Water Resources Research®

Correspondence to:

S. Hörning,
s.hoerning@uq.edu.au

Definition of Spatial Copula Based Dependence Using a Family of Non-Gaussian Spatial Random Fields

András Bárdossy[1] [iD] and Sebastian Hörning[2] [iD]

[1]Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Stuttgart, Germany, [2]Centre for Natural Gas, The University of Queensland, Brisbane, Australia

Abstract Spatial structures of natural variables are often very complex due to the different physical chemical or biological processes which contributed to the emergence of the fields. These structures often show non-Gaussian spatial dependence. Unfortunately, there are only a limited number of approaches that can explicitly consider non-Gaussian behavior. In this contribution, a very flexible way of defining non-Gaussian spatial dependence is presented. The approach is based on a kind of continuous deformation of fields with different Gaussian spatial dependence. Theoretical examples illustrate the methodology for a wide variety of non-Gaussian structures. A real-life example of groundwater quality parameters shows the practical applicability of the geostatistical model.

1. Introduction

Random fields play a very important role in the investigation of hydrological and meteorological variables. The spatial distribution of these variables is often very specific. The random fields used for the stochastic description of such fields can be divided into two parts - the univariate marginal distribution of the variable and the multivariate copula describing the spatial dependence (Bárdossy, 2006). In most cases, the copula of the multivariate Gaussian distribution is used for the description of the spatial dependence. This approach is advantageous due to its relatively simple form but has the disadvantage that many natural processes show no Gaussian types of dependence — for example, asymmetrical dependencies due to the underlying processes such as dispersion/diffusion (Guthke & Bárdossy, 2017) or advection (Bárdossy & Hörning, 2017). The importance of non-Gaussian spatial dependence structures has been demonstrated in many studies. Gomez-Hernandez and Wen (1998), for example, showed how different non-Gaussian models, all sharing the same marginals and the same covariance function, lead to different groundwater travel times. In Zinn and Harvey (2003) the authors demonstrated the effect of non-Gaussian spatial dependence on groundwater flow and mass transfer. In Haslauer et al. (2012) it was shown that spatial dependence of transmissivities cannot be adequately described using a Gaussian copula.

In Bárdossy (2006) a parametric non-Gaussian family of copulas (v-copulas) was introduced for the description of spatial variability. An interpolation and simulation approach using v-copulas was presented in Bárdossy and Li (2008). These copulas however are not very flexible and interpolation and conditional simulations are computationally very expensive, limiting its applicability in 3D cases. An alternative to describe complex spatial dependence is using vine-copulas as in Gräler (2014). The vine copula approach has the disadvantage that it can only accommodate a few points and does not produce random fields.

Non-Gaussian fields can be simulated using simulated annealing (Deutsch, 1992). The disadvantage of this is that the non-Gaussian properties have to be defined explicitly, and the subsequent annealing has a very high computational cost. Due to the very high computational effort 3 dimensional realizations can only be obtained for very small fields. Multipoint geostatistics (MPS) (Mariethoz et al., 2010; Strebelle, 2002) offers a method to reproduce non-Gaussian features derived from training images. According to Journel (2003), the major source of uncertainty in MPS is the choice of the training image and not the fluctuations between multiple realizations. Further 3 dimensional training images are not available thus 3D simulations require different treatment for example, a combination of 2D images (Huang et al., 2022).

Recently in Papalexiou et al. (2021) the authors published interesting random field constructions related to complex natural processes. Their simulations are based on Gaussian fields and subsequent geometrical transformations which are used to simulate complex patterns and motion, for example, representing advection of rainfall,

spiraling fields that resemble weather cyclones, or fields converging to a point. The authors also presented an extensive summary of the methods and application of random fields, therefore the reader interested in this topic should refer to that publication. The approach presented here complements that work from the non-Gaussian simulation perspective.

In this contribution we present a new flexible method to define non-Gaussian spatial dependence. The method is explicit — defining random fields with flexible two or three dimensional spatial copulas. Realizations of the corresponding random fields are very simple to simulate, and their conditioning to observation points requires a few iterations.

This study is structured as follows, after this introduction a brief description of copulas is presented, followed by the definition of the non-Gaussian random fields for the infinite and the finite case. In the fourth section examples of specific random fields are presented. The fifth section describes the application of the corresponding copula for selected groundwater quality parameters. The paper ends with a discussion and conclusions.

2. Copulas

In the following, the basics of copulas are briefly reviewed. For further information the interested reader is referred to Bárdossy (2006), Bárdossy and Li (2008), Joe (1997), and Nelsen (1999).

Copulas are multivariate distribution functions defined on the unit hypercube:

$$C : [0,1]^n = [0,1] \tag{1}$$

with all univariate marginals being uniformly distributed on [0,1]. Copulas are used to describe the dependence between random variables independently of their marginal distributions, thus monotonic transformations of the marginals do not influence the dependence structure. According to Sklar's theorem (Sklar, 1959) any continuous multivariate distribution $F(x_1, \ldots, x_n)$ can be represented with the help of a unique copula:

$$F(x_1, \ldots, x_n) = C\left(F_{x_1}(x_1), \ldots, F_{x_n}(x_n)\right) \tag{2}$$

where $F_{x_i}(x)$ denotes the ith one-dimensional marginal distribution of the multivariate distribution. Copulas of multivariate distributions can be extracted by taking:

$$C(u_1, \ldots, u_n) = F\left(F_{x_1}^{-1}(u_1), \ldots, F_{x_n}^{-1}(u_n)\right) \tag{3}$$

where $F_{x_i}^{-1}(x)$ denotes the ith one dimensional inverse marginal distribution.

In the geostatistical context, spatial copulas can be used to describe the joint multivariate distribution corresponding to variables that are spatially distributed in the domain of interest. As in traditional geostatistics, it is assumed that the univariate marginal distribution corresponding to each point of the domain is the same, and the spatial dependence is translation invariant. This means that for any set of points s_i in the investigation domain such that $s_i + h$ is also in the domain, the spatial copula of the multivariate distribution is defined as:

$$
\begin{aligned}
C_S(u_1, \ldots, u_k) &= P(F_Z(Z(s_1)) < u_1, \ldots, F_Z(Z(s_k)) < u_k) \\
&= P(F_Z(Z(s_1 + h)) < u_1, \ldots, F_Z(Z(s_k + h)) < u_k) \\
&= C_{S+h}(u_1, \ldots, u_k)
\end{aligned}
\tag{4}
$$

3. Methodology

Gaussian random fields are relatively simple to generate in any dimensions. If two random fields are generated using the same random numbers but different spatial variograms then the corresponding random fields are called common random fields (Guthke & Bárdossy, 2012) and can be very similar, depending on how different the variograms are. This idea can be used to define non-Gaussian random fields which are combinations of Gaussian random fields with different variograms. For the definition the formulation of Gaussian random fields using Fast Fourier transform moving average (FFT-MA) approach as suggested in Ravalec
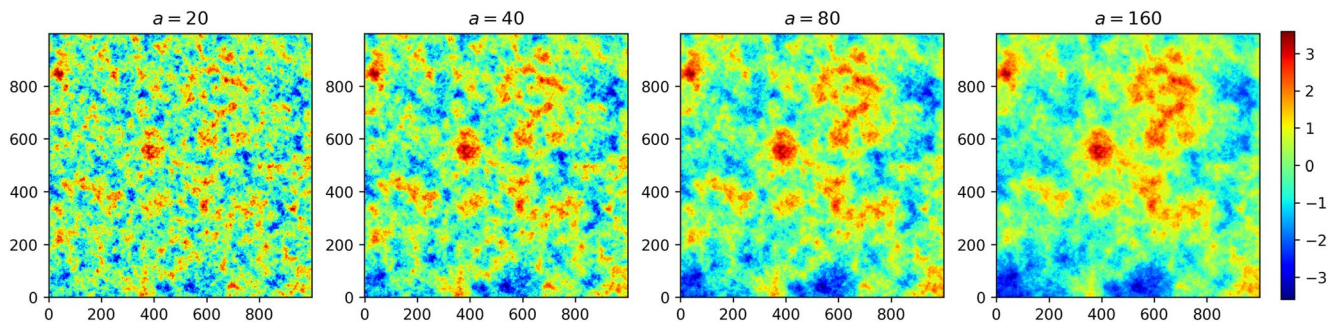
Figure 1. Common random fields simulated using Fast Fourier transform moving average. All fields share the same random numbers and an exponential variogram, however, the variogram range "a" is increasing from left to right.

et al. ([2000](#)) is used. According to FFT-MA, a Gaussian random field X with covariance function C(h) can be defined as:

$$X = \mathcal{F}^{-1}\left(\sqrt{\mathcal{F}(C(h))}\mathcal{F}(U)\right) \tag{5}$$

where U is Gaussian white noise, $\mathcal{F}$ denotes the Fourier transform and $\mathcal{F}^{-1}$ the inverse Fourier transform. FFT-MA can be used to simulate common random fields (Guthke & Bárdossy, [2012](#)). Therefore, the same Gaussianily guarantee that the marginal dist white noise, Y, is used with changing covariance functions to simulate a set of common random fields. Figure [1](#) shows an example of such a set of fields. It can be seen that the fields are similar due to the common Gaussian white noise, but the size of the structures is increasing due to the increasing variogram range.

3.1. Definition of the Non-Gaussian Field

The construction of the fields follows the same idea as presented in Bárdossy ([2023](#)). In that paper a set of multi-variate distributions are combined to form a new different distribution.

Let $C(\tau, h)$ be a set of covariance functions for $\tau \in [0, 1]$ and h being the spatial separation vector, which is continuous for each h in τ. τ is a parameter that links the different $C(\tau, h)$ and ensures a continuous transition. For each τ, a normal random field is generated using the same random numbers.

Definition: Let X_0 be a standard normal field with the covariance function $C(0, h)$. Define the field X_τ such that

$$X_\tau = \mathcal{F}^{-1}\left(\sqrt{\mathcal{F}(C(\tau,h))}\frac{\mathcal{F}(X_0)}{\sqrt{\mathcal{F}(C(0,h))}}\right) \tag{6}$$

This way a set of standard normal random fields are defined which are continuous in τ. The final field Z corresponds to the smallest root of the equation

$$X_\tau(u) = \Phi^{-1}(\tau)$$

$$Z(u) = \min\left\{\Phi^{-1}(\tau); X_\tau(u) < \Phi^{-1}(\tau)\right\} \tag{7}$$

The right-hand side of the set is not empty as for each u $X_\tau(u)$ is a continuous function and $\Phi^{-1}(0) = -\infty < X_0(u)$ and $X_1(u) < \Phi^{-1}(1) = +\infty$, where Φ^{-1} denotes the inverse of the standard normal distribution. While Equation [7](#) defines the random field $Z(u)$ *from below*, this definition could also be made *from above* which means that

$$Z(u) = \max\left\{\Phi^{-1}(\tau); X_\tau(u) > \Phi^{-1}(\tau)\right\} \tag{8}$$
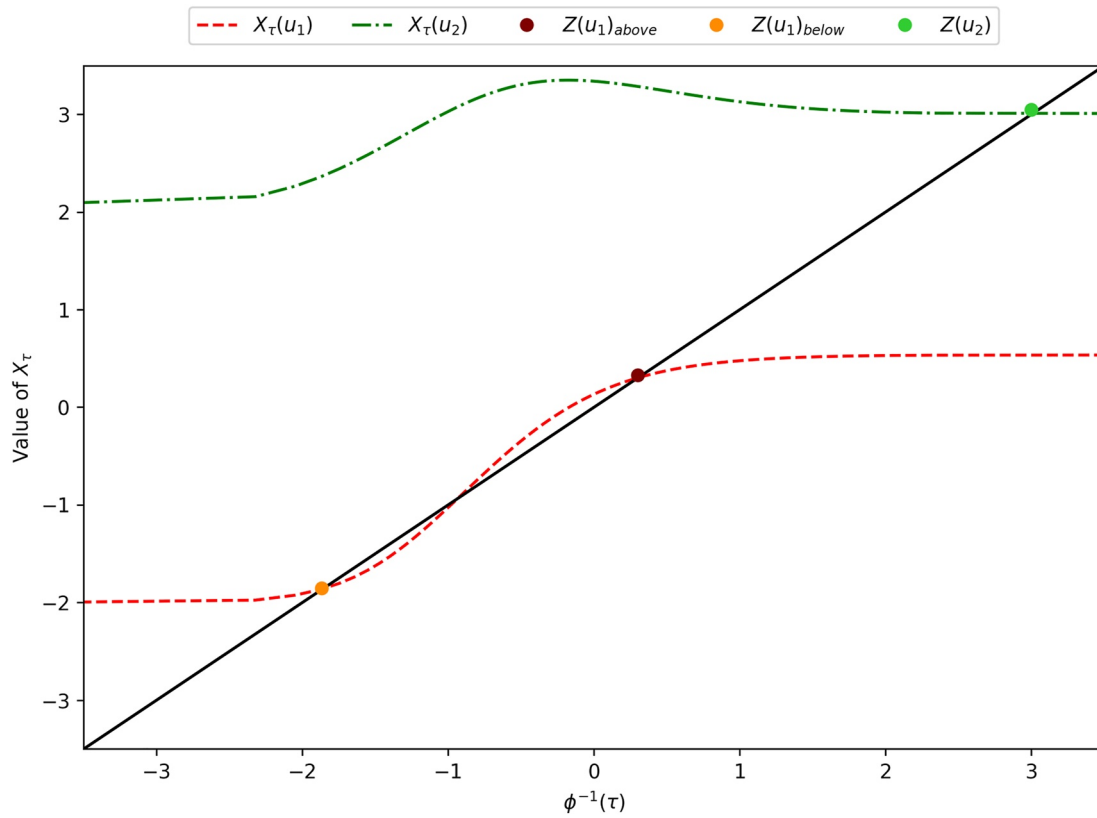
Figure 2. Definition of the random field Z.

Figure 2 illustrates the definition of the random field. For a given set of random numbers and for two selected locations u_1 and u_2, the values corresponding to the generated fields X_τ are plotted against $\Phi^{-1}(\tau)$. The line is continuous due to the continuity of the covariance function. The intersections with the diagonal represent the values assigned to $Z(u_1)$ and $Z(u_2)$, respectively. As the Figure shows the two different points get values from different X_τ fields. Further, it can be seen that location u_1 has two values $Z(u_1)_{above}$ and $Z(u_1)_{below}$ corresponding to the different definitions of the random field. If the definition in Equation 7 is applied, $Z(u_1)_{below}$ is the simulated value. If the definition in Equation 8 is applied, one gets $Z(u_1)_{above}$ as simulated value. It can be seen that there is no distinction for location u_2 as there is only one intersection with the diagonal. In general, the definition from above or from below does only affect less than 1% of all values, that is, the resulting fields are almost identical. As the definition of the random field does not necessarily guarantee that the marginal distribution of Z is normal, one can transform it to normal via $Z_N(u) = \Phi^{-1}(F_Z(Z(u)))$.

3.2. Parameters

The definition of the random field according to Equations 6 and 7 (or Equation 8) is very flexible and has a number of advantages:

1. Through the flexible selection of the covariance functions $C(\tau, h)$ very complex dependence structures can be defined - such as value-dependent geometrical anisotropies or different types of smoothness as shown in the examples in Section 4.
2. The simulation can obtain asymmetric random fields, that is, non-Gaussian random fields as shown in Section 4.
3. A large number of spatial copulas according to Equations 3 and 4 can be obtained.
4. The simulation procedure is very simple and efficient as only Gaussian random fields have to be generated to simulate the non-Gaussian field.

5. The spatial dependence can vary between fully dependent and independent.

This huge flexibility however also has some disadvantages. These are:

1. In the general form the field has an infinite number of parameters ($\tau \in [0, 1]$).
2. It is not trivial to find the appropriate covariance functions $C(\tau, h)$ based on observed data.

one way to overcome these problems is by using simple assumptions on the changes of the correlation function. The continuity of $C(\tau, h)$ can be assured by using parametric theoretical covariance functions $C(\theta, h)$ such that $\theta(\tau)$ is a continuous function.

The above definition requires an infinite number of parameters. A large number of simplifications leading to a small number of parameters are possible:

1. Linearly changing range of the variograms $\gamma(c, a, h)$

$$a(\tau) = (1 - \tau)a(0) + \tau a(1) \tag{9}$$

where c denotes the variogram sill and a denotes the variogram range.

2. Linearly changing nugget of the variograms $\gamma(c, a, h)$

$$c_0(\tau) = (1 - \tau)c_0(0) + \tau c_0(1) \tag{10}$$

3. Changing anisotropy

$$\phi(\tau) = (1 - \tau)\phi(0) + \tau \phi(1) \tag{11}$$

$$\lambda(\tau) = (1 - \tau)\lambda(0) + \tau \lambda(1) \tag{12}$$

with ϕ denoting the anisotropy ratio and λ the anisotropy angle.

4. Linearly changing variogram types $\gamma(a, h)$ with fixed range.

$$\gamma_\tau(h) = (1 - \tau)\gamma_0(h) + \tau \gamma_1(h) \tag{13}$$

The above models can be combined and many other definitions can be defined. The simulation of all these models is straightforward and very fast.

One can also define more complex structures by piece-wise linear definition of the parameters for example, instead of Equation 9 taking ranges $a(\tau_0 = 0), a(\tau_1), \ldots, a(\tau_J = 1)$:

$$a(\tau) = \frac{\tau - \tau_j}{\tau_{j+1} - \tau_j}a(\tau_{j+1}) + \left(1 - \frac{\tau - \tau_j}{\tau_{j+1} - \tau_j}\right)a(\tau_j) \quad \text{if} \quad \tau_j \leq \tau \leq \tau_{j+1} \tag{14}$$

Furthermore, then one parameter can also be changed - for example, the range and the anisotropy could both change in a linear or a piece-wise linear manner at the same time.

However, note that these kinds of models require more parameters and thus can only be assessed in the case of large data sets.

3.3. Numerical Simulation

For applications, only a finite set of different τ values have to be selected. Let:

$$0 = \tau_0 < \tau_1 < \ldots < \tau_m = 1$$

Then the definition (Equation 7) is in its finite form:

$$Z(u) = \begin{cases} X_{\tau_i}(u) & \text{if } \Phi^{-1}(\tau_{i-1}) < X_{\tau_i}(u) \le \Phi^{-1}(\tau_i) \\ & \text{and } X_{\tau_j}(u) > \Phi^{-1}(\tau_j) \, \forall j < i \\ \Phi^{-1}(\tau_{i-1}) & \text{if } X_{\tau_i}(u) \le \Phi^{-1}(\tau_{i-1}) \\ & \text{and } X_{\tau_j}(u) > \Phi^{-1}(\tau_j) \, \forall j < i \end{cases} \tag{15}$$

It is reasonable to select an equidistant spacing with $\tau_i = \frac{i}{m}$.

The above approximation allows building the distribution *from below* - starting with X_0. Note that as in the infinite definition Z_u is not dependent on the spatial structure of the fields X_τ for τ such that $Z(u) < \Phi^{-1}(\tau)$.

3.4. Conditioning

For the simulation of random fields, conditioning on observations is of great importance. Direct conditioning using conditional distributions is due to the definition of the random fields not possible. Instead, a Monte Carlo optimization-based conditioning approach as shown in the flowchart in Figure 3 can be performed. The purpose is to find a random field of the above type with $Z(u_i) = z_i$ for $i = 1, \ldots, n$ where n is the number of observations. As described in Figure 3, the procedure consists of two main steps. The first step is to simulate an initial realization with $Z(u_i) \approx z_i$. Therefore, a conditional Gaussian random field has to be simulated first. The parameters corresponding to $\tau = 0.5$ should be used for that simulation. Note that any conditional Gaussian simulation approach can be used. From that conditional field, one can calculate the corresponding random numbers Y via inverse FFT-MA. These random numbers are subsequently used to simulate the initial non-Gaussian field with $Z(u_i) \approx z_i$. In the second step, the conditioning is optimized. Therefore, the random numbers Y are updated in a window around the conditioning point locations. Note that there is no general rule for the window size, that is, it has to be adjusted for the specific problem at hand. Using the updated random numbers one can calculate the updated non-Gaussian field. From that updated field an objective function such as the sum of the squared differences is calculated. Note that other objective functions such as the root mean squared error (RMSE) could be used instead. If the objective function is less than a user defined threshold ϵ, the realization is accepted. If the objective function is greater than that threshold, the random numbers are updated again until a satisfactory match is obtained. It is worth mentioning that this conditioning approach is not exact. However, the conditioning error can be controlled using the threshold ϵ. Further, the conditioning error can be used to represent measurement uncertainties. Note that in the simplest case, the updating step is a pure Monte Carlo. However, updating can also be performed using other approaches, for example, via linear combinations of random numbers similar to the procedure presented in Hörning et al. (2019) which results in a more efficient optimization.

3.5. Interpolation and Uncertainty Estimation

The definition of these random fields is relatively complex and conditional distributions cannot be derived easily, therefore interpolation and uncertainty estimation can only be obtained using simulations. The mean and the uncertainty can be obtained using a suitably large number of conditional fields. The number of fields required to achieve a stable mean and a representative uncertainty estimation depends on the complexity of the individual case. However, as a rule of thumb, one should use at least 100 conditional realizations.

4. Examples

In this section, realizations of 2D spatial random fields demonstrate the large range of spatial dependencies covered by this family of copulas. Figures 5–14 show realizations of selected special cases and the corresponding variograms (isotropic, minor, and major directions), spatial asymmetry function, and spatial copulas. The spatial asymmetry function is a bivariate third order moment that is defined as:

$$A(h) = \frac{1}{N(h)} \sum_{u_i - u_j \approx h} \left(F(Z(u_i)) + F(Z(u_j)) - 1 \right)^3 \tag{16}$$
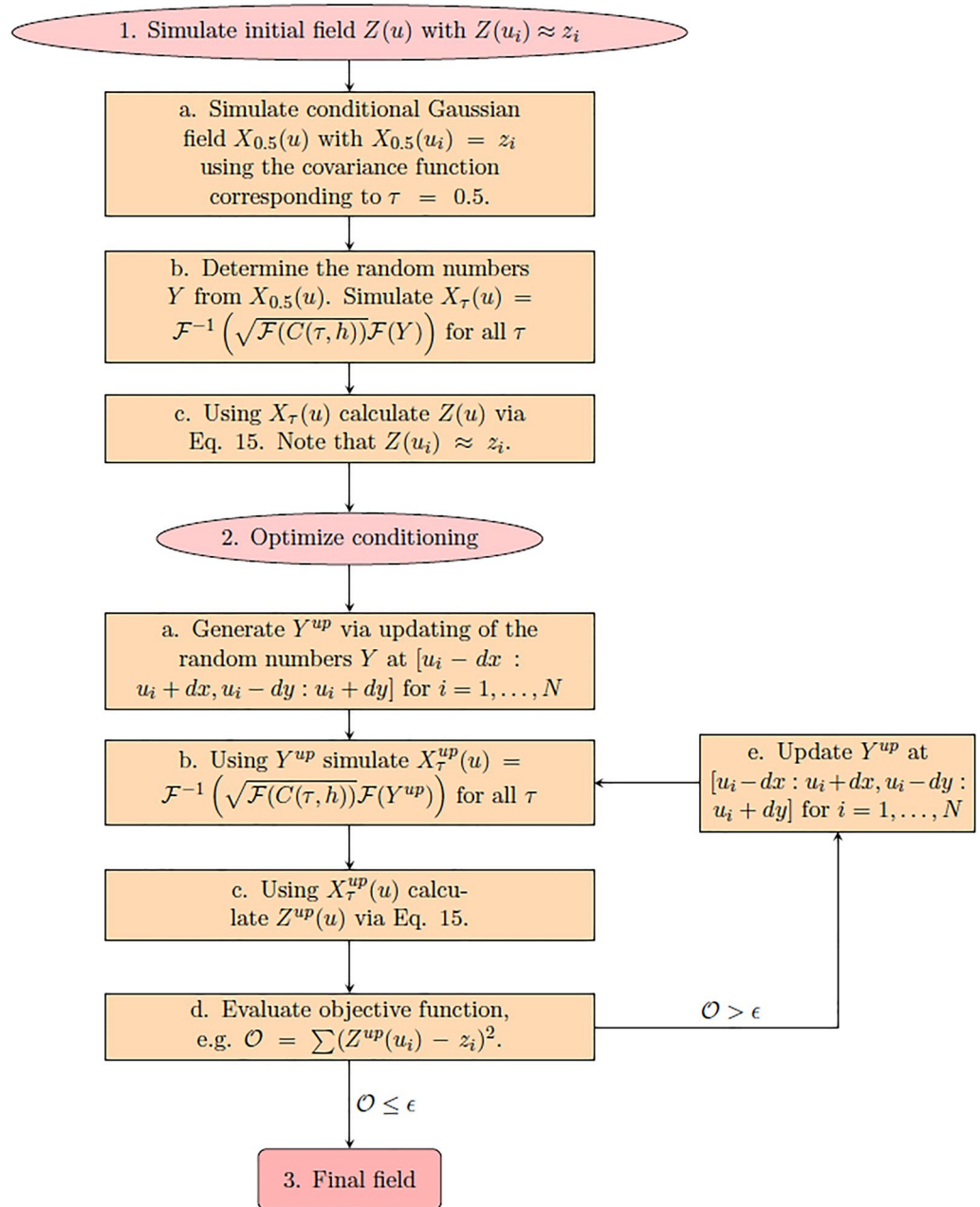
Figure 3. Flowchart for the proposed conditioning algorithm.

where F is the univariate marginal of Z and $N(h)$ denotes the number of pairs where $Z(u_i)$ and $Z(u_j)$ are separated by a distance of approximately h. The spatial asymmetry function goes beyond the average description of dependence of a variogram as it measures the different dependence between high and low quantiles. A comprehensive description of the spatial asymmetry and its interpretation can be found in Guthke and Bárdossy (2017). Another type of spatial asymmetry function is the direction-dependent asymmetry. While the function defined in Equation 16 measures order based asymmetry, its direction-dependent counterpart measures directional asymmetry. It is defined as:

$$A_d(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{u_i - u_j \approx \mathbf{h}} (F(Z(u_i)) - F(Z(u_j)))^3 \tag{17}$$
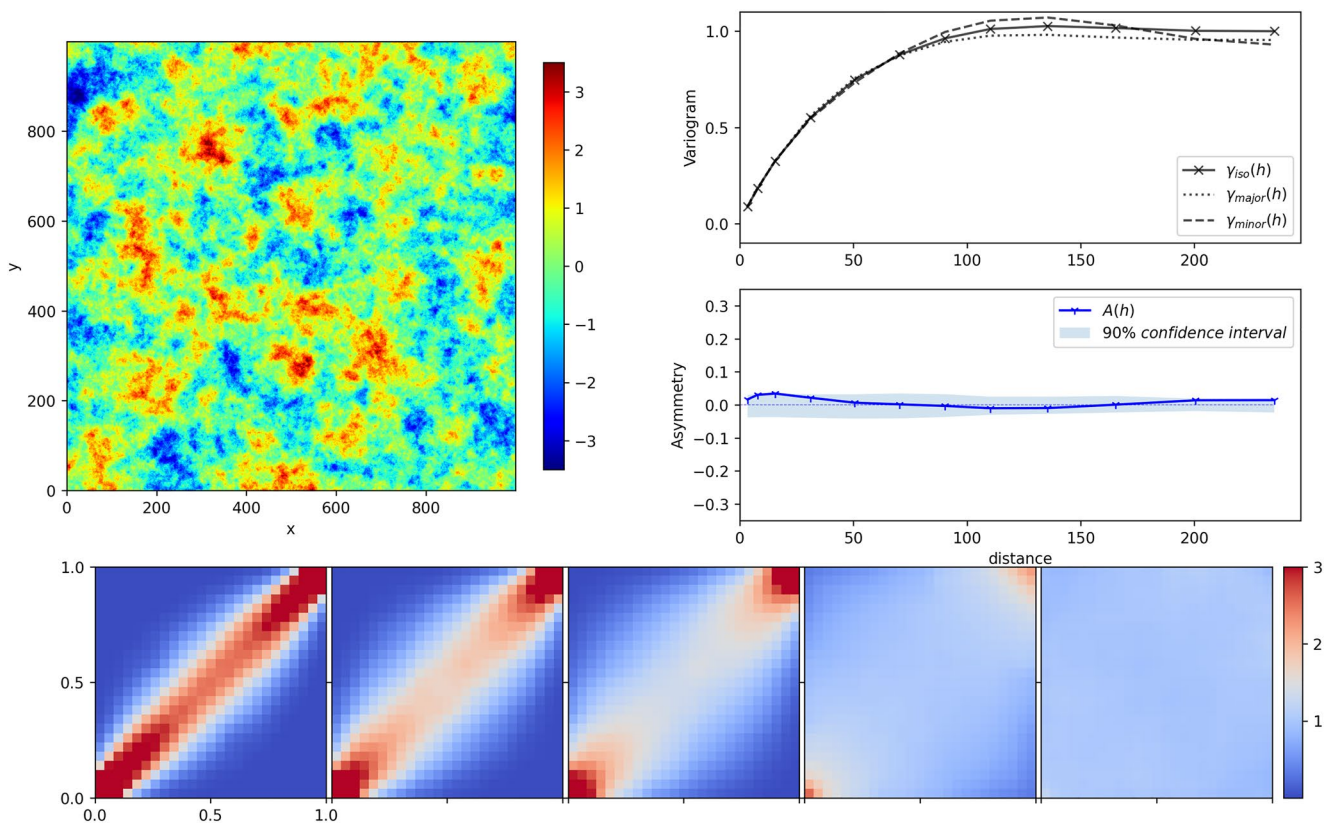
Figure 4. Base case field (isotropic Gaussian field with exponential variogram with range 40) with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry.

where $\boldsymbol{h}$ is a distance vector with a direction. Equation 17 can distinguish dependencies in different directions that could for example, arise from advection processes. In the following examples, only order asymmetry is considered but for more information on direction-dependent asymmetry the interested reader is referred to Bárdossy and Hörning (2017).

All fields were generated as common random fields (Guthke & Bárdossy, 2012), using the same random numbers for the simulation. Thus, the fields are similar but one can also see the differences depending on the corresponding model. Note, that the structures are in most cases clearly asymmetrical with different dependencies of high and low values. To demonstrate that these asymmetries are significantly non-Gaussian, each example shows the 90% confidence interval for the asymmetry that can be obtained from Gaussian fields. 100 Gaussian realizations are simulated using the variogram model corresponding to $\tau = 0.5$ and their asymmetries are calculated according to Equation 16 to determine these confidence intervals. Further note, that all fields exhibit a standard normal marginal distribution, however, as described in Section 3 any marginal distribution (parametric as well as non-parametric) can be imposed.

Figure 4 shows the base case field for comparison. It is an isotropic Gaussian spatial random field with an exponential variogram with a range of 40, simulated using FFT-MA (Ravalec et al., 2000). Figure 4 also shows the corresponding variograms, the spatial asymmetry function, and the spatial copulas for a few selected separation distances. From the spatial asymmetry function, it can be seen that the field is Gaussian as it exhibits close to zero asymmetry which lies within the 90% confidence interval. This can also be observed from the spatial copulas which show similar densities for the low and the high values, indicating symmetrical spatial dependence.

Figure 5 shows a field simulated with a linearly changing range of the variograms according to the definition in Equation 9. An exponential variogram with its range changing from 5 to 80 has been applied. It can be seen that the low values correspond to the short ranges, while the high values correspond to the long ranges. The variogram plots show that the field is isotropic. Further, apart from a small change in the variogram nugget, the
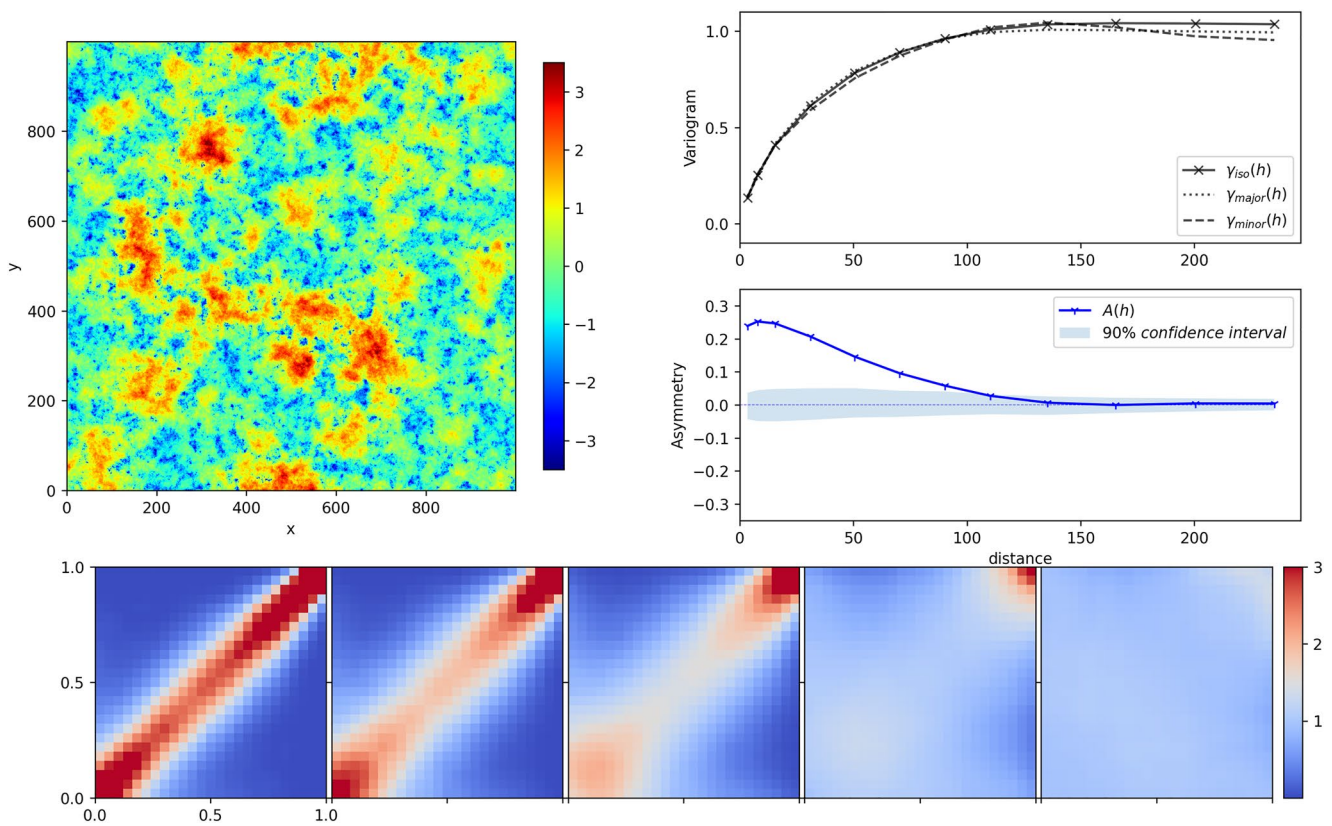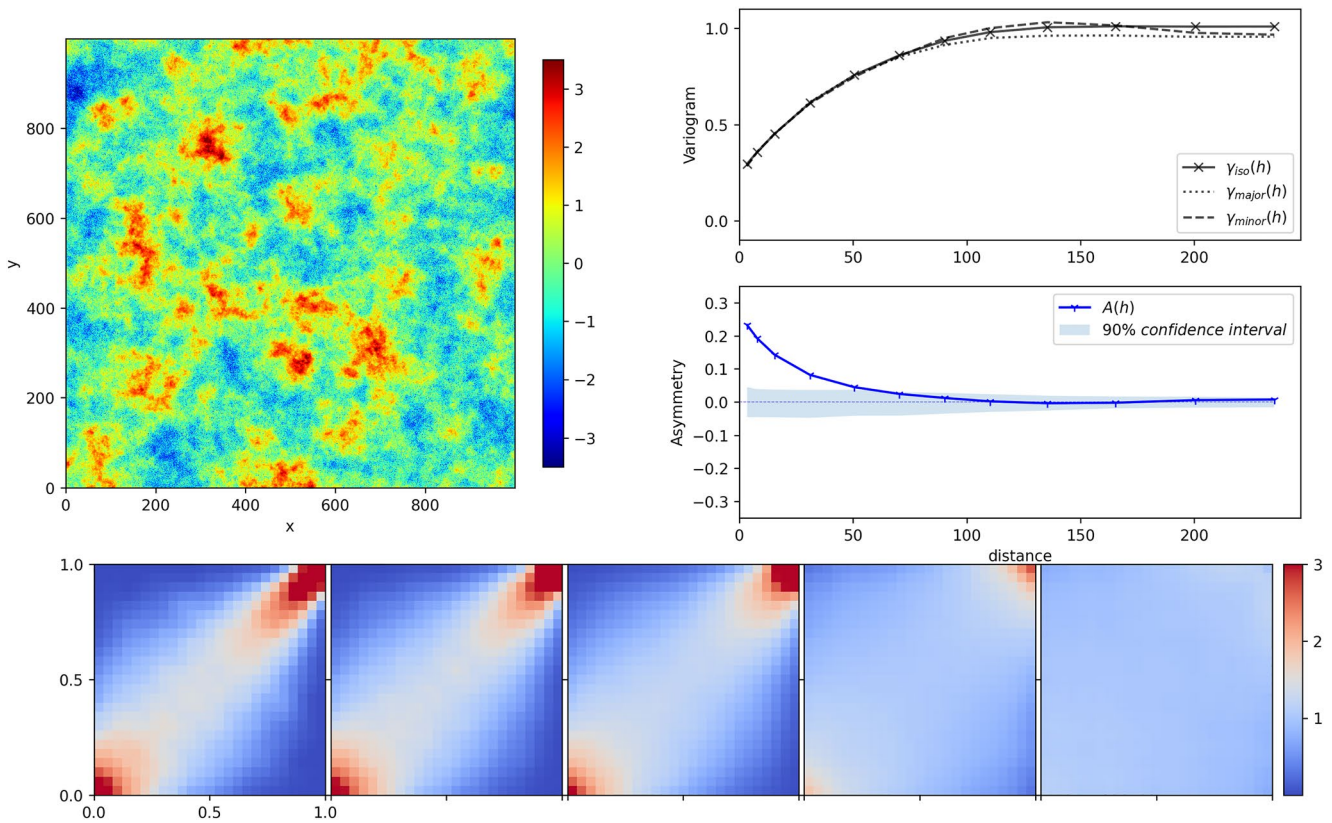
Figure 5. Simulated field with a linearly changing range of the variograms (exponential variogram with range changing from 5 to 80) according to Equation 9 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

variograms are almost identical (the average variogram range is similar to the variogram range of the base case) to the variograms corresponding to the base case shown in Figure 4. This indicates that variograms are not able to distinguish the differences between the fields. The asymmetry function, however, shows a strong positive asymmetry for separation distances up to approximately 100. It is outside the 90% confidence interval thus the field is significantly non-Gaussian. The positive asymmetry indicates that high values exhibit a stronger dependence for these separation distances compared to the low values. The high values form larger clusters while the low values form smaller clusters. This can also be seen in the spatial copulas which exhibit higher densities for high values (the upper right corner) compared to the densities for low values (lower left corner).

Figure 6 shows a field simulated using a linearly changing nugget of the variograms according to the definition in Equation 10. Again, an exponential variogram has been used. The variogram range was constant for all values of τ with a value of 50, while the nugget of the variogram linearly changed from 50% to 0%. The higher nugget has been imposed on the low values, gradually decreasing to a zero nugget for the high values. This change in the nugget can clearly be seen in the field where the low values exhibit a nugget effect while the high value have no nugget effect. The variograms again indicate that the field is isotropic and a nugget can clearly be identified. However, the variograms cannot distinguish the difference in the nugget for low and high values. The spatial copulas enable a more detailed insight. It can be seen that the densities for the lower values are more dispersed than the densities for the high values, indicating a higher nugget effect for the low values. The spatial copulas also exhibit a positive asymmetry, that is, a higher density for the high values compared to the low values. This positive asymmetry can also be seen in the spatial asymmetry function which is positive for separation distances up to approximately 70. Again, the asymmetry function is clearly outside the 90% confidence interval, indicating significant non-Gaussianity.

Figure 6. Simulated field with a linearly changing nugget of the variograms (exponential variogram with range 50, nugget changing from 50% to 0%) according to Equation 10 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

Figure 7 shows a field simulated using a combination of Equations 9 and 10, that is, with a linearly changing variogram range and a linearly changing nugget effect. Note, both the range and the nugget change according to the same parameter τ. The variogram model is again exponential with its range changing from 5 to 80 and its nugget changing from 50% to 0%. The short range, high nugget has been imposed on the low values, and the long range, zero nugget on the high values. It can be seen that the low values exhibit a short range structure as well as a nugget effect. The range gradually increases and the nugget gradually decreases with increasing values. The variograms show that the field is isotropic with a nugget effect. Again, the variograms cannot distinguish the difference in the nugget effect for the low and the high values. This difference can again be seen in the spatial copulas which exhibit a more dispersed density for the low values compared to the high values. It can also be seen that this field has a stronger positive asymmetry compared to the field in Figure 6. The copulas have higher densities for the high values and lower densities for the low values compared to the ones in Figure 6. This is also apparent from the spatial asymmetry function which is outside the 90% confidence interval with higher values up to a separation distance of approximately 100.

Figure 8 shows a field simulated with a linearly changing anisotropy according to Equation 12. The variogram is again exponential with a constant major range of 50. The anisotropy exhibits a constant angle of 25° but the ratio between the range of the major and minor axes changes from 1 to 0.2. Thus, for the high values, the variogram is isotropic (major and minor range of 50), linearly changing to a more and more anisotropic variogram for the low values. For the lowest values, the major range is 50 while the minor range is 10. This can be seen in the field where the low values exhibit very strong anisotropy and the high values are isotropic. The major and the minor direction variograms also show this anisotropy, however, the differences are not as pronounced as the variogram is an average description of spatial dependence. The change from isotropic high values to anisotropic low values also introduces spatial asymmetry. For separation distances up to approximately 100, the spatial asymmetry
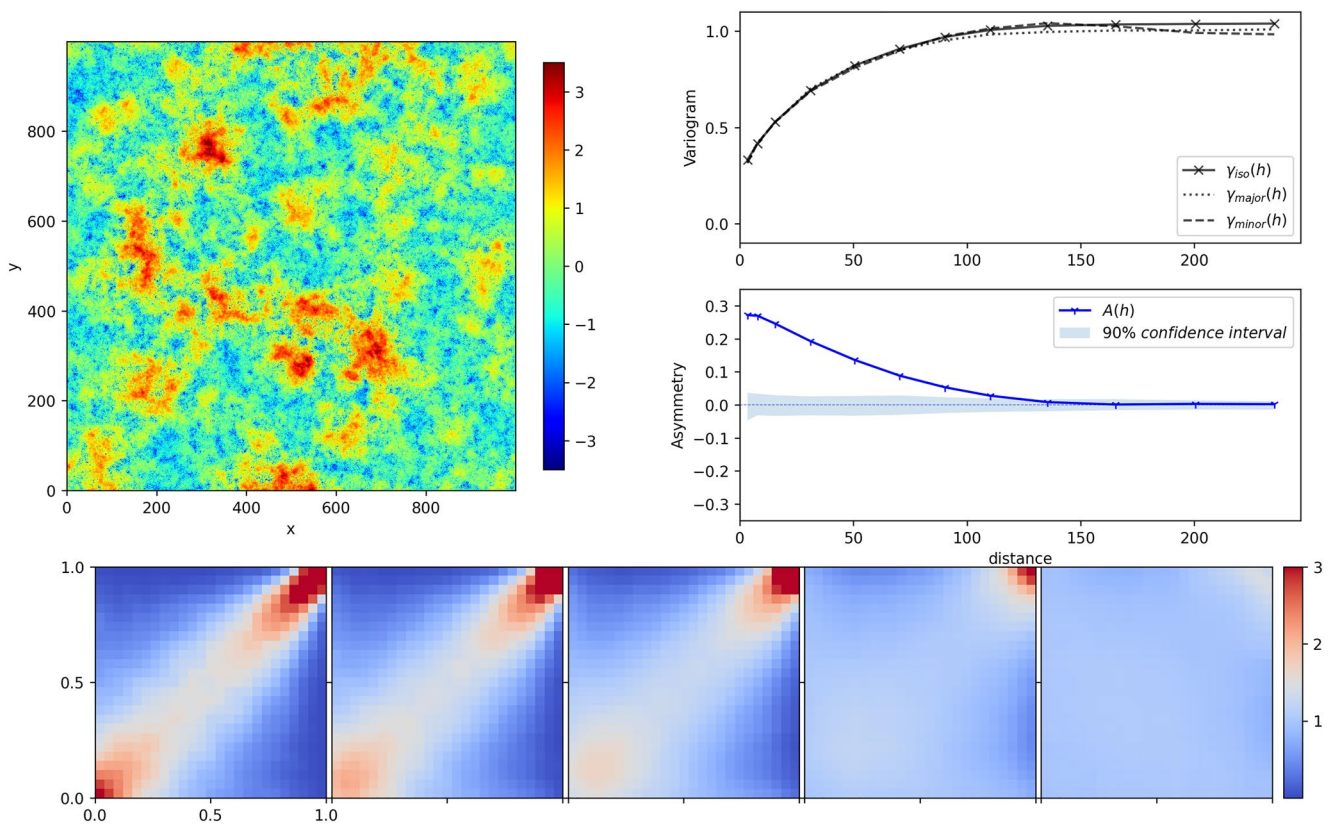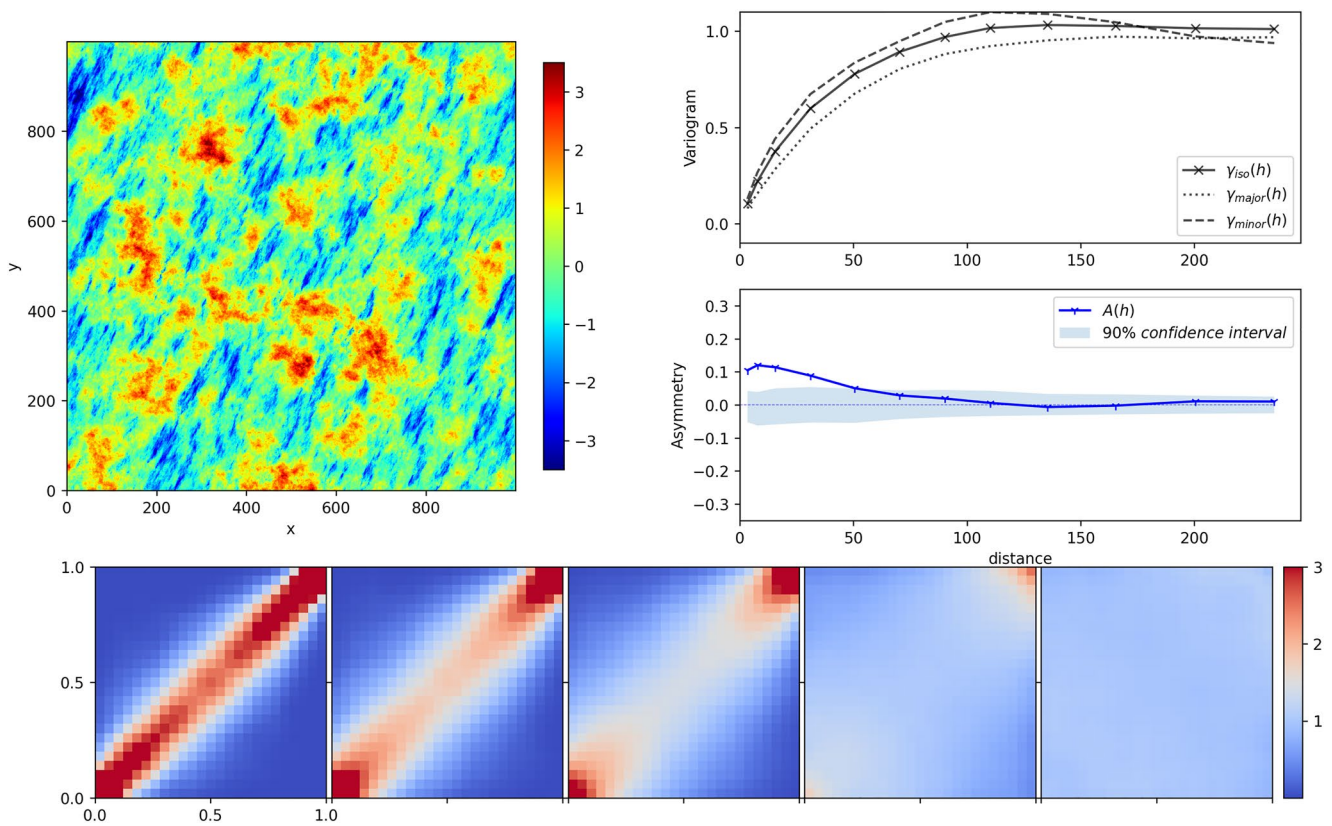
Figure 7. Simulated field with a linearly changing range and linearly changing nugget of the variograms (exponential variogram with range changing from 5 to 80 and nugget changing from 50% to 0%) according to a combination of Equations 9 and 10 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

function indicates slightly positive asymmetry which is outside the 90% confidence interval for distances up to 50. This can again also be seen in the spatial copulas which have higher densities for the high values than for the low values.

Figure 9 shows a field with a linearly changing range and a linearly changing anisotropy according to a combination of Equations 9 and 12. As in the previous examples, the variogram is exponential. It has a minor and major range of 80 (i.e., isotropic) for the high values, linearly changing to a major range of 5 and a minor range of 1 (anisotropy ratio of 0.2) for the low values. The anisotropy has again a constant angle of 25°, independent of the ratio. It can be seen that the field has large clusters of isotropic high values and small clusters of anisotropic low values. Again, due to its average nature, the variograms for the minor and major directions only show small differences. The combined change of range and anisotropy introduces a strong positive asymmetry which can be seen in the asymmetry function as well as in the spatial copulas.

Figure 10 shows another field with a linearly changing range and a linearly changing anisotropy according to Equations 9 and 12. This field, however, exhibits the opposite anisotropy to Figure 9. Here, the high values are simulated with a major range of 80 and a minor range of 16 (anisotropy ratio of 0.2) while the low values are simulated with a minor and major range of 5 (i.e., isotropic). The anisotropy angle is again 25°, independent of the ratio. As the anisotropy is in the long range part of the field, it is more distinct in the minor and major direction variograms. The asymmetry function again exhibits a positive asymmetry for separation distances up to approximately 100. This can again also be seen in the spatial copulas which show a higher density for high values than for low values.
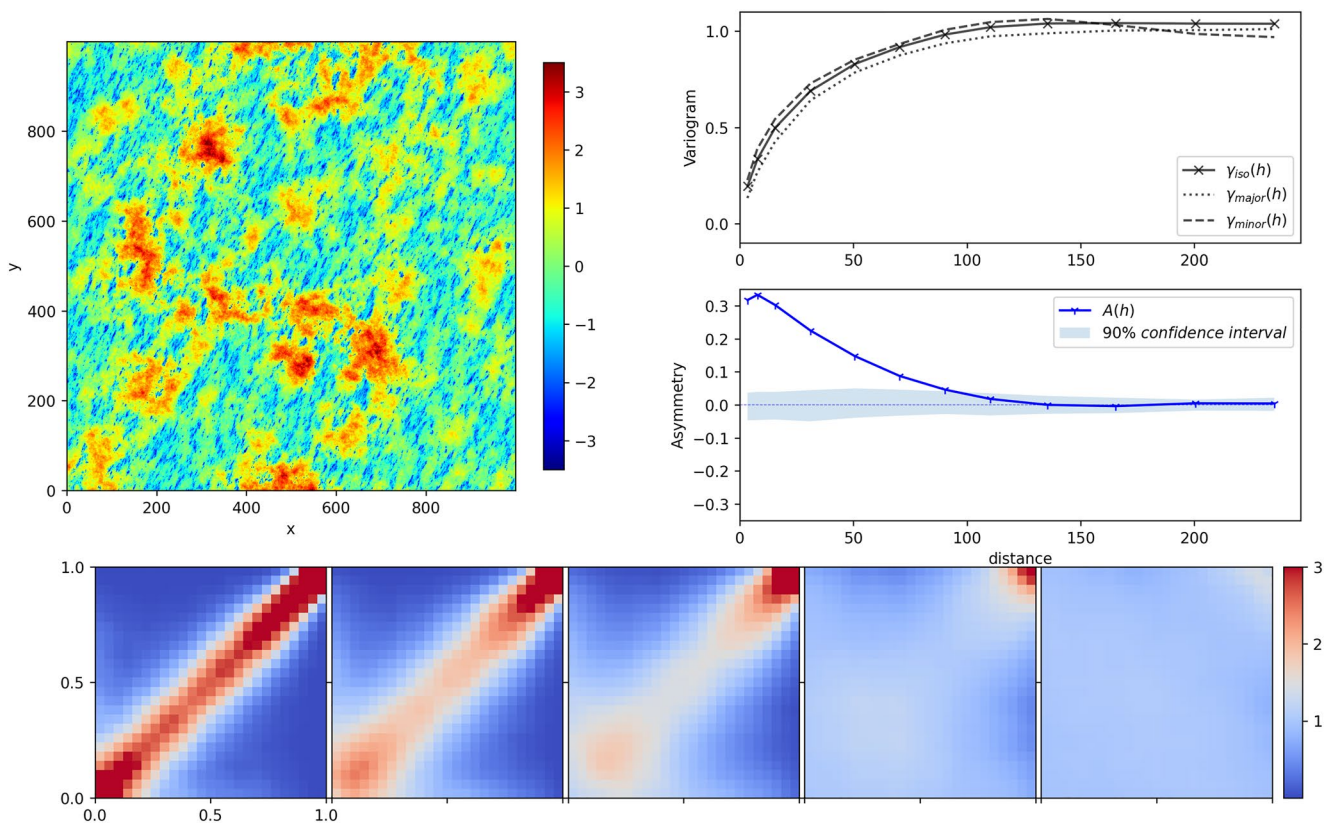
Figure 8. Simulated field with a constant variogram range and linearly changing anisotropy ratio (exponential variogram with range 50 and anisotropy ratio changing from 1 to 0.2) according to Equation 12 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

Figure 11 shows again a field with a linearly changing range and a linearly changing anisotropy according to a combination of Equations 9 and 12, but with a negative asymmetry. The field has been simulated with an exponential variogram with a major and minor range of 5 (i.e., isotropic) for the high values and a major range of 80 and a minor range of 16 for the low values (anisotropy ratio of 0.2). Again, the anisotropy angle is independent of the ratio and 25°. It can be seen that the high values form small isotropic clusters while the low values exhibit strong anisotropy with a long range in the major direction. As in the previous example, the anisotropy is clearly visible in the major and minor direction variograms. As the low values exhibit larger clusters compared to the high values, the spatial asymmetry of this field is negative as indicated by the spatial asymmetry function. The spatial copulas also show a negative asymmetry with higher densities for low values and lower densities for high values.

Figure 12 shows a field simulated with a linearly changing range, anisotropy ratio, and anisotropy angle according to a combination of Equation 9, Equation 12, and Equation 11. This example has again been simulated using an exponential variogram model. The high values have a major range of 80 and a minor range of 40 (anisotropy ratio of 0.5) with a 25° angle. The low values have a major range of 5 and a minor range of 1 (anisotropy ratio of 0.2) with a 70° angle. The difference in the angle and in the anisotropy ratio can clearly be seen in the field. The variograms for the major and the minor direction indicate anisotropy. As the high values form much larger clusters than the low values, the asymmetry function is positive for separation distances up to approximately 100. This can again also be seen in the spatial copulas which exhibit very high densities for the high values and very low and dispersed densities for the low values.

Figure 13 shows a field simulated with a linearly changing variogram model according to Equation 13. The field has been simulated using a Gaussian variogram for the low values that linearly changes to an exponential

Figure 9. Simulated field with a linearly changing variogram range and linearly changing anisotropy ratio (exponential variogram with range changing from 5 to 80 and anisotropy ratio changing from 1 to 0.2) according to Equations 9 and 12 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

variogram for the high values. Both variograms are isotropic with a range of 60. It can be seen that, as a result of the Gaussian variogram, the low values form smoother clusters while the high values are less smooth due to the exponential variogram. The variograms for the major and minor directions show no anisotropy and the asymmetry is slightly negative for separation distances up to approximately 50. This negative asymmetry is a result of the smoother clusters of low values and the less smooth high values.

Figure 14 shows a field simulated with a linearly changing variogram model and a linearly changing range according to a combination of Equation 13 and Equation 9. As in the previous example, this field has been simulated using an isotropic Gaussian variogram for the low values that changes linearly to an isotropic exponential variogram for the high values. The variogram range also changes linearly from 60 for the low values to 5 for the high values. This results in large, smooth clusters of low values, and small, less smooth clusters of high values. The minor and major variograms show no anisotropy. The large clusters of smooth, low values in combination with the small clusters of less smooth, high values lead to a very strong negative asymmetry as shown in the asymmetry function as well as the spatial copulas.

The previous examples demonstrate the very high diversity of the simulated fields. Further combinations of the parameters defined in Equations 9–13 would allow the simulation of even more complex structures.

4.1. Parameter Estimation

The definition of a non-Gaussian field according to Equation 7 is very flexible. It allows the definition of fields with a very high number of parameters. The parameter estimation for the presented approach depends on the choice of the model parametrization.
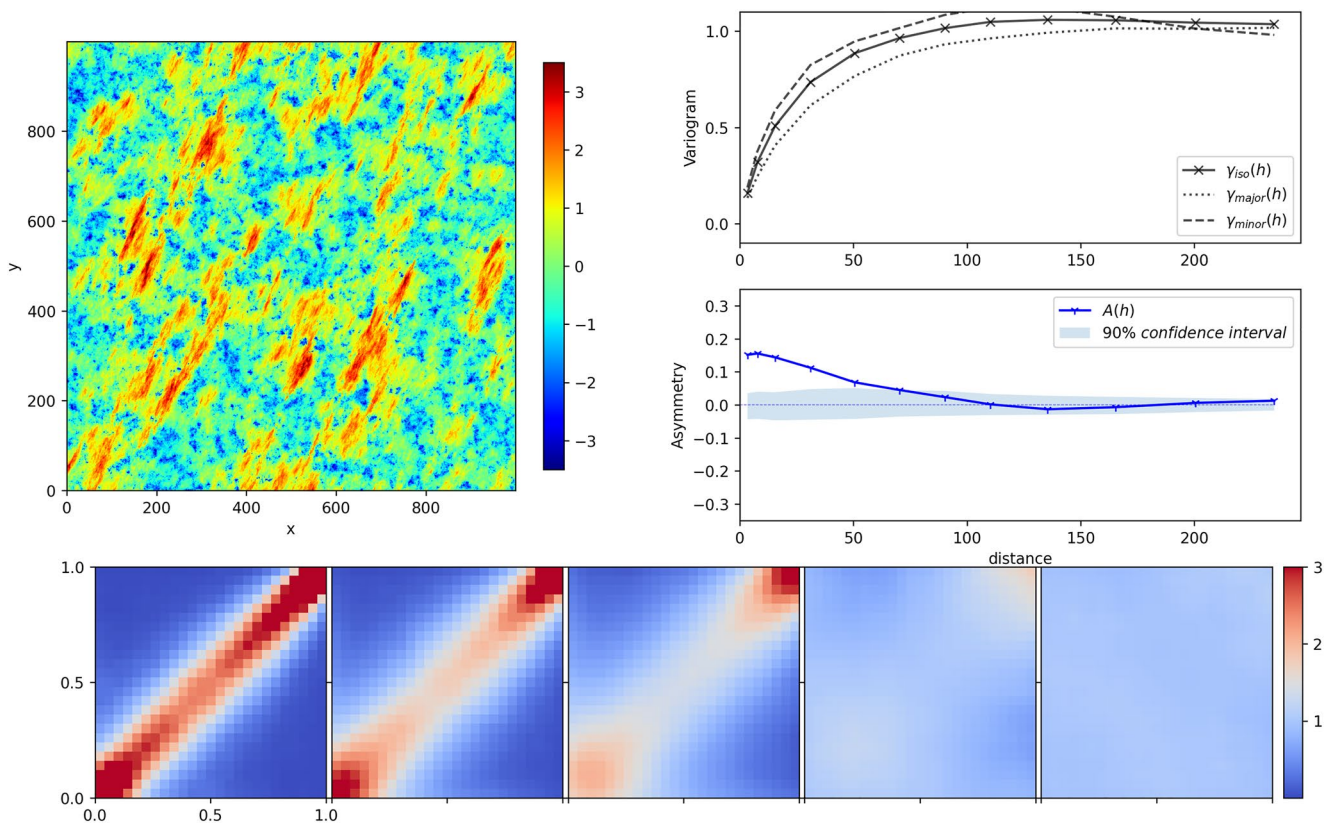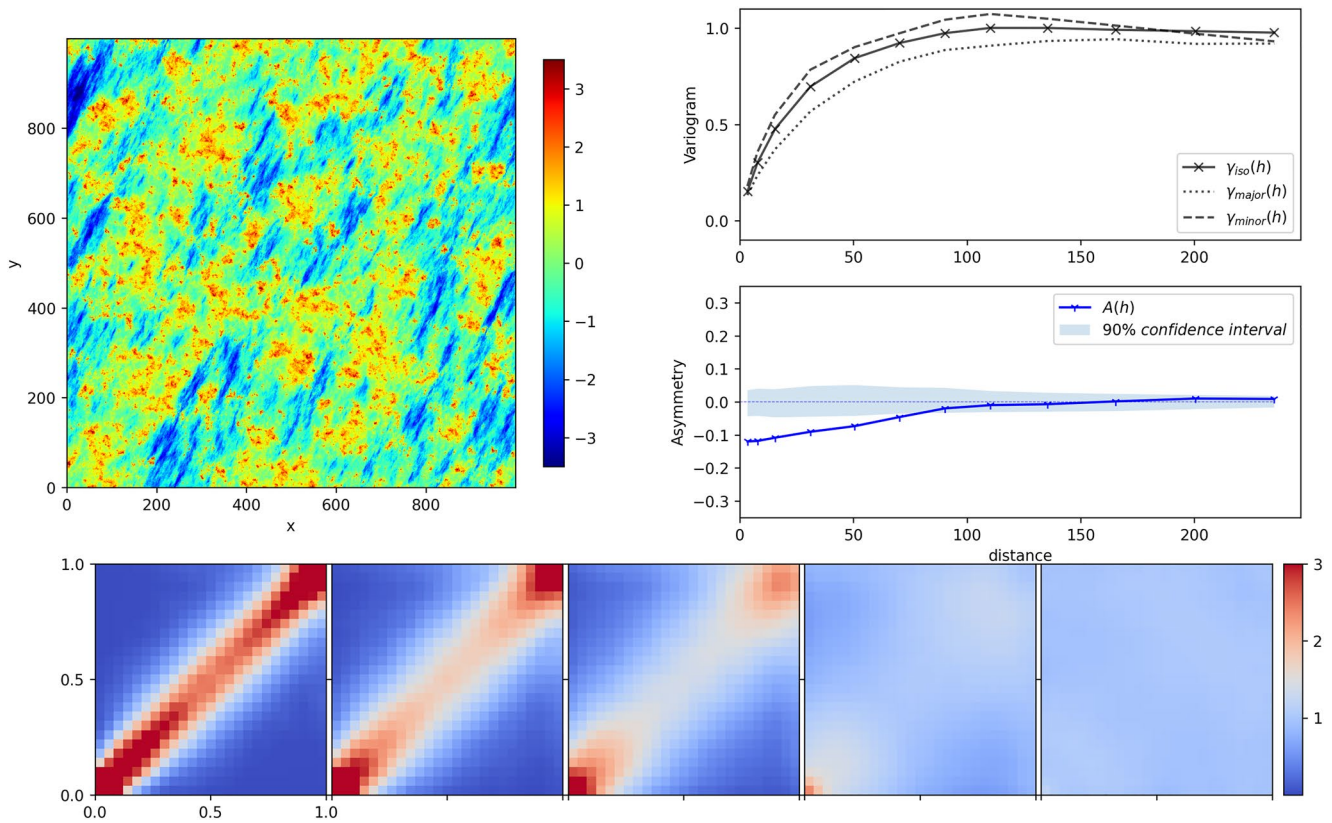
Figure 10. Simulated field with a linearly changing variogram range and linearly changing anisotropy ratio (exponential variogram with range changing from 5 to 80 and anisotropy ratio changing from 0.2 to 1) according to Equations 9 and 12 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

A reasonable method to estimate parameters is via Monte Carlo simulation. Once a model like those in Equations 9–13 is chosen, one can select a set of desired statistics such as the variogram and the spatial asymmetry function. Fields with randomly selected parameters (according to the selected model) are subsequently simulated until a set of parameters is obtained for which the observed and the simulated statistics match. The procedure is not difficult to realize as the unconditional simulation of the fields is very fast. The procedure can be summarized as:

1. The observed data are transformed to normal using a normal score transformation.
2. The experimental statistics (e.g., variogram and the spatial asymmetry function) are calculated from the observations.
3. A model is selected according to Equations 9–13. Note that combinations of these models are possible as shown in Section 4.
4. A random set of parameters of the selected model is chosen.
5. A realization with the selected model parameters is simulated.
6. The experimental statistics of the simulated field are calculated.
7. The observed and the simulated statistics are compared using a suitable objective function. In this paper, the sum of the root mean squared errors (RMSE) of the observed and simulated statistics is used but other objective functions such as the mean absolute error (MAE) could be used instead. If they fit well (e.g., the objective function value is below a predefined threshold), the parameters are accepted, otherwise, a new set of parameters is randomly generated and the procedure continues at step 5.

Figure 11. Simulated field with a linearly changing variogram range and linearly changing anisotropy ratio (exponential variogram with range changing from 80 to 5 and anisotropy ratio changing from 0.2 to 1) according to Equations 9 and 12 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

Note that one can also consider statistics of the untransformed variables for the parameter estimation. In this case, one has to transform the simulated fields (step 5) from normal back to the original marginal distribution. Further note that the above described approach can also be coupled with a suitable numerical optimization algorithm such as Powell's conjugate direction method (Powell, 1964).

For the model defined in Equation 9, the estimation can be started with the following assumption:

$$\gamma_Z(h) \approx \gamma_{X_{\frac{1}{2}}}(h) \tag{18}$$

To estimate $a(0)$ and $a(1)$, that is, the variogram ranges corresponding to $\tau = 0$ and $\tau = 1$ as defined in Equation 9, the spatial asymmetry function can be considered. One can simulate fields with the variogram shape corresponding to $\gamma_Z(h)$ with ranges from $a(0)$ to $a(1) = 2a\left(\frac{1}{2}\right) - a(0)$ and calculate the corresponding asymmetry function. The $a(0)$ for which the asymmetry function is the closest to the observed (Z) asymmetry is then selected.

A similar procedure can be selected for the model defined according to Equation 10 and a combination of the two models is also possible.

Another approach including models with anisotropy changes can be obtained by using indicator variograms corresponding to different thresholds θ:

$$I_\theta(u) = \begin{cases} 1 & \text{if } Z_{(u)} \leq \theta \\ 0 & \text{else} \end{cases} \tag{19}$$
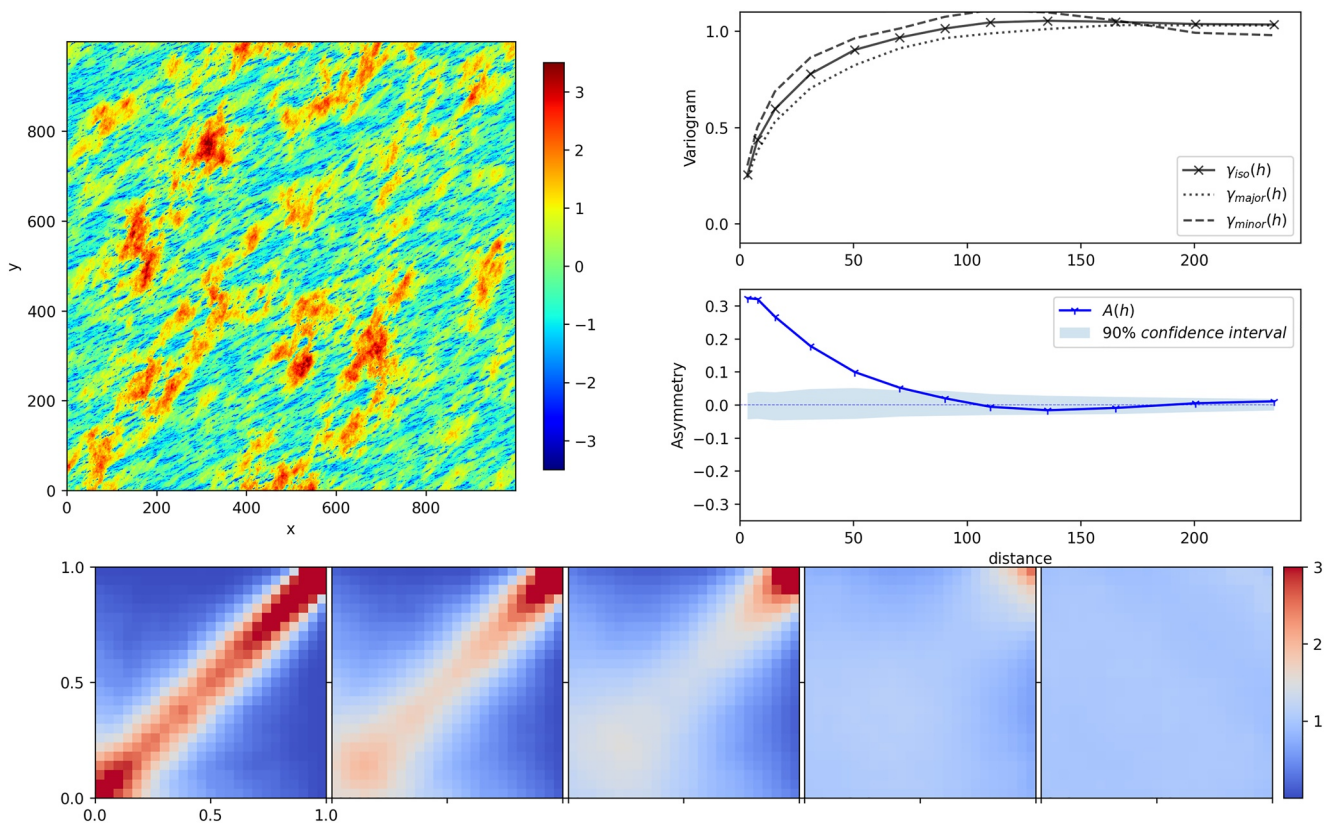
Figure 12. Simulated field with a linearly changing variogram range, linearly changing anisotropy ratio, and linearly changing anisotropy angle (exponential variogram with range changing from 5 to 80, anisotropy ratio changing from 0.2 to 0.5, and angle changing from 70° to 25°) according to Equation 9, Equation 12, and Equation 11 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

Spatial correlation functions of the indicator variables corresponding to the observed data and those from unconditional simulations can be compared, and parameters can be estimated on this basis.

4.2. Parameter Estimation Example

The above described parameter estimation approach has been applied to two of the presented fields in Section 4. The fields shown in Figures 5 and 10 have each been randomly sampled at 300 and 3,000 locations and their variograms and spatial asymmetry functions have been calculated based on these sampled values. Note that the variograms and asymmetry functions are calculated using the FFT-based approach described in Marcotte (1996) which allows the fast calculation of variogram and asymmetry maps. Thus, no separation distance lags need to be defined and the anisotropy can be obtained directly from these maps. The models according to Equations 9–13 are assumed to be known for both fields. For the field shown in Figure 5 two parameters, the two variogram ranges a (0) and a (1) need to be estimated. The results are shown in Table 1. For the field shown in Figure 10 five parameters need to be estimated. These are the two variogram ranges (a (0) and a (1)), two anisotropy ratios (ϕ(0) and ϕ(1)), and the anisotropy angle λ. The results are shown in Table 2. Note that for both examples the above described approach coupled with Powell's conjugate direction method has been applied to estimate the parameters.

The results indicate that the parameter estimation approach works well for both examples. Table 1 shows that for both cases (300 and 3,000 observations) corresponding to the field in Figure 5 the estimated parameters are close to the true parameters. For the field shown in Figure 10, Table 2 shows that the case with 3,000 observations leads to a better estimation than with 300 observations. Due to the higher complexity of this example, the case
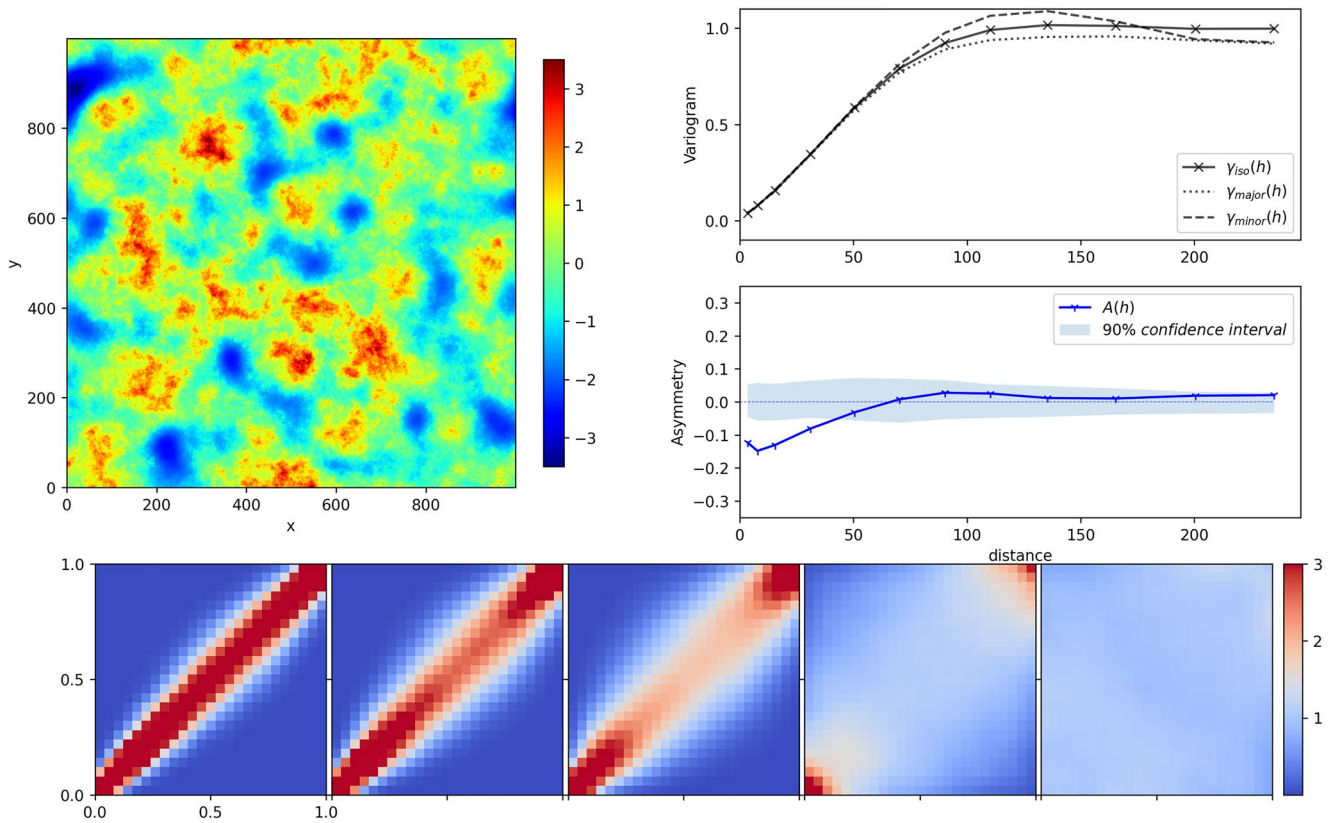
Figure 13. Simulated field with a linearly changing variogram model (Gaussian variogram changing to an exponential variogram, both isotropic with a constant range of 60) according to Equation 13 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

with more observations results in a better parameter estimation. However, the estimated parameters based on 300 observations are still satisfactory.

5. Applications

5.1. Synthetic Conditional Simulation Example

The conditioning approach described in Section 3.4 is applied to a synthetic example. First, an unconditional random field is simulated according to Equation 9 using an isotropic exponential variogram with a range of 40 for the low values, linearly decreasing to a range of 4 for the high values. This field is sampled at 50 random locations shown in Figure 15. These sampled values are subsequently used as conditioning values and the variogram model is assumed to be known. 100 conditional realizations are simulated. Figure 15 shows two conditional realizations and Figure 16 shows the ensemble mean and the ensemble standard deviation across all realizations. Note that the mean is considered to represent the interpolation while the standard deviation represents the estimation uncertainty. Figure 16 also shows the Ordinary Kriging (OK) interpolation for the sampled values together with the Kriging estimation uncertainty in terms of the standard deviation. Note that for OK the variogram used was the one corresponding to $\tau = 0.5$, that is, an exponential variogram with a range of 20. It can be seen that the uncertainty resulting from the presented approach is very different to the Kriging uncertainty. The Kriging uncertainty is only a function of the geometry of the observations while the uncertainty obtained from the presented approach is also value-based.

The proposed conditioning approach is not exact. Figure 17 shows the box plots for the 100 realizations for the 50 conditioning points. It can be seen, that, except for five outliers which are highlighted with red circles in
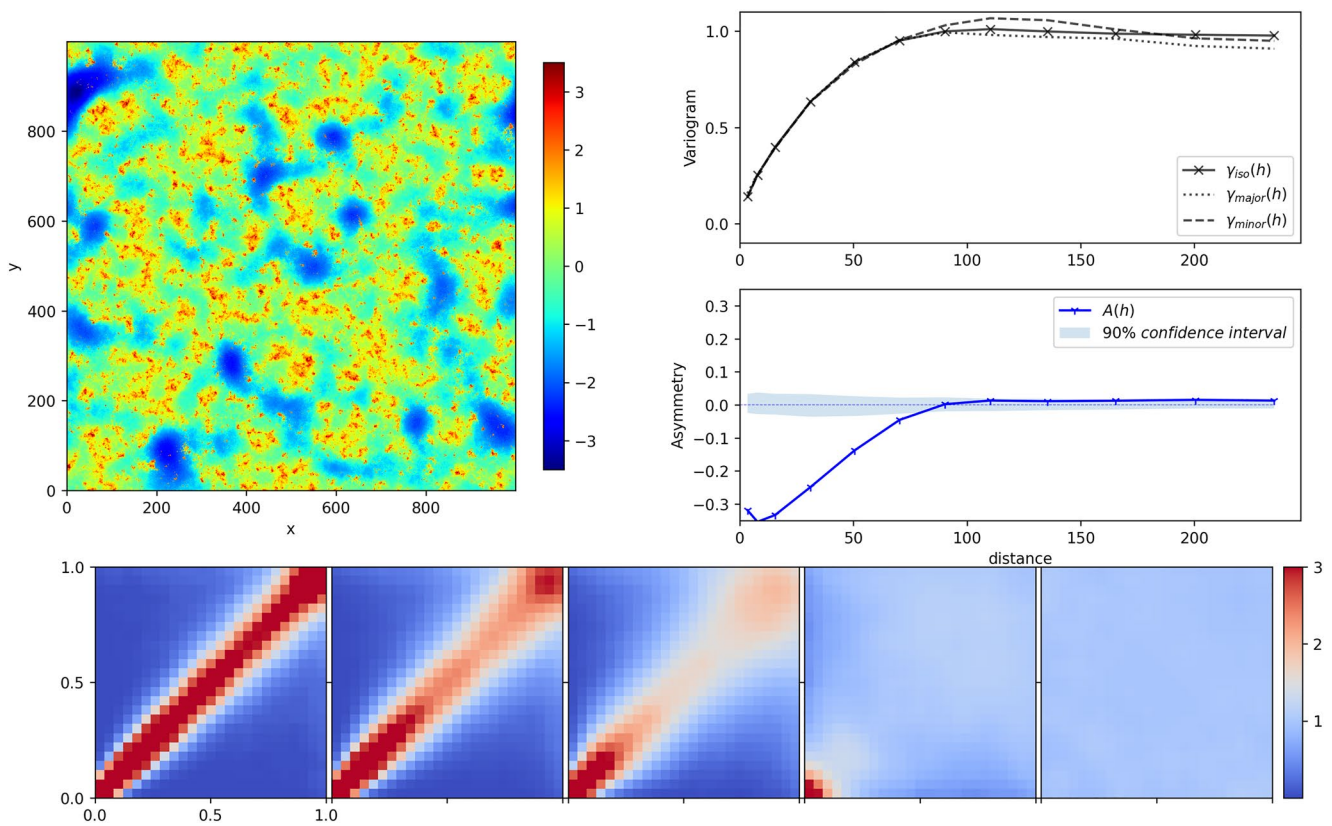
Figure 14. Simulated field with a linearly changing variogram model and linearly changing range (Gaussian variogram with range 60 linearly changing to an exponential variogram with range 5, both isotropic) according to Equations 13 and 9 with corresponding variograms, spatial asymmetry function, and spatial copulas for separation distances (from left to right) of 3, 8, 15, 50, 90. The blue shaded area represents the 90% confidence interval for the spatial asymmetry that can be observed from Gaussian fields.

Figure 17, all simulated values are close to the true conditioning values. For most of the conditioning points, the simulated values are within $\pm 1\%$ of the true values. Note, that this error margin could be adjusted to consider measurement uncertainties, for instance.

5.2. Groundwater Quality Parameter

The above concept was applied to an extensive groundwater quality data set from the state of Baden-Württemberg in South-West Germany. This data set consists of more than 2,500 measurements of groundwater quality parameters of the near-surface groundwater layer. A detailed geostatistical analysis of this data set is presented in Bárdossy (2006). The author shows that the spatial dependence structure of the investigated groundwater quality parameters is not Gaussian. In Bárdossy and Li (2008) the authors present a copula-based interpolation approach to model these non-Gaussian spatial dependence structures. Figure 18 shows the groundwater quality observation network within the state of Baden-Württemberg and the red box highlights the selected study area. The selected study area is 100 km by 100 km (discretized into 100 m by 100 m grid cells) and it contains 386 observations.

For this study, nitrate concentrations, measured in mg/L, were selected as the groundwater quality parameter of interest. A linearly changing variogram range and a linearly changing nugget according to a combination of Equations 9 and 10 was selected and the parameter estimation approach described in Section 4.1 was applied. An exponential variogram with a range of 11 km and a nugget of 0.18 for the low values, linearly changing to a range of 10 km and a nugget of 0.38 for the high values resulted in the best fit. A non-parametric marginal distribution was estimated using Kernel density estimation (KDE). Using the estimated parameters, the KDE marginal distribution, and the 386

Table 1
Estimated Parameters According to the Field Shown in Figure 5

	$a(0)$	$a(1)$
True	5	80
Estimated (300)	6.6	80.3
Estimated (3,000)	6.2	81.0

Table 2
Estimated Parameters According to the Field Shown in Figure 10

|	a (0)	a (1)	$\phi(0)$	$\phi(1)$	λ
True	80	5	0.2	1	25
Estimated (300)	74.1	7.2	0.25	0.75	23.9
Estimated (3,000)	77.4	6.2	0.28	0.81	21.5

observations, 100 realizations were simulated according to the conditional simulation approach described in Section 3.4. As an alternative, Ordinary Kriging was used for interpolation. The interpolated (mean of the 100 simulated realizations) field and the Ordinary Kriging interpolation are shown in Figure 19. Figure 20 shows the box plots for the 100 simulated realizations for the 386 nitrate observations. It can be seen that, except for a few outliers, the Monte Carlo optimization-based conditioning performs well. The standard deviations across the 100 realizations for each individual observation are less than 1.5% of the actual values. As mentioned in Section 5.1, this conditioning error can be considered to represent the measurement error of the actual nitrate concentration data. A k-fold cross-validation (with $k = 8$) was performed to test how well the model fits the observed data. The root mean squared error of the presented copula-based approach is $RMSE = 24.38$ mg/L, which is slightly lower than the Ordinary Kriging result of $RMSE = 24.43$ mg/L. As the presented simulation approach and OK provide an estimation of the full conditional distributions, these can also be verified. Therefore, the value of the conditional distribution function (i.e., $u_i^* = F_{z,x_i}(z(x_i))$) for the observation was calculated for each point of the control set in the cross-validation. The values of u_i^* should follow a uniform distribution. Figure 21 shows the distributions of u_i^* for both, Ordinary Kriging and the presented simulation approach. It can be seen that the distribution of u_i^* corresponding to the simulation approach is much closer to a uniform distribution than the distribution according to Ordinary Kriging. This indicates that the presented approach better represents the estimation uncertainty than Ordinary Kriging.

6. Discussion and Conclusions

Natural and anthropogenic processes often lead to specific spatial distributions of the related variables. The resulting fields are non-Gaussian in their spatial structure, having different spatial dependencies for high and low values. In order to capture these differences copula-based models can be used. In this contribution, a family of new copula models was introduced. The new model allows a very flexible definition of value-dependent spatial variability.

An advantage of the formulation is that the random fields can be modified by changing the generating random numbers only, and this means that the non-Gaussian properties remain unchanged. This is an advantage for the use in non-linear inversion problems as the simulation procedure is very fast. The adjustments of the U fields only influence the Gaussian X_τ fields and their subsequent combination assures that the non-Gaussian structure is fully kept.

The presented approach can provide fields with asymmetric dependence in the sense of diffusion (order based asymmetry according to Equation 16), but not asymmetric in the sense of advection (direction-dependent asymmetry according to Equation 17). Further research is required to incorporate directional asymmetry into the presented simulation approach.
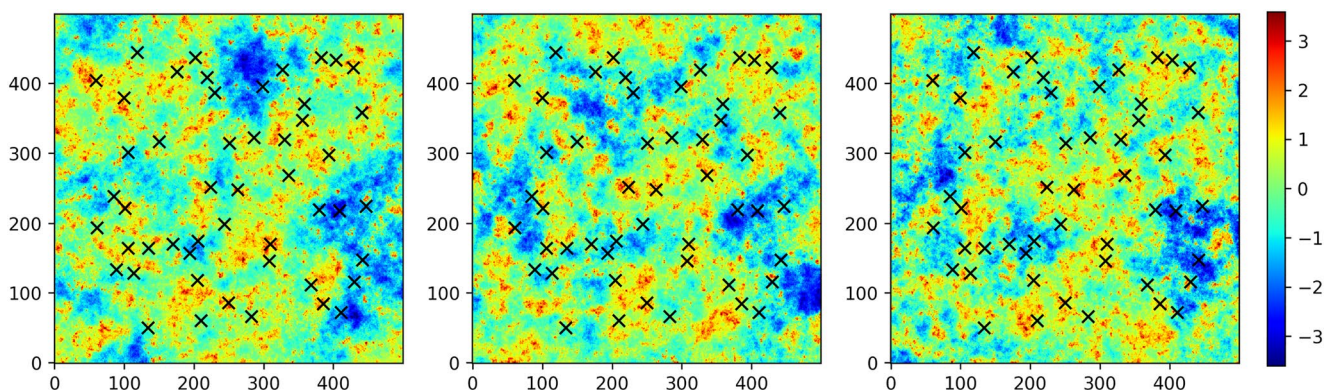
Figure 15. Left: True random field. Middle and right: Two realizations conditioned on the sampled values of the true field. "x" marking the sampling locations.
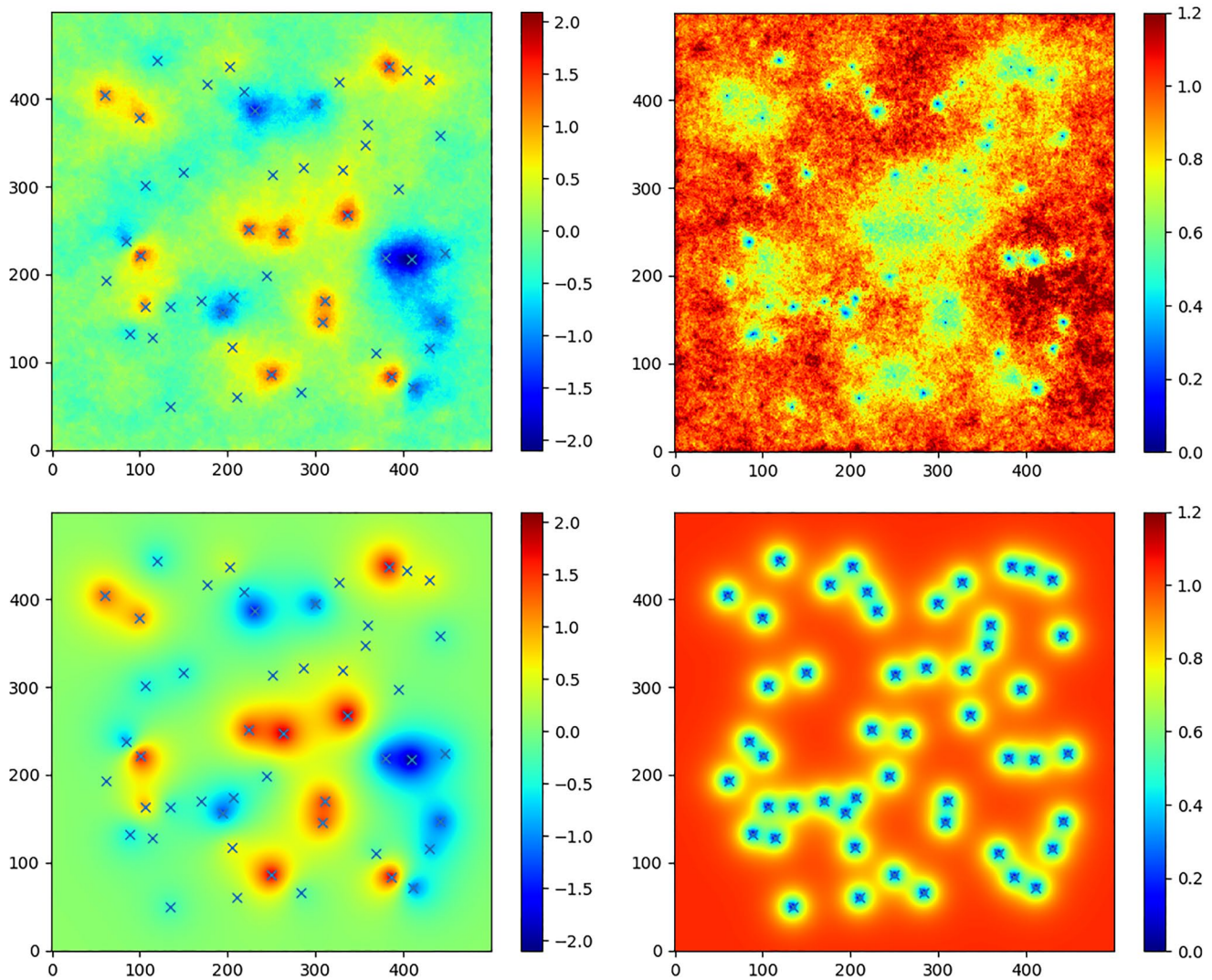
Figure 16. Upper: Mean (left) and standard deviation (right) over the 100 conditional realizations. Lower: Ordinary Kriging interpolation (left) and the corresponding estimation uncertainty in terms of the standard deviation (right).

The presented examples demonstrate the vast diversity of spatial structures resulting from the simulation approach. Further parameter combinations can generate even more complex spatial dependence structures. It has been shown that the Monte Carlo optimization-based conditioning approach performs well for both, synthetic and real-world data. The presented approach is mainly useful for unconditional and conditional simulation. The direct estimation in the sense of Kriging requires further research. At the present interpolation is obtained using a large number of conditional simulations.

The cross-validation for the nitrate concentration examples demonstrated that the presented approach had a smaller estimation error than Ordinary Kriging. Further, it was shown that the presented approach resulted in a better distribution of uncertainty than Ordinary Kriging. This also means that the simulation leads to reasonable realizations.

Changing the linear function in Equations 9–13 to a continuous $h(\tau)$ can produce additional varieties of fields and may be used to generate fields with upper or lower tail dependence.
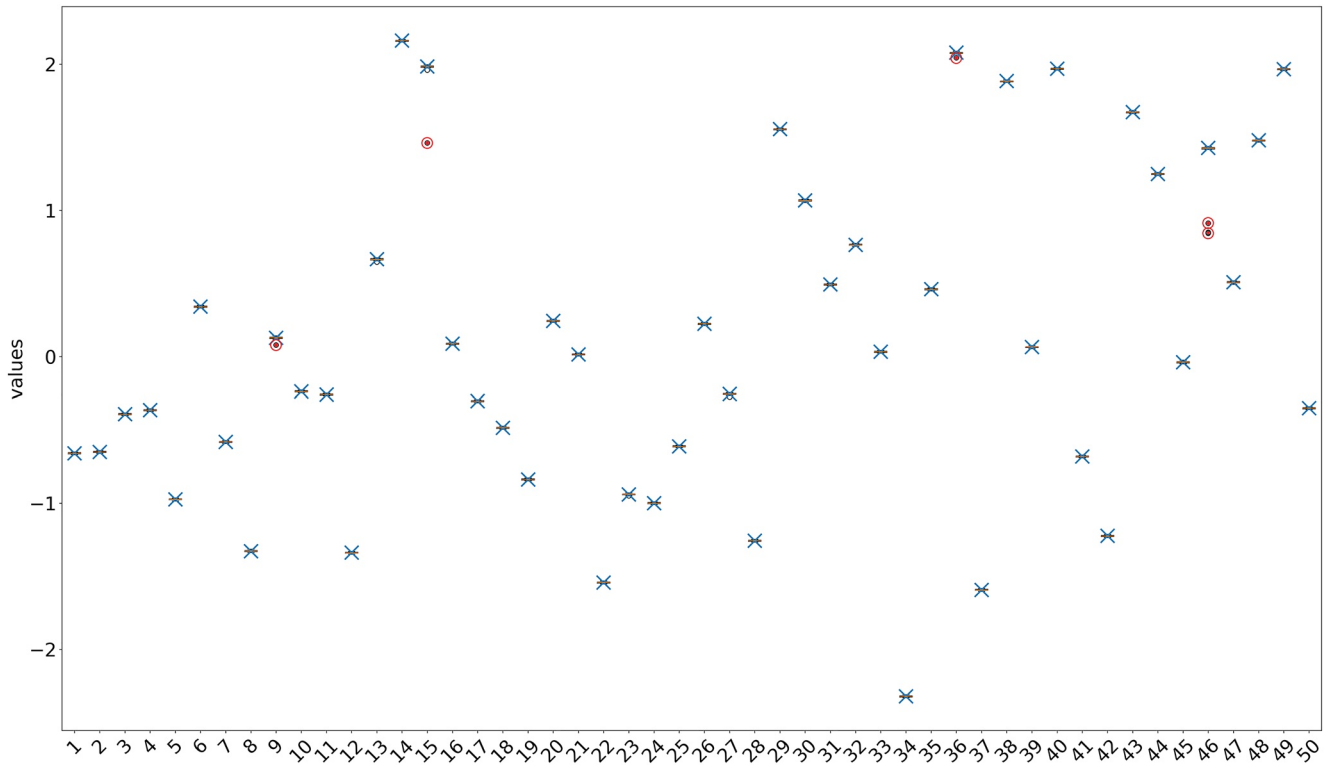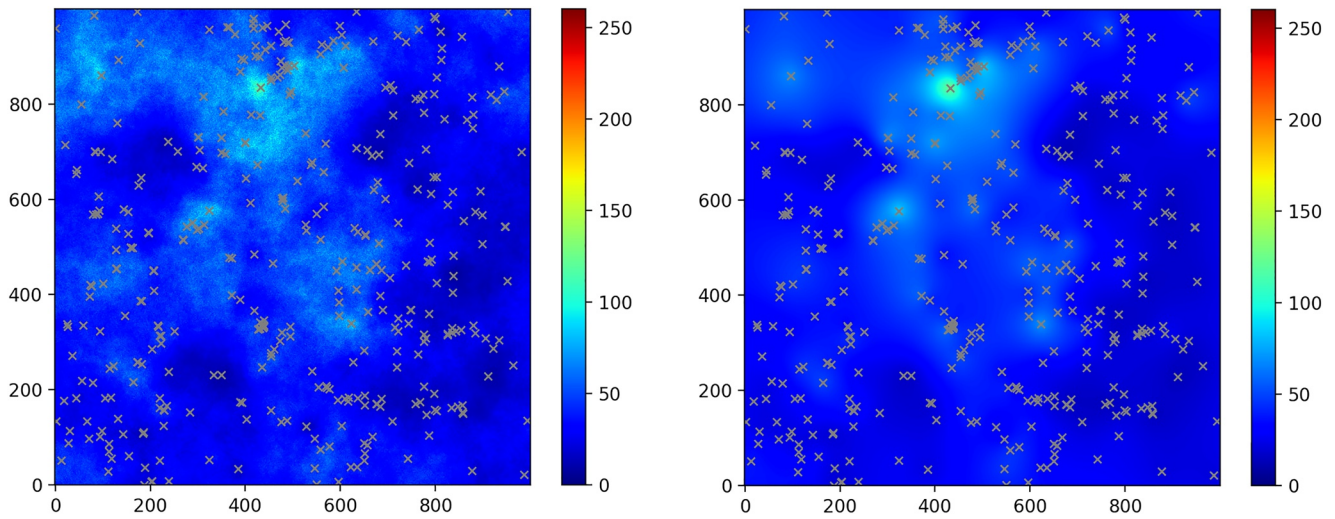
Figure 17. Box plots for the 50 conditioning points over 100 realizations where the blue "x" denotes the true value and the red circles highlight outliers. Note that there is no particular order in which the points are labeled from 1 to 50.
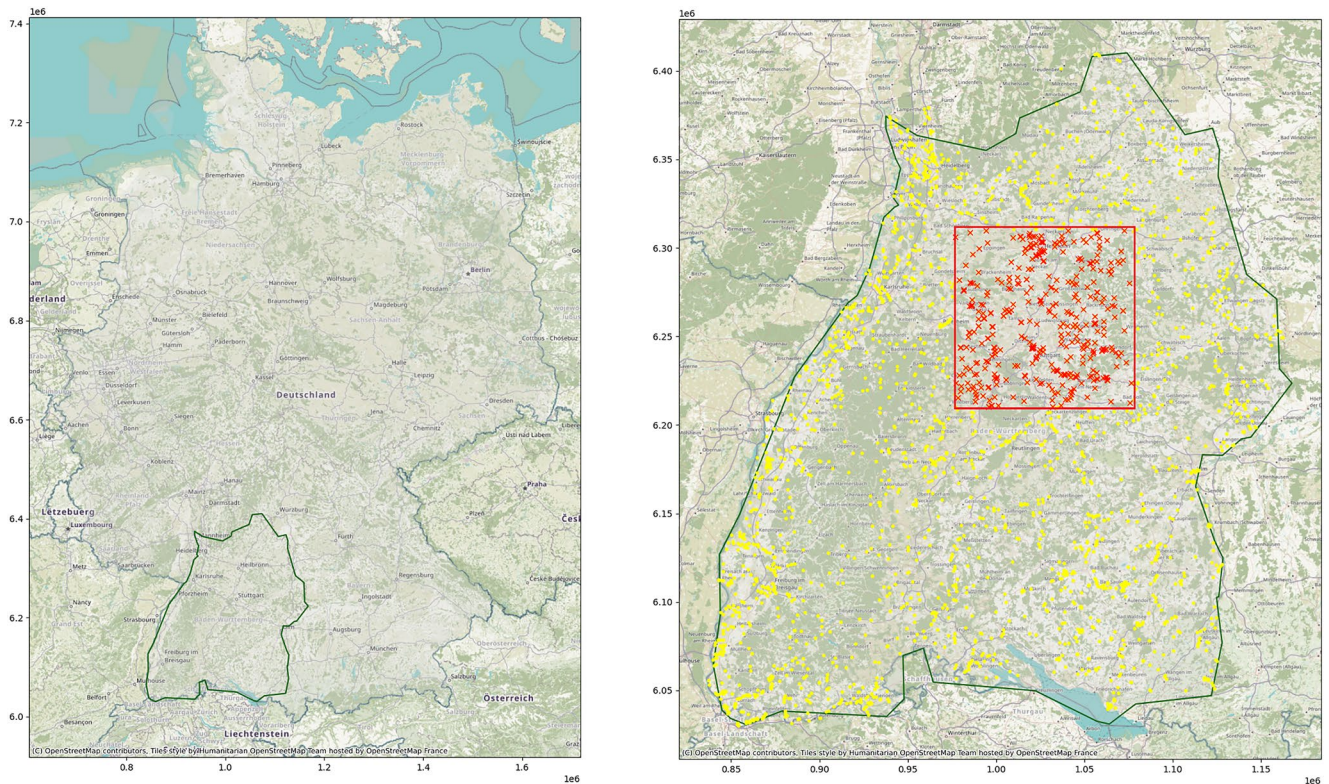

**Figure 18.** Map of Germany (left), Map of Baden-Württemberg with study area highlighted in red.

Figure 19. Mean (left) of 100 conditional realizations and Ordinary Kriging interpolation (right) of nitrate concentrations in mg/L.

Although all examples presented in this paper are 2D, it is straightforward to use the presented approach for 3D cases too. The conditioning procedure in 3D is the same as in the 2D case. This way, with relatively small computational efforts, large conditional 3D non-Gaussian fields can be simulated, orders of magnitude faster than for example, simulated annealing. Further, it is worth mentioning again, that the presented approach simulates non-Gaussian spatial structures directly rather than through a time-consuming optimization process.
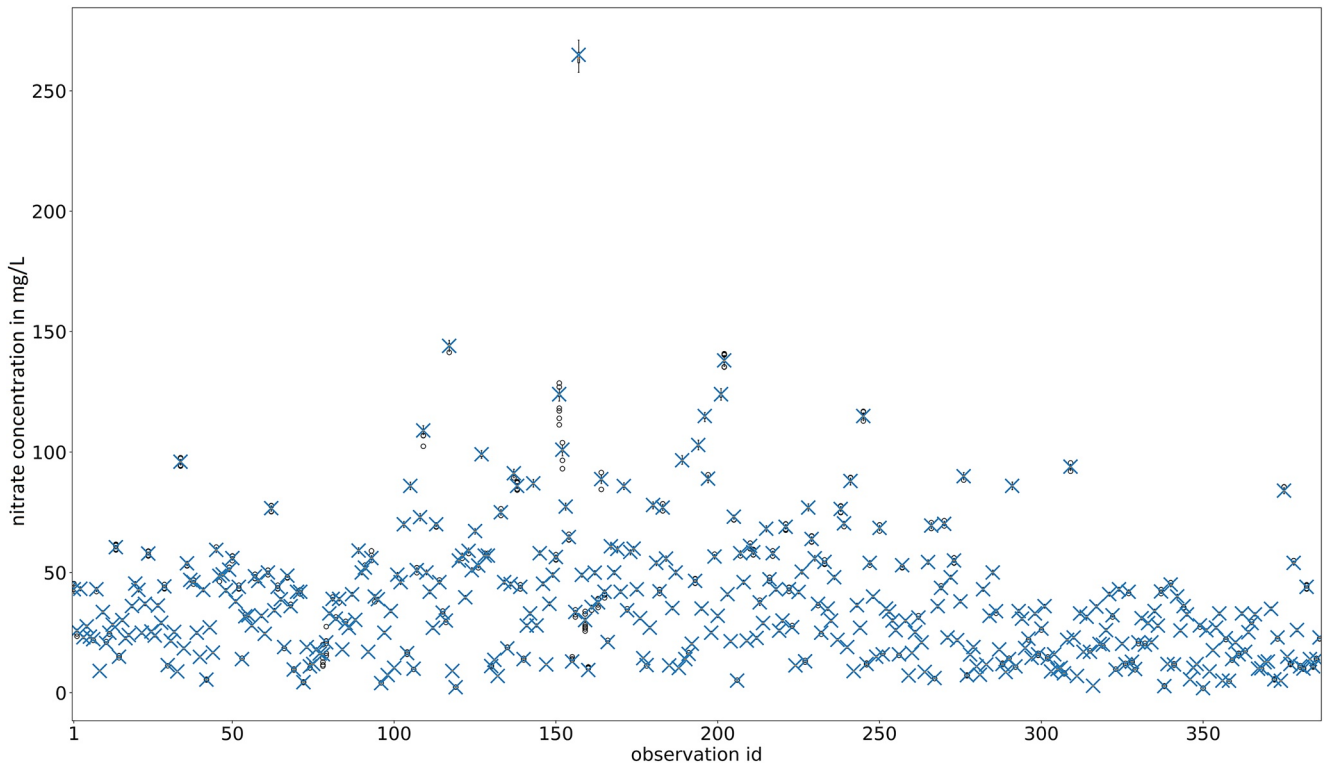
Figure 20. Box plots for the 386 nitrate conditioning points over 100 realizations where the blue x denotes the true value.
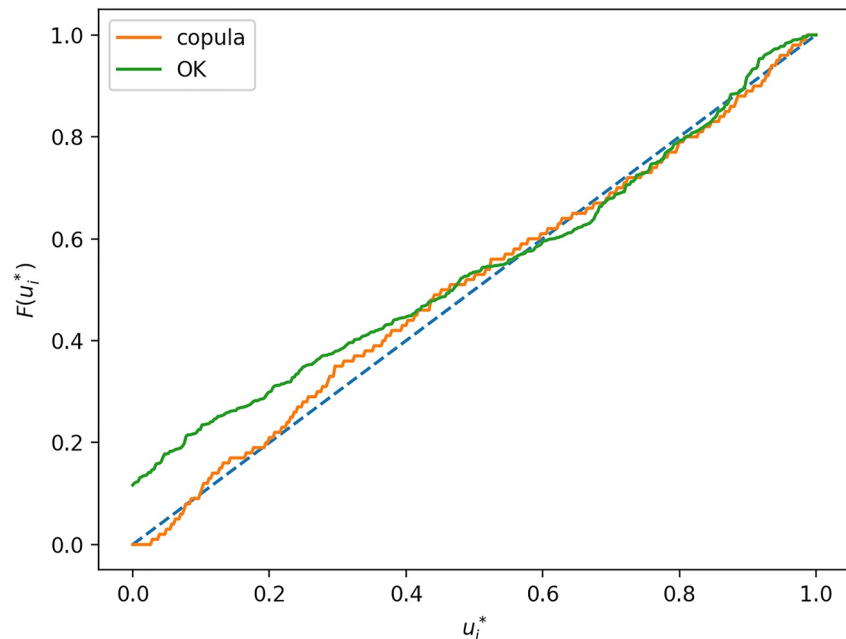
Figure 21. Distributions of $u_i^* = F_{z,x_i}(z(x_i))$ according to the presented copula simulation approach and Ordinary Kriging. The dashed blue line represents a uniform distribution.

Data Availability Statement

The data and the Python code used in the paper are available at Hörning (2023) under GNU General Public License v3.0.

References

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research*, *42*(W11416). https://doi.org/10.1029/2005WR004754

Bárdossy, A. (2023). Changing correlations: A flexible definition of non-gaussian multivariate dependence. *Stochastic Environmental Research and Risk Assessment*, *37*, 2619–2629. https://doi.org/10.1007/s00477-023-02408-1

Bárdossy, A., & Hörning, S. (2017). Process-driven direction-dependent asymmetry: Identification and quantification of directional dependence in spatial fields. *Mathematical Geosciences*, *45*(4), 871–891. https://doi.org/10.1007/s11004-017-9682-1

Bárdossy, A., & Li, J. (2008). Geostatistical interpolation using copulas. *Water Resources Research*, *44*, W07412. https://doi.org/10.1029/2007WR006115

Deutsch, C. V. (1992). Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data. *PhD Thesis*.

Gomez-Hernandez, J., & Wen, X. (1998). To be or not to be multi-gaussian? A reflection on stochastic hydrogeology. *Advances in Water Resources*, *21*, 47–61. https://doi.org/10.1016/s0309-1708(96)00031-0

Gräler, B. (2014). Modelling skewed spatial random fields through the spatial vine copula. *Spatial Statistics*, *10*, 87–102. https://doi.org/10.1016/j.spasta.2014.01.001

Guthke, P., & Bárdossy, A. (2012). Reducing the number of MC runs with antithetic and common random fields. *Advances in Water Resources*, *43*, 1–13. https://doi.org/10.1016/j.advwatres.2012.03.014

Guthke, P., & Bárdossy, A. (2017). On the link between natural emergence and manifestation of a fundamental non-gaussian geostatistical property: Asymmetry. *Spatial Statistics*, *20*, 1–29. https://doi.org/10.1016/j.spasta.2017.01.003

Haslauer, C., Guthke, P., Bárdossy, A., & Sudicky, E. (2012). Effects of non-gaussian copula-based hydraulic conductivity fields on macrodispersion. *Water Resources Research*, *48*(7). https://doi.org/10.1029/2011wr011425

Hörning, S. (2023). Sebastianhoerning/laycop: Layercopula 1.1. [Software]. Zenodo. https://doi.org/10.5281/zenodo.7508645

Hörning, S., Sreekanth, J., & Bárdossy, A. (2019). Computational efficient inverse groundwater modeling using random mixing and whittaker–shannon interpolation. *Advances in Water Resources*, *123*, 109–119. https://doi.org/10.1016/j.advwatres.2018.11.012

Huang, J., Wang, C., Wang, W., Hu, X., Feng, W., & Yin, Y. (2022). A novel method of 3d multipoint geostatistical inversion using 2d training images. *Lithosphere*, *13*. https://doi.org/10.2113/2022/5946595

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman Hall.

Journel, A. G. (2003). *Multiple-point geostatistics: A state of the art*. Stanford Center for Reservoir Forecasting.

Marcotte, D. (1996). Fast variogram computation with fft. *Computers and Geosciences*, *22*(10), 1175–1186. https://doi.org/10.1016/s0098-3004(96)00026-x

Mariethoz, G., Renard, P., & Straubhaar, J. (2010). The direct sampling method to perform multiple-point geostatistical simulations. *Water Resources Research*, *46*. https://doi.org/10.1029/2008WR007621

Nelsen, R. (1999). *An introduction to copulas*. Springer Verlag.

Papalexiou, S. M., Serinaldi, F., & Porcu, E. (2021). Advancing space-time simulation of random fields: From storms to cyclones and beyond. *Water Resources Research*, *57*(8). https://doi.org/10.1029/2020WR029466

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, *7*(2), 155–162. https://doi.org/10.1093/comjnl/7.2.155

Ravalec, M. L., Noetinger, B., & Hu, L. Y. (2000). The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning gaussian simulations. *Mathematical Geology*, *32*(6), 701–723. https://doi.org/10.1023/a:1007542406333

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris*, *8*, 229–231.

Strebelle, S. (2002). Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, *34*(1), 1–21. https://doi.org/10.1023/a:1014009426274

Zinn, B., & Harvey, C. (2003). When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate gaussian hydraulic conductivity fields. *Water Resources Research*, *39*, 1051. https://doi.org/10.1029/2001wr001146