A novel surface-based approach to represent aquifer heterogeneity in sedimentary formations

Ludovic Schorpp[1], **Julien Straubhaar**[1], **Philippe Renard**[1]

[1]Center for Hydrogeology and Geothermics, University of Neuchâtel

Key Points:

- A new surface-based stochastic facies modeling algorithm is presented
- It is flexible and relies on few parameters to produce a variety of geological settings
- The proposed method can efficiently reproduce observed fluvio-glacial structures

Corresponding author: Ludovic Schorpp, ludovic.schorpp@unine.ch

Abstract

Sedimentary formations that compose most aquifers are difficult to model as a result of the nature of their deposition. Their formation generally involves multiple processes (alluvial, glacial, lacustrine, etc.) that contribute to the complex organization of these deposits. Representative models can be obtained using process-based or rule-based methods. However, such methods have several drawbacks: complicated parametrization, large computing time, and challenging, if not impossible, conditioning. To address these problems, we propose a new hierarchical surface-based algorithm, named EROSim. First, a predefined number of stochastic surfaces are simulated in a given order (from older to younger). These surfaces are simulated independently but interact with each other through erosion rules. Each surface is either an erosive or a deposition surface. The deposition surfaces represent the boundaries of depositional events, whereas the erosive surfaces can remove parts of the previously simulated deposits. Finally, these surfaces delimit sedimentary regions that are filled with facies. The approach is quite simple, general, flexible, and can be conditioned to borehole data. The applicability of the method is illustrated using data from fluvio-glacial sedimentary deposits observed in the Bümberg quarry in Switzerland.

1 Introduction

Groundwater flow and solute transport processes including dispersion, mixing, or chemical reactions are heavily influenced by geological heterogeneity occurring at multiple scales (Kitanidis, 2015; Chiogna et al., 2015; Bennett et al., 2017; Soltanian et al., 2020; Wallace et al., 2021). To investigate the impact of geological heterogeneity on these processes, multiGaussian random functions are often used (Dagan, 1989; Rubin, 2003; Chiles & Delfiner, 2012; Geng et al., 2020; Zech et al., 2021) because they provide a parsimonious but flexible mathematical framework. To incorporate more geological concepts and knowledge in these analyses and to account for different types of connectivity, alternative models (from process-based to structure-imitating approaches) have also been developed for a wide range of geological environments (Koltermann & Gorelick, 1996; de Marsily et al., 2005).

In this paper, we propose a new model to represent the geological heterogeneity produced by sedimentary processes in unconsolidated fluvio-glacial environments. We focus on this geological setting because it contains the most heavily exploited groundwater resources in Switzerland (and many other countries) for drinking water supply and shallow geothermic. Being close to the surface, these aquifers are also prone to anthropogenic contamination, and describing their internal heterogeneity is important for the analysis of contaminant transport. These formations are the result of a rich and complex sedimentological history (A. Miall, 1996). Outcrop observations and geostatistical analysis show that fluvio-glacial sediments are structured in a hierarchical manner (A. Miall, 1996; Heinz et al., 2003; Ritzi et al., 2004; Bayer et al., 2011). Therefore modeling approaches aiming at studying the impact of this type of heterogeneity should also include these hierarchical relationships. One way to achieve this aim is to construct directly a hierarchical multiGaussian model (Neuman et al., 2008). To integrate more geological concepts, Scheibe & Freyberg (1995) and Ramanathan et al. (2010), use sophisticated object-based methods where sedimentary structures are created hierarchically following sedimentological rules. Webb (1994) or Pirot et al. (2015) proposed also a hierarchical approach, where multiple geomorphological surfaces are stochastically generated and stacked together to define the major units. These units are then filled with facies using a deformation process (Pirot et al., 2015) or based on an estimation of the Froude number (Webb, 1994). While these methods provide models that exhibit realistic geological features, they are difficult to constrain to field data and borehole observations. Comunian et al. (2011) or Bennett et al. (2017) decompose the problem and model a set of surfaces that delimit volumes that can then be filled with other facies simulation techniques. Zuffetti et al. (2020) describe in detail the limitations of most of the previous methods and show that they do not fully account for the stratigraphical hierarchy. Following these observations, Zuffetti et al. (2020) introduced a generic framework to overcome these limitations by defining how sub-units should be modeled into larger units at multiple scales. Based on these concepts, Schorpp et al. (2022) proposed the ArchPy approach that is capable of handling the hier-

archical relations when constructing a 3D geological model. This approach has been coupled with geophysical and hydrogeological inversion and applied successfully to synthetic data (Neven et al., 2022) and to characterize the northern area of the upper Aare fluvio-glacial aquifer in Switzerland (Neven & Renard, 2023).

In this paper, we propose to go a step further and develop the EROSim method that allows filling the stratigraphic units with detailed facies models while ensuring conditioning. The proposed method belongs to the family of surface-based methods (SBM) that emerged in the early 2000s (Pyrcz & Deutsch, 2014) with the pioneering work of Xie et al. (1999, 2001). SBMs consider that the different geological features (layers, architectural elements, facies, etc.) can be separated by surfaces (Pyrcz et al., 2015; Titus et al., 2021; Jo et al., 2020). These methods also integrate the notion of time during which geological objects are deposited. The different surfaces are stacked on top of each other and delimit the geological units or sediment types. The surfaces can either be deterministic or stochastic. Compared to pixel-based methods, SBM can maintain complex geometries throughout the modeling process.

The general idea of the proposed approach consists of generating multiple stochastic surfaces and combining them to delimit the different rock types or lithofacies. This strategy is highly flexible and allows performing conditional simulations even with a complex and realistic sedimentary structure. We show in the paper how this model is capable of simulating and extending detailed sedimentary structures that are directly observable in outcrops.

The paper is organized as follows. We first introduce the proposed simulation methodology in section 2. Section 3 illustrates the sensitivity of the method to its parameters and shows how it compares to other facies simulation techniques in simple cases. Then, we illustrate how this model can be used to represent and extend detailed information collected on outcrops in a gravel pit in the upper Aare Valley in Switzerland. The relevance of the results and the different advantages and limitations of the method are discussed in section 4.

2 A surface-based approach to represent aquifer heterogeneity

This section describes the EROSim approach and its implementation. First, the general workflow is introduced, we then present the notations and definitions that are used in the following to describe precisely each step of the method.

2.1 General principle

The principle of the method is to decompose the simulation domain into multiple regions (or volumes) using stochastic surfaces. Each region for each simulation corresponds to a single categorical value representing a lithology, a facies, or a unit, in a finite 2D or 3D domain. The three main steps of the method are therefore the following.

1. *Surface simulation*. A set of surfaces are stochastically simulated and modified according to erosion rules. These surfaces represent sedimentological events (deposition or erosion).
2. *Region delimitation*. The ensemble of surfaces simulated in step 1 forms a tessellation of the simulation domain, where each tile is individualized as a region. A graph of the spatial relationships between the regions is constructed.
3. *Facies assignment*. During this step, a facies is assigned to each of the regions defined in step 2 while accounting for the spatial continuity of the facies.

2.2 Notations and definitions

The simulation domain is denoted $\Omega \subset \mathbb{R}^n$ where $n = 2$ or 3 is its spatial dimension. We then consider a finite set of lithofacies $\mathbf{K} = \{K_1, K_2, ..., K_k\}$, where k is the number of facies to simulate. The goal is to obtain a stochastic process f that can map any location to a certain facies such that $f : \Omega \to \mathbf{K}$.

We also consider a number of ordered stochastic surfaces S_t that delimit different regions V_i stored in a set $\mathbf{V}$. Each volume can only take one value in $\mathbf{K}$, i.e. f is constant on each volume V_i. The subscript $t \in \mathbb{N}$ in S_t represents the simulation time step and can be seen as analogous to geological time. These surfaces are stochastic processes of dimension $n-1$, which means, for example, that if the simulation domain is in 3D ($n=3$), S_t are 2D stochastic surfaces. The regions V_i are objects of dimension n each corresponding to a single connected component, they do not intersect each other. Furthermore, the ensemble of all the V_i fills Ω. We can also refer to the V_i as areas if $n=2$ or volumes if $n=3$. The boundaries S_t are ordered by age (from younger to older) in a list $\mathbf{S}$. The order of the surfaces is important because it represents the sedimentological history of the simulation domain and has consequences on the interactions that can exist between the surfaces through erosion rules.

An important aspect of stochastic geological models is their ability to be conditioned by borehole data. In this study, each borehole is assumed to be 1D and vertical. The list of boreholes is denoted $\mathbf{B}$. A borehole contains a sequence of contiguous intervals. For a simulation in 2D (resp. 3D), a borehole is located by a position $\mathbf{x} = (x)$ (resp. $\mathbf{x} = (x,y)$). The facies encountered along the borehole are defined with a sequence of elevations $z_1 < \cdots < z_m$ and a sequence of facies $k_1, \ldots k_{m-1} \in \mathbf{K}$. The i-th interval between the bottom and top elevations z_i and z_{i+1}, is filled with the facies k_i, i.e. $f(\mathbf{x},z) = k_i$ if $z_i \leq z < z_{i+1}$ (and $(\mathbf{x},z) \in \Omega$), for $i = 1, \ldots, m$. Hence, a borehole B is expressed as $B = (\mathbf{x}, \{z_1, \ldots, z_m\}, \{k_1, \ldots, k_{m-1}\})$ or as $B = (\mathbf{x}, ([z_1, z_2[, k_1), \ldots, ([z_{m-1}, z_m[, k_{m-1}))$.

2.3 Unconditional simulation

The unconditional simulation algorithm is summarized in algorithm 1. The steps are described in detail below.

Algorithm 1 Unconditional algorithm

Require: Parameters

N : integer - number of simulated depositional surfaces
$\gamma_0, \ldots, \gamma_{N-1}$: covariance (or variogram) models for each surface
$\mu_0, \ldots, \mu_{N-1}$: mean elevations for each surface
ξ : in $[0,1]$ - proportion of eroding surfaces
p_{global} : target proportions of the facies, over the whole domain
α : in $[0,1]$ - clustering parameter

1: Order all the surfaces in $\mathbf{S}$ by their means (μ_i)
2: Set surface index $t = 0$
3: **while** $t < N$ **do**
4: Determine if S_t is erode or onlap, given ξ
5: Unconditional simulation
6: **if** S_t is *onlap* **then**
7: $t = t + 1$
8: Apply erosion-deposition Rules (EDR, Eq. 3 - 4)
9: Define regions V_i as described in section 2.3.3
10: Assign facies to V_i using Algorithm 2 (depending on p_{global}, α)

2.3.1 Surface ordering

Assuming that the parameters are defined, and before generating the surfaces, the first step (line 1) consists of checking the input parameters and ordering the surfaces by their mean elevation (from low to high).

This sorting is used to represent the evolution of geological time, with each event (a surface) occurring one after the other. Proceeding that way, we assume that the geological processes gradually increase upward due to the gradual accumulation of the sediments in the system.

2.3.2 Surface simulation

The second step consists of simulating N depositional surfaces through the domain Ω. Each time a new surface is simulated, from oldest to youngest, erosion-deposition rules (EDR) are applied. We consider here only the depositional surfaces because the erosion events are assumed to deposit no sediments, as explained below.

Let S_t be the simulated surface at time t, before applying the EDR. In the following examples, S_t is modeled as a Gaussian Random Field (GRF) following a specified mean μ_t and a specified covariance (or variogram) model γ_t. But, the mathematical model used to simulate the surfaces could be different. We use GRFs for convenience and simplicity. GRFs are easy to simulate (Chiles & Delfiner, 2012) and to constrain (useful for the conditional case). They are also flexible as they can handle non-stationary mean or covariance parameters. To generate our GRFs, we used the Geone[1] python library that provides a set of common geostatistical, Multiple Point Statistics modeling and image analysis tools.

In the following, for illustration purposes, we generally consider that all the surfaces follow the same covariance model (all γ_i are identical) but differ from each other by their mean. Moreover, the stochastic processes are assumed stationary, i.e. the mean and covariance are constant spatially.

Every surface S_t can be expressed as a function $S_t = S_t(\mathbf{x})$, defined for spatial locations $\mathbf{x} \in \mathbb{R}^{n-1}$. Let us then denote by S_t^* the surface at time t modified by the application of the EDR.

At time $t = 0$, the surface S_0^* is initialized as

$$S_0^*(\mathbf{x}) \longleftarrow S_0(\mathbf{x}). \tag{1}$$

For the following time steps ($t > 0$), we first determine if the event is an erosional event or a depositional event. The decision is randomized based on the probability ξ given by the user. This probability represents the fraction of erosive events among all the geological events (deposition and erosion) that are simulated. Then we generate a new surface.

$$S_t = \mathrm{GRF}(\mu_t, \gamma_t). \tag{2}$$

If the event is depositional, we apply the depositional rule:

$$S_t^*(\mathbf{x}) \longleftarrow \max(S_t(\mathbf{x}),\, S_{t-1}^*(\mathbf{x})). \tag{3}$$

and increment the value of t for the next iteration. For depositional events (Eq. 3), if the simulated surface is below a part of a previously simulated surface, there is no deposition and the elevation at that location is equal to the latter.

Otherwise, if the event is erosive, the computed surface S_t erodes the previously deposited formations.

$$S_k^*(\mathbf{x}) \longleftarrow \min(S_t(\mathbf{x}),\, S_k^*(\mathbf{x})), \quad \text{for } k = 0, \ldots, t-1 \tag{4}$$

All previously simulated surfaces that are above the simulated erosional surfaces are updated (Eq. 4). The time t is not incremented in that situation.

A schematic representation of these rules is shown in Figure 1A-B. In this example, the first three surfaces are depositional and the last (S_3) is erosive. The surfaces S_0, S_1, and S_2 are first simulated and adjusted according to the first two equations (Eqs. 1 and 3), and the surface S_3 erodes them where S_3 is simulated below the others (Eq. 4).

[1] `http://www.github.com/randlab/geone`

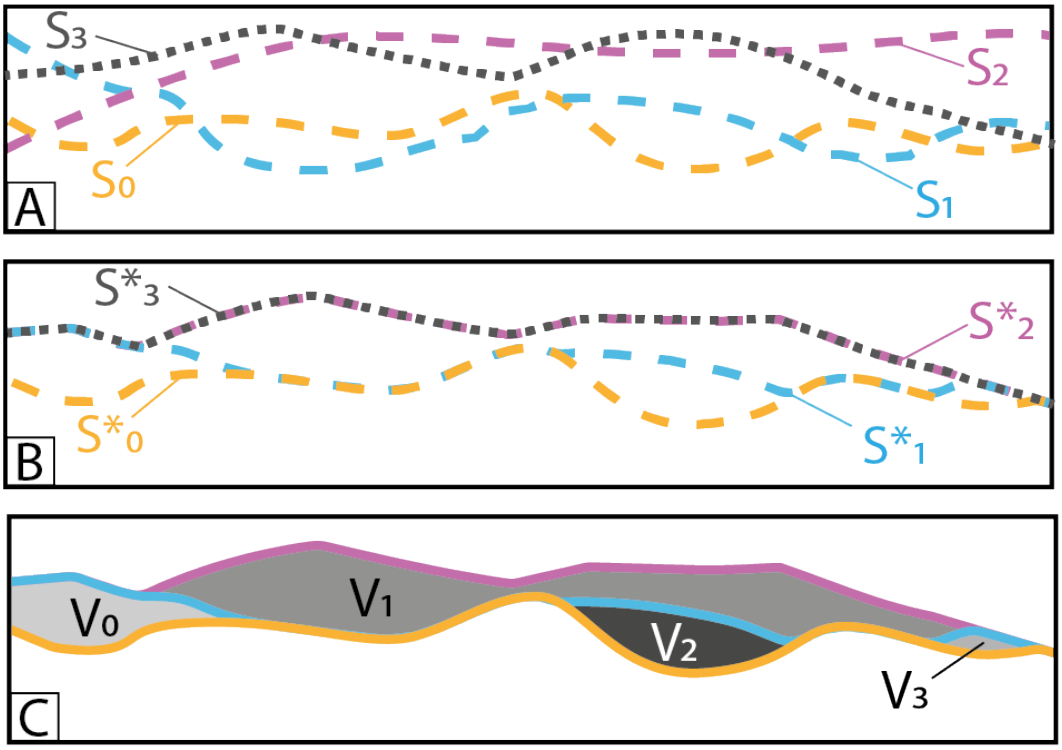

Figure 1. Schematic representation of the delimitation procedure for the regions. **A**: four surfaces have been simulated: yellow, blue, pink, and grey. The last one (grey) is eroding while the others are onlap. **B**: Situation after the application of the deposition and erosion rules. **C**: distinct regions are differentiated.

2.3.3 Region delimitation

Once the surfaces have been simulated and the EDR applied (Algorithm 1, lines 1 to 10), the simulation domain Ω is divided into distinct regions. All of these regions V_i are defined by exactly two successive (in simulation time) onlap surfaces (two depositional events), one delimiting its top and the other its base. A region can only exist where its top surface is strictly above its bottom surface. It does not exist where the two surfaces are exactly at the same elevation. This implies that regions are delimited on the sides when the top and bottom surfaces meet. If multiple distinct regions occur between two surfaces, they will be treated as completely different and independent regions. This means that later in the simulation different facies may be assigned to these regions, even if they are defined by the same surfaces. This is in contrast to previous studies, where each simulated surface delimits one sedimentological entity (Webb 1994; Pirot et al. 2015; Pyrcz et al. 2005). In practice, the individual regions are identified as the connected components of the sets $\{(\mathbf{x},z) \in \Omega : S_t^*(\mathbf{x}) \leq z < S_{t+1}^*(\mathbf{x})\}$, $t = 0,\dots,N-1$, where N is the total number of onlap surfaces simulated. Finally, two supplementary regions, made up of all points of Ω respectively below the surface S_0^* and above the last surface S_{N-1}^*, are added, to cover the entire simulation domain. Alternatively, Ω can be reduced to the domain between S_0^* and S_{N-1}^*. An example of region delimitation in 2D is illustrated in Figure 1C.

2.3.4 Facies attribution

The aim is to assign a facies to each of the regions V_i determined in the previous step. To do so, we propose a simple algorithm that is summarized in Algorithm 2.

Algorithm 2 Graph-based indicator simulation

Require: Parameters
p_{global} : global proportions of the facies, over the whole domain
α : in $[0,1]$ - clustering parameter
$\mathbf{V}$: set of regions covering the whole domain
G : a spatial graph G representing the connections of the regions

1: $V_i \leftarrow$ select a random region in $\mathbf{V}$
2: $p_{target} \leftarrow$ compute target proportions using Eq. 6
3: **if** $\alpha < 1$ **then**
4: $\quad \mathbf{V}_{\mathscr{J}} \leftarrow$ get the neighbours of V_i using G
5: $\quad p_{neig} \leftarrow$ compute the local proportions using Eq. 7
6: **else**
7: $\quad p_{neig} = 0$ (unused)
8: $p_{V_i} \leftarrow$ compute facies probabilities for V_i using Eq. 8
9: Draw a facies in K according to p_{V_i} and assign it to V_i
10: Update the global current proportions p_{cur}
11: Go to 1 until all regions are filled

We are assuming that based on borehole or outcrop data, the user can provide an estimate of the target global facies proportions p_{global} in the study area, e.g. 50 % of sand, 20 % of gravel, and 30 % of clay. In addition, we want to provide a simple parametrization allowing the user to control the spatial continuity of the facies. For this, we introduce a clustering parameter α that the user can adjust.

The attribution then follows a simple method with adaptive target proportions to respect as well as possible the global facies proportions and spatial continuity. The idea is to sequentially populate the regions with random facies according to a probability distribution adapted for each region. At each step, the target facies probabilities are first computed accounting for the global target proportions p_{global}, the current proportions p_{cur} over the already simulated regions, and also the local proportions derived from the surrounding regions.

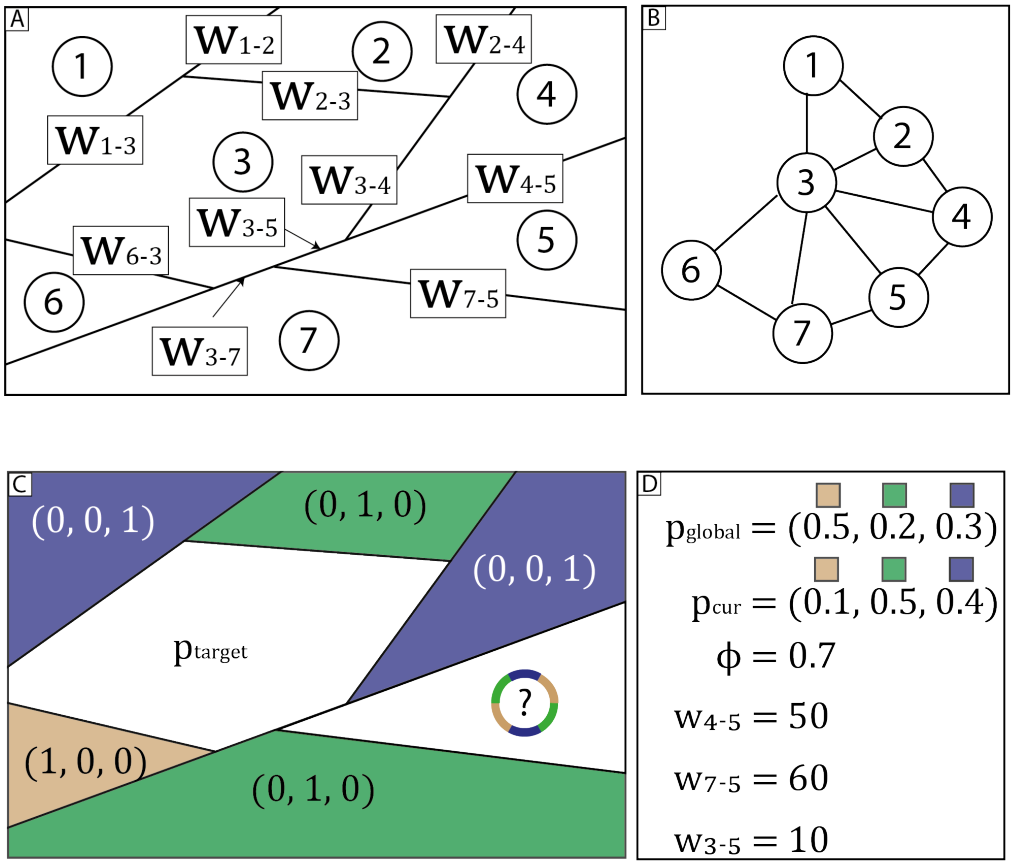

Figure 2. Conceptual representation of a 2D simulation domain separated in seven regions (**A**) and its associated spatial graph (**B**). The w_{i-j} correspond to the length of contact between V_i and V_j, they are used as weights for the edges in the graph. **C** represents a situation during the attribution of the facies (three in total) for volume 5. The facies proportions are written in each region (note that region 3 is not defined at this point of the simulation and the modified global proportion p_{target} is used). **D** details the variables required to compute the probability mass used to draw a facies in region 5 (see text for details).

First, target proportions p_{target} are calculated such that

$$\phi\, p_{cur} + (1-\phi) p_{target} = p_{global}, \tag{5}$$

where ϕ is the ratio of the total volume of the already simulated regions over the volume of Ω. Equation (5) means that if the entire non-simulated volume was filled with proportions p_{target}, the final proportion over Ω would be p_{global}. The target proportions are explicitly expressed as

$$p_{target} = \frac{p_{global} - \phi\, p_{cur}}{(1-\phi)}. \tag{6}$$

If p_{target} for a certain facies becomes negative, it is set to 0 and the proportions for the other facies are rescaled so that the sum of p_{target} is equal to 1. This situation can occur when a current proportion is too high ($\phi\, p_{cur}(i)$ exceeds $p_{global}(i)$ for some component i). It is important to understand that p_{target} is computed before the facies attribution of each region, and then it can vary since it depends on p_{cur} and ϕ. Therefore, the global target proportions for the unsimulated part of the domain are continuously corrected until the simulation is finished.

Secondly, to introduce some spatial continuity, the local proportions p_{neigh} are computed from the regions around the simulated one. This is done by using a spatial graph G describing the regions and their contacts. Each vertex in G refers to a region, and an edge links two vertices if the two corresponding regions are in contact (adjacent). The size of the contact (length if $n = 2$ or area if $n = 3$) is the weight attached to the edge. An example in 2D is shown in Figure 2A-B. Note that the graph is constructed prior to the facies assignment procedure.

We consider V_i the region in which a facies has to be attributed (currently simulated region) and $\mathscr{J}$ the set of indices j of the regions adjacent to V_i, i.e. $j \in \mathscr{J}$ when the vertices corresponding to V_i and V_j are linked in G by an edge, of weight w_{ij}. Then, the local facies probabilities p_{neig} for V_i are defined based on the neighbors V_j as follows:

$$p_{neig} = \frac{\sum_{j\in\mathscr{J}} w_{ij} \cdot p_{V_j}}{\sum_{j\in\mathscr{J}} w_{ij}} \tag{7}$$

where p_{V_j} is the proportions of facies in the neighbor region V_j. The vector p_{V_j} is of length k where k is the number of facies and it is defined as follows. If a facies K_l has already been assigned in V_j, then all the components of p_{V_j} are set to 0 except the l-th component that is set to 1. If V_j is a region not yet simulated (with no assigned facies), then p_{V_j} is set to the target proportions p_{target} computed above.

Finally, the probability distribution, p_{V_i}, used to draw a facies in the region V_i is obtained by combining probabilities p_{target} and p_{neigh} (see equations 5 and 6) via a log-linear pooling operator (Allard et al., 2012):

$$p_{V_i} = (p_{target})^{\alpha} (p_{neig})^{1-\alpha} \tag{8}$$

where α is a user-defined parameter ranging from 0 to 1 that controls the clustering of the regions of identical facies. If $\alpha = 0$, the same facies are more likely to be adjacent and, in contrast, if $\alpha = 1$, the facies are drawn only according to p_{target} (p_{neigh} is ignored).

The main advantage of this method is that it is simple and requires only the global target proportion p_{global} and the parameter α as input. Note that the graph-based approach presented here is applied to a domain divided into tiles and then is independent of the first part of the EROSim methodology. More complex methods could be considered, such as those using rules to guide the position of the facies (for example, by constraining some facies to appear more frequently in small regions).

To illustrate the procedure, let us consider the 2D case shown in Figure 2C, based on the domain delimitation in Figure 2A. Here the facies in V_5 have to be drawn from p_{V_5} using the equation 8. First, we need to calculate p_{target} which depends on p_{cur} and ϕ (Eq. 6). We obtain $p_{target} = (\frac{43}{30}, -\frac{1}{2}, \frac{1}{15})$ which after setting the second probability to 0 and rescaling becomes:

$p_{target} = (\frac{43}{45}, 0, \frac{2}{45}) \approx (0.96, 0, 0.04)$. Then we need to compute p_{neig}, which depends on the facies in the regions surrounding V_5, which are V_3, V_4, V_7 according to the graph. Applying Eq. 7 with values in Figure 2 gives: $p_{neig} = (0.08, 0.50, 0.42)$ which is very different from p_{target}. We see that the most probable facies according to neighbors are, in fact, the green and blue facies, as they share the longest contact length with V_5. We can also note that since the facies proportion is undefined in region 3, p_{target} is taken for that region. Finally, we can apply Eq. 8 and obtain p_{V_5} that can range from p_{neig} to p_{target}, depending on the α parameter chosen. In this particular case, α has a strong impact as if it is close to 0, priority is given to the neighbors, and the beige facies has poor chances of being drawn. In contrast, if α is close to 1, priority is given to the adjusted global proportions, which gives a great chance for the beige facies to be drawn. This is due to the over-representation of green and blue facies in Figure 2C compared to the global target proportions p^0_{global}.

2.4 Conditional algorithm

We now consider the problem of conditioning the simulations. The conditioning data are facies intervals in boreholes as defined in Section 2.2. Considering any borehole,

$$B_i = (\mathbf{x}_i, \{z_1, \ldots, z_m\}, \{k_1, \ldots, k_{m-1}\}), \tag{9}$$

conditioning consists of ensuring that at least one surface must pass through each of the geological interfaces $(\mathbf{x}_i; z_j)$ observed in each borehole. Then, to condition the regions V_i with the proper facies, it is sufficient to ensure that no region covers two intervals having different facies.

The method that we propose to ensure the conditioning is described in detail in Algorithm 3. The method is direct and requires no iteration, but it requires checking at each step that it does not create situations that will lead to a violation of the conditioning data. This is the reason why the algorithm is complex. In the following, we explain the main principles of the algorithm.

The surfaces are simulated from the bottom to the top. Surfaces can be onlap deposits or erosive as for the unconditional algorithms and they define regions. Because of their erosional capabilities, they can remove or cut parts of the regions that have been defined by previous depositions. The general aim of the algorithm is to ensure that in the end, no region covers two intervals having different facies in the boreholes.

This problem can happen in two situations: if the newly simulated region covers two (or more) intervals with different facies within the *same borehole*, or if the new region covers two (or more) intervals in *different boreholes*.

The first situation is avoided by forcing at least one surface to pass through each of the facies interface in each borehole as indicated above.

The second situation is more complex. It requires analyzing the position of the regions and intervals encountered in the different boreholes. As some surfaces have the potential to remove older ones by erosion, some situations that are compatible with the conditioning data at a certain time step may become incompatible with the data later on. The inverse is also possible. To solve this problem, the conditioning algorithm proceeds by generating the surfaces sequentially from oldest to youngest and uses inequality data (upper or lower bounds) to constrain the simulation. Three rules are defined and implemented in the algorithm:

- R1 For each interface point along each borehole with a facies transition, attribute a surface in $\mathbf{S}$ that has to go through that point; store this information in a data structure (a python dictionary) referred to as d_s. Note that one surface $\mathbf{S}$ can be attributed to one or several interfaces (in one or several boreholes).
- R2 When a facies interval is constrained (attributed to a region), we consider it untouchable and no *future* changes can alter it. This is done by imposing lower bounds (LB) along the borehole when simulating future erosive surfaces.
- R3 Onlap surfaces have to avoid creating (temporary) regions overlapping two different facies intervals to prevent the apparition of conditioning errors in the following steps.

Algorithm 3 Conditional algorithm

Require: Same parameters as the Algorithm 1
Require: List of boreholes **B**
Order all the surfaces in **S** by their means
$d_s \leftarrow$ attribute a surface to each borehole facies transition and store it
Set surface index $t = 0$
$LB \leftarrow$ initialize lists for lower bounds conditioning
while $t < N$ **do** ▷ loop over the surfaces
    if S_t in d_s **then** ▷ If the surface is attributed
        $EP, UB \leftarrow$ Initialize lists for conditioning points (equality points and upper bounds)
        $B_i \leftarrow$ get borehole(s) and facies interval(s) associated with S_t
        Check situation at B_i according to Figure 3
        if Situation 1 **then**
            Set S_t to *onlap*
            $EP \leftarrow$ set an equality point to top of facies interval(s)
            $UB \leftarrow$ set upper bounds at other near boreholes locations to prevent connecting non-identical facies (given Figure 3A)
        else if Situation 2 **then**
            Set S_t to *erode*
            $EP \leftarrow$ set an equality point to top of facies interval
        Compute one conditional surface with conditioning points EP, UB and LB
        Add EP to list of LB
        Remove facies interval of d_s
    else
        Determine if S_t is *erode* or *onlap*, given ξ
        if S_t is *onlap* **then**
            Set LB and UB to prevent connecting two different facies as shown in Figure 3C
            Conditional simulation with LB and UB
        else
            Conditional simulation with only LB
    if S_t is $onlap$ **then**
        $t = t + 1$ ▷ Increment time of deposition
    Apply Erosion Rules (Eq. 3 - 4)
Define regions V_i as described in section 2.3.3
Assign facies to V_i using Algorithm 2

To start, we generate N values within a uniform distribution between the minimum and maximum altitudes of the zone of interest. These values are ranked (step 1, in Algo. 3) from minimum to maximum and will correspond to the mean altitudes of the N surfaces delimiting the regions.

Following rule R1, one surface is attributed to each interface (step 2, in Algo. 3): for a given interface at elevation z, a surface S_t in $\mathbf{S}$ is chosen randomly according to a probability computed with a Gaussian distribution $\varphi_{\mu_t,\sigma_t^2}(z)$ of mean μ_t and variance σ_t^2, with μ_t being the prescribed mean elevation of S_t and σ_t^2 the prescribed variance of S_t. $\varphi_{\mu_t,\sigma_t^2}(z)$ is the probability that the surface S_t takes the value z at any location. Using this approach ensures that there is a reasonable chance that these transitions will be represented by a surface that is likely to be present at that depth. In addition, surface allocation ensures that surfaces associated with intervals in a borehole are of increasing index. This is to avoid the unrealistic situation where a younger interval is constrained before an older one. To do this, surface indices are randomly drawn, as described above until they are higher than the index previously drawn for the lower interval. At the end of the process, only a subset of the surfaces S_t are attributed to the interfaces observed in the boreholes. A surface can be assigned to several intervals of identical facies (below the interface). This is forbidden if the facies are different because there is a high risk of creating a region that would connect them.

The surfaces are then simulated successively.

If the current surface is attributed to an interface, several situations can arise. Figure 3 shows some of these situations. In the figure, the surfaces that have been simulated previously are represented in black and the intervals of the borehole data that have already been constrained by the conditioning algorithm and that should therefore not be perturbed anymore are highlighted with a red rectangle (rule R2). The new surface to be simulated is represented in green. In situations 1 and 2, the green rectangle highlights the current interval with a given facies that needs to be constrained during that iteration.

In situation 1 (Fig. 3A), an onlap surface has to be considered. The current simulation elevation (last surface elevation, in black) at the well location (left well) is below the attributed interface. The green circle represents the position of the interface attributed to that surface (during step 2). It is a conditioning data, or equality point (EP) for the GRF simulation of this surface (step 12, in Algo. 3). But the facies located above the already constrained interval in the second well (right well) is blue and therefore different from the facies in the left well (brown). To avoid creating a region that would connect these different facies, we impose in the right well an inequality (upper bound or UB) for the simulation of this surface (step 13, in Algo. 3). We then use a Gibbs sampler (Freulon & de Fouquet, 1993) to constrain the conditional GRF simulation (step 17). After applying the erosion-deposition rules (step 29), the new region will cover only the brown interval in the left well in that case.

In situation 2 (Fig. 3B), the current simulation elevation at the left well location is above the attributed interface. In this situation, an erosive surface has to be considered to ensure the removal of the previously simulated deposits. We add an EP at the interface (step 16). Then, to avoid breaking an interval that has already been constrained in other wells, we impose an LB constraint on the top of the intervals that were previously constrained (see right well with red rectangle in Fig. 3B). We then use the GRF simulation technique with these new constraints to generate the conditional simulation and apply the erosion-deposition rules as described above (steps 17 and 29).

The interval being now constrained, we can add the EP data to LB to prevent them being eroded by subsequent erosive surfaces (step 18) and interface is no longer attributed (step 19).

If the current surface is not attributed to an interface, We decide with a probability ξ if it is onlap or erosive (Step 21 in Algo. 3). To prevent to connect different facies intervals, we add several constraints to onlap surface, as illustrated in situation 3 (Fig. 3C). In this example the surface is forced to pass in the blue interval in the right borehole while avoiding the brown interval in the left borehole. The rest of the procedure is the same as before. Practically, this is done by randomly choosing a facies interval among the next to be constrained in each borehole and forcing the surface to pass through it using UB and LB, while avoiding other nonidentical facies intervals by applying UB. This prevents the connection of different facies.

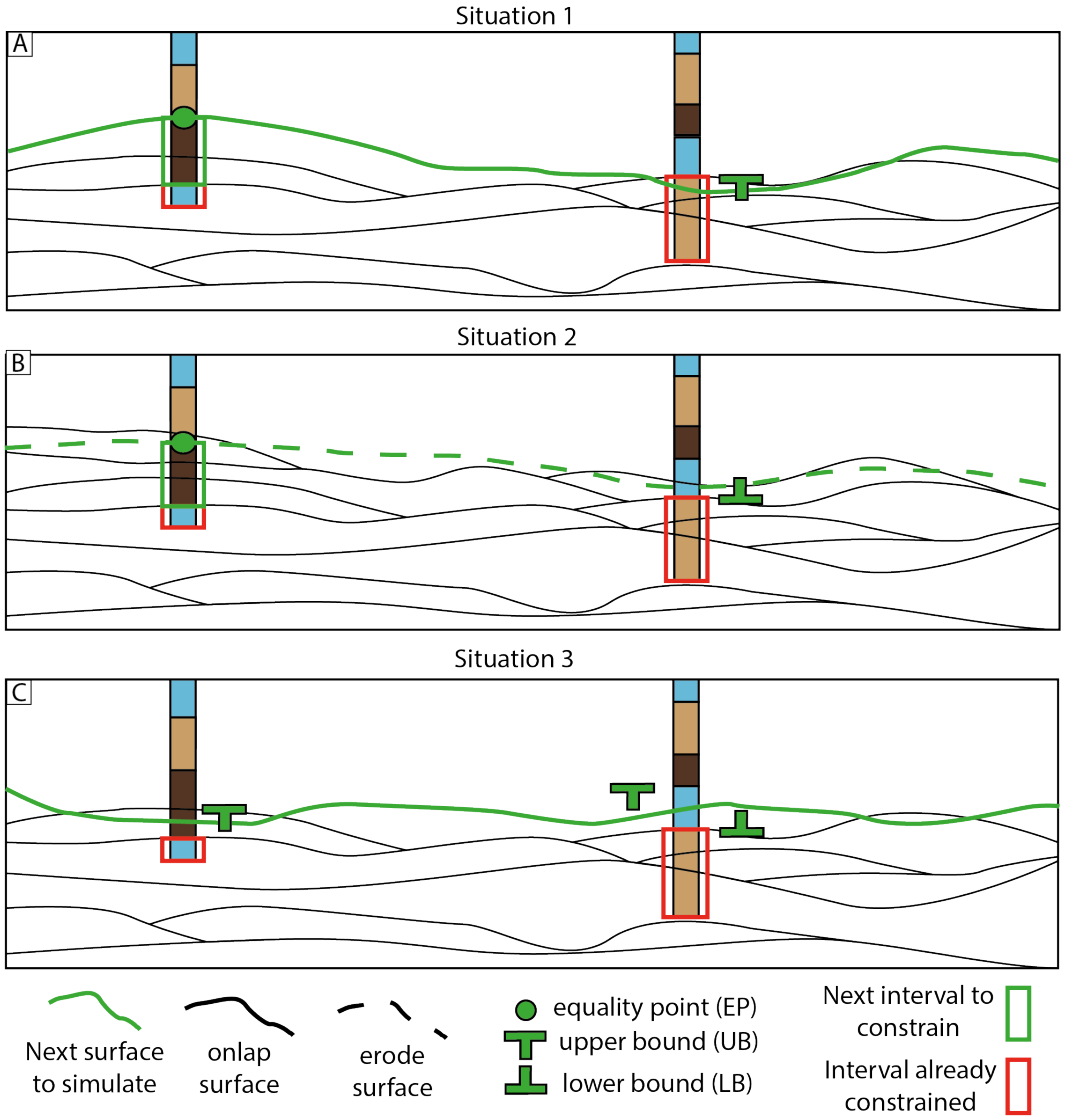

Figure 3. Three different situations can arise during the simulations of the surfaces. Situations 1 and 2 (**A** and **B**) concern a case where an interval has to be constrained while situation 3 (**C**) concerns onlap surfaces not constrained by an interface. See text for details. The horizontal bar of the "T" of the upper and lower bounds (UB and LB) is positioned at the altitude where the bound is applied. Note that these symbols are slightly offset to the left or right for better visibility. But they still apply to the borehole position.

The remaining steps (volume definition and facies assignment) are almost identical to the unconditional case. Before applying Algorithm 2, we simply attribute the facies from the boreholes to the regions that intersect them.

2.5 Hierarchical workflow

Hierarchical structures are very common in Quaternary deposits (A. Miall, 1996). To model these features, one can apply EROSim within a hierarchical workflow. The hierarchical levels can be seen as the results of sedimentological processes at different spatial and temporal scales (or levels).

We consider these levels as sets of surfaces that separate not only lithofacies but also distinct sedimentological units. The basic principle is that a surface of a specific hierarchical level can only affect surfaces that have an equal or lower hierarchical level. For example, a specific sedimentological body delimited by surfaces of a higher rank, cannot be eroded by a surface of a lower rank. From a computational point of view, this implies that we can simulate these deposits sequentially by hierarchical level from higher to lower levels.

For example, let us consider a case with two hierarchical ranks: one defined by surfaces having a large extent and delimiting large sedimentological units (or regions), and another one defined on a smaller scale with numerous surfaces frequently intersecting each other and producing regions of smaller sizes. To model this system, it is possible to decompose the simulation in two steps: first, high-order surfaces are simulated to delimit the boundaries of the different sedimentary units. In the second step, EROSim simulations are performed inside these units. Practically, this implies constraining the upper and lower boundaries of the simulation domain of EROSim with the simulated surfaces of the first. This can be extended to as many hierarchical levels as needed.

Using such a hierarchical approach, it is also possible to include more constraints in the model by setting different simulation parameters (α, ξ, N, surfaces interpolation parameters) for each unit or hierarchical level. But it also implies inferring a large number of parameters.

The usefulness of the hierarchical approach is presented in section 4.

3 Parameter sensitivity

This section illustrates the EROSim capability to simulate a wide variety of sedimentary structures. We first present some unconditional 2D examples, as well as the effect of the different parameters (α, ξ, etc.) on the simulations (e.g. shape and size of the regions). We then demonstrate that the algorithm can be conditioned to well-data. Some 3D simulations are presented in section 4 with a case study.

3.1 Unconditional simulations

For most of the examples shown in this section, we consider a 2D vertical slice (x,z) of dimension 60×30 m^2 and with a spatial resolution of 0.33×0.15 m^2. We simulate a total of four facies ($k = 4$) in equal proportions. To keep things simple, we used the same covariance for all surfaces in a single EROSim simulation. However, there is no difficulty in using different covariance models. For example, erosive surfaces could be assimilated to more energetic events and therefore could have smoother surfaces. In all simulations, the surface means are drawn from a uniform law between the top and bottom of the simulation domain. This allows distributing the surfaces uniformly over the domain and obtaining stationary simulations, but this is not a fixed feature of the algorithm.

Figure 4 shows one unconditional simulation with a stationary cubic covariance with a range of 45 m and a total variance of 5 m^2. The simulation parameters are $N = 100$ surfaces, 10 % of erosive surfaces ($\xi = 0.1$), and the facies simulation being only driven by the marginal target probability ($\alpha = 1$).

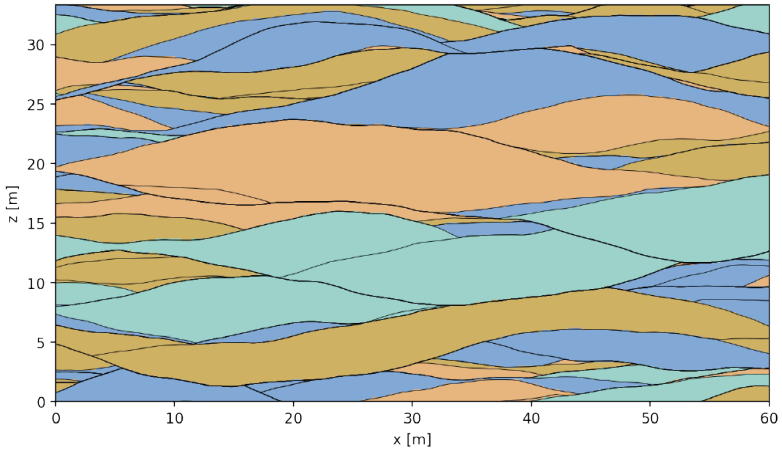

Figure 4. Example of an EROSim unconditional simulation with 100 surfaces. The four different colors represent four different facies. The black lines delimit the regions.

Figure 5 illustrates the effects of the different simulation parameters. As expected the range of the covariance controls the geometry of the surfaces. A shorter range leads to rougher surfaces. But it also controls the size of the regions. A shorter range creates more intersections between the surfaces and leads to smaller regions (Figs. 5A to 5C). Logically, the number of regions also increases with the range. The proportion of erosional surfaces (ξ) does not modify significantly the size of the regions, but it changes rather their shapes (Figs 5D - 5F). When there is no erosion ($\xi = 0$) the boundaries of the regions are mostly concave, but as ξ increases, more and more convex boundaries appear. This is the effect of erosion, which removes some parts of the regions that are filled later. As expected, the number of surfaces (N) influences the thickness of the regions (Figs. 5G to 5I), and their numbers. This is mainly due to our choice of drawing the surface means in a uniform law. The influence of the α parameter is also clearly visible (Figs. 5J to 5L). When α is equal to 0 (Fig. 5J), the regions of the same facies tend to be clustered through the whole simulation domain, while when α is equal to 1, they are evenly distributed.

Different covariance models can also affect the shape of the regions. For example, Fig. 6B shows a simulation with a spherical covariance model with a small range. The boundary between the regions are rough, and they are more individual regions than with a larger range.

Figures 6A, 6C, and 6D illustrate the possibility of using non-stationary means or variograms through the simulation domain. In Fig. 6A, the means are following a sinusoidal trend. In Fig. 6C the means follow linear trends with varying slopes. And finally, in Fig. 6D the sill and range of the variograms are progressively increasing to the right. These examples, even if they are a bit theoretical, show that if some information is known about the non-stationarity of the sedimentary structure, EROSim is capable of handling this information.

3.2 Conditional simulations

In this section, we illustrate the capabilities of EROSim to produce conditional simulations based on borehole data. The number of facies is reduced to three in equal proportions. For all the examples, we used the same simulation parameters ($N = 100$, $\xi = 0.1$, and $\alpha = 0.5$). The surface means were randomly drawn from a uniform distribution between the minimal and maximal elevations. All surfaces are simulated with the same cubic covariance model with a sill of 5 m^2 and a range of 15 m or 55 m.

Figure 7 shows different conditional EROSim simulations using a range of 15 m (Fig. 7A, B) or a range of 55 m (Fig. 7C, D). The borehole intervals are correctly respected without any apparent

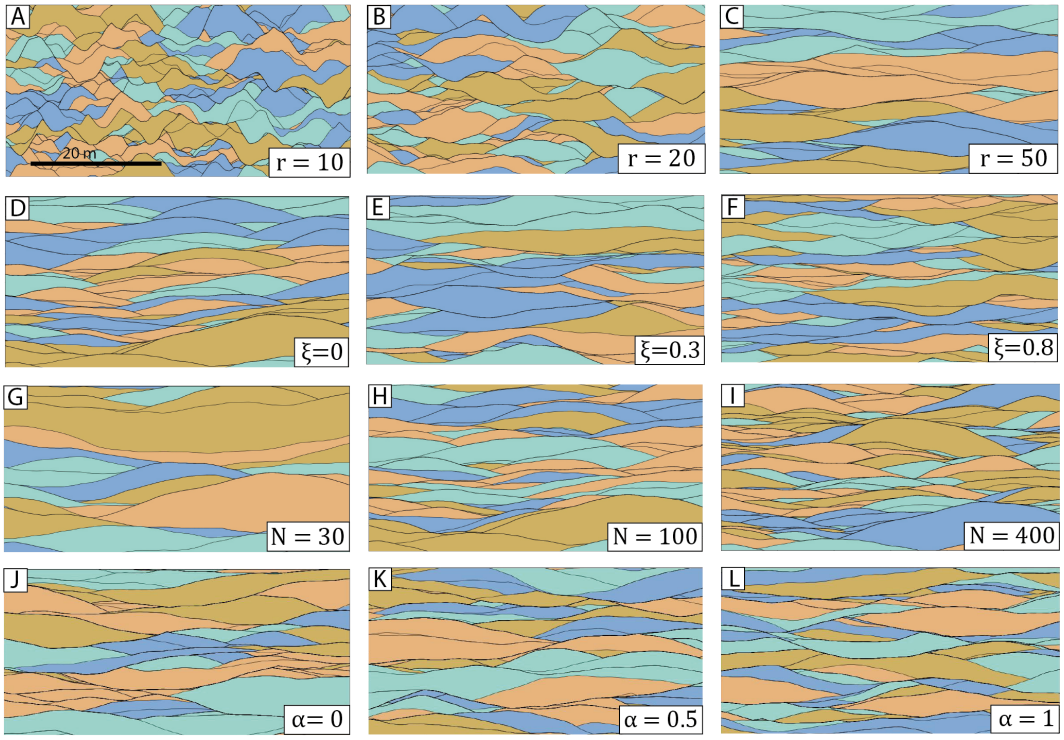

Figure 5. Different EROSim simulations where one parameter has been changed on each. **A** - **C**: Simulations where the variogram range increases from 10 to 20 and then to 50. **D** - **F**: Simulations where the ratio of erosive layer (ξ) increases from 0 to 30% and then to 80%. **G** - **I**: Simulations where the number of surfaces (N) is modified, from 30 to 100 and then to 400. **J** - **L**: Simulations where the α has been modified, from 0 to 0.5 and then to 1.

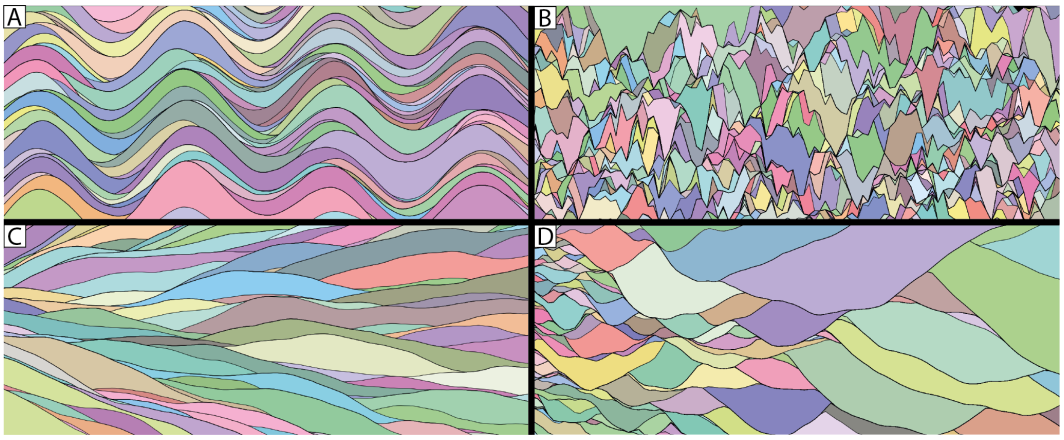

Figure 6. Realizations showing the capabilities of EROSim by varying different modeling parameters such as the mean (**A**, **C**), the variance (**D**) or the covariance model (**B**, spherical). The different colors of the regions do not represent anything in particular and are just used to distinguish them.

deformation of the regions. Note that despite using the same value for the clustering parameter α (=0.5) for both cases, the facies regions are distributed differently. Realizations obtained with a variogram range of 15 m display more variability in the types of contact between the regions than the ones made with a variogram range of 55 m, since using the lower range results in a larger number of regions.

We used Sequential Indicator Simulations (SIS, Journel, 1983) to compare EROSim capabilities over one of the most standard facies modeling tools. Indicator variograms were estimated using EROSim simulations and used to produce SIS realizations. Four of those simulations are shown in Figure 8. SIS respects the conditioning data, proportions, and input variograms but does not succeed in reproducing some of the patterns (sharp contacts) proposed by EROSim. SIS produces more noisy and pixelized simulations with irregular facies boundaries.

If we compare the probability maps obtained by averaging 100 realizations (Fig. 9), we observe strong similarities between the two methods. Indeed, facies probabilities seem practically identical. But if we compare Shannon's Entropy (Shannon, 1948), we see that SIS results are generally more uncertain (higher entropy) between the boreholes than EROSim. This is visible between the second and third boreholes (between $x = 25$ m and $x = 55$ m) which have nearly identical facies logs, implying a possible connection between the two. While EROSim is almost certain that a continuous connection exists between the two blue facies, SIS is not. The same comment can be made for the orange facies but the difference is not blatant.

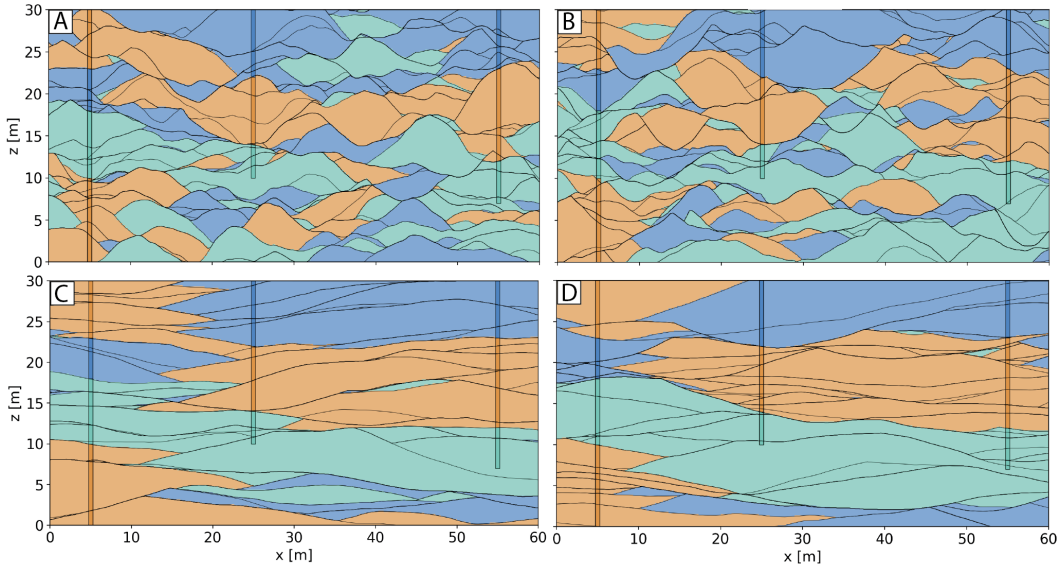

Figure 7. EROSim conditional simulations based on borehole data. **A** and **B** are two equiprobable realizations using a cubic covariance model with a range of 15 m and **C** and **D** were made with a cubic covariance with a range of 55 m.

4 Application to the upper Aare valley

In this last section, we illustrate the ability of EROSim to simulate structures similar to the ones observed in the field.

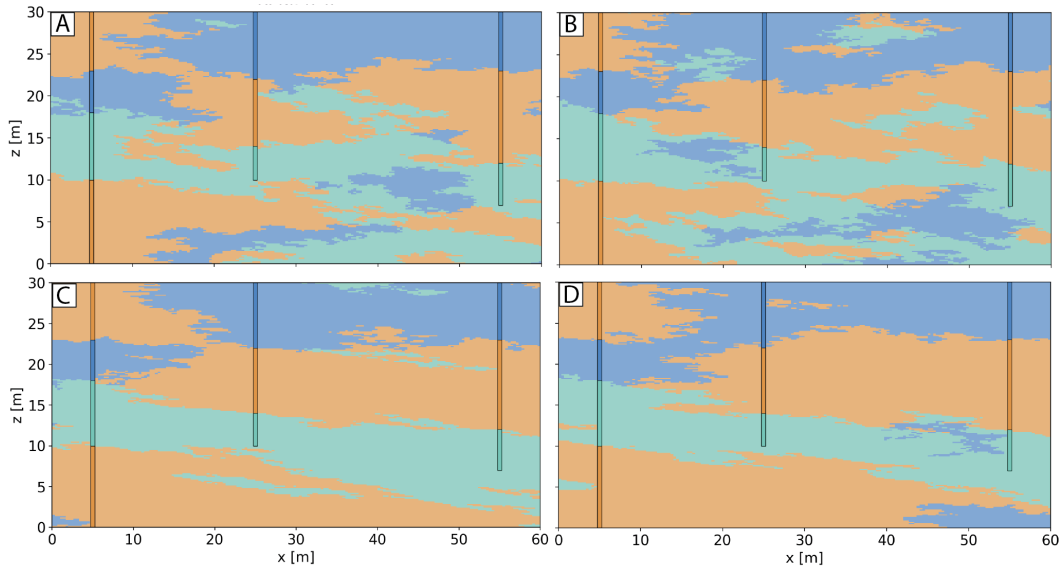

Figure 8. SIS simulations where each indicator variogram has been inferred from the corresponding EROSim simulations. Indicator variograms used to generate **A** and **B** were inferred on EROSim simulations with a range of 15 m (fig. 7 A, B). Identical procedure was applied to generate **C** and **D** but on EROSim simulations with a range of 55 m (fig. 7 C, D).

4.1 Study site and field data

The study site is located in the Bümberg quarry, in the canton of Bern, in Switzerland. The quarry walls show fluvio-glacial Quaternary sediments made up of different sand and gravel facies that show complex relationships. This type of sedimentary architecture can be observed in many other queries in Switzerland and abroad. The Bümberg quarry has been identified as a suitable sedimentological analog for the Upper Aare Valley aquifer. To characterize its heterogeneity, the sedimentological structures, hierarchical levels, and deposits of the quarry walls were manually interpreted based on a field survey and a high-resolution photogrammetric UAV survey to obtain 3D ortho-normal images (Menga, 2021). This enabled the digitization of the delimitation and the characterization of the different stratigraphic boundaries and unit extents. In this quarry, Menga (2021) analyzed two walls, one oriented North-South (180 m long and 13 m high) and one oriented East-West (115 m long and 12 m high) to characterize the possible anisotropy of the sedimentological structures.

Several different lithofacies have been identified and described according to the classification proposed in A. D. Miall (2013) that takes into account the grain size and the sedimentological structures of the deposits. The different grain sizes are described by a capital letter: C (cobbles), P (pebbles), G (gravels), S (sands), and L (silts). The arrangement and/or the structure of the sediments are described by a lowercase letter: o (open framework), i (imbricated), h (horizontal stratification), n (normally graded), r (reverse graded), l (low angle stratification), p (planar-cross stratification), t (trough-cross stratification), s (draping troughs), x (cross-stratification), m (massive).

Menga (2021) regrouped them into seven facies groups and for each group a sedimentological origin interpretation has been proposed. These groups and interpretations are shown in Table 1 and Figure 10B shows the spatial organization of these facies on the East-West wall of the quarry. Gravel-dominated facies are the predominant groups of facies, whereas sand-dominated facies are quite dispersed and only present locally. Gravel facies are distinguished into four groups: bedload sheet aggradation (yellow), transverse bar migration (blue), scour fills (brown), and gravelly dunes

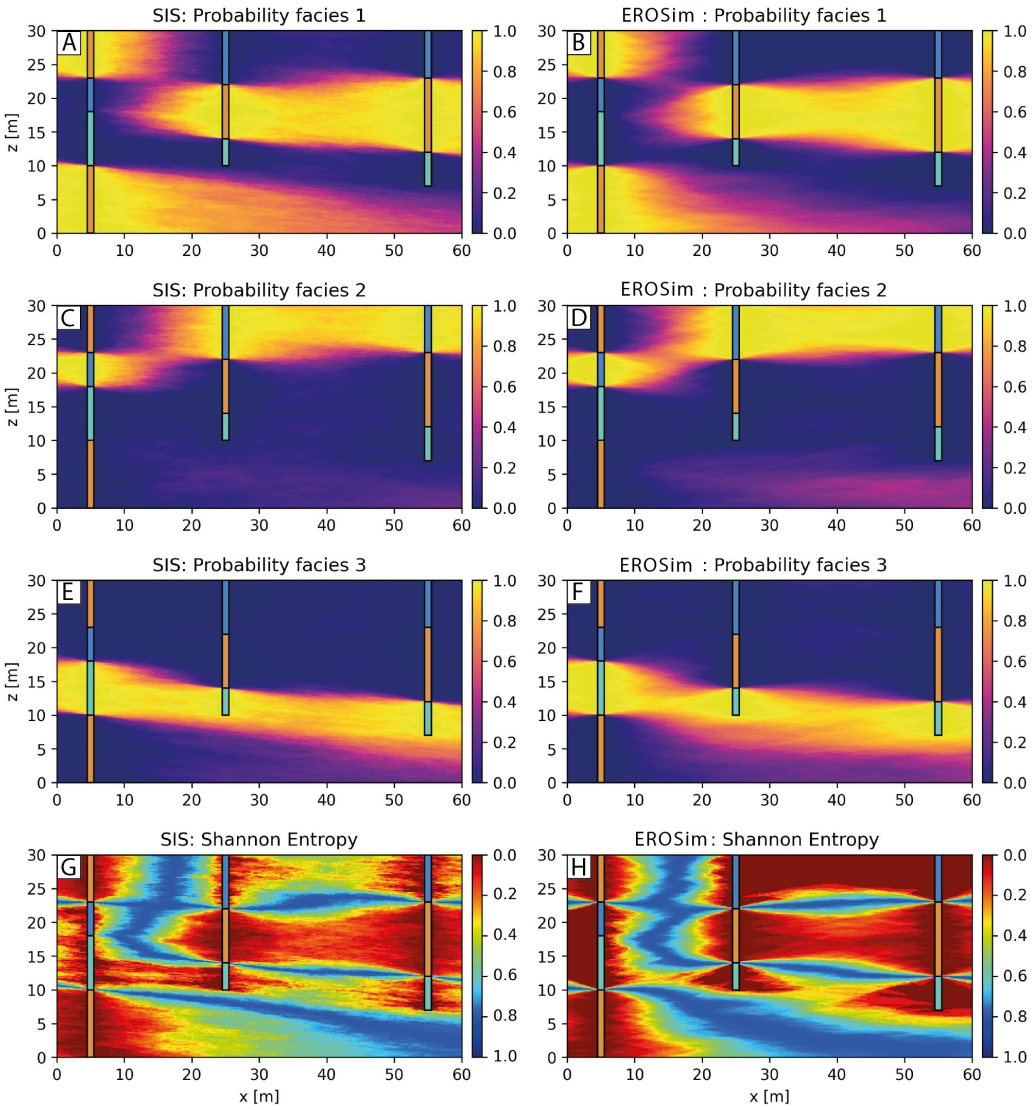

Figure 9. Probability of occurrence for each facies (computed over 100 realizations) considering a variogram range of 55 for EROSim simulations (fig. 7C, D) and SIS simulations made with variograms inferred on EROS simulations (fig. 8C, D). **A**, **C**, **E** are the probability maps for facies 1 to 3 for SIS method and **B**, **D**, **F** are the probability maps for facies 1 to 3 for EROSim method. **G** and **H** are the related Shannon entropies to probability maps for SIS and EROSim, respectively.

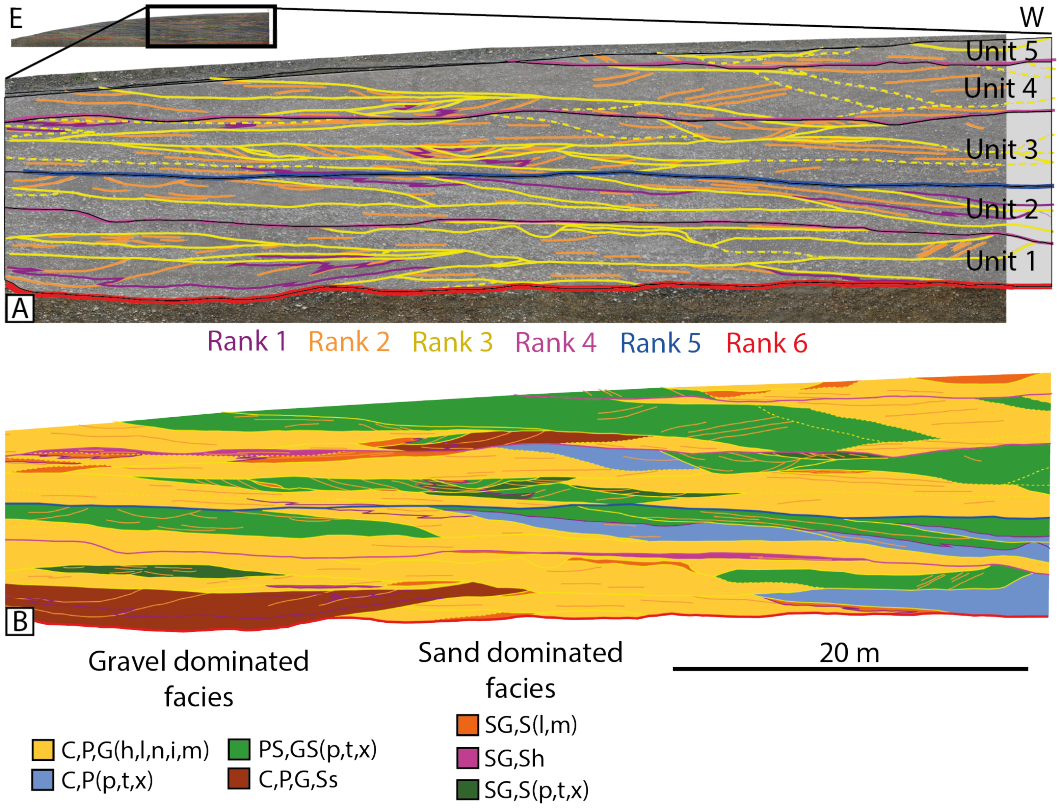

Figure 10. Interpretation of a part of the East-West wall of the Bümberg quarry realized by Menga (2021). **A** is the reconstructed image of the wall acquired by the drone as well as the sedimentological surfaces that separate the subfacies bodies. The hierarchical rank of each surface is given by its color. The five stratigraphical units, delimited by surfaces of rank 4 or higher are also shown. **B** shows the corresponding colorized facies groups interpretation. For readability and clarity, only half of the wall is shown. Figures and data are taken and modified from Menga (2021)).

(green), for a more comprehensive understanding of these sedimentological formations, please refer to A. Miall (1996).

Table 1. Facies code and their interpretation in the Bümberg quarry. The colors used to represent them are also given.

Facies code	Interpretation	Facies color
C,P,G(h,l,n,i,m)	Bedload sheet aggradation	Yellow
C,P(p,t,x)	Transverse bar migration	Blue
C,P,G,Ss	Scour fill	Brown
PS,GS(p,t,x)	Gravelly dunes	Green
SG,S(p,t,x)	Sandy-gravelly dunes	Dark Green
SG,Sh	Sandy wedges	Pink
SH,S(l,m)	High flow regime sandy levels	Orange

In addition, six levels of sedimentological hierarchies have also been recognized, and named *Rank 1* (lowest rank) to *Rank 6* (highest rank). Figure 10A shows the interpreted surfaces on the wall of the Bümberg quarry. The surfaces of *Rank 1* to *Rank 2* are likely the result of very local processes ($\sim$ 1 - 5 m) while *Rank 3* surfaces have a larger extent ($\sim$ 5 - 60 m). Higher-order surfaces (*Rank 4 - Rank 6*) exceed the size of the domain and can be treated equally in this situation. There are a total of six of these surfaces on the two walls, which delimit five sedimentological bodies (Units in Figure 10A) that differ in terms of facies proportions and structures.

4.2 Model setup and parameters

To account for the hierarchical relations observed on the study site, we consider the hierarchical approach proposed in section 2.5. At the lowest rank, the regions are filled with facies while surfaces of higher ranks are simulated independently to delimit stratigraphic units. For the upper Aare valley, the aim is to produce a model of lithofacies. Based on the field observations, we consider that it is reasonable to set the lowest rank to *Rank 3* surfaces (Fig. 10A, yellow lines) and to fill the delimited regions with the facies that are mostly differentiated by these surfaces. Figure 10B shows that these are gravel-dominated facies. The sand facies (dark green, pink, and orange) are more dispersed and often delimited by surfaces of *Rank 1* or *2*. Therefore, we must consider as well the hierarchical levels above *Rank 3* (pink, blue and red surfaces). Furthermore, the regions delimited by these higher rank surfaces have different characteristics in terms of sedimentary body sizes and facies proportions. Finally, we only simulate the four dominating facies (mainly gravels). Orange and pink facies have been regrouped into the yellow facies, while dark green facies has been regrouped with green facies.

As explained in section 2.5, the approach is decomposed in two steps. First, the geostatistical parameters of the higher-rank surfaces are required to define the extent of each unit. From base to top in this quarry, Menga (2021) recognized a first unit bounded by a surface of *Rank 6* and *Rank 4*, a second unit bounded at the top by a *Rank 5* surface, a third, fourth and fifth units, bounded at the top by *Rank 4* surfaces (fig. 10A). Consequently, it is required to simulate three *Rank 4* surfaces and one *Rank 5* surface. Therefore, for each wall, the variograms for *Rank 4* and *Rank 5* were estimated on the available surfaces as well as the mean altitudes of each of the surfaces. The parameters to model those surfaces are provided in tables 2-3 in the appendix. *Rank 6* surface was not modeled and was simply considered to be the bottom of the simulations.

In this phase, we undertake the simulation of *Rank 3* surfaces within each distinct unit. EROSim requires that we infer the following parameters: the number of surfaces (N), a distribution for the mean altitudes of the surfaces, the variogram models of the surfaces (μ_i and γ_i), the proportion of erosive surfaces (ξ), the proportions of the facies (p_{global}), and lastly, the clustering parameter (α).

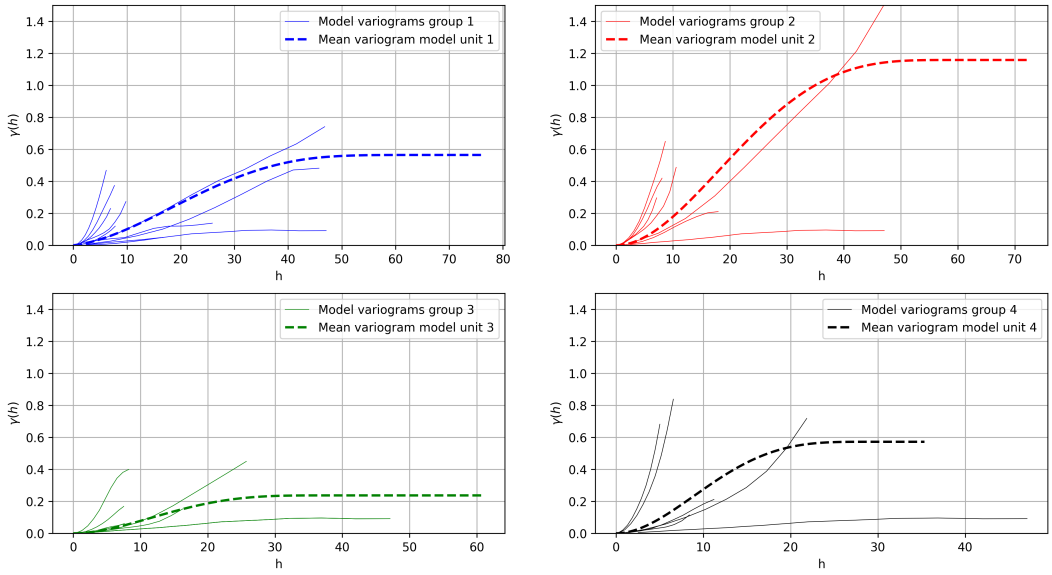

Figure 11. Experimental and fitted model variograms realized on the EW wall of the Bümberg quarry for each different unit.

The γ_i were estimated using the interpreted surfaces, drawn by Menga (2021), by computing the experimental variogram on each independent line (or surfaces) and by adjusting a variogram model on them. Note that we only considered lines with a minimal length of 5 m to obtain representative statistics. The spatial statistics of these surfaces for the EW Bümberg wall are summarized in Fig. 11 for each unit. Unit 5 is not represented, as there was not enough surfaces to infer proper variogram models. In this case, we used the mean values of the parameters, averaged on all the variogram models. Finally, for each unit, the inferred parameters of the variogram models were averaged to keep only one set of parameters.

The other parameters (N, ξ, p_{global}, α) were estimated by trial and error and are given in Table 4 in the Appendix.

4.3 Results

We first compare 2D simulations with actual observations on the quarry walls and then present results from simulated 3D blocks.

2D cross-sections

Figures 12 and 13 show EROSim simulations of the quarry walls and the observed geology for the same locations. In these figures, the lines delimiting the regions are not shown for clarity. Generally, the simulations reproduce rather well the shapes of the regions as compared to the reference. A more quantitative analysis of the results has been conducted on an ensemble of 100 realizations for the two sections. It shows that the indicator variograms computed for each facies for each EROSim simulation are well distributed around the reference indicator variograms, as illustrated in Figures 14A to 14C for N-S wall. The proportions are also well respected (Fig. 14D). Similar results are obtained with the E-W wall.

3D models

EROSim can also be used in 3D. Figure 15 shows 3D simulations using the sedimentological statistics inferred on both walls of the Bümberg quarry. Standard parameters (tab. 4) have been averaged between the two walls, as well as the altitudes of the surfaces of the higher ranks. The

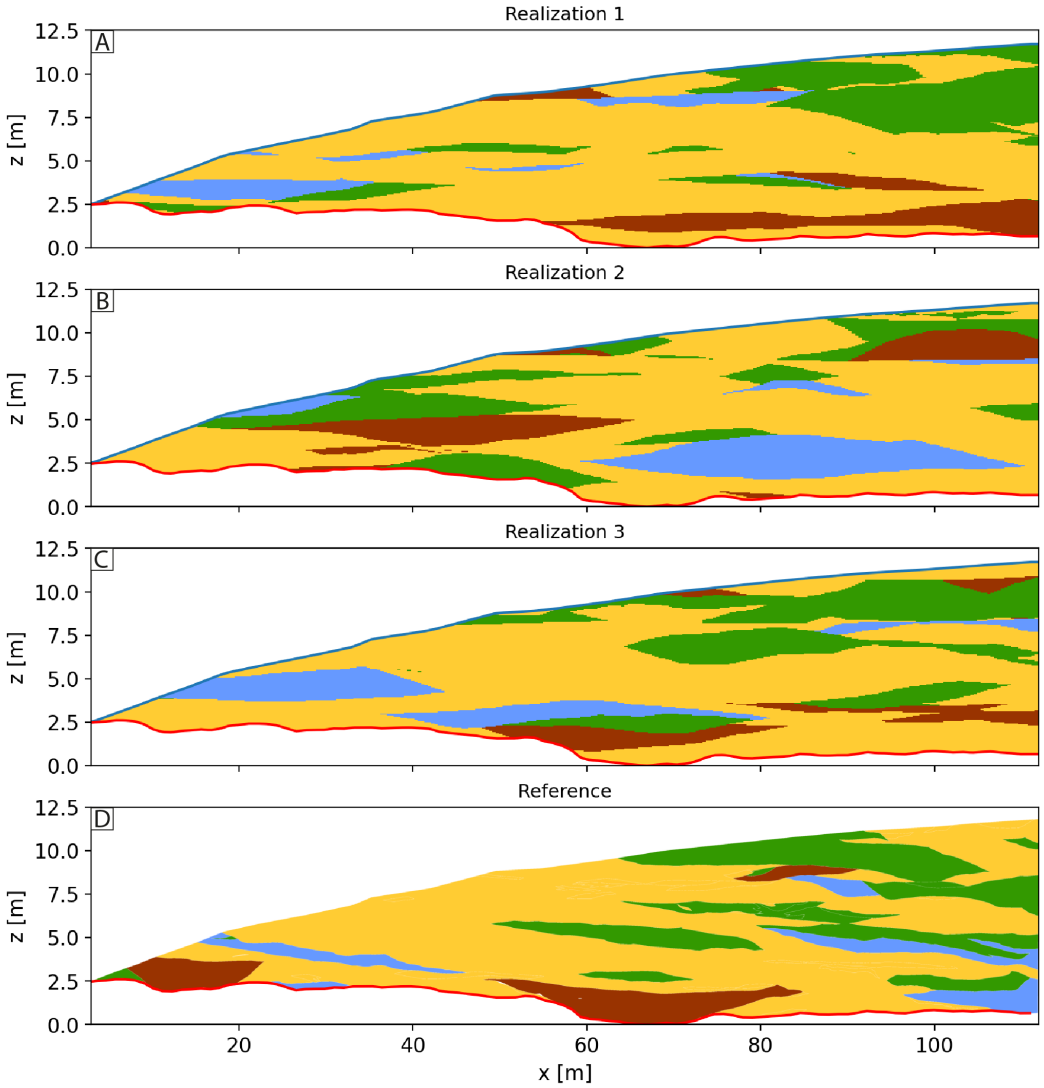

Figure 12. Three EROSim simulations of the East-West wall of the Bümberg quarry (**A - C**) and the interpreted wall made by Menga (2021) considered as the reference (**D**). The sedimentological lines that separate the regions are not shown.

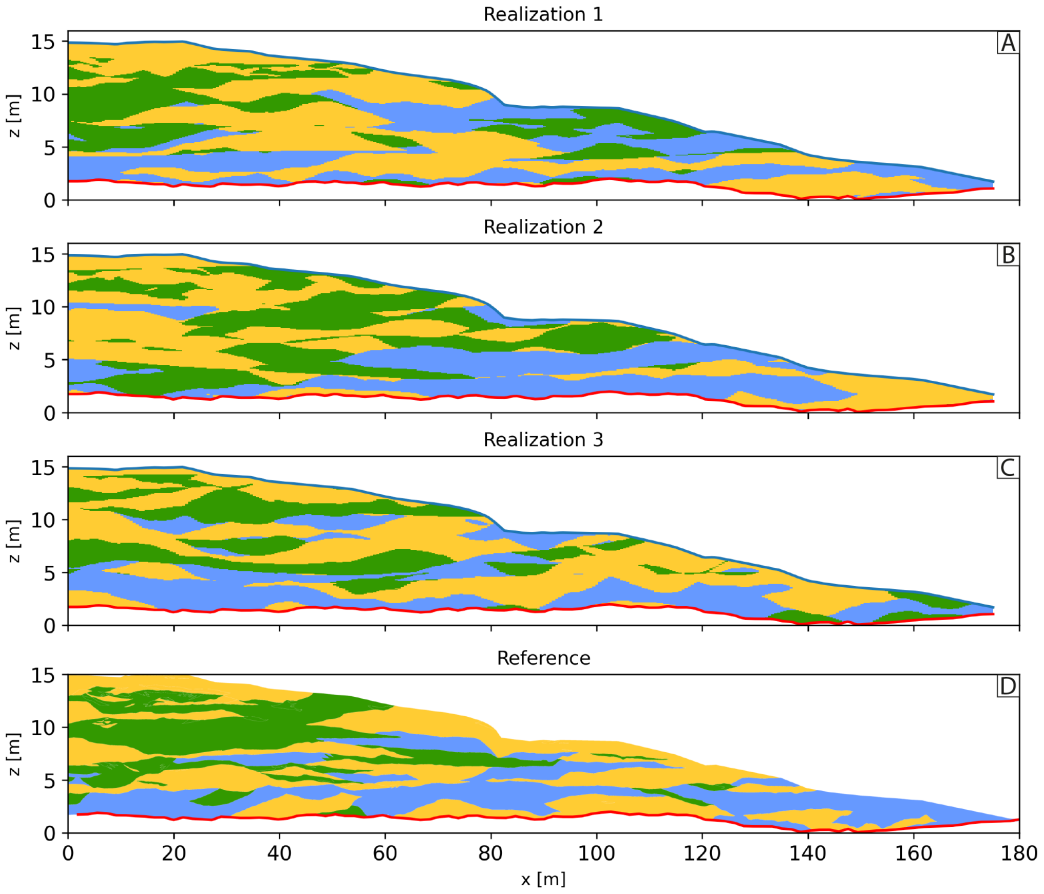

Figure 13. Three EROSim simulations of the North-South wall of the Bümberg quarry (**A - C**) and the interpreted wall made by Menga (2021) considered as the reference (**D**). The sedimentological lines that separate the regions are not shown.

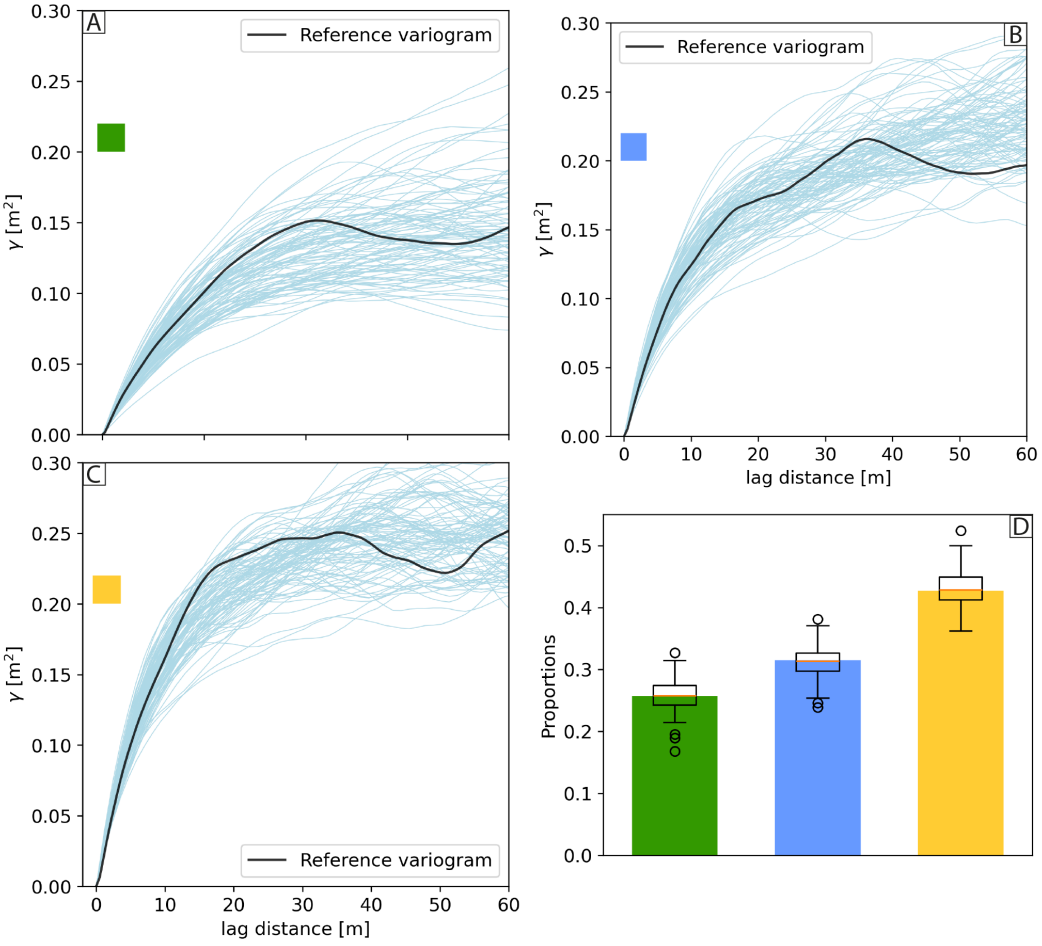

Figure 14. Comparison of proportions (**D**) and indicator variograms (**A**-**C**) for each facies between the references and 100 realizations of the N-S wall.

1D variograms were also combined to obtain 2D variograms for the surface simulations. This was done assuming that the spatial statistics inferred on the walls represent the statistics on the major and minor axes of anisotropy of the 2D variogram. The variogram ranges on the NS wall were assigned to the range on the Y axis of the 2D variograms, and the variogram ranges on the EW wall were assigned to the range on the X axis. Sills were averaged. The grid used for the visualization has a cell size of $0.85 \times 0.85 \times 0.14$ m^3 and contains $100 \times 100 \times 100$ cells.

Figure 15 shows significant variability between the realizations because of the absence of conditioning. It also shows that EROSim can reproduce the expected shapes of the sedimentological structures while exploring a broad variety of plausible configurations. The sedimentological lines bounding the regions are not shown here for the sake of visibility, but it is important to remember that this additional information is available. Regions can be individualized and treated differently in successive modeling steps (e.g. simulation of physical properties).

5 Discussion

In this paper, we presented EROSim, a novel surface-based simulation method and a conditioning algorithm that can model sedimentological heterogeneity that is frequently observed in fluvio-glacial systems.

The application of this method with the data from the Bümberg quarry site shows that the results obtained with this method represent well the main sedimentological features observed on the quarry walls. Slight visual differences remain in both the shape of the regions and the distribution of the facies. This is most probably due to using one unique variogram to model all the internal surfaces. Indeed, Figure 11 shows a large spectrum of estimated variogram models from the field data, regardless of the unit in which we are. The same thing is also observed on the NS wall. Therefore, the use of a single variogram model cannot capture this variability. However, the observed variability is related to the estimation of variograms on very short lines which may create artifacts, whereas EROSim uses these variograms to simulate surfaces over the whole simulation domain. It is also important to remember that erosion processes influence the shape of these lines, which necessarily influences the spatial statistics of these objects. Modeling the interfaces with different variograms would be possible but it would make the parametrization of EROSim much more complex and we argue that it would not necessarily be useful or critical for groundwater applications.

As compared to other variogram-based facies simulation methods EROSim provides more control over the geometry of the geological interfaces for the user. The simulated models include a representation of internal structures within single facies such as orientations of the deposition patterns, and a hierarchical structure. The orientation of the local patterns, given by the simulated surfaces, can be used to orient the local anisotropy of permeability. The definition of these surfaces permits as well defining thin and continuous structures such as clay plugs which may have an important influence on flow and transport and which are difficult to simulate with other variogram-based approaches.

Compared to object-based methods (OBM) or Multiple Point Statistics (MPS), EROSim is not designed to generate specific geological shapes, such as meanders, deep-water lobes, etc. However, EROSim does not require sophisticated inputs, such as a training dataset (TD) for MPS or the definition of geological objects and their spatial relations for OBM. Determining the parameters of EROSim from outcrop data is simpler than determining all the input required for MPS or OBM. In addition, more complexity could be added in EROSim relatively easily by including for example some non-stationarity as shown in Figure 6 or by having a more complex parametrization. While we presented a brief comparison with SIS in terms of conditioning in this paper, a more comprehensive comparison involving other methods and including the impact of the sedimentary structures on flow and transport has still to be done. It will be the topic of future research.

A few limitations and possible extensions of the proposed approach can be identified. The conditional algorithm requires imposing rules on the simulation of surfaces. This is permitted thanks to the use of inequality data when modeling Gaussian Random Fields, and this is one of the main

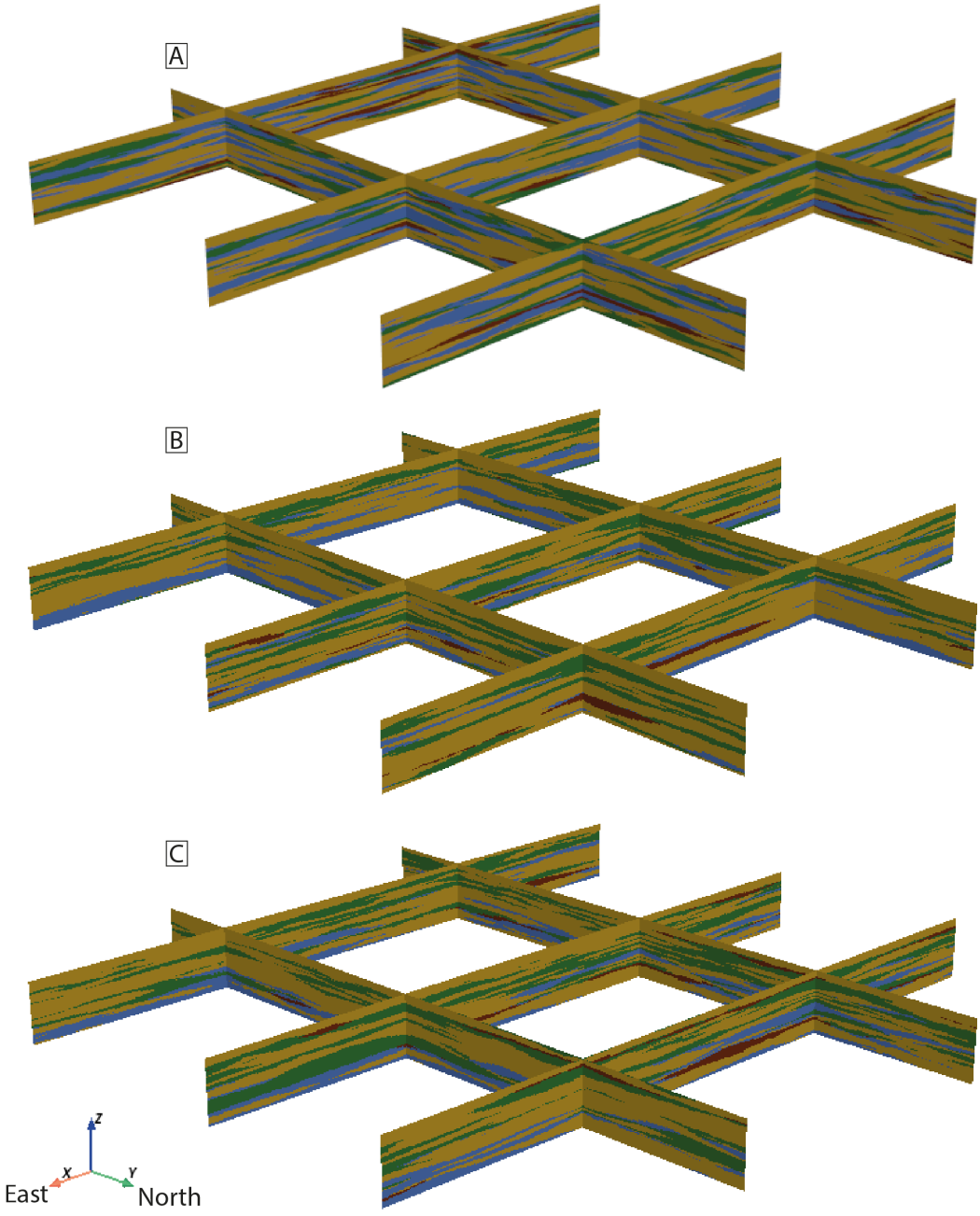

Figure 15. Three different realizations (**A - C**) of EROSim 3D simulations of the Bumberg quarry. The model dimensions are 170 m x 170 m x 14 m, there is no vertical exaggeration.

novel contributions of this paper. In most cases, the conditioning does not perturb the shape of the regions but it may distort them if the conditioning data are very dense (high number of borehole data). Input parameters must be compatible with the data. If too few surfaces are defined, the algorithm may not be able to respect all the interfaces within the boreholes. But this limitation is not specific to EROSim and applies to all geostatistical methods in general.

There is also room for improvement to integrate other kinds of conditioning data (seismic data, locally varying facies proportions, etc.). For the simulation of the facies, we decided to keep the parametrization extremely simple. This is a potential limitation and more sophisticated rules involving transition probabilities between each facies, similar to what is done in TPROGS (Carle, 1999), could be accounted for. We could also account for the volume or shape of the region when estimating the probability of affecting a facies to a certain region. However, once again, we argue that the simplicity of the proposed model is probably sufficient for many applications and we leave this possible extensions to future works. In the Bümberg case study, we identified the parameters by computing individual variograms and averaging them. The other parameters were mainly identified by trial and error. Further research should be devoted to developing a more automated technique to infer the parameters.

6 Conclusion

In this paper, we presented a new approach, EROSim, to model geological heterogeneity by simulating litho-facies models. The method is designed to represent sedimentary structures typically present in fluvio-glacial Quaternary deposits which are the most frequently used aquifers in Switzerland. EROSim offers a new perspective in the field of facies modeling by initially creating regions and subsequently populating them with facies. Moreover, it seamlessly integrates geological principles via erosion-deposition rules, introducing a significant degree of flexibility into its realizations and rendering it suitable for diverse sedimentological contexts. The litho-facies are assigned to the regions using a graph-based approach accounting for global proportions over the entire domain and for local proportions derived from adjacent regions. The conditioning algorithm employs inequality data derived from the borehole observations and the sedimentological rules.

The capacity of this model to represent different geometries and sedimentary patterns has been illustrated with several unconditional and conditional examples. Through these examples, we show the influence of the model parameters (e.g., α, ξ, etc.) on the resulting simulations. Furthermore, EROSim is applied to a real field site from a gravel quarry, where the spatial statistics of the sedimentological surfaces are inferred and used to parameterize the simulations. The resulting simulations show spatial patterns closely similar to those observed on the quarry walls. A numerical comparison of the proportions of the facies and the indicator variograms of the facies confirms this similarity.

The main advantages of this approach are: its simplicity for the parameterization, its capacity to generate realistic 3D simulations from data acquired in 2D, its hierarchical structures, the possibility to condition the simulations by borehole data, and finally the availability of the code.

Further research is still needed to analyze the impact of the structures generated by EROSim on flow and solute transport and to compare the performances of this model with other facies modeling techniques.

Open Research Section

The github repository of the code as well as dataset used can be found on `http://www.github.com/randlab/erosim`.

Acknowledgments

The authors wish to express their thanks to the sedimentological team of the University of Milano, Riccardo Bersezio, Chiara Zuffetti, Alessandro Comunian and most particularly Ilaria Menga for

her meticulous work on the geology of the Aar Valley, without which this work likely would not have been done.

References

Allard, D., Comunian, A., & Renard, P. (2012). Probability aggregation methods in geoscience. *Mathematical Geosciences*, *44*, 545–581.

Bayer, P., Huggenberger, P., Renard, P., & Comunian, A. (2011). Three-dimensional high resolution fluvio-glacial aquifer analog: Part 1: Field study. *Journal of Hydrology*, *405*(1-2), 1–9.

Bennett, J. P., Haslauer, C. P., & Cirpka, O. A. (2017). The impact of sedimentary anisotropy on solute mixing in stacked scour-pool structures. *Water Resources Research*, *53*(4), 2813–2832.

Carle, S. F. (1999). T-progs: transition probability geostatistical software, version 2.1. *Department of Land, Air and Water Resources, University of California, Davis*.

Chiles, J.-P., & Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty* (Vol. 713). John Wiley & Sons.

Chiogna, G., Cirpka, O. A., Rolle, M., & Bellin, A. (2015). Helical flow in three-dimensional nonstationary anisotropic heterogeneous porous media. *Water Resources Research*, *51*(1), 261–280.

Comunian, A., Renard, P., Straubhaar, J., & Bayer, P. (2011). Three-dimensional high resolution fluvio-glacial aquifer analog–part 2: Geostatistical modeling. *Journal of hydrology*, *405*(1-2), 10–23.

Dagan, G. (1989). *Flow and transport in porous formations*. Springer Science & Business Media.

de Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., & Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal*, *13*, 161–183.

Freulon, X., & de Fouquet, C. (1993). Conditioning a gaussian model with inequalities. *Geostatistics Tróia'92: Volume 1*, 201–212.

Geng, X., Michael, H. A., Boufadel, M. C., Molz, F. J., Gerges, F., & Lee, K. (2020). Heterogeneity affects intertidal flow topology in coastal beach aquifers. *Geophysical Research Letters*, *47*(17), e2020GL089612.

Heinz, J., Kleineidam, S., Teutsch, G., & Aigner, T. (2003). Heterogeneity patterns of quaternary glaciofluvial gravel bodies (sw-germany): application to hydrogeology. *Sedimentary geology*, *158*(1-2), 1–23.

Jo, H., Santos, J. E., & Pyrcz, M. J. (2020). Conditioning well data to rule-based lobe model by machine learning with a generative adversarial network. *Energy Exploration & Exploitation*, *38*(6), 2558–2578.

Journel, A. G. (1983, June). Nonparametric estimation of spatial distributions. *Journal of the International Association for Mathematical Geology*, *15*(3), 445–468. doi: 10.1007/BF01031292

Kitanidis, P. K. (2015). Persistent questions of heterogeneity, uncertainty, and scale in subsurface flow and transport. *Water Resources Research*, *51*(8), 5888–5904.

Koltermann, C. E., & Gorelick, S. M. (1996). Heterogeneity in sedimentary deposits: A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research*, *32*(9), 2617–2658.

Menga, I. (2021). *Analysis of late pleistocene aquifer analogues to characterize heterogeneity of the post glacial aquifers of the aare valley (switzerland)* (Unpublished master's thesis). Universita' degli studi di Milano, facoltà di Scienze e Tecnologie, Via Luigi Mangiagalli, 34, 20133 Milano MI, Italy.

Miall, A. (1996). *The geology of fluvial deposits. sedimentary facies, basin analysis, and petroleum geology*. Springer.

Miall, A. D. (2013). *The geology of fluvial deposits: sedimentary facies, basin analysis, and petroleum geology*. Springer.

Neuman, S. P., Riva, M., & Guadagnini, A. (2008). On the geostatistical characterization of hierarchical media. *Water Resources Research*, *44*(2).

Neven, A., & Renard, P. (2023). A novel methodology for the stochastic integration of geophysical

and hydrogeological data in geologically consistent models. *Water Resources Research*, *59*(7), e2023WR034992.

Neven, A., Schorpp, L., & Renard, P. (2022). Stochastic multi-fidelity joint hydrogeophysical inversion of consistent geological models. *Frontiers in Water*, *4*, 989440.

Pirot, G., Straubhaar, J., & Renard, P. (2015). A pseudo genetic model of coarse braided-river deposits. *Water Resources Research*, *51*(12), 9595–9611.

Pyrcz, M. J., Catuneanu, O., & Deutsch, C. V. (2005). Stochastic surface-based modeling of turbidite lobes. *AAPG bulletin*, *89*(2), 177–191.

Pyrcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. Oxford university press.

Pyrcz, M. J., Sech, R. P., Covault, J. A., Willis, B. J., Sylvester, Z., & Sun, T. (2015). Stratigraphic rule-based reservoir modeling. *Bulletin of Canadian Petroleum Geology*, *63*(4), 287–303.

Ramanathan, R., Guin, A., Ritzi Jr, R. W., Dominic, D. F., Freedman, V. L., Scheibe, T. D., & Lunt, I. A. (2010). Simulating the heterogeneity in braided channel belt deposits: 1. a geometric-based methodology and code. *Water Resources Research*, *46*(4).

Ritzi, R. W., Dai, Z., Dominic, D. F., & Rubin, Y. N. (2004). Spatial correlation of permeability in cross-stratified sediment with hierarchical architecture. *Water Resources Research*, *40*(3).

Rubin, Y. (2003). *Applied stochastic hydrogeology*. Oxford University Press.

Scheibe, T. D., & Freyberg, D. L. (1995). Use of sedimentological information for geometric simulation of natural porous media structure. *Water Resources Research*, *31*(12), 3259–3270.

Schorpp, L., Straubhaar, J., & Renard, P. (2022). Automated hierarchical 3d modeling of quaternary aquifers: the archpy approach. front. *Earth Sci*, *10*, 884075.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379–423.

Soltanian, M. R., Behzadi, F., & de Barros, F. P. (2020). Dilution enhancement in hierarchical and multiscale heterogeneous sediments. *Journal of Hydrology*, *587*, 125025.

Titus, Z., Heaney, C., Jacquemyn, C., Salinas, P., Jackson, M., & Pain, C. (2021). Conditioning surface-based geological models to well data using artificial neural networks. *Computational Geosciences*, 1–24.

Wallace, C. D., Tonina, D., McGarr, J. T., de Barros, F. P., & Soltanian, M. R. (2021). Spatiotemporal dynamics of nitrous oxide emission hotspots in heterogeneous riparian sediments. *Water Resources Research*, *57*(12), e2021WR030496.

Webb, E. K. (1994). Simulating the three-dimensional distribution of sediment units in braided-stream deposits. *Journal of Sedimentary Research*, *64*(2b), 219–231.

Xie, Y., Cullick, A. S., & Deutsch, C. V. (2001). Surface-geometry and trend modeling for integration of stratigraphic data in reservoir models. In *Spe western regional meeting.*

Xie, Y., Deutsch, C. V., & Cullick, A. (1999). A short note on surface-based modelling for integration of stratigraphic data in geostatistical reservoir models. *CCG Annual Report One, Edmonton: University of Alberta*.

Zech, A., Attinger, S., Bellin, A., Cvetkovic, V., Dagan, G., Dentz, M., … Teutsch, G. (2021). A comparison of six transport models of the made-1 experiment implemented with different types of hydraulic data. *Water resources research*, *57*(5), e2020WR028672.

Zuffetti, C., Comunian, A., Bersezio, R., & Renard, P. (2020). A new perspective to model subsurface stratigraphy in alluvial hydrogeological basins, introducing geological hierarchy and relative chronology. *Computers & Geosciences*, *140*, 104506.

Appendix

Simulation parameters

Table 2. Covariance models parameters (C: contribution (sill) and r: range). Subscripts *cub1* and *cub2* indicate two different cubic covariance models. Contributions that were lower than 0.01 were discarded.

Quarry wall	Rank surfaces	r_{cub1} [m]	C_{cub1} [m^2]	r_{cub2} [m]	C_{cub2} [m^2]
NS	3 in unit 5	-	-	32	0.35
	3 in unit 4	15	0.27	30	0.39
	3 in unit 3	-	-	25	0.64
	3 in unit 2	-	-	35	1.06
	3 in unit 1	28	0.41	23	0.38
	4	-	-	90	0.26
	5	68	0.65	30	0.1
EW	3 in unit 5	-	-	44	0.44
	3 in unit 4	-	-	29	0.77
	3 in unit 3	-	-	35	0.55
	3 in unit 2	-	-	60	1.35
	3 in unit 1	70	0.51	45	0.29
	4	-	-	90	0.25
	5	25	0.12	60	0.15

Table 3. List of high-order surfaces with their rank and mean altitude used for the simulations.

Quarry wall	Surface ID	Rank	Mean altitude [m]	Bottom of ...	Top of ...
NS	1	4	13.68	unit 5	unit 4
	2	4	10.99	unit 4	unit 3
	3	5	7.22	unit 3	unit 2
	4	4	4.43	unit 2	unit 1
EW	1	4	10.59	unit 5	unit 4
	2	4	8.23	unit 4	unit 3
	3	5	5.75	unit 3	unit 2
	4	4	3.90	unit 2	unit 1

Table 4. EROSim parameters used for the simulation of the quarry walls.

Quarry wall	Unit group	α	N	ξ	p_{global} (blue, brown, green, yellow)
NS	unit 1	0.8	10	0.2	(0.5, 0, 0.1, 0.4)
	unit 2	0.8	15	0.2	(0.32, 0, 0.31, 0.37)
	unit 3	0.5	10	0.2	(0.20, 0, 0.35, 0.45)
	unit 4	0.9	20	0.2	(0, 0, 0.50, 0.50)
	unit 5	0.9	3	0.2	(0, 0, 0, 1)
EW	unit 1	1	20	0.2	(0.10, 0.15, 0.10, 0.70)
	unit 2	1	15	0.2	(0.08, 0, 0.12, 0.8)
	unit 3	1	10	0.2	(0.05, 0, 0.22, 0.73)
	unit 4	0.9	10	0.2	(0, 0.10, 0.50, 0.40)
	unit 5	0.9	3	0.2	(0, 0, 0.10, 0.90)