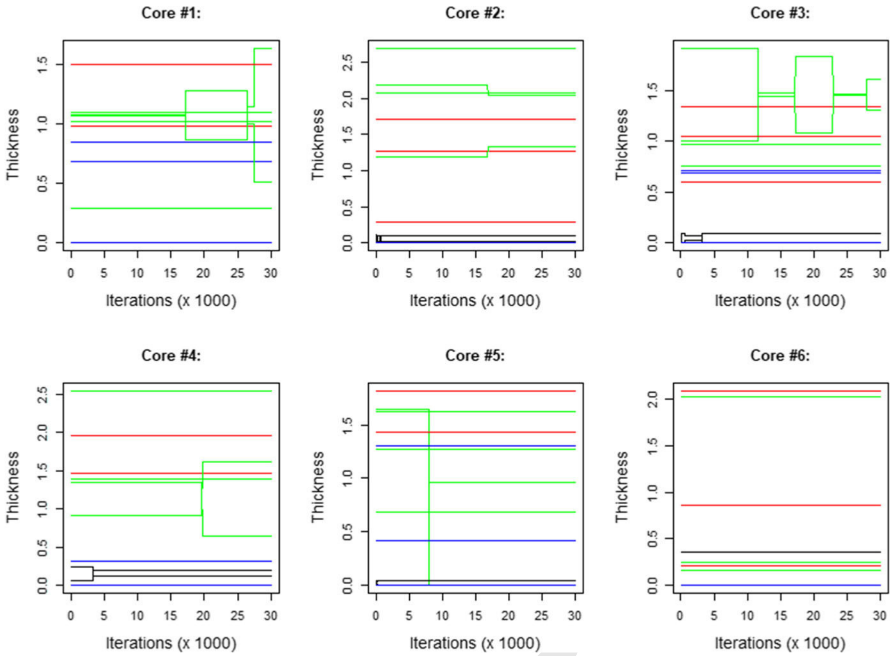


Fig. 7 Thickness of different layers in synthetic boreholes ♯1 to ♯6 as a function of iterations. Layers are represented according to the color of the category they belong to

ness can be shared between the two layers, or it can be attributed to one layer only, the other one being zero.

Figure 8 shows the thickness of layers ♯6 to ♯11 as a function of iterations for each borehole intersecting the layer. It is the dual representation of Fig. 7. Some layers have constant thickness across all boreholes, as it is the case for the `Red` layer ♯7, which intersects 9 out of the 12 boreholes. On the three others, the conditioning does not make it possible to *Merge* or *Split* the layer. In layers ♯10 and ♯12, the situation is quite the opposite. Since the total thickness must remain constant, variations on layers ♯10 and ♯12 are complementary for `Green`. These layers are part of the [`Green-Blue-Green-Blue`] sequence from layer 10 to layer 13 already mentioned. This representation offers a complementary view of the variations of this layer.

5.3 Conditional Simulations

Two ingredients are necessary in order to perform a simulation conditional on the observed data. First, one needs all observed sequences to be coherently completed in accordance with the parent sequence. Second, the simulation requires parameters for μ, β, p and α. These must be jointly sampled from the posterior distribution in a coherent way. Independent and identically distributed sets of augmented sequences and estimated parameters are accessible by sampling from independently MCMC runs after the burn-in period. Alternatively, one can sample from the same MCMC run if the number of iterations between two samples is large enough. The exact number

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

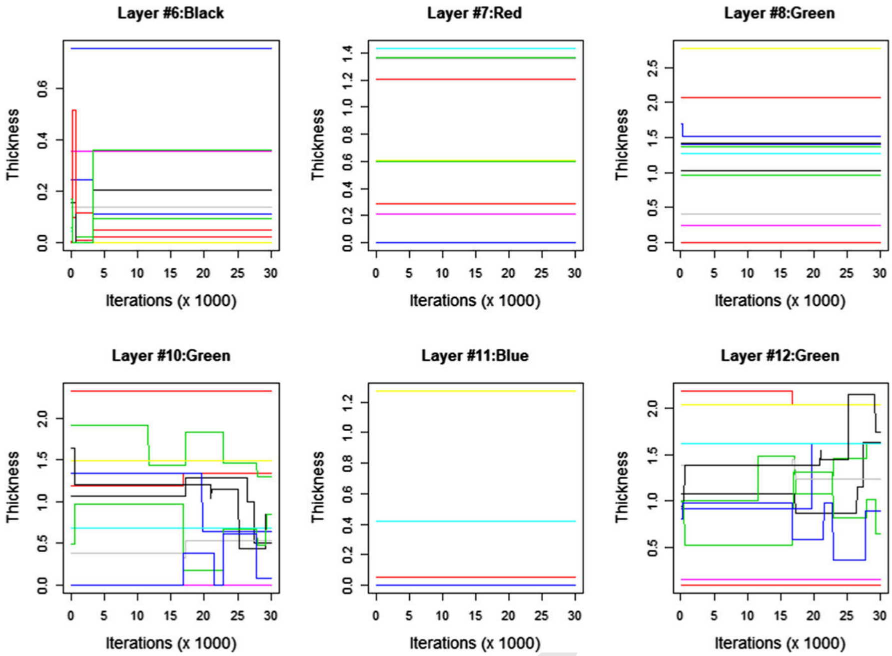


Fig. 8 Thickness of layers ♯6 to ♯12 as a function of iterations. Each borehole is represented with a different color

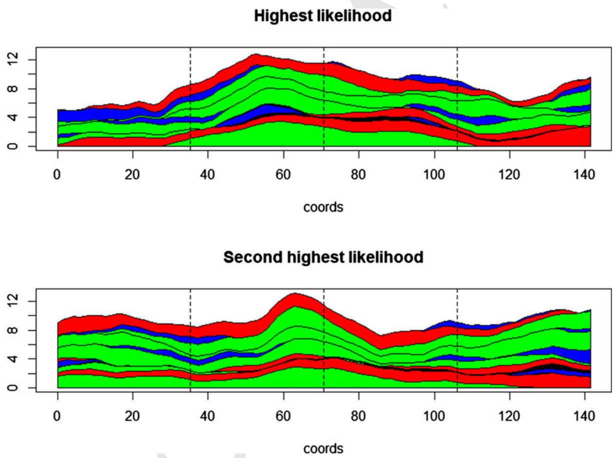


Fig. 9 Two conditional simulations. The completed sequences and the posterior parameters correspond to the most likely configuration of the MCMC run

depends on the mixing properties of the MCMC algorithm. In practice, allowing a number of iterations larger than the burn-in period is a safe enough option. The set of parameters, together with the completed sequences corresponding to the highest likelihoods recorded, have been selected for conditional simulations. They are depicted in Fig. 9. Both simulations honor perfectly the data at the boreholes (dashed vertical

Journal: **11004** Article No.: **9875** TYPESET DISK LE CP Disp.:**2020/6/8** Pages: **29** Layout: **Small-X**

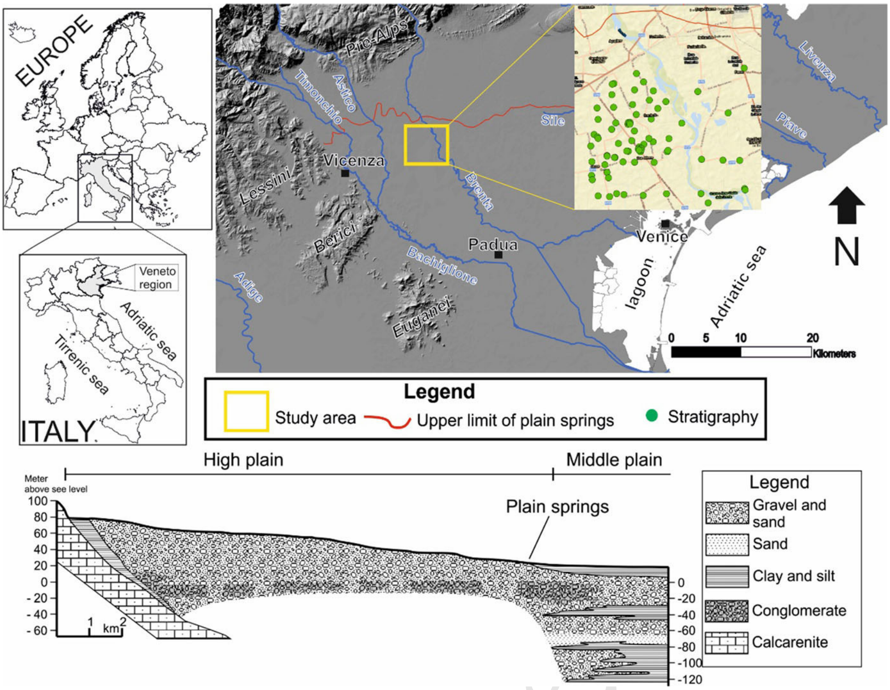


Fig. 10 The study area of the real data example and the stratigraphy

lines), but they show significantly different behaviors away from the conditioning data.

6 A Case Study: Deposition of Materials on an Aquifer

6.1 Study Area and Dataset Description

The study area (Fig. 10) is in the central part of the Venetian Plain (Italy), on the Brenta megafan (principally on the right bank of the actual Brenta River) of the Northern Padua district. In such an area, several rivers (Bacchiglione, Brenta, Astico and Timonchio) are responsible for the deposition of a significant portion of the material, hundreds of meters thick, which forms the subsoil of the Venetian Plain. Along the piedmont belt of the plain, fans from adjacent rivers laterally penetrate gravelly alluvial fans. The result is entirely gravelly subsoil throughout the thickness of the high Venetian Plain. Because deeper fans often invade further areas of the high plain from the undifferentiated gravel cover, the terminal parts of the fans extend downstream for various distances, producing an alluvial cover that is no longer uniformly gravelly, but is instead composed by alternating layers of gravel and silty clay of swampy, lagoon or marine origin (Fabbri et al. 2016).

The data-set contains 24 boreholes drilled in a 5 km × 6 km region, with a minimum distance between boreholes of 0.23 km (Fig. 11, top-left panel). Since the maximum

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

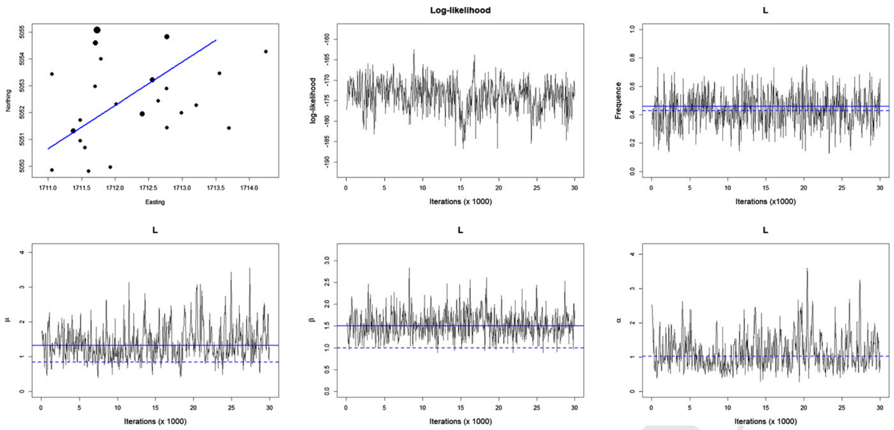


Fig. 11 Location of the 24 boreholes analyzed in the Veneto dataset (top left); diameter is proportional to the number of thicknesses recorded (from 2 to 4); thick blue line: cross-section for conditional simulation. Then, from top to bottom and from left to right: total likelihood, p, μ, β and α as a function of iterations for category L. Continuous lines: posterior medians. Dashed lines: initial values

Table 3 Empirical estimates of presence, average thickness and initial values for τ and μ

	L	S	G	A	Overall
Number of records	22	18	12	3	55
Proportion of presence, $p_j(0)$	0.46	0.75	0.25	0.13	0.38
Average thickness (in m), $\bar{T}_j$	0.73	2.25	3.89	1.10	1.94
Initial value, $\tau_j(0)$	0.10	-0.67	0.67	1.15	–
initial value, $\mu_j(0)$	0.96	2.06	6.52	2.21	–

depth of the boreholes is highly variable, a depth window between the surface (from 35 m to 40 m above sea level) and 25 m above sea level is selected. There are four categories L(imo) (Silt), S(abbia) (Sand), G(hiaia) (Gravel), A(rgilla) (Clay) and the parent sequence, containing six layers, is: [L-S-G-L-A-G]. Notice that, since there is only one layer for S and A, the associated thicknesses on the boreholes are known without ambiguity when present, which is not necessarily the case for the thicknesses associated to L and G. From two to four layers are observed on each borehole. One borehole contains an observed sequence of length 4 and five boreholes contain an observed sequence of length 3. The empirical estimates of the presence and the average thicknesses are shown in Table 3. The most observed categories are S followed by L as measured by the proportion of presence (for L, $\rho_j(0) = 22/(2 \times 24) = 0.46$. The less observed category is A, with three records only.

6.2 Model Setting

The empirical estimates are transformed into initial values for τ_j and μ_j, by setting initial values for β_j to $\beta_j(0) = 1$. Thus, for each category j

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

$$\mu_j(0) = \frac{\bar{T}_j p_j(0)}{\phi(\tau_j(0))}, \quad \text{with } \tau_j(0) = \Phi^{-1}(1 - p_j(0)).$$

Preliminary tests (not reported here) showed that the likelihood computed with a Matérn covariance function is almost always significantly larger with a smoothness parameter $\nu = 1/2$ than with $\nu = 3/2$ or $\nu = 5/2$. Therefore, the parameter ν is set to 1/2, corresponding to an exponential covariance function, even though this covariance function corresponds to continuous but non differentiable random surfaces. This point will be further discussed in Sect. 7. Initial values for the range are set to 1 km.

In this dataset, sequences are highly incomplete. As a consequence, the MCMC algorithm needs to have good mixing properties in order to explore the many possible augmented sequences that are compatible with the observations. Proposals follow a random walk with flat uninformative priors similar to that of Sect. 4 for p_j and β_j. PC priors were used for μ_j and α_j, with $\epsilon_\alpha = \epsilon_\mu = 0.01$ and $(\alpha_0, \mu_0) = (0.25, 10)$. Algorithm 2 is run for 30,000 iterations, after a burn-in period of 2,500 iterations. The values of the parameters are then sampled every 50 iterations, so that $m = 600$ posterior samples are collected. The proposals in the Metropolis-Hasting steps follow a uniform random walk with increments in $[-0.4, 0.4]$ for μ_j and β_j, in $[-0.15, 0.15]$ for p_j and in $[-0.2, 0.2]$ for the range α_j. With these choices, the acceptance ratio for the parameters was around 0.8. Although it is higher than recommended, it does not appear to have a negative impact on the estimation procedure. Instead the acceptance ratio of new thickness configurations was equal to 0.22 due to the incompleteness of this data set. Figure 11 shows the values of the parameters p, μ, β and α as a function of iterations after burn-in, for category `L`. It is quite clear that the chain is stationary with good mixing. Notice the difference between the initial values and the posterior medians. Similar results have been obtained for the other categories.

6.3 Results

6.3.1 Analysis of Thicknesses

When data belonging to the categories `L` and `G` are observed on the boreholes, the recorded thickness might belong to a single layer or to two layers. For these categories, the posterior thickness distribution might therefore look different from the observed one. Figure 12 (left) shows how thicknesses of the first layer `L` in borehole #1 vary along iterations thanks to the *Split*, *Merge* and *Displace* moves of the MCMC. On this borehole, the observed sequence is [`L-A-G`]. The measured thickness for `L` is equal to 0.4. Since the parent sequence is [`L-S-G-L-A-G`] this thickness could correspond to the first layer only (case I), to the fourth layer only (case II), or it could be shared between the two layers (case III). Figure 12 (right) represents the posterior histogram of the thickness in the first layer. Case I corresponds to 0.4, case II to 0 and case III to any value in the interval (0, 0.4). Frequencies computed along the iterations reveal that case III is the most likely case, with an estimated probability of 0.47. The probabilities of case I and case II are equal to 0.42 and 0.11, respectively. A similar analysis can be performed easily on other boreholes and categories.

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

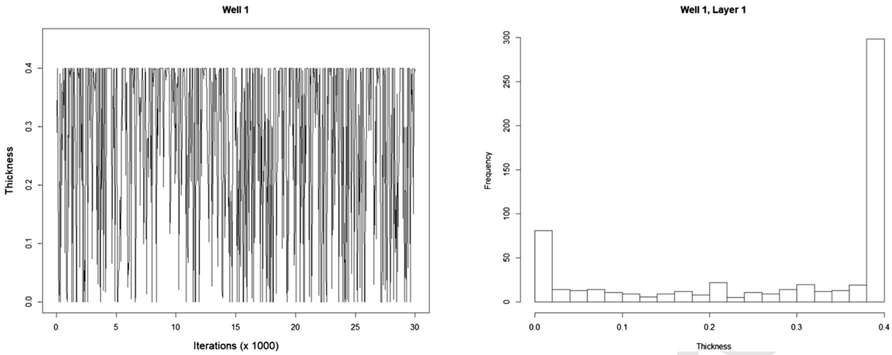


Fig. 12 Thickness of the first layer L in borehole # 1. Left: as a function of iterations. Right: posterior histogram

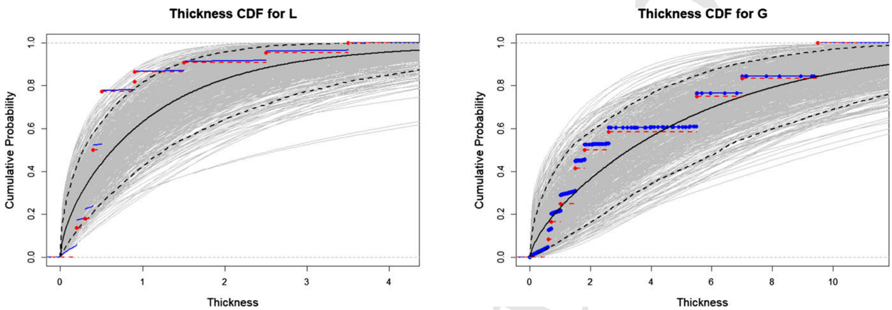


Fig. 13 Thickness cumulative distributions (TCD). In gray: MCMC samples of the posterior theoretical TCD according to (16); black continuous curve: pointwise posterior median TCD; black dashed curves: pointwise posterior 0.05 and 0.95 posterior quantiles. Red dashed curve: TCD of the original data; blue curve: TCD of the MCMC samples. Left: category L; right: category G

For a given category (for simplicity the index j is dropped), and for given parameters (p, μ, β), the theoretical thickness cumulative distribution (TCD) is

$$P(Z \leq z \mid p, \mu, \beta) = \int_{\tau}^{\tau+(z/\mu)^{1/\beta}} \frac{\phi(y)}{p} dy = \frac{\Phi\left(\tau + (z/\mu)^{1/\beta}\right) - \Phi(\tau)}{p}, \quad (16)$$

with $\Phi(\tau) = 1 - p$. The parameters are sampled every 50 iterations of the MCMC, thereby mitigating the correlation between successive samples. At each recorded iteration $k = 1, \ldots, m$, the posterior samples $p(k)$, $\mu(k)$ and $\beta(k)$ make it possible to compute a posterior theoretical TCD according to (16). Those are represented in gray on Fig. 13 for categories L and G. The ensemble of m posterior TCDs allows us to compute pointwise median and the pointwise quantiles $q_{0.05}$ and $q_{0.95}$, which are represented with black continuous and dashed lines, respectively.

Empirical posterior TCD can alternatively be computed from the thickness values recorded along the sampled iterations $k = 1, \ldots, m$. In principle, empirical and theoretical TCDs should match. Figure 13 shows the original and posterior TCDs, respectively in red and blue. Thanks to the *Split*, *Merge* and *Displace* movements, the

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

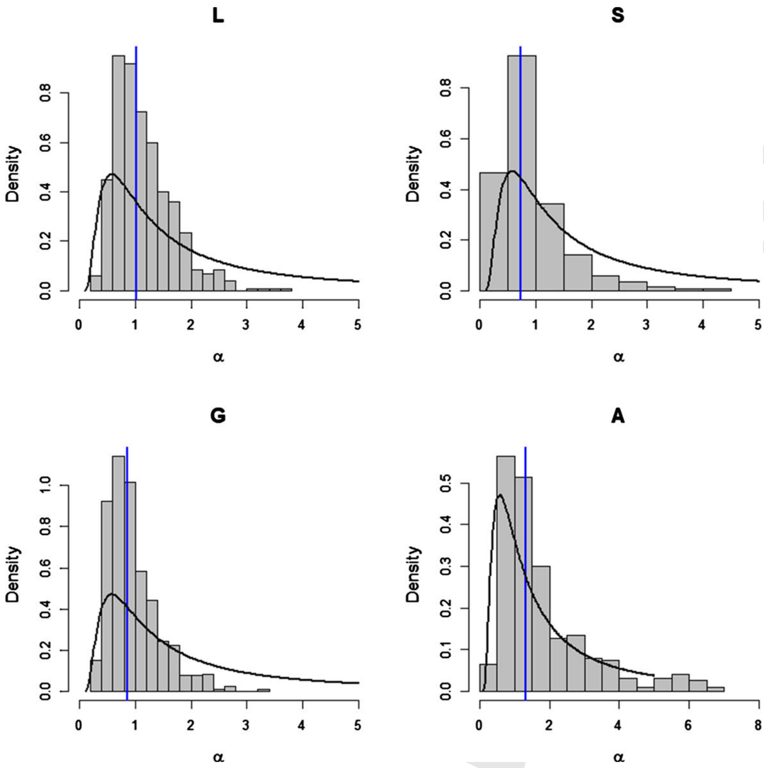


Fig. 14 For each category, posterior histogram of the spatial range and prior distribution (continuous line). Blue vertical line: posterior median

posterior TCD is slightly smoother than the original one since values intermediate to the observed ones are simulated.

Overall, the match between the empirical and the theoretical TCD is very satisfactory since the empirical curve is fully included in the envelope of the MCMC samples for category `G` and is mostly included in the envelope for category `L`.

6.3.2 Spatial Analysis and Conditional Simulation

Figure 14 shows the posterior histograms of the spatial range for the four categories, with the prior density also shown. This figure indicates that the prior has a heavy weight on the posterior distributions for each unit. However, when a category is well informed (`L` and `G`), the posterior distribution is more concentrated around the posterior median (indicated with a vertical blue line), equal to 1.03, 0.73 and 0.85 for categories `L`, `S` and `G`, respectively. On the contrary, category `A` has only three records. Since the likelihood contains very little information, the posterior distribution is very close to the prior one. The result of this analysis is that there is indeed a significant amount of spatial correlations in the random fields modeling the thickness of the layers for all categories but `A`.

Journal: **11004** Article No.: **9875** TYPESET DISK LE CP Disp.:**2020/6/8** Pages: **29** Layout: **Small-X**

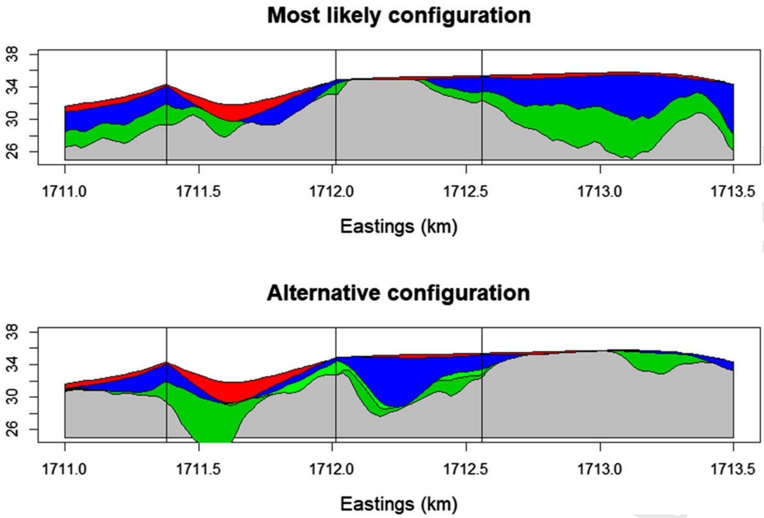


Fig. 15 Two cross-sections along the line shown in Fig. 11 (top left). Notice that there are two different layers for G in the bottom cross-section

Figure 15 shows two conditional simulations performed along the cross-section depicted in Fig. 11 (top left). This cross-section has been chosen because it is close to three conditioning boreholes (shown with black vertical lines on Fig. 15) with incomplete observed sequences that allow different thickness configurations in category G. The color code is the following: red for L, blue for S, green for G and black for A. The gray color corresponds to undefined lithofacies below the last recorded layer. The first cross-section corresponds to iteration 8,900 after burn-in, for which the likelihood was the highest along the whole MCMC (log-likelihood is equal to -162.5). Here, the G thickness is entirely in layer # 6. The second cross-section corresponds to a configuration where the G thickness is now shared between the two layers. Different shades of green have been used to distinguish the two layers. This second configuration corresponds to the most likely configuration with shared thicknesses between the two G layers (log-likelihood is equal to -171.3). Notice that it is significantly less likely than the first configuration, indicating that the data is orders of magnitude less likely with the second configuration than with the first one. Notice also that the cross-sections are quite different when moving away from the conditioning boreholes. The parameters corresponding to these two configurations are reported in Table 4.

7 Concluding Remarks

In this paper a new rule-based approach for simulating depositional sequences of surfaces conditionally to lithofacies thickness data has been presented. A distinctive feature of this approach is that it takes into account in a coherent way the different amount of information along horizontal and vertical dimensions that are usually con-

Table 4 Parameters corresponding to the two configurations shown in Fig. 15

	First configuration Log-likelihood = −162.5				Second configuration Log-likelihood = −171.3			
	L	S	G	A	L	S	G	A
p_j	0.40	0.81	0.23	0.05	0.45	0.64	0.46	0.48
μ_j	1.29	1.80	6.99	2.20	1.98	2.02	5.01	11.03
β_j	1.54	1.59	1.01	0.50	1.42	1.41	1.39	1.76
α_j	1.25	0.29	0.78	0.71	2.03	0.54	0.43	3.58

tained in borehole datasets: few cores and, consequently, few horizontal information but complete information along the depth.

This is achieved by supposing that there exists a common lithological sequence of facies that is compatible with the observed data. Moreover the sequence is supposed to be known in advance. The facies thickness, which is non-negative, is modeled by means of a truncated and transformed stationary Gaussian field. In principle, other non-negative random fields could be considered, but this choice made it possible to exploit the flexibility of Gaussian random fields in the selection of the covariance functions with different degrees of smoothness. The evaluation of the likelihood is made possible thanks to the Gaussian framework for which well known methods and efficient computing tools are available.

A data augmentation algorithm, coupled with a MCMC algorithm, is employed for learning the parameters of the stochastic model from borehole data. A very interesting feature of the proposed algorithm is that the exploration of all different configurations that are compatible with the available data is possible. Thanks to the MCMC approach and the Bayesian framework, it associates a likelihood to each of the possible realizations corresponding to a set of parameters. From those, as shown in Sect. 6.3, one can assess an empirical probability for each different configuration, select the most likely configurations and compute many other statistics of interest to the user. The algorithm requires multiple (to the order of $M \times n$) evaluations of the joint probability of a Gaussian vector being below a given threshold. The current implementation in R uses the `mvtnorm` package (Genz et al. 2019) that handles vectors with a few dozens of coordinates rather easily. It starts to slow down quite significantly around 100 coordinates and is unable to cope with more than 1,000 coordinates. Further research is thus required if the number of boreholes goes from moderate to high or very high. One possible choice could be the approximation proposed in Martinetti and Geniaux (2017), but the impact of using a less precise approximation remains to be evaluated.

A too small dataset entails difficulties in specifying the regularity and the range of the covariance function, as was shown with category A that has only three records. It was found in the present work that parameters were reasonably well estimated with 15 records per category. On the other hand, as the data set gets larger and denser (e.g. when the horizontal distance between nearest neighbor boreholes becomes a small fraction of the range parameter) the likelihood will get more peaked around local maxima, thereby decreasing the mixing of the MCMC. In this case, exploring all

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

configurations coherent with the parent sequence is likely to become more difficult. Longer chains and multiple chains starting from very different initial configurations will probably be necessary.

Several assumptions and restrictions have been made in this work, which can be lifted in order to generalize this work. The stationarity assumption, which has proved appropriate here, could be relaxed and the parameters could be easily modified to take covariates into account. Only a few half-integer values of the smoothness parameters have been considered, and the fitting of this parameter was done outside the MCMC machinery. In principle, the smoothness parameter could be different for different facies and it could be estimated in the Bayesian framework, just as any other parameter. Estimating simultaneously the three parameters of the Matérn covariance in a Bayesian context is known to be extremely difficult. When there are only few data, this was made possible thanks to the PC priors (Fuglstad et al. 2019). Currently, to the best of our knowledge, the simultaneous PC prior for (ν, α, σ^2) for Matérn covariance is unknown. Finding such PC priors is left for further research.

Currently, independent MCMCs are launched, one for every possible value $\nu \in \{1/2, 3/2, 5/2\}$. The one with the highest likelihood and the best mixing is selected and ν is fixed at that value. When analyzing the data from the Venetian plain, it was found that $\nu = 1/2$ was best, despite the fact that the associated thicknesses (and thus surfaces) are mean-square continuous but not differentiable. One could have imposed $\nu = 3/2$, but at the cost of a very short spatial range implying almost no spatial correlation. Whether one should let the data speak or impose a model for the regularity is a debate. Here, a data-driven approach was chosen.

Finally, the function that transforms the Gaussian values to thicknesses was chosen to be a power function, but any other positive function could be used.

One information that is often available in real applications and on much more points than boreholes is the nature of the facies on the surface. It is possible to incorporate such information at the cost of small changes in the method. At a given location s where this information is available, one could consider that the facies of the upper layer, say facies j, is known and has a positive thickness. The conditioning data would therefore be that $W_{\text{upper}}(x) > \tau_j$. This conditioning can easily be handled within our MCMC procedure. At this location, there would be no conditioning for the other layers.

The proposed approach depends on the existence and the knowledge of a common lithological sequence of facies that is compatible with the observed data. If the sequence is unknown, it is possible to derive it from the data, possibly by imposing some restriction, such as minimum length. This problem has not been tackled here, since it has been considered beyond the scope of this work. However it is worth mentioning that the approach presented here can be modified to account for several different parent sequences with their associated prior probabilities.

Acknowledgements This work was initiated during a visit of the first author to Ca' Foscari University of Venice. He acknowledges the support of that institution. We wish to thank two anonymous reviewers for their in-depth and detailed reading of the first version of the manuscript. Their many valuable comments helped us to improve the manuscript.

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

Appendix A: A Longer Example of Incomplete Sequence

See Table 5.

Table 5 A longer and more complex example of a parent sequence $\mathbf{C}$=[Blue-Red, Blue-Green-Blue-Red-Green-Blue] with respect to a recorded sequence $\mathbf{C}^o$ and $\mathbf{T}^o$

Parent	Recorded		Compatible augmented sequences								
$\mathbf{C}$	$\mathbf{C}^o$	$\mathbf{T}^o$	$\mathbf{C}^a$	$\mathbf{T}^a$	$\mathbf{Z}^a$	$\mathbf{C}^a$	$\mathbf{T}^a$	$\mathbf{Z}^a$	$\mathbf{C}^a$	$\mathbf{T}^a$	$\mathbf{Z}^a$
Blue	Blue	T_1^o	Blue	T_1^o	T_1^o	Blue	T_1^o	T_1^o	Blue	T_1^o	T_1^o
Red	Red	T_2^o	Red	T_2^o	$T_2^o - T_1^o$	Red	T_2^o	$T_2^o - T_1^o$		T_1^o	0
Blue	Green	T_3^o		T_2^o	0		T_2^o	0		T_1^o	0
Green	Blue	T_4^o	Green	T_3^o	$T_3^o - T_2^o$		T_2^o	0		T_1^o	0
Blue	–	–	Blue	T_4^o	$T_4^o - T_3^o$		T_2^o	0		T_1^o	0
Red	–	–		T_4^o	0		T_2^o	0	Red	T_2^o	$T_2^o - T_1^o$
Green	–	–		T_4^o	0	Green	T_3^o	$T_3^o - T_2^o$	Green	T_3^o	$T_3^o - T_2^o$
Blue	–	–		T_4^o	0	Blue	T_4^o	$T_4^o - T_3^o$	Blue	T_4^o	$T_4^o - T_3^o$
			Blue	T_1^o	T_1^o	Blue	T_1^o	T_1^o	Blue	T_1^o	T_1^o
			Red	$\tilde{T}$	$\tilde{T} - T_1^o$	Red	T_2^o	$T_2^o - T_1^o$		T_1^o	0
				$\tilde{T}$	0		T_2^o	0	Red	T_2^o	$T_2^o - T_1^o$
				$\tilde{T}$	0	Green	$\tilde{T}$	$\tilde{T} - T_2^o$	Green	T_3^o	$T_3^o - T_2^o$
				$\tilde{T}$	0		$\tilde{T}$	0	Blue	$\tilde{T}$	$\tilde{T} - T_3^o$
			Red	T_2^o	$T_2^o - \tilde{T}$		$\tilde{T}$	0		$\tilde{T}$	0
			Green	T_3^o	$T_3^o - T_2^o$	Green	T_3^o	$T_3^o - \tilde{T}$		$\tilde{T}$	0
			Blue	T_4^o	$T_4^o - T_3^o$	Blue	T_4^o	$T_4^o - T_3^o$	Blue	T_4^o	$T_4^o - \tilde{T}$
			Blue	$\tilde{T}$	$\tilde{T}$	Blue	$\tilde{T}$	$\tilde{T}$	Blue	$\tilde{T}$	$\tilde{T}$
				$\tilde{T}$	0		$\tilde{T}$	0		$\tilde{T}$	0
				$\tilde{T}$	0	Blue	T_1^o	$T_1^o - \tilde{T}$	Blue	$\tilde{\tilde{T}}$	$\tilde{\tilde{T}} - \tilde{T}$
				$\tilde{T}$	0		T_1^o	0		$\tilde{\tilde{T}}$	0
			Blue	T_1^o	$T_1^o - \tilde{T}$		T_1^o	0	Blue	T_1^o	$T_1^o - \tilde{\tilde{T}}$
			Red	T_2^o	$T_2^o - T_1^o$	Red	T_2^o	$T_2^o - T_1^o$	Red	T_2^o	$T_2^o - T_1^o$
			Green	T_3^o	$T_3^o - T_2^o$	Green	T_3^o	$T_3^o - T_2^o$	Green	T_3^o	$T_3^o - T_2^o$
			Blue	T_4^o	$T_4^o - T_3^o$	Blue	T_4^o	$T_4^o - T_3^o$	Blue	T_4^o	$T_4^o - T_3^o$

Only nine compatible augmented sequences are reported

References

Allard D, Froidevaux R, Biver P (2006) Conditional simulation of multi-type non stationary Markov object models respecting specified proportions. Math Geol 38(8):959–986

Allard D, D'Or D, Froidevaux R (2011) An efficient maximum entropy approach for categorical variable prediction. Eur J Soil Sci 62(3):381–393

Allcroft DJ, Glasbey CA (2003) A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. J R Stat Soc Ser C Appl Stat 52(4):487–498

Armstrong M, Galli A, Beucher H, Loc'h G, Renard D, Doligez B, Eschard R, Geffroy F (2011) Plurigaussian simulations in geosciences. Springer, Berlin

Baxevani A, Lennartsson J (2015) A spatiotemporal precipitation generator based on a censored latent Gaussian field. Water Resour Res 51(6):4338–4358

Benoit L, Allard D, Mariethoz G (2018a) Stochastic rainfall modeling at sub-kilometer scale. Water Resour Res 54(6):4108–4130

Benoit N, Marcotte D, Boucher A, D'Or D, Bajc A, Rezaee H (2018b) Directional hydrostratigraphic units simulation using MCP algorithm. Stoch Environ Res Risk Assess 32(5):1435–1455

Bertoncello A, Sun T, Li H, Mariethoz G, Caers J (2013) Conditioning surface-based geological models to well and thickness data. Math Geosci 45(7):873–893

Beucher H, Galli A, Le Loc'h G, Ravenne C, Group H et al (1993) Including a regional trend in reservoir modelling using the truncated Gaussian method. In: Soares (ed) Geostat Tróia'92. Kluwer, pp 555–566

Carle SF, Fogg GE (1996) Transition probability-based indicator geostatistics. Math Geol 28(4):453–476

Chilès J-P, Delfiner P (2012) Geostatistics: modeling spatial uncertainty, 2nd edn. Wiley, New York

Comunian A, Renard P, Straubhaar J (2012) 3D multiple-point statistics simulation using 2D training images. Comput Geosci 40:49–65

Comunian A, Jha SK, Giambastiani BM, Mariethoz G, Kelly BF (2014) Training images from process-imitating methods. Math Geosci 46(2):241–260

Cressie N (1993). Statistics for spatial data, revised edition. Wiley, New York

Fabbri P, Piccinini L, Marcolongo E, Pola M, Conchetto E, Zangheri P (2016) Does a change of irrigation technique impact on groundwater resources? A case study in Northeastern Italy. Environ Sci Policy 63:63–75

Fuglstad G-A, Simpson D, Lindgren F, Rue H (2019) Constructing priors that penalize the complexity of Gaussian random fields. J Am Stat Assoc 114(525):445–452

Galli A, Beucher H, Le Loc'h G, Doligez B, Group H (1994) The pros and cons of the truncated Gaussian method. Geostatistical simulations. Springer, Berlin, pp 217–233

Gelfand AE (2000) Gibbs sampling. J Am Stat Assoc 95(452):1300–1304

Genz A, Bretz F (2009) Computation of multivariate normal and t probabilities. Lecture notes in statistics, Springer, Heidelberg

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2019) mvtnorm: multivariate normal and t distributions. R package version 1.0-11

Le Blévec T, Dubrule O, John CM, Hampson GJ (2017) Modelling asymmetrical facies successions using pluri-Gaussian simulations. Geostatistics Valencia 2016. Springer, Berlin, pp 59–75

Le Blévec T, Dubrule O, John CM, Hampson GJ (2018) Geostatistical modelling of cyclic and rhythmic facies architectures. Math Geosci 50(6):609–637

Liu L, Shih Y-CT, Strawderman RL, Zhang D, Johnson BA, Chai H et al (2019) Statistical analysis of zero-inflated nonnegative continuous data: a review. Stat Sci 34(2):253–279

Marcotte D, Allard D (2018) Gibbs sampling on large lattice with GMRF. Comput Geosci 111:190–199

Mariethoz G, Caers J (2014) Multiple-point geostatistics: stochastic modeling with training images. Wiley, New York

Martinetti D, Geniaux G (2017) Approximate likelihood estimation of spatial probit models. Region Sci Urban Econ 64:30–45

Matheron G, Beucher H, De Fouquet C, Galli A, Guerillot D, Ravenne C et al (1987) Conditional simulation of the geometry of fluvio-deltaic reservoirs. In: SPE annual technical conference and exhibition. Society of Petroleum Engineers

Pyrcz MJ, Sech RP, Covault JA, Willis BJ, Sylvester Z, Sun T, Garner D (2015) Stratigraphic rule-based reservoir modeling. Bull Can Pet Geol 63(4):287–303

Sartore L, Fabbri P, Gaetan C (2016) spMC: an R-package for 3D lithological reconstructions based on spatial Markov chains. Comput Geosci 94:40–47

Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH et al (2017) Penalising model component complexity: a principled, practical approach to constructing priors. Stat Sci 32(1):1–28

Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 34(1):1–21

Syversveen AR, Omre H (1997) Conditioning of marked point processes within a Bayesian framework. Scand J Stat 24(3):341–352

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

Tanner MA (1996) Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions. Springer, Berlin
Zhang H (2004) Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J Am Stat Assoc 99(465):250–261

Journal: 11004 Article No.: 9875 TYPESET DISK LE CP Disp.:2020/6/8 Pages: 29 Layout: Small-X

Journal: 11004
Article: 9875

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details required	Author's response
1.	Please check and confirm if the authors and their respective affiliations have been correctly identified.	
2.	Please confirm if the corresponding author is correctly identified.	