Advances in Water Resources 51 (2013) 168–196

Contents lists available at SciVerse ScienceDirect

Advances in Water Resources

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

Connectivity metrics for subsurface flow and transport

Philippe Renard [a,*], Denis Allard [b]

[a] Centre of Hydrogeology and Geothermics, University of Neuchâtel, 11 Rue Emile Argand, CH-2000 Neuchâtel, Switzerland
[b] Biostatistique et Processus Spatiaux (BioSP), INRA, Site Agroparc, 84914 Avignon, France

ARTICLE INFO

Article history:
Available online 13 December 2011

Keywords:
Review
Heterogeneous media
Connectivity
Percolation
Effective permeability
Euler number

ABSTRACT

Understanding the role of connectivity for the characterization of heterogeneous porous aquifers or reservoirs is a very active and new field of research. In that framework, connectivity metrics are becoming important tools to describe a reservoir. In this paper, we provide a review of the various metrics that were proposed so far, and we classify them in four main groups. We define first the static connectivity metrics which depend only on the connectivity structure of the parameter fields (hydraulic conductivity or geological facies). By contrast, dynamic connectivity metrics are related to physical processes such as flow or transport. The dynamic metrics depend on the problem configuration and on the specific physics that is considered. Most dynamic connectivity metrics are directly expressed as a function of an upscaled physical parameter describing the overall behavior of the media. Another important distinction is that connectivity metrics can either be global or localized. The global metrics are not related to a specific location while the localized metrics relate to one or several specific points in the field. Using these metrics to characterize a given aquifer requires the possibility to measure dynamic connectivity metrics in the field, to relate them with static connectivity metrics, and to constrain models with those information. Some tools are already available for these different steps and reviewed here, but they are not yet routinely integrated in practical applications. This is why new steps should be added in hydrogeological studies to infer the connectivity structure and to better constrain the models. These steps must include specific field methodologies, interpretation techniques, and modeling tools to provide more realistic and more reliable forecasts in a broad range of applications.

1. Introduction

Over the last 50 years, conceptualizing and modeling aquifer heterogeneity has evolved in three main phases. In the first phase, most models considered an ensemble of regions having constant equivalent properties. These regions were generally drawn on the base of geological mapping and their properties were estimated either from typical average values within a geological formation or by means of model calibration. In the second phase, small scale variability has been considered as a key feature and geostatistics was intensively used to model the spatial variability [1–3]. Most models were then based on the multi-Gaussian assumption that has the advantage to be extremely parsimonious. In this framework, the main parameters controlling the degree of heterogeneity were the variance of the logarithm of the hydraulic conductivity and its correlation length. An extremely broad range of important results have been obtained using this model in the stochastic hydrogeology literature [3–6]. But several authors raised the point that the multi-Gaussian model was too restrictive and could not describe the full range of connectivity patterns that one finds in nature [7–10]. That can be considered as the start of the third phase; nowadays, in addition to the variance and the correlation length, the connectivity structure of the heterogeneity is understood as a property that strongly influences groundwater flow and solute transport in aquifers. In parallel to this evolution, a broad range of stochastic models has been developed to represent geological structures such as channels, lenses, deltas, with the aim to better reproduce the expected connectivity of aquifers [11,12]. But surprisingly, it is only since the last 10 years that a significant number of papers have been published on the quantification of connectivity and its relations with the physical properties of aquifers (Fig. 1). It is worth noticing that a similar evolution occurred simultaneously in the neighboring fields of surface hydrology [13–20], geomorphology [21–24], landscape ecology [25–28], pore scale or soil physics [29–34].

A common issue that appears in the literature is that, even if the concept of connectivity refers to a rather intuitive notion, there is not a single mathematical definition that is adopted by the whole community. The word connectivity is often used as a broad concept with various definitions related for example to the efficiency with which runoff moves from source areas to streams [35] or to the presence of an unsaturated zone between a river and an underlying

* Corresponding author.
E-mail addresses: philippe.renard@unine.ch (P. Renard), allard@avignon.inra.fr (D. Allard).

0309-1708/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2011.12.001

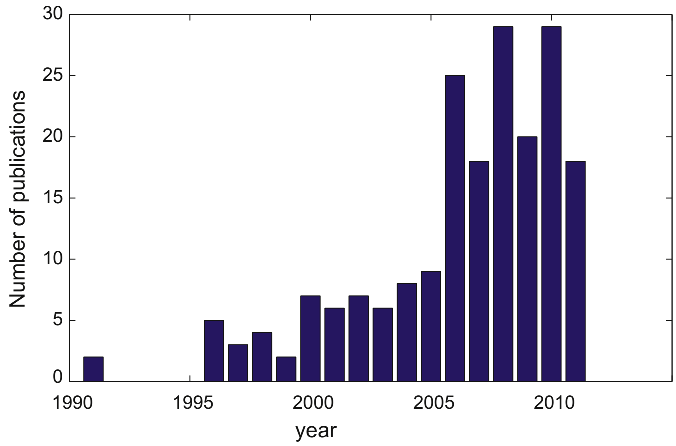

Fig. 1. Evolution of the number of articles having the word *connectivity* in their title in the top five journals in groundwater hydrology according to ISI Web of Knowledge. No paper mentions the concept in its title before 1990. The first papers in the 1990s relate to fractured media. It is only after 2000 that regular publications are found.

aquifer [36]. Therefore, as pointed out by several authors [37,18,19], there is a need to adopt some standards and to share them.

However, providing a standard single short definition of connectivity is probably an impossible task. In the context of groundwater hydrology, connectivity is a broad concept related to the existence of a path enabling fast flow and transport from one location to another. As described in the following paragraphs, the concept of connectivity is quantified using both static and dynamic connectivity metrics.

The static connectivity metrics are only a function of the spatial distribution of the parameters (lithology or directly permeability). They are statistical measures describing the connectivity of a field without considering a specific physical process. Because the most common statistics used to describe the spatial distribution of a field are based on two-point statistics, such as covariances, variograms, or spatial entropy, it is important now to remind the reader that they are not able to describe the connectivity. Two-point statistical metrics describe only the probability of having a certain value at a certain location knowing the value at another location. This is not influenced by the structure in between and by the possibility of having or not a connection through a complex path [10,39,13,40–42]. This is illustrated in Fig. 2 where two stochastic simulations are compared. The one on the left (Fig. 2(a)) displays channels of high permeability. The one on the right (Fig. 2(b)) was generated in order to reproduce the histogram and variogram of the first image. The histograms (Fig. 2(c)) and experimental variograms (Fig. 2(d)) of the two fields are thus not distinguishable from a statistical point of view. Obviously, if only punctual observations are available at a small number of locations from one of those field and if one carries on a statistical analysis of the data, it will not be possible to distinguish if the field belongs to the category on the right or to the category on the left, or to another one, while their response to flow and transport will be very different. This is the reason why other statistical metrics, such has those developed in the framework of percolation theory [43,44], and already widely applied to fractured media [45–48], should be used. With the emergence of new heterogeneity models involving high-order statistics [40,49,50], the static connectivity metrics are becoming an important tool to assess the quality of the simulations or to rank them [51–54].

The dynamic metrics are more diverse since they depend on the type of processes. Usually, dynamic connectivity metrics represent better the physics but they depend both on the geometry and on additional physical parameters such as the type of boundary conditions or the state of the system. For example, in unsaturated media the flow properties of the system will depend on its degree of saturation and on the spatial variability and connectivity of the saturated zone [13,29,55,36]. It implies that the dynamic metrics can vary with time. Dynamic and static connectivity metrics are linked through complex relations [29,42,56–59]. Dynamic metrics are easier to measure in the field.

Another important way to classify connectivity metrics is to distinguish the global metrics from localized ones. Global metrics of connectivity, also called geobody or sand-body connectivity [60], are related to the overall structure of a media and provide information which is not related to a specific location. We will review several of these metrics in the following, but to make the point clear an example of this type of metrics is the probability that two distinct points taken randomly in the domain belong to a connected pattern. Localized metrics, also referred to as reservoir-to-well connectivity or simply reservoir connectivity [60], are related to one or several specific positions in the domain of interest. For example, one might be interested in the connected volume at a given well location. Or a tracer test can provide the indication that two specific points in the study domain are connected or are not connected. This is a localized deterministic metric taking only a binary value: 1 = connected, 0 = disconnected. On a stochastic model, similar information relating two points can be expressed in statistical term. One can for example compute the probability that a certain location in a model connects with another one [48]. Note the semantic subtle distinction; these metrics are said to be localized, because they are related to specific locations. However, as it will be made clear in the next sections, it is not a local property in the sense that the whole media is potentially involved in the connectivity metric.

In terms of heterogeneity modeling, the connectivity metrics are usually not explicit input parameters but are rather implicit derived properties of the medium generated by a given simulation technique. In practice, often the connectivity of several realizations of a stochastic model obtained with the same input parameters will have a range of connectivity properties. The statistical relations between the connectivity metrics and the input parameters of various stochastic models have been analyzed [61–63] and can be used as a guide to select an appropriate model or a set of parameters when required.

The aims of this paper are to highlight the importance of connectivity on flow and transport and to provide a review of the most important connectivity metrics and their use in groundwater hydrology. From a general perspective, the major questions that we will cover in this work are the following: (1) How to define static connectivity? (2) How to define dynamic connectivity and how to relate it to static metrics? (3) How to infer connectivity in the field? (4) How to constrain models of heterogeneous media with a given connectivity? and (5) What do we gain if we properly use connectivity information? In the next section we discuss the importance of connectivity on flow and transport, and how it emerged from different studies. In Section 3, we review several static connectivity metrics. We first recall their definition as rigorously as possible; we then present their most important properties and how they are related to subsurface flow and transport. Section 4 is concerned with dynamic connectivity metric and Section 5 shows how static and dynamic metrics are related. In Section 6 we discuss the practice of stochastic simulations of subsurface structures when connectivity information is available and must be respected. We also discuss how this issue is related to solving inverse problems. We close this paper with a discussion of avenues for future work in Section 7. Appendix A provides some details on how to compute the connectivity metrics with available software.

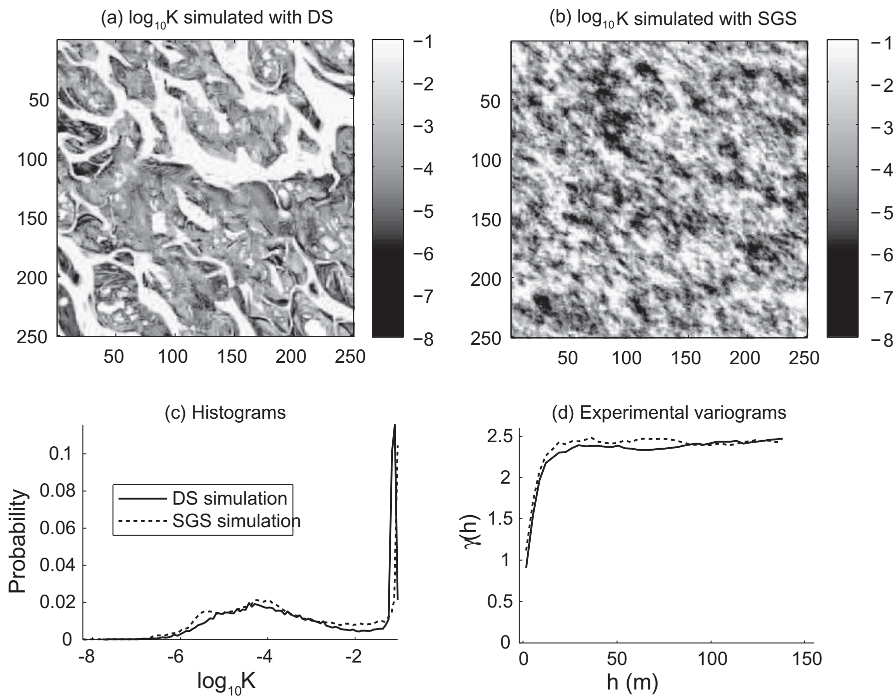

Fig. 2. Two heterogeneous fields showing different connectivity patterns but identical histograms and variograms. From Mariethoz [38].

2. The importance of connectivity on flow and transport

The importance of connectivity was early recognized for flow in fractured rocks [64,65]. This triggered a vast amount of research [45,66–69,47,48,70]. Fractured media and their connectivity having already been the topic of several review papers [71–74], we focus this review only on porous media.

In these aquifers or reservoirs, the role of connectivity was already analyzed quantitatively with a Boolean 2D model of sand lenses in a shally matrix representing fluvial sediments by Allen [75,76]. He showed that the proportion of connected sand lenses is growing very rapidly with the sand proportion if it is greater than 50%. In the mid 1980s, Fogg [77] did a detailed 3D modeling study of the Wilcox aquifer in Texas. He showed that the connectivity of sand lenses is a critical factor influencing flow and transport at a regional scale. According to his results, the interconnectedness between the lenses is a factor more important than the permeability of the lenses themselves. In his paper, Fogg also stated that the available piezometric measurements were not sufficient to locate and identify the connectivity between sand lenses in 3D while the knowledge of the connectivity is essential for accurate transport simulations and to evaluate properly the prediction uncertainty. This conclusion demonstrates the importance of having other sources of information such as a prior geological knowledge, geophysical observations, or tracer observations in order to infer the connectivity.

At the Hanford site (USA), Poeter and Townsend [78] investigated contaminant transport and built several models of heterogeneity based on two different conceptual geological models having different spatial connectivity patterns. The resulting sequential indicator simulations were used to estimate the critical flow paths and to show their influence on the overall solute transport behavior.

At the Lawrence Livermore National Laboratory site, LaBolle and Fogg [79] observed that high permeability channels embedded in less permeable floodplain deposits dominate the solute transport. Ronayne et al. [80] used the probability perturbation inverse method [81] and multiple-point statistics [40] to analyze on the same site the anomalous pressure signals observed in six observation wells during a pumping test. They showed that those data can be explained if high permeability channels connecting some piezometers with the pumping well are accounted for in the model.

These type of observations, highlighting the influence of connectivity on flow and transport processes, have been reported at various scales and on various geological environments [82–84].

In addition, several authors [85–90] have reported observations of anomalies of transport behavior and scale effects that are difficult to explain with traditional multi-Gaussian models of heterogeneity. To explain the observation that the values of the permeabilities estimated at various scales apparently increase with the dimension of the sample (from laboratory experiment to regional scale models), Sánchez-Vila et al. [9] constructed a series of models with different degrees of connectivity and computed their equivalent conductivities. They showed that indeed a model with a higher degree of connectivity than the multi-Gaussian model could explain at least partly the observations. A rather similar investigation was conducted by Wen and Gómez-Hernández [91] on the transport properties of non-multi-Gaussian media. Their study showed that high connectivity of the high permeable regions leads to faster arrivals and longer tailing, clearly characteristic of non-Gaussian transport responses. Wen and Gómez-Hernández [91] suggested that a careful evaluation of the connectivity of an aquifer should be conducted prior to apply a given type of model. This is also described in detail in Gómez-Hernández and Wen [10].

Zinn and Harvey [92] went a step further in the systematic analysis of the impact of connectivity. They constructed an ensemble of permeability fields displaying different connectivity structures as did Sánchez-Vila et al. [9] or Wen and Gómez-Hernández [91]. But here, they built the transmissivity fields using a series of transformations of an initial multi-Gaussian field that ensures that all the fields have identical pdfs, and nearly identical variograms but very different connectivity structures. This transformation will be detailed in Section 3.2. In the initial multi-Gaussian field (Fig. 3(a)) representing the logarithm of the transmissivity, the high and low values are isolated in lenses having a size proportional to the correlation length and only the intermediate values are well connected. This is a well known feature of the multi-Gaussian model [7,8]. In the transformed fields, one can chose to connect low values (Fig. 3(b)) or high values (Fig. 3(c)), thus leading to what will be called the disconnected or connected field models in the rest of this paper.

By simulating flow and transport through a large ensemble of those fields and by varying the variance, Zinn and Harvey [92] showed how the equivalent conductivity K_{eff} of the connected or disconnected fields depart from the multi-Gaussian case (Fig. 3(d)). As a direct consequence, solute transport is much faster in the connected fields (Fig. 3(e)) and the deviation increases as a function of the variance. But the difference in the breakthrough curves between the three configurations is not only due to the differences in effective conductivity (which controls the mean velocity of the tracer), it also relates to the presence of mass transfer between the mobile and less mobile water phases. As a consequence, the overall breakthrough curve through the highly connected media displays a non-Fickian behavior with faster arrivals than the standard model and a longer tailing due to the mass transfer process (Fig. 3(f)). The apparent macrodispersivity α_l does not follow any more the standard asymptotic behavior $\alpha_l = \sigma_Y^2 I_Y$ observed in multi-Gaussian fields (where σ_Y^2 is the variance of the logarithm of the hydraulic conductivity K, and I_Y is the correlation length of ln$K$). For the poorly connected media, the behavior is opposite. More generally, Zinn and Harvey [92] showed that isotropic media with a log-normal distribution of the local conductivity values can behave like stratified media if the high conductivity values are highly connected. The transport is not Fickian and the equivalent hydraulic conductivity is higher than the geometric mean.

Neuweiler and co-authors [93,55,94] extended this work to unsaturated media and two-phase flow problems. Again, they showed that the connectivity structure has a strong impact on the equivalent relative permeability curves. They also indicated that simple estimates of effective parameters can reproduce the typical time scales of the flow processes if the connectivity information is correctly incorporated in the estimates.

Another important aspect is that inverse methods which assume that the underlying heterogeneity is multi-Gaussian may not perform properly if the unknown reality is not multi-Gaussian. To test that assumption Kerrou et al. [95] generated a synthetic channelized aquifer model and used it as a virtual reality to study if the multi-Gaussian model could be sufficiently robust to provide correct forecasts of the total flux through the medium and capture zone delineation under those conditions. The results showed that when a large number of head and transmissivity measurements are available, the inverse method is heavily constrained by the underlying multi-Gaussian assumption and the prediction uncertainty is very small but the forecasts were biased. This was attributed to the fact that the Sequential Self Calibration method (like

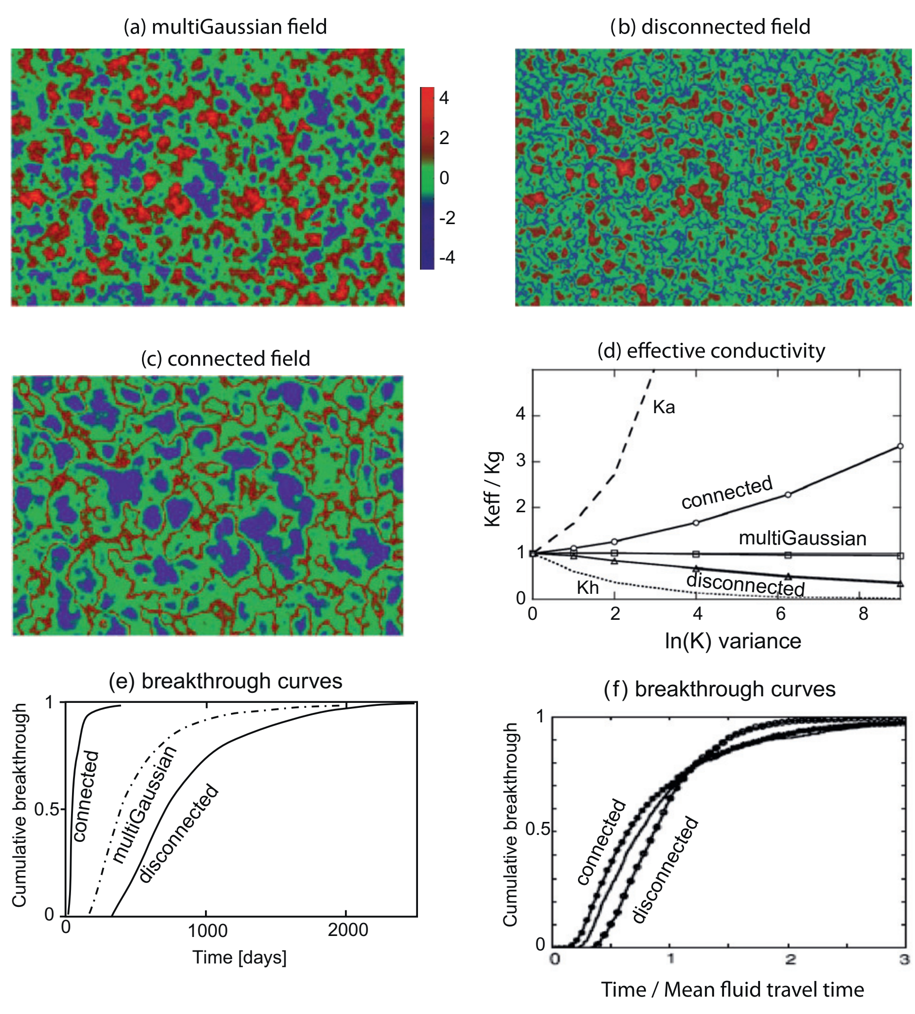

Fig. 3. Illustration of the procedure used by Zinn and Harvey [92] to construct various types of fields and some results concerning their flow and transport properties. K_{eff} is the effective conductivity, K_g is the geometric mean of the local conductivity values, K_a the arithmetic mean and K_h the harmonic one. Cumulative breakthrough is the total mass of tracer recovered at the outlet (modified from Zinn and Harvey [92]).

any other method based on a multi-Gaussian model) reproduced very well the variogram of the transmissivity data, their local values, the head values at the observation nodes, but not the connectivity structure as shown by comparing the connectivity functions (see definition below) of the reference and the results of the inversion technique. Recently Zhou et al. [96] showed that these problems do not occur when using ensemble Kalman filters (EnKF) even when the underlying prior model is incorrect. The method consists in generating an initial ensemble of models using multiple-point statistics and to iteratively update those fields using the standard or a modified version of the EnKF (allowing to deal with non multi-Gaussian distributions). In the case that is investigated, Zhou et al. [96] are able to reconstruct the structure of the channels in a very efficient manner. This is a very interesting and promising result probably due to the fact that EnKf do not ensure in general that the simulated fields honor the prior model and that the transient data are sufficient in this case to identify the presence of the channels.

To conclude this section, all these results indicate that inferring the connectivity of a field is extremely important to allow producing reasonable forecasts. Blindly assuming that aquifers are multi-Gaussian is not a conservative assumption [92].

3. Static connectivity metrics

3.1. Grids, neighborhoods and clusters

Aquifers and reservoirs are in general modeled on regular grids on which neighborhood relations are defined. The most common grid in 2D is the square grid with 4 neighbors (4-connectivity); in 3D it is the regular cubic grid with 6 neighbors (6-connectivity). Other lattices with other connectivities could alternatively be defined. For example, the 8-connectivity square grid (the 4 corners are added in the neighborhood), the hexagonal grid with 6 neighbors or the honeycomb grid with 3 neighbors are other possible grids in 2D, see Fig. 4 for two examples of regular grids. All connectivity characteristics discussed in this paper are based on the definition of neighbors. They thus depend on the chosen lattice and neighborhoods. From now on, we will always consider regular cubic grids (regular square grids in 2D) equipped with the 6-connectivity (4-connectivity in 2D) as they are the most frequently used.

The grid will be denoted $\mathcal{G}$, its spatial dimension d and L will be the number of grid nodes along one of its dimension. We will define a point or a cell on this grid by its location x which is a vector of d spatial coordinates. The total number of cells of $\mathcal{G}$ is thus L^d. In each cell, there are only two types of porous media represented by an indicator variable $I(x)$ that can either take the value 1 (the cell is highly permeable) or 0 (the cell is impermeable). $I(x)$ can be deterministic, or can be one realization of a random function. Let us denote X the subset of $\mathcal{G}$ representing all cells in which $I(x) = 1$. X^c will be denoted its complement set. Two points x and y of X are said to be connected if there exists a sequence in X of neighboring points between x and y. When x and y are connected, we will denote $x \leftrightarrow y$. An obvious consequence of this definition is that $x \leftrightarrow y$ and $y \leftrightarrow z$ entails $x \leftrightarrow z$. A subset A of X is said to be connected if $x \leftrightarrow x'$ for any x and x' of A. The connected components of X are the largest connected subsets of X.

Far from being a bivariate characteristic, the connectivity involves not only the entire grid, but also the neighborhood relationship defined on $\mathcal{G}$. The concept of connected component is well known in percolation theory under the terminology of *clusters* [43,44]. The same concept is known in petroleum engineering literature as a *geobody* [51,60,57]. We further define the cluster identification function $C(x)$ that identifies with a unique value each cluster and is equal to 0 for all cells not in X (Fig. 5). The details of an efficient computer implementation of the calculation of $C(x)$ from $I(x)$ are given for example in Hoshen and Kopelman [97] or Newman and Ziff [98] (see Appendix A for additional information about available codes). Once the function $C(x)$ is defined, testing the connectivity between two cells x and y is equivalent to testing that $C(x)$ is equal to $C(y)$. This approach is computationally efficient to test the connectivity between many cells.

At this stage, it is important to recall that the connectivity of the complement set X^c is related to the connectivity of the set X in a way which involves the lattice with the dual connectivity. Consider the set represented in Fig. 6 by black grid nodes. For the 4-connectivity square grid it is composed of two clusters. The impossibility of a flow between to distinct clusters means implicitly that the background is connected. On the square grid, the dual of the 4-connectivity is thus the 8-connectivity. In contrast to this, the hexagonal grid with 6 neighbors is auto-dual. The 6-connectivity neighborhood applies to X and its complement X^c.

3.2. General probabilistic framework and three important random sets

We need to precise a little bit the general set-up considered. Many properties of the connectivity metrics studied in this section involve probabilities. The meaning of probabilities is straightforward when X is a random set. In this case, they can then be estimated through simulation studies. In the rest of this paper, all simulated models are stationary, i.e. translation invariant. The consequence is that the connectivity between two points x and y depends only on the vector $h = y - x$. The meaning of a probability is less obvious when studying a specified aquifer or a reservoir. In this case, making stationarity and ergodicity assumptions will allow us to compute some connectivity metrics by counting on $\mathcal{G}$ the number of repetitions of events depending on some distance vector h. More details will be given below.

Many numerical studies were conducted for describing and quantifying connectivity metrics [99,62,61,60,57] based on stochastic simulations of random sets.

Among models of interest, the Boolean model is by far the most frequently studied, see e.g. [61,62,60,57]. We recall briefly its

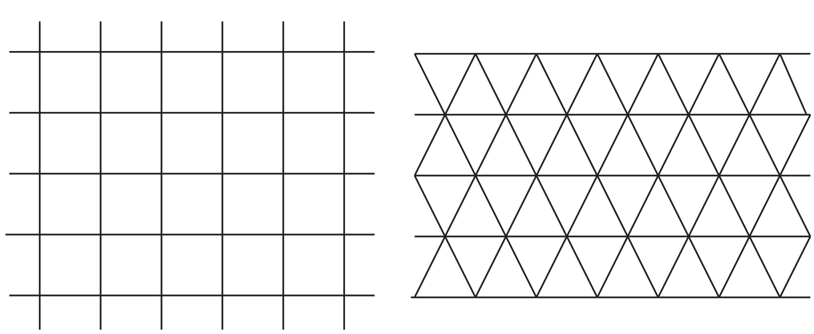

Fig. 4. Two examples of regular grids in 2D. Left: regular square grid with four neighbors. Right: regular hexagonal grid with six neighbors.

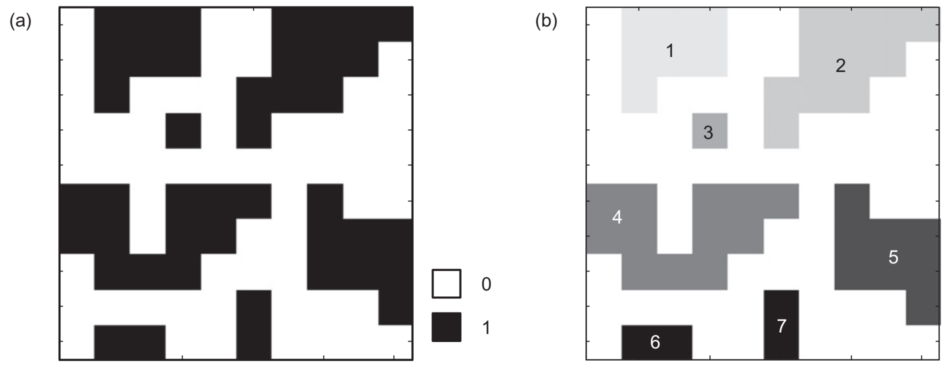

Fig. 5. Identification of the connected components of an indicator variable (a) using the cluster identification function $C(x)$ (b).

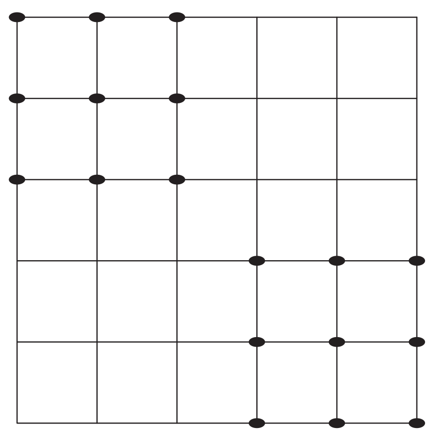

Fig. 6. The black clusters are not connected for the 4-connectivity; the background is thus connected for the 8-connectivity.

definition. More details can be found for example in Lantuéjoul [100]. In a Boolean model, objects are located at points drawn uniformly in a given domain. The objects can have random or fixed shape, size and orientation. Here, we consider a discrete version of the Boolean model. Squares of size l are centered at each grid node with a probability $p_0 \in [0,1]$. The probability of being in the permeable phase, called the proportion and denoted p is related to p_0 according to $p = 1-(1-p_0)^{l^2}$. Note that $l = 1$ corresponds to Bernoulli grids, i.e. random sets X such that each grid node is in the permeable phase independently to all other grid nodes, while $l \to \infty$ with L/l kept constant tends to a continuous Boolean model of squares of size L/l in the plane.

Another model which will be extensively illustrated in this paper is the truncated Gaussian model. For this model, a standard multi-Gaussian random function $Y(x)$ is first simulated on $\mathcal{G}$. The random set X is then defined as the set of points x above a threshold t. Since $p = P(Y > t) = 1 - G(t)$, where G is the cpf of a Gaussian random variable, there is a one-to-one correspondence between the proportion and the threshold. The random function $Y(x)$, and thus the random set X for a given proportion, is fully characterized by the covariance function $c(h)$ of $Y(x)$. The regularity of $c(h)$ at the origin is a parameter of crucial importance for the continuity of X and some of its connectivity characteristics. If $c(h)$ is twice differentiable at the origin, i.e. if $c''(h = 0) < \infty$, which is for example the case for the Gaussian covariance model, X will display very regular boundaries [100,101]. On the continuous plane, X will have a finite perimeter in bounded domains. On the contrary, if $c''(h = 0) = \infty$, which is for example the case for the exponential or spherical covariance models, X will display very erratic boundaries. For these models, the perimeter of X in a bounded domain of the continuous plane is infinite.

As a third family of model illustrated in this paper, we will consider the transformation of multi-Gaussian fields proposed in Zinn and Harvey [92] and Vogel [102]. First, an initial classical multi-Gaussian field $Y_0(x)$ having a mean of zero and a unit variance is generated (see Fig. 3(a)). To connect high or low values, the absolute value of the multi-Gaussian field is taken. The minimum values are now connected. By stretching the univariate distribution using a normal score transform, a field Y_{di} with a Gaussian pdf and connected low values (or disconnected high values) is obtained (Fig. 3(b)):

$$Y_{\mathrm{di}}(x) = \sqrt{2}\ \mathrm{erfinv}\left(2\ \mathrm{erf}\left(\frac{|Y_0(x)|}{\sqrt{2}}\right) - 1\right), \tag{1}$$

where erf and erfinv are the error function and its inverse. This transformation alters the covariance of the field (but not the overall variance) and therefore a spatial stretching is applied to ensure that the new field has a variogram identical to the initial multi-Gaussian field. To obtain a field Y_{di} with connected high values, the distribution is simply reversed (Fig. 3(c)) $Y_{\mathrm{co}} = -Y_{\mathrm{di}}$. The latter transformation leads to the so-called connected fields.

In all the binary models, the proportion p can vary from 0 to 1. For a given realization, the set X_p describing the permeable phase will thus increase with p in the following way: if a cell x belongs to X_p for a given p, it also belongs to all $X_{p'}$ for all $p' \geqslant p$. This can be achieved by tuning the threshold t for truncated Gaussian models and for any thresholded continuous field, or by tuning the probability p_0 for Boolean models or the external proportion for other models. On an aquifer, X_p would for example correspond to the thresholding of the permeability. In the context of oil reservoirs, p is the net-to-gross parameter.

3.3. Scalar indicators

Among the many indicators that can be defined to characterize the connectivity, we first present the scalar indicators that have been the most studied in the literature. We will use two major sources: (i) from integral geometry we will borrow a topological functional known as the Euler characteristic which will be related to the number of clusters; (ii) using the percolation theory paradigm, we will detail how some connectivity metrics such as the size of the largest cluster or the connectivity at long distance vary with the proportion. In particular, we will show how these quantities change dramatically around a specific proportion, known as the percolation threshold.

3.3.1. The Euler characteristic and the number of clusters

One of the most important result of integral geometry [103,104] states that there is only a very limited number of functionals describing the geometry of sets which are motion invariant and additive in the sense that $g(A \cup B) = g(A) + g(B)$ whenever A and B are disjoint subsets of $\mathcal{G}$. Up to the multiplication by a constant, there is in fact only one functional per dimension $d' \leqslant d$. In 3D, the four functionals are, in decreasing order of dimension, the volume, surface, mean curvature and the Euler characteristic. This last characteristic, dimensionless, provides a number which is only related to the shape (the topology) of the set. It is the unique integer valued functional ϕ defined on subsets A of X satisfying $\phi(A) = 0$ if $A = \emptyset$ and $\phi(A) = 1$ if A is a cube with all cells in the permeable phase. In 2D, the Euler characteristic is simply the number of clusters minus the number of holes in the clusters. In 3D, it is the number of clusters, minus the number of "handles" plus the number of holes. Thus, for example, the Euler characteristic of a snowball is 1 (one cluster, no hole, no handle); that of a tennis ball is 2 (one cluster, no handle, one hole); that of coffee cup is 0 (one cluster, no hole, one handle).

It can be shown [105] that on a grid the Euler characteristic of a set X can be computed according to:

$$\phi(X) = \sum_{i=0}^{d} (-1)^i \# e_i(X), \tag{2}$$

where e_i are the elementary facets of dimension i on the graph of $\mathcal{G}$, i.e. $\#e_0(X)$ is the number of sites of X and $\#e_1(X)$, $\#e_2(X)$, $\#e_3(X)$ its number of edges, faces of 4 adjacent sites and volumes of 8 adjacent sites. Eq. (2) provides a very efficient algorithm for computing the Euler characteristic (see Appendix A for available codes). Note that Eq. (2) shows that ϕ is nothing but a local 8-point statistics averaged on the entire grid. It is thus a local characteristic. Explicit formula for the specific Euler characteristic $\phi(X)/|\mathcal{G}| = \phi(X)/L^d$ is available for the Boolean model of fixed size squares described above [99]. For the truncated Gaussian model on $\mathcal{G}$, theoretical values can be obtained by computing numerically the multiple integrals corresponding to Eq. (2). On the continuous plane or 3D space, explicit formulas are also available for the Boolean model and for the truncated Gaussian model, provided the covariance function is twice differentiable at the origin [106,107], see Section 3.7.

Contrarily to $\phi(X)$, the number of clusters, $N(X)$, is a non-additive, global characteristic for which there exists no analytical formula. Allard [99] showed that for stationary random sets, the specific number of clusters can be related to the mean size of the cluster containing the origin provided the origin is in the permeable phase:

$$\lim_{L \to \infty} N(X)/L^3 = pE_X[\#C_0 | I(0) = 1], \tag{3}$$

where $E_X[]$ stands for the mathematical expectation with respect to the distribution of the random set X and $\#C_0$ denotes the number of cells of the cluster containing the origin. There is thus a strong relation between the number of clusters and their size through Eq. (3).

Let us now assume that the proportion increases as presented in the introduction of this section and let us denote the set X_p to emphasize its dependency to p. By convention, $X_0 = \emptyset$ and $X_1 = \mathcal{G}$. This situation is illustrated in Fig. 7 for 2D Boolean models of fixed size squares, for different square sizes l, while keeping the ratio grid size/ grain size constant. At low p, the clusters are small and well spaced. As p increases but still remains low enough, existing clusters increase in volume and new ones are created in the many empty spaces. As a result, both $\phi(X_p)$ and $N(X_p)$ increase. Furthermore, for low values of p, there is no room for "holes", and $N(X_p) \simeq \phi(X_p)$ (Fig. 7(a), left). Then, as p continues to increase, clusters that were different will merge. Creation of new clusters and merging of already existing ones are antagonistic effects which find a balance for a particular proportion where $N(X_p)$ and $\phi(X_p)$ reach a maximum. For larger proportions, $N(X_p)$ and $\phi(X_p)$ will decrease. The Euler characteristic decreases faster because the merging of clusters tends to create "holes" (Fig. 7a, middle). At some point, the Euler characteristic becomes negative, because the number of holes is larger than the number of clusters. At a particular proportion, the clusters become dominated by a single large cluster, spanning the whole region and joining opposite sides of the grid. The set X_p is said to percolate. For very large proportions almost all clusters are absorbed by the largest one (Fig. 7(a), right). The number of clusters tends to one, indicating that all clusters have merged into a single one. In this case, all points of the permeable phase are connected. The Euler characteristic reaches a minimum. Eventually the holes get filled and $\phi(p)$ increases again, up to its limit value $\phi(1) = 1$.

The number of clusters and the Euler characteristic are thus indicators of the connectivity computed on the whole grid, but with very different meanings. $\phi(X)$ is a local characteristic summed over the whole grid, while $N(X)$ is a truly global one. Although very similar for low p, they behave very differently when $p \gg 0$. However, unless $N(X)$ is very small, they are not direct indicators of the size of the existing clusters, and they do not provide information about how likely two points of the permeable phase are connected. Therefore, further indicators are needed and will be described in the following paragraphs.

3.3.2. Percolation

Percolation is the transition from many disconnected clusters to a very large spanning cluster as p increases. On finite grids, several percolation metrics can be defined.

(1) A first metric is the ratio of the volume of the largest cluster to the total volume, denoted $\Theta(p)$. This quantity corresponds to the first geobody connectivity defined in [60,57]. Fig. 8 illustrates this metric for 2D truncated Gaussian models with different covariance functions.

(2) A second metric, denoted $\Gamma(p)$ is the proportion of the pairs of cells (distinct or not) that are connected amongst all the pairs of permeable cells:

$$\Gamma(p) = \frac{1}{n_p^2} \sum_{i=1}^{N(X_p)} n_i^2, \tag{4}$$

where n_p is the total number of permeable cells in X_p and n_i is the number of cells within cluster C_i. Note that by construction the first moment of the distribution of the cluster size, $n_p^{-1} \sum_{i=1}^{N(X_p)} n_i$, is always equal to 1. $\Gamma(p)$ is its second order moment, and the moment of order 0 is $N(X_p)$. This quantity is the second geobody connectivity studied in [57], in which it is shown that it behaves very similarly to $\Theta(p)$.

(3) A third quantity of interest is the indicator function $T(X_p)$ equal to 1 when two opposite sides of the field are connected and equal to 0 otherwise. We shall call the particular proportion defined as the lowest value p such that $T(X_p) = 1$ the percolation transition of X. It will be denoted $p_t(X)$.

Percolation theory [43,44] studies connectivity, mostly on infinite Bernoulli grids. Its fundamental theoretical result is that, on these grids, there exists a proportion p_c, called percolation threshold, with $p_c \neq 0$ and $p_c \neq 1$ such that if $p > p_c$, there will be, with probability equal to one, a unique cluster with infinite volume, denoted C_∞. If $p < p_c$ this event has probability zero. This theoretical result was generalized in [99] to any random set with finite range, i.e. random sets whose centered covariance function, $K(h)$, is such that there exists $a > 0$ with $K(h) = 0$ whenever $|h| > a$. In summary, as $L \to \infty$:

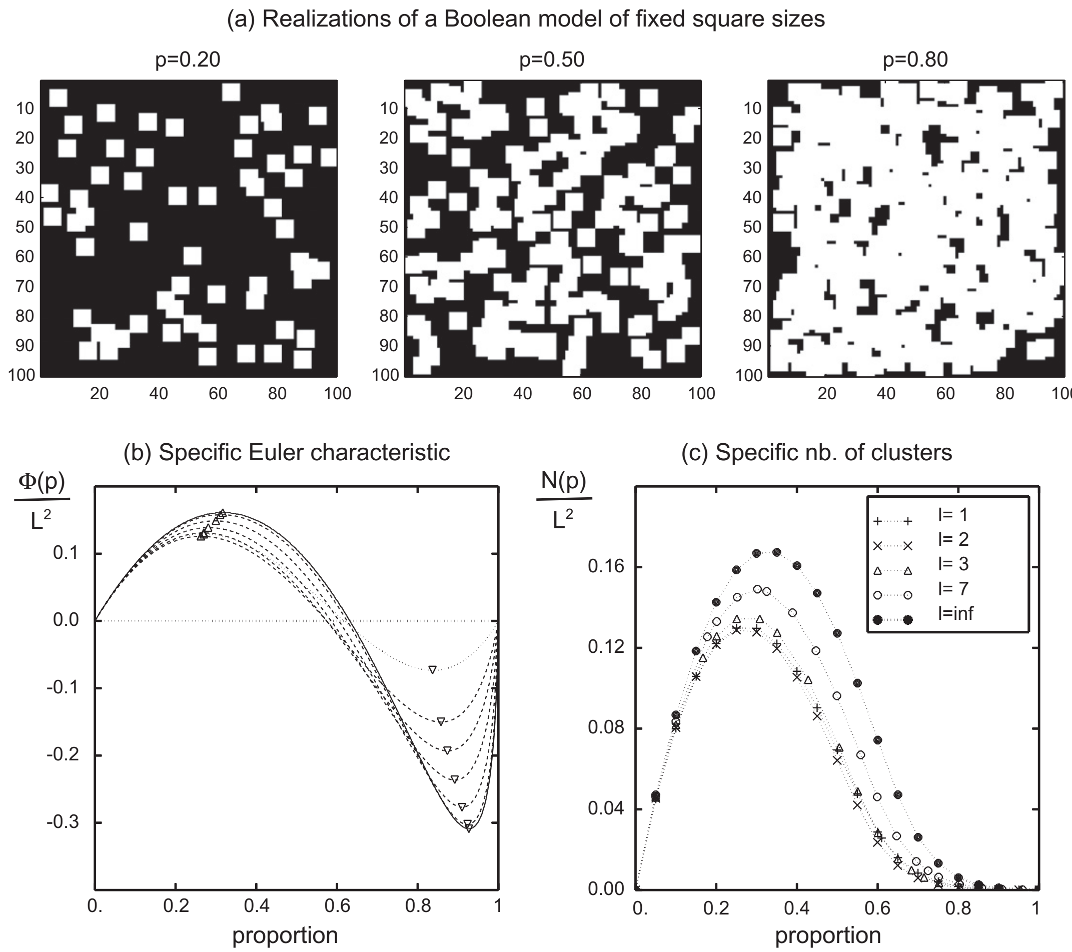

Fig. 7. Specific Euler characteristic (b) and specific number of clusters (c) for a 2D Boolean model of fixed size squares. (b): from grain size $l = 1$ (dashed line) to continuous model $l \to \infty$ (solid line). (c): grain size is denoted l. In both case, the ratio L/l is kept constant.

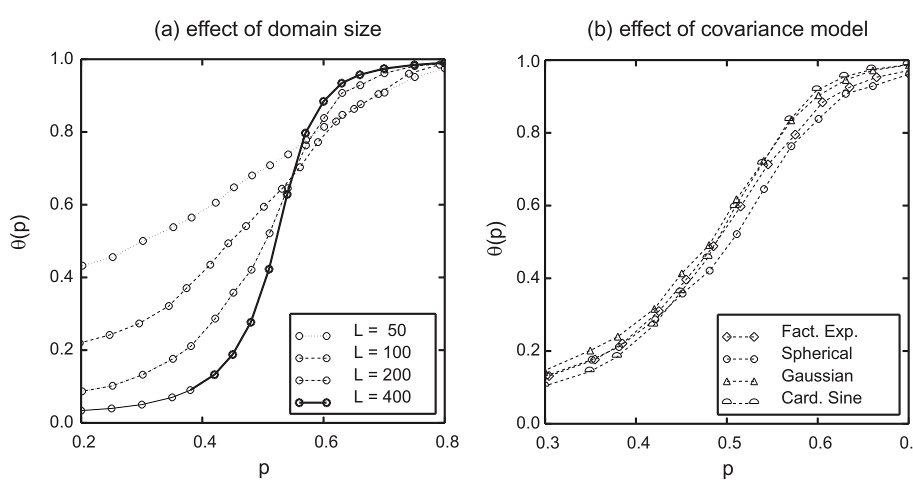

Fig. 8. Connectivity metric $\Theta(p)$ for the truncated Gaussian model (average computed on 200 realizations). Left: spherical covariance of fixed range and varying size of the domain. Right: Different covariance models on the same grid with identical ratio L/l.

$$\Theta(p) \to P(x \in C_\infty), \quad \Gamma(p) \to \Theta(p)^2, \quad p_t \to p_c, \tag{5}$$

and

$$\begin{aligned} p < p_c: \quad & \Theta(p) = 0; \quad \Gamma(p) = 0; \quad T(X_p) = 0, \\ p > p_c: \quad & \Theta(p) > 0; \quad \Gamma(p) > 0; \quad T(X_p) = 1. \end{aligned} \tag{6}$$

Going back to finite grids, instead of a sharp transition at p_c, the functions $\Theta(p)$ and $\Gamma(p)$ follow a sigmoid shape illustrated in Fig. 8 (S-shape, for short), typical of a phase transition phenomenon located at a proportion equal to the percolation transition $p_t(X)$. Many numerical studies were conducted for describing and quantifying this S-shape curve [99,62,61,60,57] based on stochastic simulations of random sets. King [61] illustrates the concepts of percolation theory on 2D Boolean models of rectangles and 3D cubes of fixed size and orientation. A similar study was conducted in 2D and 3D in Allard and Heresim Group [62] and Allard [99] with an emphasis on the variability of the percolation threshold on finite grids and on the analysis of the connectivity range (see definition later: Eq. (12)). In the same papers a truncated Gaussian model is studied, and a comparison between the two models is made. Larue and Hovadik [60] and Hovadik and Larue [57] performed an extensive simulation study which includes Boolean models, truncated Gaussian models and SIS simulations, with interesting extensions to dynamical measures of connectivity. These studies showed clearly that the *S*-curve describing percolation could be observed on all tested situations. The main results are briefly summarized:

- *Space dimension:* Percolation holds from values between $p = 0.55$ and $p = 0.65$ in 2D, down to values roughly between $p = 0.25$ and $p = 0.35$ in 3D. There is thus a dramatic effect of the dimension of the space.
- *Variability as a function of p:* The variability of $\Theta(p)$, from one realization to the other is maximum at those proportions where $\Theta(p)$ increases the fastest, i.e. for proportions near the percolation transition. The slope of the curve at $p_t(X)$ is thus an excellent indicator of the variability at the percolation.
- *Object and grid size:* The variability of $\Theta(p)$ depends also on the grid size compared to a typical length of the random set, denoted l (we will give a precise definition of this typical length in Section 3.5). As the ratio L/l increases, two observations can be made. Firstly, the transition is sharper, implying that the slope of the *S*-shape curve increases. This effect is clearly visible on the left panel of Fig. 8. Secondly, the connectivity metrics $\Theta(p)$, $\Gamma(p)$ or $p_t(X)$ are less variable from one realization to the other. This point is very well illustrated in [60]. From percolation theory considerations [44,43,61], a universal scaling law was derived:

$$\Theta(p;L/l) = (L/l)^{-\beta_d/\nu_d} F\left[(p-p_c)(L/l)^{1/\nu_d}\right], \tag{7}$$

  where β_d and ν_d are universal exponents depending only on the dimension d. The function F is a universal S-curve, but its precise form is unknown. The right panel of Fig. 8 shows clearly that the curves $\Theta(p)$ are very similar for truncated Gaussian models with different covariance functions when the ratio L/l is kept constant.
- *Discretization effect:* When increasing l while keeping a constant ratio L/l, the variability of all percolation metrics remains constant, while the Euler characteristic and the number of clusters may vary a lot. Euler characteristic and number of clusters depend highly on the regularity at the origin of the covariance of X. For Boolean models and for truncated Gaussian models with twice differentiable covariance functions at the origin (Gaussian, cubic, etc.) the boundary of the random set is very regular. This has two consequences: first, there is a low number of clusters; second, connectivity metrics do not change much with the discretization as soon as a reasonable one is reached. On the contrary, for truncated Gaussian models with exponential or spherical covariances, there is a massive number of very small clusters not participating to the connectivity at long distance. Percolation quantities such as $\Theta(p)$ and $\Gamma(p)$ do not vary with the discretization since they do not depend on small clusters.

3.4. Connectivity function

The connectivity function is defined as the probability that a cell x in X is connected with (i.e. belongs to the same cluster as) another cell of X located at y [62,13]:

$$\tau(x,y) = P(x \leftrightarrow y | x,y \in X) = P(C(x) = C(y) \neq 0). \tag{8}$$

When the random set X is stationary, which we always assume in this paper, we more simply consider the function of the lag vector h

$$\tau(h) = P(x \leftrightarrow x+h | x, x+h \in X) = P(C(x) = C(x+h) \neq 0). \tag{9}$$

It is a global characteristic since connections involve the entire grid.

On single realizations, this quantity can be computed by dividing the occurrences of the event $\{C(x) = C(x+h) \neq 0\}$ by the occurrences of $\{C(x) \neq 0 \text{ and } C(x+h) \neq 0\}$ on the grid, as a function of h. From now on, we will consider (9) as the definition of the connectivity function.

Since $\{x \leftrightarrow x+h\}$ necessitates $\{x, x+h \in X\}$, the unconditional probability $P(x \leftrightarrow x+h)$ can easily be decomposed as:

$$\begin{aligned} P(x \leftrightarrow x+h) &= P(x \leftrightarrow x+h | x, x+h \in X) P(x, x+h \in X) \\ &= \tau(h) K(h), \end{aligned} \tag{10}$$

where $K(h)$ is the non-centered covariance function of X. We recall that $K(0) = p$, where $p = E[I(x)]$ is the proportion of the random set and that $\lim_{|h|\to\infty} K(h) = p^2$. The probability that any two grid points is connected is thus naturally factorized into the product of the non-centered covariance and the connectivity function. This probability $P(x \leftrightarrow x+h)$ is denoted $C_2(h)$ and named the 2-point cluster function by Torquato et al. [108], Matheron [101]. A slightly different definition is used by Stauffer and Aharony [43]. They consider the probability $P(x \leftrightarrow x+h | x \in X) = P(x \leftrightarrow x+h)/p$.

It has been observed empirically [99,109] that for many random sets, $\tau(h)$ is anisotropic for the Euclidean metric at short distances, in particular because of the anisotropy of the neighborhood definition, and that is isotropic at medium to long distance despite the anisotropy of the cubic lattice and the potential anisotropy of the random sets.

On Fig. 9a the connectivity function $\tau(h)$ is reported as a function of p for a fixed separation vector h. The existence of a sharp transition very similar in nature to the *S*-functions in Fig. 8 is clearly visible. For large distances, there is a clear separation between the proportions for which $\tau(h) \simeq 0$ from those where $\tau(h) \gg 0$.

The way $\tau(h)$ decreases with $|h|$ depends on the proportion p. It is illustrated in Fig. 9b. For proportions below the percolation threshold, $\tau(h)$ decreases rapidly and converges towards 0 for large $|h|$ because all connected components are small. For proportions above the percolation threshold, $\tau(h)$ tends to a sill whose value is nothing but $P(x \in C_\infty)^2$, where C_∞ is the largest cluster (the percolating cluster, of infinite size when $L \to \infty$). On finite grids, $P(s \in C_\infty)^2 \simeq \Theta(p)^2$, the approximation being sharper as the grid size increases.

The connectivity function of random sets with regular boundaries is almost parabolic near the origin, while it is near to linear at the origin for irregular random sets such as truncated Gaussian models.

In the process of writing this review, we found an interesting (and so far unread) relationship between the overall sum of the connectivity function $\tau(h)$ computed in X_p and the global index $\Gamma(p)$ defined in Eq. (4):

$$\begin{aligned} \sum_h \tau(h) &= n_p^{-1} \sum_{x\in X_p} \sum_{x+h \in X_p} P(x \leftrightarrow x+h) = n_p^{-1} \sum_{x \in X_p} |C_x| \\ &= n_p^{-1} \sum_{i=1}^{N(X_p)} \sum_{x_i \in C_i} |C_i| = n_p^{-1} \sum_{i=1}^{N(X_p)} |C_i|^2 = \Gamma(p) n_p, \end{aligned} \tag{11}$$

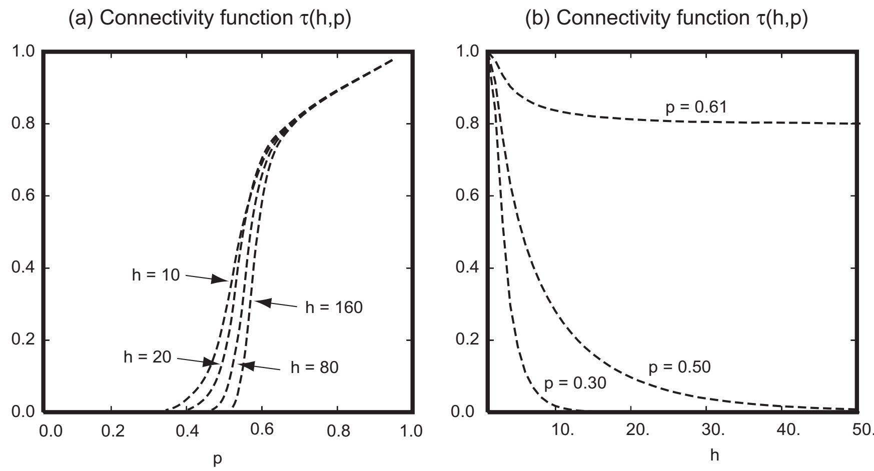

Fig. 9. Connectivity function for a truncated Gaussian model: (a) as a function of p, for increasing values of $|h|$ from left to right; (b) as a function of $|h|$, for different values of p.

where $C_1, \ldots, C_{N(X_p)}$ denotes the clusters of X_p and $|C_i|$ their size. This results deserves some comments.

- When $p < p_c$, the connectivity function tends rapidly to 0. The sum $\sum_h \tau(h)$ is thus bounded away from n_p, entailing a connectivity index $\Gamma(p)$ to the order $O(n_p^{-1})$.
- When $p > p_c$, $\tau(h)$ converges towards a non-null value as $|h| \to \infty$. The sum $\sum_h \tau(h)$ is thus to the order $O(n_p)$, implying that $\Gamma(p)$ is to the order $O(1)$.

The quantity $\Gamma(p)$ can thus be used a criterion for a first, quick assessment of the global connectivity. If it is larger than say, 20%, there is a very good chance that a very large connected component dominates, spanning the whole domain. In this case percolation holds almost surely. We will provide some justification supporting this value in Section 3.7.

Let us now study the connectivity function $\tau(h)$ in more details. To emphasize the dependency of the connectivity function to the proportion, it will be noted $\tau_p(h)$ in the rest of this section. It can be proved [44], and it has been confirmed on numerous studies [61,99], that the decreasing part of the connectivity function behaves as an isotropic exponential function for large distances h, i.e. for $p < p_c$

$$\tau_p(h) = \exp\{-|h|/\xi(p)\}, \quad \text{for } |h| \gg 1, \tag{12}$$

where $\xi(p)$ is a length parameter, called correlation length in percolation theory [44]. Since this denomination can be confused with the range parameter of $K(h)$, it was renamed *connectivity range* in [99,62]. It is a "typical" length of the clusters and it depends obviously on p. Connectivity at distances $|h|$ larger than $3\xi(p)$ are very unlikely. From this equation the value $\xi(p)$ that best fits the empirical connectivity equation can be estimated. Alternatively, since $\int_h \tau(h)dh = \omega_d \xi(p)^d$ for a constant ω_d depending on the dimension d ($\omega_2 = 2\pi$ and $\omega_3 = 8\pi$), an excellent approximation of $\xi(p)$ is given by Eq. (11):

$$\xi(p)^d \simeq \omega_d^{-1} \Gamma(p) n_p. \tag{13}$$

This equation provides an equivalence, up to some multiplicative constant, between the global connectivity index and the dth power of the connectivity range.

For proportions above the percolation threshold, the connectivity function can be decomposed as the sum of two terms, one corresponding to the infinite cluster C_∞ (the largest cluster on finite grids), the other one corresponding to connections between finite size clusters:

$$\begin{aligned}\tau_p(h) &= P(x \leftrightarrow x+h | x, x+h \in X) \\ &= P(C(x) = C(x+h) = C_\infty) + P(C(x) = C(x+h) \neq C_\infty) \\ &= \Theta(p)^2 + \tau_p^f(h),\end{aligned}$$

where $\Theta(p) \simeq |C_{max}|/|X|$ on finite grids and $\tau^f(h)$ is the connectivity function restricted to other clusters of X than the largest one, behaving as in Eq. (12).

3.5. Power laws

We have seen that a great number of numerical studies indicate that many percolation metrics behave according to S-curves. Theoretical considerations, already described in Section 3.3.2 and in Eq. (7), indicate further that up to some scaling, they can universally be described by some abstract function F. The precise shape of F is unknown, but it can be usefully summarized by a location and a scale parameter: the location parameter, e.g. the point of inflexion of the curve $\Theta(p)$, is related to the percolation transition; the scale parameter, e.g. the slope of the curve at the inflexion point is related to the variability at the percolation transition. Note that the cumulative probability function (cpf) of the percolation transition $p_t(X)$, obtained from several realizations of the same random set model, follows a S-curve very similar to those of $\Theta(p)$ and $\Gamma(p)$. This third percolation metric can thus be summarized by the average and variance of the values $p_t(X)$. For the rest of this section, we will denote $\bar{p}_f$ the average (on many realizations) of any of these location parameters and σ_f will denote its standard deviation.

In [62,99] it is shown that the Boolean model and the truncated Gaussian model can be unified in a unique framework if the "object" volume is defined as the integral range, A. We recall that the integral range is the normalized integral of the centered covariance function [100]:

$$A = \frac{1}{K(0)} \sum_{y \in \mathcal{G}} \sum_{x \in \mathcal{G}} \left[K(y-x) - p^2\right]. \tag{14}$$

Closed form expressions for A exist for some models, including the Boolean model and truncated Gaussian model [99]. For the connected field model, A can be computed numerically. The quantity l in Eq. (7) can then be set to the dth-root of A: $l = A^{1/d}$.

From the scaling law (7) and from empirical considerations [62,99], one can now establish that, if L/l is large enough (say $L/l > 5$):

$$\sigma_f \simeq a_d (L/l)^{-1/\nu_d}; \quad p_c - \bar{p}_f \simeq b_d (L/l)^{-1/\nu_d}; \quad \xi(p)/l \simeq c_d (p_c - p)^{1/\nu_d}, \tag{15}$$

where a_d, b_d and c_d are constants depending on the dimension d. These scaling laws are illustrated in Fig. 10 for 2D random sets. On the left panel, the linear relationship between the log of the standard deviation of the percolation transition and the log of the ratio L/l for the two models is a strong evidence in support of the power law relating the variance of the percolation transition on finite grids to the ratio L/l. The right panel depicts the proportion p as a function of $(\xi/l)^{-1/\nu_2}$ for the Boolean model and for the truncated Gaussian model. The intersection with the y axis, corresponding to $\xi \to \infty$, provides an estimate of the percolation threshold on an infinite grid of the considered random set.

Numerous empirical studies [43,44,61,62] agree on $\nu_2 = 4/3$ and $\nu_3 = 0.875$. For the regular infinite cubic grid, the same authors found that the percolation threshold is $p_c = 0.311$ in 3D and $p_c = 0.59$ in 2D for Bernoulli grids. For other random sets, p_c must be estimated from Eq. (15). Allard [99] found $a_2 = c_2 = 0.45$ and $b_2 = 0.23$. Values for $d = 3$ are not known. Note that in Eq. (15), $\xi(p)$ is the expected value of the connectivity range among many realizations, whereas in Eq. (13) it is computed for the particular realization under study. Comparison between the two values provides thus a fast and efficient way for deciding whether the realization is less or more connected than on average.

Put together, Eq. (15) state that the proportion interval [0, 1] can be approximately divided in three domains:

- $p < p_c - (L/l)^{-1/\nu}$: percolation is very unlikely; the connectivity function is almost null at distances $|h| > 3\xi(p)$.
- $p_c - (L/l)^{-1/\nu} \leqslant p \leqslant p_c + (L/l)^{-1/\nu}$: percolation and connections at long distance are likely but not certain. From a connectivity point of view, it is a domain of high variability for all global metrics. Several clusters of quite similar size can coexist. Note that the range of this interval decreases if the grid size increases or if the range decreases. To the limit, if $L \to \infty$, the interval is reduced to a single proportion equal to the percolation threshold.
- $p > p_c + (L/l)^{-1/\nu}$: percolation holds almost certainly. A single very large cluster dominates, spanning the entire domain. Any two points are almost certainly connected.

Broadly speaking, there is thus only three connectivity behaviors, corresponding to the three domains described above: below, around or above the percolation threshold.

3.6. Binary mixture of permeable media

All the tools described above for binary media do not consider the fact that the two phases can be permeable. We assumed that X^c, the complementary phase of X, was completely impermeable. A single pixel belonging to X^c separating two permeable bodies was sufficient to consider that these two bodies were disconnected.

However, most often in practice, the geological medium is made of two permeable phases. Let us assume, without any loss of generality, that X is the high permeability phase, and X^c the low permeability phase. Because the flow has to go through X^c, the connectivity of both phases play an important role and we cannot consider only X.

In 2D and for infinite media, the two phases play a symmetrical role: if the low permeability phase is percolating, the high permeability phase is disconnected and vice versa. This is not true in 3D. But, even in 2D, the connectivity metrics of one phase are not uniquely related to their equivalent connectivity metrics for the complementary phase. One has therefore to compute the connectivity metrics of the two phases to get a complete vision of the topology of the medium. In the following, we denote all those additional metrics by simply adding a subscript c to the name of the variables that we already introduced. For example, Γ_c is the probability of having two points of the low permeability phase connected, or ϕ_c is the Euler characteristic of the low permeability phase. Using this extended set of metrics one can already get a good characterization of binary media.

However, this may not be sufficient to describe accurately the conductive properties of the medium that are strongly influenced by the proximity between high permeable inclusions [110]. To quantify that effect, Knudby et al. [110] compute the ratio of the average distance (within X^c) between the inclusions to the average distance between the center of the inclusions by analyzing the position and size of the inclusions along the flow direction.

A more general analysis of the proximity of the inclusions can be conducted using the two fundamental operations of mathematical morphology: the erosion and dilation of a set [105]. To introduce this tool, we have to define a small set that is called the *structuring element*. It will be denoted B in the following. It is characterized by its shape which is often very simple such as a square or circle in 2D or a sphere or a cube in 3D. The size and shape of the structuring element are defined by the user.

The *erosion* of X by B is the set of points x of $\mathcal{G}$ such that the structuring element B_x centered on each point x is entirely included in X:

$$X_{erod} = X \ominus B = \{x \in \mathcal{G} : B_x \subseteq X\}. \tag{16}$$

The result is a new set that has been eroded by a layer whose width is the radius of B. The erosion removes all the isolated clusters whose breadth is smaller than the diameter of B, i.e. those part of X such that B does not fit in.

The *dilation* of X by B is the set of points x of $\mathcal{G}$ such that the structuring element B_x intersect X:

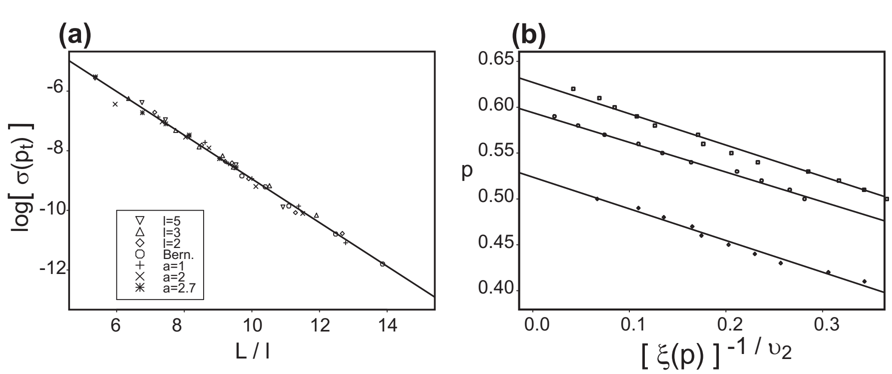

Fig. 10. Power laws for a 2D Boolean model (squares of size l) and a 2D Truncated Gaussian model (exp. covariance with range a). Left: log $(\sigma(p_t))$ vs. log (L/l). Right: p vs. $(\xi(p)/l)^{-1/\nu_2}$.

$$X_{dilat} = X \oplus B = \{x \in \mathcal{G} : B_x \cap X \neq \emptyset\}. \quad (17)$$

This operation adds a layer of pixels to the set X. The width of this layer is equal to the radius of the structuring element B. Applying a dilation fills the holes and bays which are narrower than the diameter of B.

We have now defined the basic tools that we need to generalize the previous connectivity definitions and investigate more precisely the connectivity structure of a binary field with two permeable phases.

Let us consider a simple structuring element B that is an elementary cross (1 central block, plus one single block on each side of it and in all directions). We pose $X_0 = X$ and build two series of sets. The first is obtained by a series of dilations:

$$X_k = X_{k-1} \oplus B, \quad k \in 1, \ldots, M, \quad (18)$$

where $X_M = \mathcal{G}$; the whole grid is filled after M successive dilations. The second series is obtained by a series of erosions,

$$X_{-k} = X_{-(k-1)} \ominus B, \quad k \in 1, \ldots, N, \quad (19)$$

until the set X_0 is completely eroded for $k = N$. By combining the two, we get a series of successive concentric sets X_k with $k = [-N, \ldots, -1, 0, 1, \ldots, M]$. k is a distance to the boundary of the initial set. A convenient representation of this ensemble of sets is to represent in any point of the domain the value k corresponding to the set with the lowest index k to which that point belongs. This can be seen as a map of oriented distances to the boundary of the initial set X. Positive distances are those corresponding to points located outside of X and negative distances are for the points within X. By computing scalar connectivity metrics defined previously on the ensemble of sets, we obtain a series of characteristic curves $\Gamma(k)$, $\phi(k)$, etc.

The behavior of these curves allows to identify precisely if the metric computed on the initial set is stable: for example a stable value of Γ is reflected by a constant value around $k = 0$. Sudden drops in the curve indicate the presence of critical thresholds at a certain average distance.

Fig. 11 (left column) illustrate the behavior of those curves for three different binary media having the same proportion p of inclusions and the same values for Γ and ϕ. However, the horizontal flux through these three media is different. For the medium represented in Fig. 11(h), the horizontal flux is higher than for the medium represented in the two other figures since all the high permeability inclusions are almost aligned. The medium will almost behave like a stratified one. The middle column shows for the three media the distance function. Finally, the right column shows the characteristic curves $\Gamma(k)$ and $\Gamma_c(k)$ for the corresponding media. We see clearly that even if the connectivity indicators of the three medium are identical, their characteristic curves are very different. In particular, for the medium with inclusions placed on a regular grid, we see that the connection of all inclusions occurs abruptly for a distance between 35 and 45 (Fig. 11(c)). This shows that the distance between the inclusions has little variability. On the opposite, in the case of the random location of the inclusions, both functions $\Gamma(k)$ and $\Gamma_c(k)$ display some complex variations showing the variability in distances between the inclusions (Fig. 11(g)).

On the last line, we see clearly that for a small dilation distance (lower than 10), the inclusions rapidly merge and the probability of having two cells connected remain for some time at a value of 0.5 (Fig. 11(j)) indicating that we have two big inclusions (the two disconnected layers of inclusions). At a distance of about 50 these two connected components merge to become a single one ($\Gamma = 1$).

Note that anisotropic distances and directional effects can be accounted for. One has simply to use a structuring element B such as a segment with a given orientation to define the distance only along a certain direction. Applying a vertical and an horizontal distance and weighting them is also a possibility.

Finally, instead of using a simple erosion or dilation operation which are only a function of distance, one can use an opening or closing morphological operation which are a function of the size of the objects (see [111] for a definition). Vogel et al. [112] applied such a technique to relate the connectivity with the pore size distribution. Similar application could be made in hydrogeology to relate the connectivity with the channel size distribution.

3.7. Connectivity of continuous fields

The next step is to generalize the static connectivity metrics to random fields of continuous variables. These fields can be obtained through direct imaging of a medium via remote sensing, tomography, geophysics, etc. or they can be produced by simulation techniques. We will denote $Y(x)$ the continuous variable that we want to characterize. Typically it will be the logarithm of the hydraulic conductivity.

The procedure follows rather closely what has already been described for binary media. It is based on the decomposition of the continuous field into a series of sets. For each set, the static connectivity metrics are computed. There are different ways to decompose the field. The simplest is to apply a threshold t to the continuous variable:

$$X_t = \{x : Y(x) \geqslant t\}. \quad (20)$$

In this way, we obtain a series of sets such that $X_{t_1} \subset X_{t_2}$ if $t_1 > t_2$. The threshold is chosen to vary between the minimum and maximum value of Y. By computing the scalar connectivity metrics on each of these sets, one can build a set of characteristic curves for the medium: $\phi(t)$,$\Gamma(t)$. Many authors applied that technique for the Euler characteristic $\phi(t)$ [29,102,113,106,34,94]. Here, we argue that one should also use $\Gamma(t)$ and consider the connectivity metrics of the complementary phase: $\phi_c(t)$,$\Gamma_c(t)$.

To illustrate how the different characteristic curves vary as function of the threshold t, we first consider a multi-Gaussian medium such as the one shown in Fig. 12(a). When the range parameter is small as compared to the grid size L, Mantz et al. [106] give the analytical expression of the Euler number as a function of the threshold in 2D:

$$\phi(t) = \frac{L^2 k^2}{\sqrt{2\pi\sigma^2}} (t - E[Y]) \exp\left[\frac{(t - E[Y])^2}{2\sigma^2}\right], \quad (21)$$

where k^2 is proportional to the second derivative of the centered covariance function $c(h)$ of $Y(x)$:

$$k^2 = -\frac{c''(h = 0)}{2\pi\sigma^2}. \quad (22)$$

Note that, as already discussed in Section 3.2, Eq. (21) only applies to random fields with regular covariance functions, i.e. covariance functions with finite second derivative at the origin. For other models, such as exponential covariance models, the Euler number is infinite. This theoretical result tells us that $\phi(t)$ should be symmetrical around the mean (in our example $E[Y] = 0$). When the range parameter is not small compared to the grid size, the symmetry around the mean value is broken $\phi(t - E[Y]) \neq -\phi_c(t - E[Y])$ and instead of having a single critical value corresponding to a unique percolation threshold, we can define a series of 5 critical points (Fig. 12(c)). This is due to (1) the fact that the grid is not symmetrical in terms of 4-connectivity as described in Section 3.1 and (2) finite size effects.

In Section 3.3 we showed how connectivity metrics such as ϕ behave statistically (i.e. on average on many realizations) as a function of p. Here, we will describe those 5 critical stages step by step on a single realization. It is perhaps useful to recall that the propor-

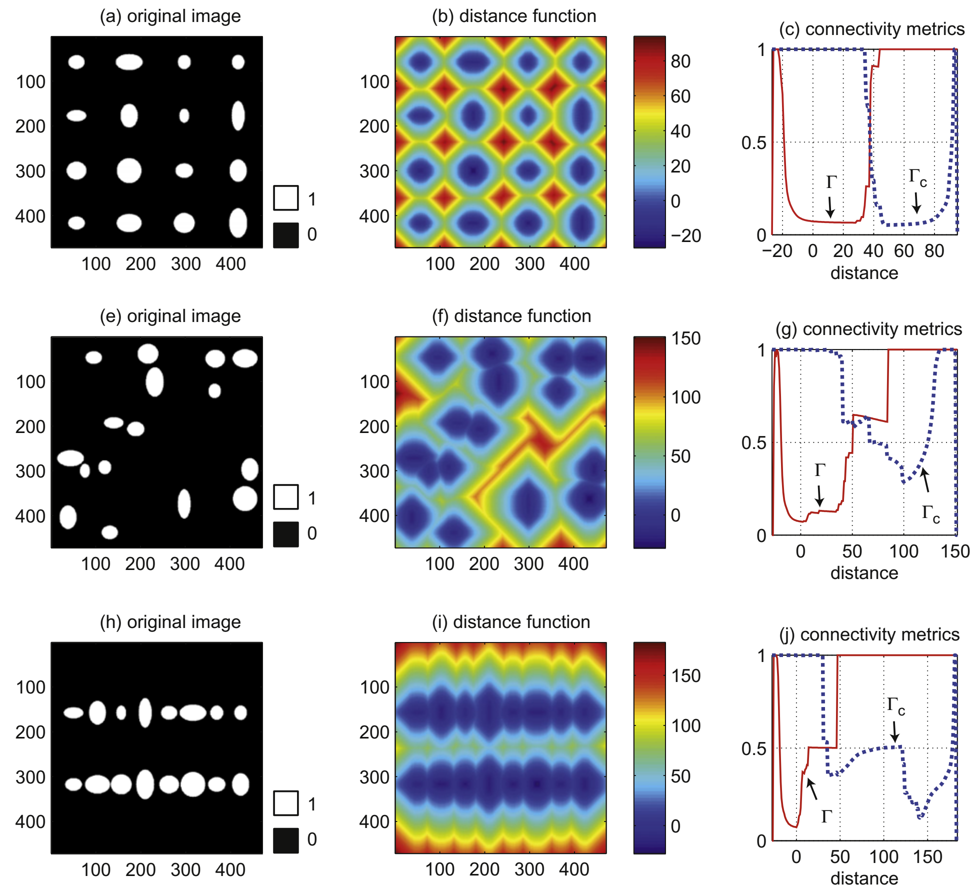

Fig. 11. Illustration of the use of morphometric distances to better quantify the topology of a mixture of permeable media. The media displayed in the left column have been borrowed in Knudby et al. [110].

tion can be deduced from the threshold with $p = 1 - G(t)$, where $G(t)$ it the cpf of a $(0,1)$ Gaussian random variable. For very low values of t, the set X_t is the complete grid $\mathcal{G}$, all the values in the field are above the first threshold. The Euler characteristic $\phi(t)$ is equal to 1, as well as Γ is equal to 1 because all the pixels are connected. The complementary phase is empty, hence $\phi_c(t) = 0$. When t increases, we start to isolate clusters of low values (Fig. 12). $\Gamma(t)$ remains equal to 1 since all the high values are connected together in a single cluster. $\phi_c(t)$ increases because the number of inclusions of low permeability increases. On the opposite $\phi(t)$ decreases because there are more and more holes in the high permeability unique connected component. When we reach a value of t around -1.5, we see that $\phi_c(t)$ starts to decrease and $\phi(t)$ starts to increase. This is because the lenses of low permeability start to coalesce and connect together to create a smaller number of larger connected components. The change is rather sudden and abrupt around $t = -1$, corresponding to $p = 0.16$. At this point $\Gamma(t)$ starts to become lower than 1. We start to have isolated clusters of high permeability disconnected to the larger cluster which still span the whole domain and connect one side to the other (Fig. 12e shows two clusters for $t = 0.7$). The large cluster is divided in two pieces when $\phi(t) = 0$, which occurs at $t = -0.295$. Before, $\phi(t)$ was negative because there were more holes than connected components, now the situation will reverse, there will be more isolated lenses than holes. The transition corresponds to a new threshold. But, at that point the low permeability part is not yet completely connected as revealed by the fact that $\phi_c = 0.17$ (Fig. 12f shows several clusters for $t = -0.295$, i.e. $p = 0.61$). At $t = 0$, none of the two phases are well connected. We have large clusters of high and low permeability (Fig. 12e and f for $t = 0$). When we continue to increase t, we follow in reverse order the same behaviors that were described for $t < 0$. When $t = 0.15$, corresponding to $p = 0.44$, $\phi_c(t) = 0$, a large cluster of low permeability connects all the faces of the bloc. When $t = 0.7$, i.e. $p = 0.24$ the low permeability part becomes fully connected and $\Gamma_c = 1$. The zones with permeabilities lower than t constitute a unique connected component $\Gamma_c = 1$. The permeabilities higher than t belong to isolated clusters.

All this detailed description highlights the fact that there is not a unique percolation threshold for fields of finite size with a typical correlation length that is not very small as compared to the size of the field. There is clearly a range of thresholds around $t = 0$ where none of the two phases percolate. Instead, it is more useful in that case to define a range of values $[t_1, t_2]$ which defines the most connected part of the field. To estimate $[t_1, t_2]$ we suggest to identify the values such that none of the complementary phases are almost fully connected:

$$[t_1, t_2] = \{t : \alpha < (\Gamma(t) \quad \text{and} \quad \Gamma_c(t)) < 1 - \alpha\}, \tag{23}$$

for some value α.

It is interesting to see how such intervals can be related to the power laws (Eq. (15)) which apply statistically when L is large with respect to the correlation parameter of the multi-Gaussian random field. The integral range of Y is $A_Y = \pi\lambda^2$, with $\lambda = 10$ in Fig. 12 and $\lambda = 20$ in Fig. 13. The range parameter of the random set at $t = 0$ is approximately $A_X = 0.69\ A_Y$ [62]. The characteristic length l is thus $l = A_X^{1/2} = 1.47\lambda$. According to the power laws in Eq. (15), the transition domain of proportions is [0.37, 0.65] for the simulation in Fig. 12 which is to be compared to the critical points $p = 0.44$ and $p = 0.61$ described above (Table 1). On the simulation represented in Fig. 13 one gets [0.44, 0.58]. Note that

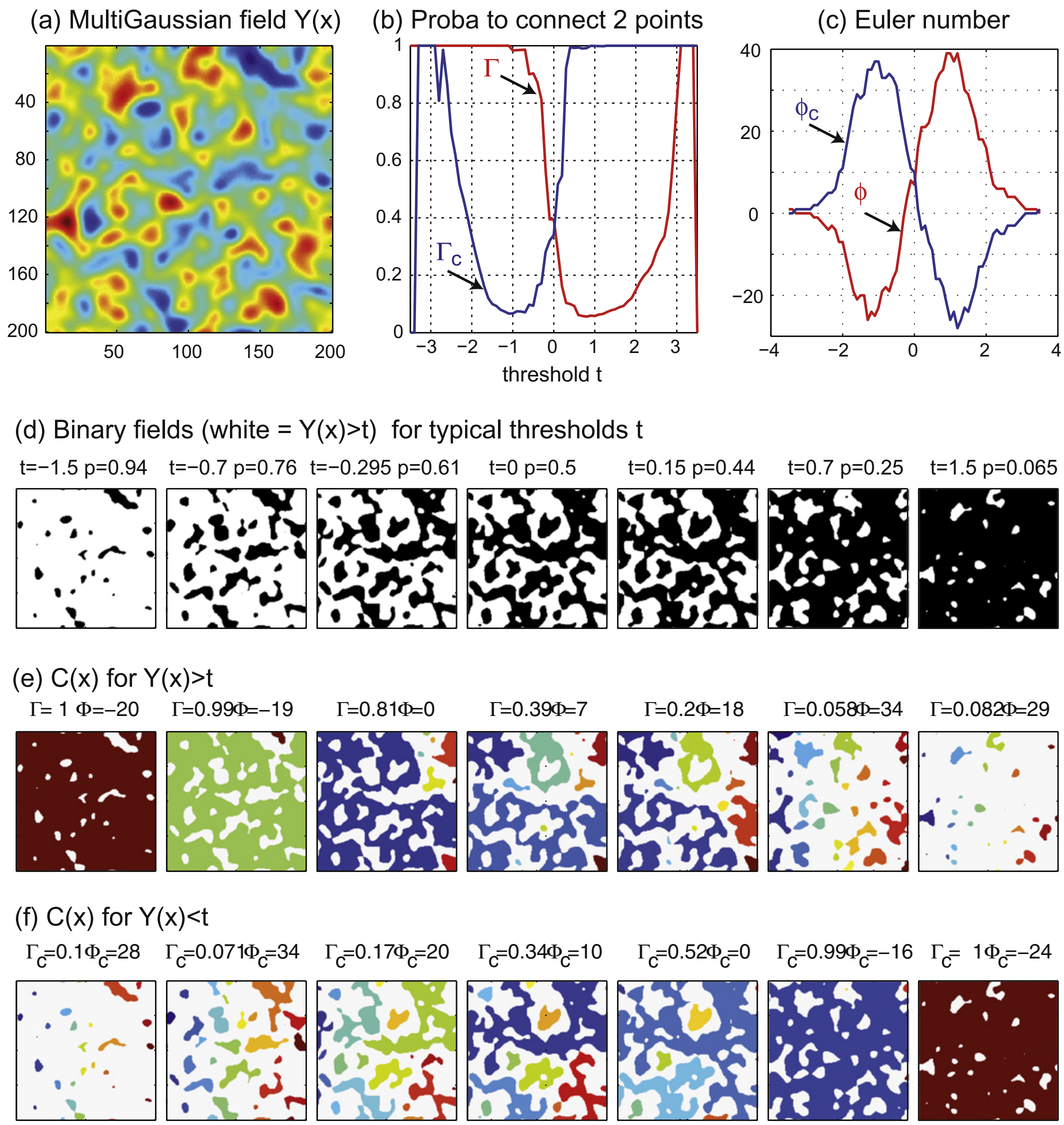

Fig. 12. Illustration of the evolution of the characteristic curves as a function of the threshold value for a multi-Gaussian medium.

the midpoint of the intervals is $p_c = 0.51$, the percolation threshold of the truncated Gaussian model with a Gaussian covariance function. These proportions correspond to the threshold intervals $[-0.33, 0.39]$ and $[-0.15, 0.20]$, respectively. Reading Γ and Γ_c values corresponding to these thresholds leads to Table 1, from which we can roughly set $\alpha \simeq 0.2$. There is thus a very good agreement between the two approaches for the multi-Gaussian random field.

To illustrate how the characteristic curves are influenced by the type of connectivity, Fig. 13 compares the characteristic curves of a multi-Gaussian field, a connected field having the same univariate distribution, and a multiple-point simulation with a different histogram. The most striking feature for the connected field is a shift of the thresholds t_1 and t_2 toward the high values. For t below $t_1 = 0$, there are a relatively small number of holes in a unique cluster $\phi < 0$. Above t_1 the unique cluster starts to be divided in a small number of components: $\Gamma(t) < 1$ and $\Gamma_c(t) > 0$. The crossing of the Euler characteristic $\phi(t)$ with the horizontal axis occurs largely above the mean at a value $t_0 \approx 1.8$. Above this point, a very high number of tiny isolated clusters of high conductivity appear within the high permeability channels. This is reflected by the strong rise of $\phi(t)$. For $t > t_2 = 2$, the complementary phase is completely connected ($\Gamma_c = 1$). The range of thresholds leading to connected fields is therefore between 0 and 2 for this medium instead of between -0.4 and 0.4 for the multi-Gaussian field. The second feature is that the transition is far less sharp than for the multi-Gaussian field, despite the fact that both fields have the same histogram and the same correlation length.

Finally, the characteristic curves for the multiple-point simulation (Fig. 13(g)) display an interesting behavior. The Euler characteristic (Fig. 13(i)) is less dissymmetric than for the connected field (Fig. 13(f)). The probability of connection shows a sudden fall for values larger than -3.5, (Fig. 13(h)) and then it raises slowly, while the probability of having the complementary component connected is constant $\Gamma_c = 0.4$. At a value of $t = -1.5$, $\Gamma(t)$ falls to a value of 0. There is a broad range of high values that are well connected, but there are different spatial components in these high values. They can be seen in Fig. 13(g): there are long and connected channels of high values which are coexisting with small ellipsoidal shapes of high values as well. These two patterns do not affect much the Euler characteristic curve, but they really influence the probability of connection of the two phases.

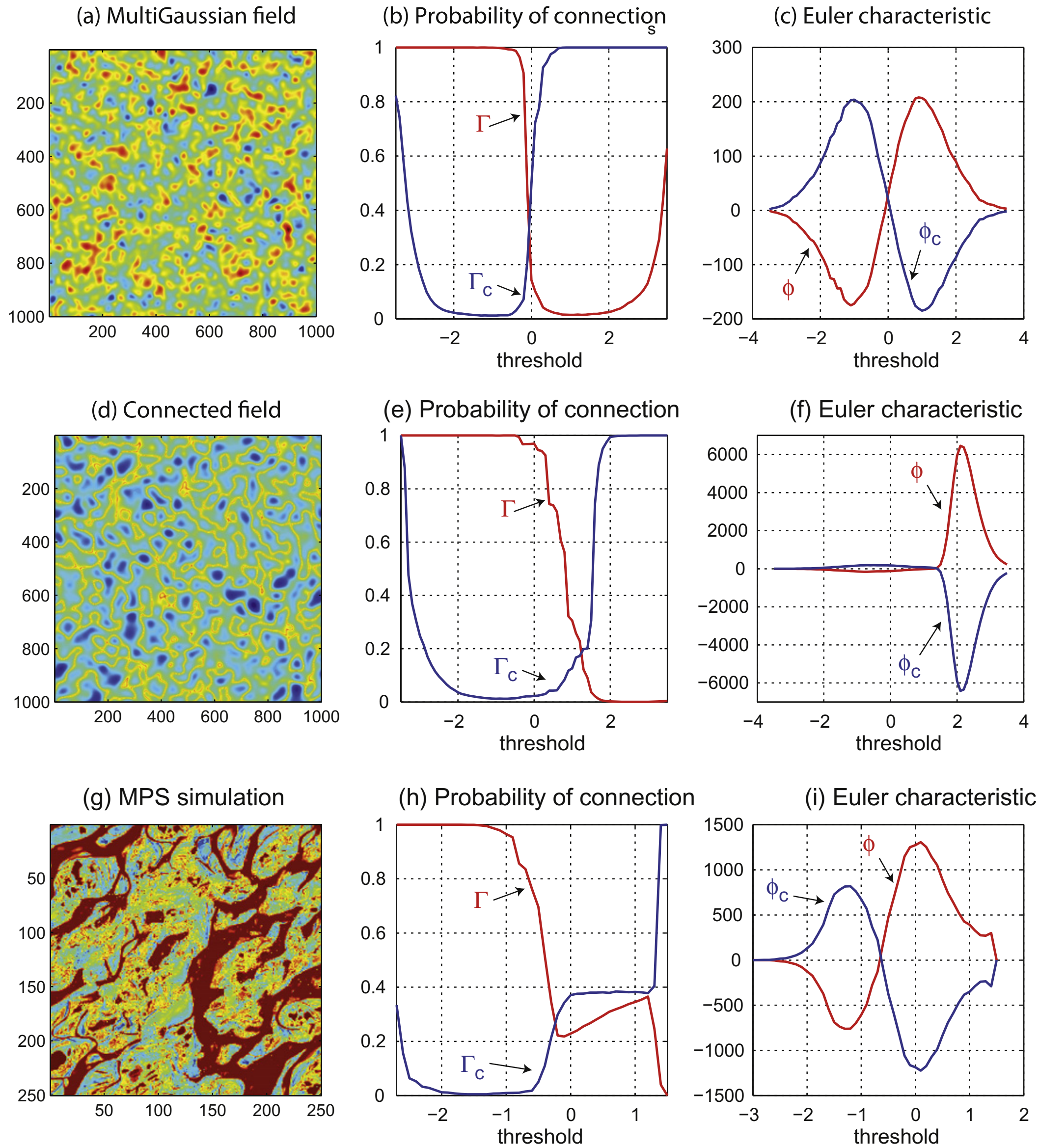

Fig. 13. Connectivity of continuous fields.

Table 1
Connectivity values corresponding to power law intervals.

	$L = 200$		$L = 1000$	
p	0.37	0.65	0.44	0.58
t	−0.33	0.39	−0.15	0.20
Γ	0.93	0.14	0.57	0.07
Γ_c	0.14	0.86	0.21	0.82

As a final remark, a simple thresholding technique is used to decompose the continuous field into a set of binary images. An alternative approach is to segment the image in connected regions having similar values of the continuous variables. A recent technique to do so was proposed by Soille [114]. It uses two parameters: the maximum difference between the values of two adjacent pixels who belong to the same connected component, and the range of the maximum differences acceptable between the values of pixels inside a given connected component. This definition allows to define a unique partition of a continuous field on which one could compute the connectivity metrics.

4. Dynamic connectivity metrics

We have seen in Section 2 that differences in connectivity patterns influence qualitatively the flow and transport properties of underground reservoirs. In particular, the total flux flowing through a porous medium and the distribution of the travel times are controlled by the connectivity of the high permeability areas (channeling) or by the presence of hydraulic barriers. To quantify the relations between the static connectivity and the response of the medium to flow and transport, one can define a series of dynamic metrics which are based on the response of the medium to an experiment related to a given physical process. These different metrics are described in this section.

Behind the theoretical interest in understanding the behavior of complex heterogeneous system, some dynamic metrics can be measured in the field during an experiment, and therefore if a theory shows that a relation exists between the dynamic metric and some static connectivity metrics then one can use this relation in an inverse manner to infer static connectivity information from field experiment. This constraints should then help making better and more accurate forecasts.

In the search for good dynamic connectivity metrics, the aim is therefore to define relevant and simple metrics that can ideally be inferred in the field and which relate to static connectivity properties. In the following, we have classified these different metrics first as a function of the physical setup (type of flow conditions) and in each subsection we refine the classification as a function of the physical process.

4.1. Uniform flow

Uniform flow occurs under natural conditions in the absence of punctual sources or springs: typically, in the middle of an aquifer far away from abstraction wells, punctual discharge or recharge zones. Those conditions can be described as the limiting case of the flow that would occur in an infinite statistically homogeneous medium under a uniform head gradient. In this situation, the flow velocity and head gradient are stationary around a mean value. Their fluctuations are only controlled by the local heterogeneity. It is the configuration that is most often used to investigate effective properties of heterogeneous materials. In practice, it can be approximated in numerical experiments by imposing a uniform head gradient on the boundary of a domain. Ideally, the boundary conditions should be such that they allow to be as close as possible to the infinite medium situation and can for example assume a periodicity of the medium. If the heterogeneities are sufficiently small as compared to the size of the domain, the boundary effects due to the selection of one specific type of boundary conditions becomes negligible even on a numerical experiment on a block of finite size. Under those very general conditions, one can study the effective flow and transport properties of a heterogeneous media and use the results of those experiments to quantify its effective behavior. By defining specific criteria which compare the effective behavior of a medium with what could be expected under a standard situation (for example for a multi-Gaussian one), it is possible to infer the degree of connectivity.

4.1.1. Effective hydraulic conductivity

For 2D flow in steady state, Knudby and Carrera [42] consider three candidate metrics: (1) the value of the power used in power averaging effective hydraulic conductivity, (2) the ratio of the effective conductivity K_{eff} [m/s] to the geometric mean K_G [m/s], and (3) the ratio of the critical path conductivity to the geometric mean. In their numerical investigation they show that the second and third metrics provide similar results for high variances, but the second is easier to evaluate from field observations. The first one does not scale properly with the variance of the log of the hydraulic conductivity and therefore they recommend to use the second metric. Following this work, we limit ourselves to the description of that metric, and we will denote it CF. It is defined as follows:

$$CF = \frac{K_{eff}}{K_G}. \tag{24}$$

K_{eff} is evaluated by solving numerically the uniform flow problem through the medium with permeameter type boundary conditions. Constant heads are prescribed on the upstream and downstream faces of the block, and no-flow boundary conditions are prescribed on the other faces:

$$K_{eff} = \frac{Q}{A} \times \frac{l}{h_1 - h_2}, \tag{25}$$

where Q [m^3/s] is the total discharge flowing through the medium, A [m^2] the area crossed by the flow, l [m] the length of the block in the direction parallel to the flow, and $h_1 - h_2$ [m] is the prescribed head difference.

Since the effective conductivity of a heterogeneous medium is always bounded by the harmonic K_H and arithmetic K_A means of the local values, CF is always positive and bounded by the following values:

$$\frac{K_H}{K_G} \leqslant CF \leqslant \frac{K_A}{K_G}. \tag{26}$$

The greater the value of CF, the most permeable is the medium and one can therefore expect that the high permeability values are highly connected.

However, one must be careful with the interpretation of the value of CF. In this connectivity metric, K_G was chosen to represent the theoretical effective conductivity of a multi-Gaussian medium. When the effective conductivity K_{eff} is greater than K_G, $CF > 1$ and the medium should therefore be more connected than a multi-Gaussian medium. On the opposite, if $CF < 1$ the medium should be less connected than a multi-Gaussian one. This interpretation is correct only when applied in two dimensions for isotropic media (as it was correctly done in Knudby and Carrera [42]) since the geometric mean is the effective property of such a media [115]. However, in 3D the effective hydraulic conductivity of a multi-Gaussian isotropic medium with a log normal distribution is well approximated by a power average with a power of one third [116]:

$$K_{eff} = \left[\frac{1}{V}\int K(x)^{1/3} dV\right]^3. \tag{27}$$

This value is larger than the geometric mean. Using Eq. (24) in this situation may be misleading since a value of CF greater than 1 would not mean that the medium is more connected than a 3D multi-Gaussian one. This issue occurs for example in the recent paper of Bianchi et al. [117] who showed a value of CF slightly greater than 1 for a 3D bloc that was obtained by Sequential Gaussian Simulation. This should not be misinterpreted, as it does not allow to conclude that this bloc is more connected than expected.

To make the definition of CF more general, and applicable in any dimension and for anisotropic media, K_G should be replaced with a value K_{MG} of the theoretical effective hydraulic conductivity accounting for these dimensionality or anisotropy effects in multi-Gaussian fields. Such expressions are available in the literature and have been reviewed in several papers [118–120]. We propose for example to replace K_G by the simple approximation K_{MG}^{ii} for the effective conductivity in the principal direction of anisotropy ii proposed by Ababou [121]:

$$K_{MG}^{ii} = K_G \exp\left(\sigma^2_{\ln K}\left[\frac{1}{2} - \frac{1}{d}\frac{\lambda_h}{\lambda_i}\right]\right), \tag{28}$$

where d is the space dimension (1, 2 or 3), λ_i is the correlation length of the multi-Gaussian field in the principal direction i, λ_h is the harmonic mean of all the correlation lengths, and $\sigma^2_{\ln K}$ is the variance of the logarithm of the hydraulic conductivity. The previous expression is a conjecture that is exact in most asymptotic situations. For example, in 3D it is equal to Eq. (27) for a multi-Gaussian medium. It has been tested positively against a wide range of numerical simulations. Extending CF in this manner implies also that it must be defined along specific directions, allowing to consider directional differences in flow connectivity which are likely to occur in natural media.

For binary mixtures, K_G is the theoretical effective conductivity only in the special case where the medium is two dimensional, isotropic and that the two phases can be exchanged without a change of the statistical properties of the medium. This is for example true in a random isotropic mixture having exactly 50% of each phase and in which there is not a statistical differences in patterns made by the two phases. In the general binary case, K_G should be therefore considered only as a normalizing factor and whether CF is greater or lower than 1 will highly depend on the proportions of the two phases but will not be directly interpretable in terms of relative connectivity structure as compared to a standard. Similarly for 3D or for a complex mixture of several discrete facies, the choice of the normalizing factor in Eq. (24) still requires some additional work.

4.1.2. Apparent hydraulic diffusivity

Instead of considering the total flux through the medium in steady state, one can consider the temporal evolution of the flux in transient regime and use it to estimate the apparent hydraulic diffusivity to quantify the flow connectivity [56]. The diffusivity is related to the speed of a pressure wave traveling through the medium. The underlying concept is that the fastest the pressure wave the most connected is the medium. The experimental setup is a rectangular domain, with initial conditions at rest (hydraulic head = 1 everywhere in the domain). The head is fixed on one side of the bloc ($h = 1$), the two perpendicular sides are no flow boundary conditions, and suddenly at a time $t = 0$ the head is lowered from 1 to 0 on the last face of the bloc. This creates a pressure drop that propagates through the medium as a function of time. One can then analyze the temporal evolution of the inflow $Q(t)$ through the boundary where the head remained constant ($h = 1$). The simulation is conducted until a steady state flux Q_{SS} is reached. An analytical solution exists for such a problem in a homogeneous medium:

$$\frac{Q(t)}{Q_{SS}} = \left(1 + 2\sum_{n=1}^{\infty}(-1)^n e^{-n^2\pi^2 Dt/L_x^2}\right), \tag{29}$$

where $D = K/S_S$ [m^2/s] is the hydraulic diffusivity, i.e. the ratio between the hydraulic conductivity K [m/s] and the specific storage S_s [1/m]. In heterogeneous transmissivity fields, $Q(t)$ has a different slope than in homogeneous ones because a part of the pressure wave arrives earlier and because the steady state is reached later. To get a quantitative estimation of the highly connected path, the idea of Knudby and Carrera [56] is to use the early increase of the discharge signal and to fit Eq. (29) on it to obtain an apparent diffusivity. More precisely, they identify the time t_5 at which 5% of the relative flux $\frac{Q(t_5)}{Q_{SS}} = 0.05$ has been reached and solve the following equation to estimate D_a:

$$\left(1 + 2\sum_{n=1}^{\infty}(-1)^n e^{-n^2\pi^2 D_a t_5/L_x^2}\right) = 0.05. \tag{30}$$

Finally, the value of D_a is normalized by $D_G = K_G/S_s$:

$$D_R = \frac{D_a}{D_G}. \tag{31}$$

Knudby and Carrera [56] note that D_a is always greater than $D_{eff} = K_{eff}/K_G$ for a heterogeneous media, and therefore D_R will always be greater than CF.

4.1.3. First arrivals of solute

For solute transport under uniform flow conditions, Knudby and Carrera [42] propose a global dynamic connectivity metric CT. It is defined as the ratio between the average arrival time t_a and the early arrival time $t_{5\%}$ of the first 5% of particles flowing through the medium:

$$CT = \frac{t_a}{t_{5\%}}. \tag{32}$$

The travel times are obtained from the results of an advective transport simulation using a particle method. The average travel time is given by dividing the travelled distance L_x by the mean velocity $K_{eff}/\Phi \times (h_1 - h_2)/L_x$ through the medium:

$$t_a = \Phi\frac{L_x^2}{K_{eff}(h_1 - h_2)}. \tag{33}$$

with Φ is the porosity, and $h_1 - h_2$ is the prescribed head difference between the two faces of the bloc. The effective hydraulic conductivity K_{eff} is the one defined in Eq. (25). A high value of CT corresponds to a strong tailing effect and fast first arrival time, characteristic of the connected media.

In their numerical study, Knudby and Carrera [42] show that CT allows identifying fields having a higher connectivity and therefore they recommend to use that metric. They note also that the transport connectivity CT and the flow connectivity CF metrics are weakly correlated. This is one of their major argument to conclude that connectivity is a process dependent concept requiring to define a series of different metrics. In a second series of numerical experiments, involving the comparison of CF, CT and D_R, Knudby and Carrera [56] show that the apparent hydraulic diffusivity D_R provides more information on transport than CF, but it is better correlated to CF than CT. The logarithm of D_R is well correlated to the logarithm of the product CT and CF. Generally speaking, D_R is a good indicator of flow channeling which strongly influences the first arrivals of solute.

4.1.4. Flow channeling

The connectivity structure of the hydraulic conductivity field has a direct impact on the structure of the flow field. When high conductivity areas are well connected, one can see the emergence of continuous channels of high velocities in a uniform flow. Such channels will highly control the presence of fast first arrivals as described in the previous paragraphs. To quantify the degree of channeling, some authors use indicators focusing on the consequences of channeling on the head and velocity statistics [122,123], or on its impact on transport indicators [124]. Here, following Le Goc et al. [125], we focus on the description of the channels themselves, and describe first the two indicators D_{ic} and D_{cc} introduced by these authors.

They consider a uniform flow in steady state. They divide the flow field into n flow tubes, each carrying the same proportion Q/n of the total flux through the medium. D_{ic} is then defined as follows:

$$D_{ic} = \frac{L}{nS_2(W_n)}, \tag{34}$$

where L is the size of the domain (square), and $S_2(W_n)$ is the participation ratio of the width Wn of the various flow channels. The participation ratio provides a statistical measure of the distribution of the channel width biased toward the higher values. It is defined as the ratio of the square of the first (non-centered) moment $M_1(W_n)$ by the product of the zeroth $M_0(W_n)$ and second $M_2(W_n)$ moments:

$$S_2(W_n) = \frac{M_1(W_n)^2}{M_0(W_n)M_2(W_n)}, \tag{35}$$

where the moments are computed on a grid having N cells of volume V_i with:

$$M_k(W_n) = \sum_{i=1}^{N} W_n(i)^k V_i. \tag{36}$$

The motivation for using the participation ratio is to obtain a characteristic length of the larger channels in the flow field without having to predefine a specific threshold in their statistical distribution. D_{ic} can be interpreted as a characteristic distance between the highly conductive channels which tend to be very narrow and to concentrate a large part of the flow. It can also be seen as a measure of the size of the obstacles in the direction perpendicular to the flow.

The second indicator D_{cc} focuses on the persistence of the channels in the flow direction:

$$D_{cc} = L'[1 - S_2(q')], \tag{37}$$

where q' is the Lagrangian derivative of the flow rate along the flow line, and L' is the average length of the flow lines. This definition is based on the identification of the regions of high changes in fluxes corresponding to the extreme points of a flow channel. D_{cc} is a characteristic length of the channels in the direction of the flow. These two metrics have been tested numerically [125] for a wide range of 2D continuous or fractured media and have shown that they can efficiently quantify the degree of channeling (Fig. 14). D_{ic} reflects the channel density, while D_{cc} reflects the extension of the channels.

4.2. Radial flow

Radial flow conditions occur around a well when it is pumped or when a fluid is injected. Under those conditions the flow field is convergent (divergent) and the interaction between the flow and the spatial heterogeneity of the aquifer is strongly affected by those specific conditions. Because pumping wells are used in a broad range of aquifer characterization techniques and engineering applications, the definition of connectivity metrics for those situations is very important.

4.2.1. Apparent storativity

Trinchero et al. [58] introduce a localized connectivity metric for radial flow conditions around a well. Their technique is based on previous studies by Meier et al. [126] who have shown on field data (Fig. 15(a)) and on numerical experiments that an observation well reacting sooner than another one located at the same distance of the pumping well is better connected to the pumping well [126]. This behavior has been investigated analytically by Sánchez-Vila et al. [127]. To quantify this effect, one has to interpret the drawdown data using the classical Jacob's method. It consists in fitting a straight line to the late time drawdown data in semi-logarithmic scale (Fig. 15(b)). The time t_0 [s] at which the straight line intersect the horizontal axis (drawdown = 0) allows to estimate the storativity coefficient S_{est}:

$$S_{est} = \frac{2.25 T t_0}{r^2}, \tag{38}$$

with T [m^2] the transmissivity of the aquifer estimated from the slope a [m] of the Jacob's straight line, $T = 0.183\ Q/a$ with Q [m^3/s] the pumping rate, and r [m] the distance between the pumping well and the observation well. The sooner an observation well reacts to pumping, the smaller is the time t_0 and the smaller is the estimated storativity S_{est}. By comparing the estimated storativity at different locations, one can therefore compare how these locations are connected to the well. For the multi-Gaussian field displayed in

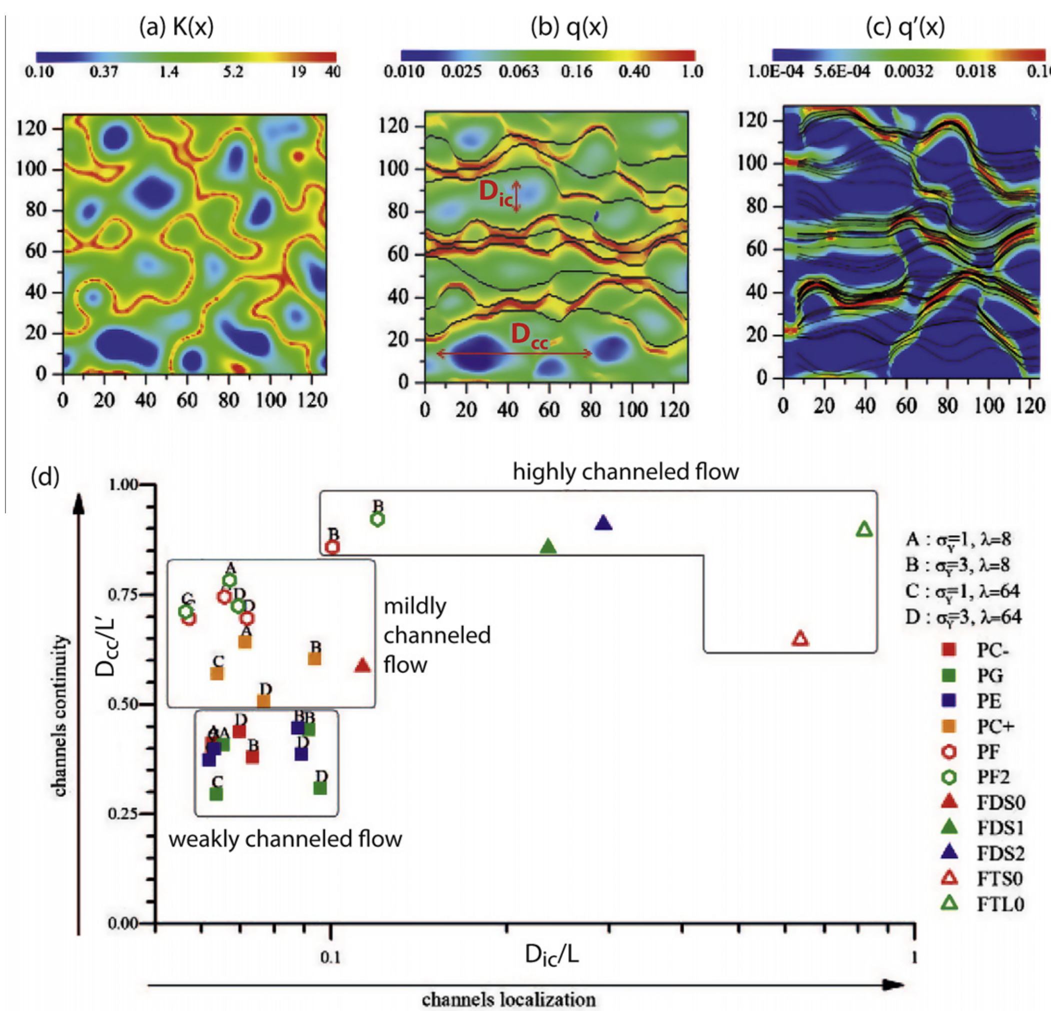

Fig. 14. Indicators of channeling: illustration of D_{ic} and D_{cc} for a connected field (a). The color represents the intensity of the Lagrangian Darcy velocities (b) and their derivative (c). The graph (d) shows how several flow experiments can be classified according to these two metrics. L represents the domain size and L' the mean length of the flow lines (Modified from Le Goc et al. [125]).

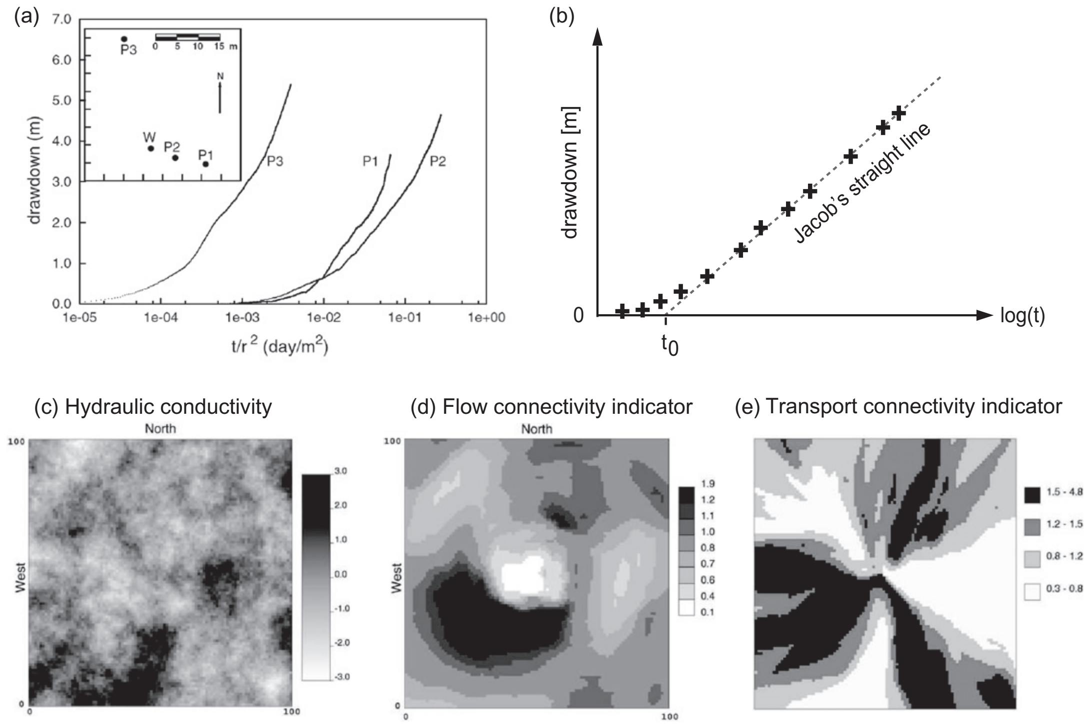

Fig. 15. Dynamic connectivity under radial flow conditions. (a) Results of a pumping test in a fractured media. The box represents the locations of the pumping well W and 3 observations wells. The drawdown curves for the three observation wells are plotted against time normalized by the squared distance between the pumping well and the observation well. (b) Interpretation of a drawdown curve using Jacob's straight line to estimate the flow connectivity. (c) A multi-Gaussian field used for the numerical investigation. (d) map of flow connectivity metrics. (e) map of transport connectivity metrics. Modified from Trinchero et al. [58].

Fig. 15(c), and for a pumping well located in the center of the figure, one can compute the flow connectivity metric (Eq. (38)) at each location. The resulting map (Fig. 15(d)) confirms that the regions having a low conductivity values (in white in Fig. 15(c)) and not well connected to the well (in the center of the field) have indeed a high value of S_{est}. The main advantage of this metric is that it is easily measurable in the field.

A theoretical relation between S_{est} and the hydraulic conductivity field has been derived analytically using a perturbation technique by Sánchez-Vila et al. [127]. It shows that the estimated storativity at a given location can be expressed as a weighted moving average of the values of the transmissivities. The weights are maximum along the line going between the pumping well and the observation well but the final average is also largely influenced by the values around the two wells. This explains rather well the fact that the zones of equal values of S_{est} in Fig. 15d are mainly concentric with a strong lateral continuity.

4.2.2. Apparent porosity

For solute transport, under radial flow conditions, Trinchero et al. [58] propose a localized connectivity metric. They define it in the framework of a converging tracer test towards a well. They consider the average travel time t_a [s] between the injection well and the pumping well. Then, they normalize it by the advective travel time in radial flow conditions to get a porosity estimate Φ_{est} [−]:

$$\Phi_{est} = \frac{Q_W t_a}{\pi(r_i^2 - r_w^2)}, \tag{39}$$

where r_i [m] is the radial distance between the pumping well and the injection well, r_w [m] is the radius of the well, and Q_w [m^2/s] the pumping rate per unit of thickness. In this setup, the average travel time t_a is estimated as the first temporal moment of the breakthrough curve, $C_b(t)$:

$$t_a = \frac{\int_0^\infty t C_b(t)dt}{\int_0^\infty C_b(t)dt}. \tag{40}$$

When there is a high permeability path connecting the injection well and the pumping well, Φ_{est} is small (and lower than the true porosity), while on the opposite situation Φ_{est} is high (and larger than the true porosity). Using a series of transport simulations, they can map the value of this metric and identify the regions which are highly connected to the pumping well (Fig. 15(e)). The comparison of this map with the flow connectivity metric (Fig. 15(d)) reveals very clearly the different processes. The white areas are the zones that are well connected to the pumping well while the black areas are poorly connected. The geometry of these regions is essentially radial and follows areas delimited by streamlines converging toward the well. Using a perturbation approach, Trinchero et al. [58] obtain an analytical approximation of Φ_{est} that clearly shows that it is related to two terms. The first depends on the transmissivity values located along the streamline between the injection well and the pumping well. The second accounts for the overall hydraulic response of the system and involves the radial flow connectivity metric S_{est}.

4.3. Producer and injector

Another typical flow configuration that is encountered in practical applications is the doublet. At least two wells are considered, one is an injector and the other one is a producer. This is the basic pattern, but often more wells are involved and a complex system can be setup with groups of injectors and producers. From a theoretical point of view, these systems can most of the time be decomposed into smaller units including only one doublet and this is why we will just consider the case of two wells in the following. Doublets have many applications. They are used in the oil industry for enhanced oil recovery, in geothermics to extract heat in deep systems, in mining to leach the uranium in place and extract it, for decontamination problems or more simply for aquifer storage and recovery of fresh water. Often, the connectivity between the injector and the producer is a key issue for practical purposes. In the oil industry, if an injector is too well connected to the producer (for example because of the presence of a karst conduit) the oil recovery in the producer will be very inefficient since all the water injected will directly flow to the producer without pushing the oil in place. Similarly, if a geothermal well is badly connected to the reservoir in place and to the pumping well, it will be very difficult to exchange enough heat and the system will not be efficient. All those considerations explain that there is a significant amount of literature related to the connectivity between producer and injector wells and that there is not a single metric that has been widely adopted since there are different problems covered by the same terminology.

4.3.1. Recovery efficiency at 0.5 PVI

A simple indicator used in the oil industry is the *recovery efficiency at 0.5 pore volume injected (PVI)*. It is the amount of oil recovered normalized by the total amount of oil in the reservoir, after having injected a water volume of half of the pore volume of the reservoir. This quantity can be estimated by running a two-phase flow model and by computing the oil recovery curve at the producer. On that curve, one can directly read the recovery efficiency. Hovadik and Larue [57] shows that for channelized reservoir, this indicator correlates very well with a static connectivity metrics accounting for the volume of reservoir connected to the well.

5. Relations between static and dynamic connectivity metrics

In this section, we provide an overview of some relations that have been proposed to relate the static and dynamic connectivity metrics.

5.1. Effective conductivity of binary media

We have seen that the effective conductivity of a medium is directly related to the dynamic connectivity metrics *CF*. The effective properties of binary media have been studied for many years and several approaches have been developed to estimate them from a statistical description of the medium. These methods are reviewed in Renard and Marsily [119] or Sánchez-Vila et al. [120]. But most existing formulas do not account for the static connectivity metrics.

Noticeable exceptions were obtained in the framework of percolation theory. The proposed power laws for the effective conductivity account for the percolation threshold p_c. In that line of research, one of the most interesting result is the formula proposed by Bernabe et al. [128] for a binary mixture. Below p_c the inclusions are predominantly disconnected and therefore they use the lower bounds of Hashin and Shtrikman [129] to estimate the overall hydraulic conductivity of the medium:

$$p < p_c, \quad K_{eff} = K_0 + \frac{p}{\frac{1}{K_1 - K_0} + \frac{1-p}{3K_0}}, \tag{41}$$

where K_0 [m/s] is the smaller conductivity and K_1 [m/s] the larger one. Above p_c the inclusions are predominantly connected and therefore they use the upper Hashin and Shtrikman bound [129] but they account for the fact that the matrix still contains a proportion p_M of permeable inclusions. This is estimated in two steps. First they use a power law derived from percolation theory which states that around the percolation threshold, the volume fraction of high permeability material belonging to the percolation cluster, p^*, follows a power law:

$$p^* = p\left(\frac{p - p_c}{1 - p_c}\right)^b. \tag{42}$$

b is equal to 0.14 in 2D and 0.41 in 3D. They assume that this expression can be applied all the way between p_c and 1. The proportion of mixed material above p_c can therefore be expressed as follows:

$$p_M = 1 - \frac{1-p}{1-p^*}, \tag{43}$$

and the equivalent conductivity K_M [m/s] of the mixture is expressed using again the lower bound of Hashin and Shtrikman:

$$K_M = K_0 + \frac{p_M}{\frac{1}{K_1 - K_0} + \frac{1-p_M}{3K_1}}. \tag{44}$$

Finally, the effective conductivity of the medium above p_c is taken as the upper bound of Hashin and Shtrikman assuming a binary medium made with a proportion p^* of the high conductivity K_1 cells connected to a large cluster and a proportion $1 - p^*$ of inclusions of mixed conductivity K_M:

$$p \geqslant p_c: \quad K_{eff} = K_1 + \frac{1-p^*}{\frac{1}{K_M - K_1} + \frac{p^*}{3K_1}}. \tag{45}$$

Their formula compares well against numerical simulations and experiments (Fig. 16(a)), but one can observe that there is a range of values of p around p_c such that the effective conductivities obtained by numerical experiments is very large. The proposed formula is only able to estimate the mean value in this situation. One can therefore expect that including an additional static metric of the connectivity of a specific realization within the previous formula could provide a better forecast for these situations.

For 2D media, Knudby and Carrera [42] have compared the effective conductivity of a series of 2D hydraulic conductivity fields having various types of connectivity structure. On those fields, they also computed the connectivity range ξ of the connectivity function $\tau(h)$ of the high permeability areas. They first show that ξ does not correlate at all with the *CT* metric which is proportional to the effective conductivity. This result is due to the fact that the effective conductivity of a heterogeneous media is not simply a function of the size of the high conductivity area but is also highly controlled by the presence of barriers of low permeability. This is well illustrated by the formula that was proposed by Knudby et al. [110] to approximate the effective conductivity of binary media:

$$\frac{1}{K_{eff}} = \left[\frac{1}{K_a} - \frac{1}{K_H}\right]\frac{\left(1 - D'_{norm}\right)^2 - p^2}{1 - p^2} + \frac{1}{K_H}, \tag{46}$$

where p is the proportion of high permeable cells, and D'_{norm} $[-]$ is the ratio of the mean distance between the inclusions to the mean distance between their centers, with both distances taken in the direction of the average flow. In that formula, D'_{norm} is a geometrical parameter related to the size of the inclusions and their spatial arrangement. The results of the numerical experiments of Knudby et al. [110] show that Eq. (46) performs much better than most of the formulas that were proposed earlier for these types of fields (Fig. 16(b)). McKenna et al. [130] has extended this formula in a statistical framework in the case of truncated multi-Gaussian fields.

5.2. Effective conductivity of continuous media

To our knowledge there are only a few attempts to relate the effective conductivity of continuous fields with their static connectivity metrics.

Samouelian et al. [113] proceeds as follows. First, they normalize the field linearly between the minimum and maximum values:

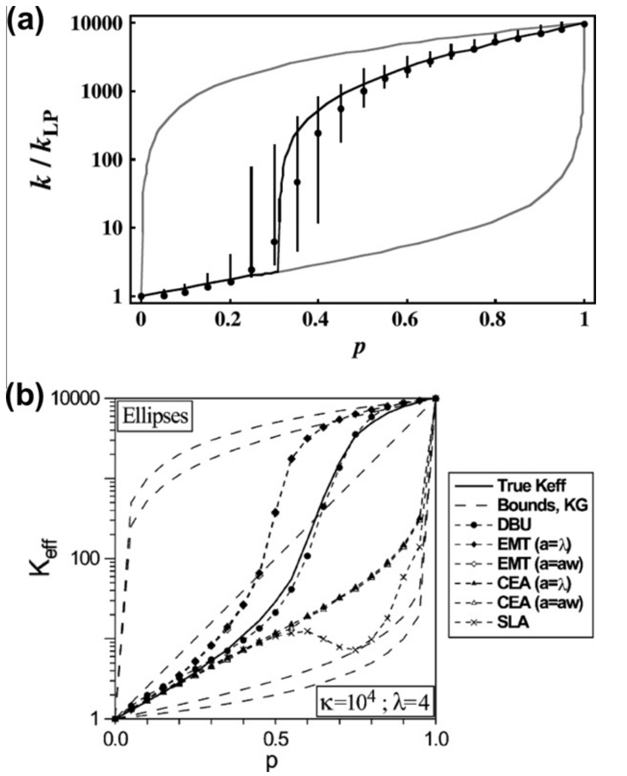

Fig. 16. Comparison of numerical experiments with analytical expressions for binary media. (a) $10 \times 10 \times 10$ networks, numerical experiments against Eqs. (41)–(45). k_{LP} is the permeability of the low permeability phase. [128] (b) 2D ellipses. Comparison of several methods against numerical results (black line). Eq. (46) is denoted DBU on the legend [56].

$$k(x) = \frac{K(x) - K_{\min}}{K_{\max} - K_{\min}}. \tag{47}$$

Then they compute the Euler characteristic curve $\phi(t)$ of the normalized field $k(x)$ to identify the value t_0 such that $\phi(t_0) = 0$. Following the suggestion of Mecke and Wagner [131], they consider t_0 as a percolation threshold. In a second step, they extract from the univariate distribution of k the critical mass fraction m_c which is the probability of having a value larger than t_0:

$$m_c = P(k \geqslant t_0). \tag{48}$$

If this value is small, it indicates according to Samouelian et al. [113] that the continuous path must be straight and hence more efficient, a decreasing value of m_c is interpreted as an increase in connectivity. Therefore, the effective conductivity should increase with the inverse of m_c. They combine these two topological indicators in a factor q:

$$q = a\left[\frac{t_0}{m_c}\right]^b - 1, \tag{49}$$

to be used in the classical power averaging estimation of the effective conductivity:

$$K_{eff} = \left[\frac{1}{V}\int K(x)^q dV\right]^{\frac{1}{q}}. \tag{50}$$

The factors a and b are adjusted by non-linear regression so that the formula reproduces as accurately as possible a training data set. They obtain consistently some values for a in the order of 0.85 and for b on the order of 0.40.

Instead of considering only the critical value t_0, Neuweiler et al. [94] divide the univariate distribution of the hydraulic conductivities in classes. One of the classes corresponds to the connected matrix and the other classes correspond to inclusions. The selection of the thresholds t_1 and t_2 defining the connected interval is assumed to be known. The hydraulic conductivity of each class is taken as the arithmetic average of the values within the class. To illustrate the methodology, they use the same types of connected, multi-Gaussian, and disconnected fields as Zinn and Harvey [92], and they fix different thresholds for the three configurations. By approximating the medium as a three or two components medium, they use Maxwell's approach [55,94] and get an estimate of the conductivity which is very simple. In 2D, for a three media decomposition their formula is the following:

$$K_{eff} = K_1 \frac{1 - A}{1 + A}, \tag{51}$$

$$A = p_2 \frac{K_2 - K_1}{K_2 + K_1} + p_3 \frac{K_3 - K_1}{K_3 + K_1}, \tag{52}$$

where K_1 is the hydraulic connectivity of the background connected material, K_2 and K_3 are the hydraulic connectivities of the high or low conductivity lenses isolated in the background material, and p_i is the volume percentage of materials i. By repeating similar computations for the wetting and non-wetting phase and for different capillary pressures they obtain the effective mobility function for their multiphase problem. This approximation is compared to numerical results for buoyant counter flow problem of DNAPL and water and shows to perform reasonably well for the different types of media that are considered.

5.3. Transport processes

Willmann et al. [132] have investigated the relations between the transport behavior of a medium and its connectivity metrics. Unfortunately, they only considered dynamic connectivity metrics such as *CF* and *CT* introduced above. In their study, they consider five types of media with different structures and types of connectivity: multi-Gaussian with a small integral scale, stationary with two nested variograms, non-stationary power variogram, non-stationary with conditioning data to create a channel, and a connected field built by transformation of a multi-Gaussian one. For all those fields, they simulated transport using an advection dispersion equation at the local scale. They analyzed the breakthrough curves and in particular the non-Fickian tailing that typically occur in connected fields. To model this type of behavior they used memory functions and characterized them with the slope of the late time breakthrough curve as a function of the logarithm of the time. They show the existence of a relation between this slope and the connectivity metrics *CT* and *CF*. The slope decreases with both metrics and stabilizes toward a value of 2.

6. Generating random media with a given connectivity structure

In this section, we discuss the practice of stochastic simulations of subsurface structures when connectivity information is available and must be respected. We shall make the distinction between soft and hard connectivity information. Soft information corresponds to global knowledge about the connectivity structures, such as being above the percolation threshold or imposing the possibility of long distance connectivities. This amounts to honor statistically the connectivity metrics described above. The theoretical considerations seen in Section 3 provide useful guidelines for performing simulations respecting global connectivity information and we will discuss below how this can be done in practice. Hard information corresponds to localized precise information about connections. For example a tracer test was able to prove that two wells are connected. Such binary information, either true or false, is more difficult to respect and calls for specific techniques detailed below.

6.1. Honoring soft, global connectivity metrics

The first and easiest way to honor global connectivity metrics is to tune the parameters of the model so as to be in adequation with the target connectivity behavior. Earlier, we have seen that power laws describe how the expectation of the percolation threshold on a finite grid and its variance depend essentially on two driving factors controlling percolation: the difference $p_c - p$ between the percolation threshold and the proportion, and the ratio L/l between the size of the grid and the size of the objects (or the range of the covariance). In Section 3.5, we have shown that there is only three connectivity behaviors, corresponding to three domains of the proportion interval: below, around or above the percolation transition. We also saw that discretization can play a role, but of second order only. Larue and Hovadik [60] and Hovadik and Larue [57] review a series of other factors shifting the "around percolation domain" either to the lower proportions or to the higher proportions:

- Anisotropy has a major impact. Generally speaking, the higher the anisotropy, the higher the percolation threshold. Imposing almost parallel channels in object-based models or sheeted reservoirs with very high vertical non-stationarity of the intensity of the point process driving the object model lead to a dramatic shift of the *S*-curve to the right [60]. In essence, 3D reservoirs are made "quasi 2D" reservoirs, thus shifting the percolation threshold from the low 3D value (around 0.32) towards the much higher 2D value (around 0.59).
- Imposing attraction or repulsion between objects as in [133] will also have a major impact on connectivity. Object models with repulsion, called compensational stacking of channels in [60], will discourage intersection between objects, thus yielding to relatively high proportions with poor connectivity. Attraction between objects would lead to the inverse effect.
- Mixed objects with non-permeable phase, or multi-type object models with non-permeable objects which could possibly erode permeable ones will have more complex effects on connectivity. On the one hand, presence of non-permeable facies will destroy connections. But on the other hand more objects are needed for a given proportion, thus creating new connections. The final effect will depend of the relative proportions of the facies, size and shape of the objects etc.

Another approach can be to use multiple-point statistics [40,49,50,134] with a training image that displays the connectivity patterns that are required in order to generate stochastic simulations having the same patterns. However, the connectivity metrics of the resulting simulations are strongly affected by the parameters of the simulation algorithm (such as the size of the neighborhood) [54] and the method does not ensure a priori that the connectivity metrics of interest will be reproduced. A systematic study of the effect of the parametrization of the method on the reproduction of the connectivity metrics has still to be carried out.

When tuning the parameters is not sufficient to honor the target connectivity metrics, a post-processing of the prior simulations can be undertaken. Simulated annealing is a flexible stochastic global optimization method that is suited to incorporate information from different sources, both statistical or measured. We shall not describe this optimization algorithm here; readers are invited to refer to the original paper by Kirkpatrick et al. [135] or to classical textbooks such as Robert and Casella [136]. Schlüter and Vogel [34] used a simulated annealing algorithm to constrain 2D simulations to honor the Euler characteristic and chord length distributions as functions of the threshold which transforms a continuous simulation into a set of binary ones. The constrained values are those computed on training images which are either multi-Gaussian or a connected field according to Zinn and Harvey [92]. It is shown that matching only the Euler characteristic leaves too much freedom for the size of the cluster. The long connected bands are not reproduced. This result confirms the theoretical distinction made between $\phi(X)$ and $N(X)$ as connectivity metrics: the Euler characteristic is in essence a local property of X, not a global one. Reproducing the chord length in four directions (horizontal, vertical and diagonals) in both the permeable and impermeable phases was able to reproduce the elongated morphology of the connected clusters, but their morphology appeared to be too rectilinear and too fragmented. Considering either less directions or one phase only lead to poor reconstruction. Respecting both statistics simultaneously improved very much the reconstruction. This approach proved to be very efficient in honoring other connectivity metrics, such as percolation threshold and connectivity function. Furthermore, breakthrough curves of a conservative solute in the vadose zone under steady flow conditions were adequately reproduced. Jiao et al. [137] include in the simulated annealing algorithm a constraint to reproduce in addition the connectivity function $C_2(h) = \tau(h)K(h)$. They show that the connectivity function is a key characteristic that must be accounted for to reconstruct textures that are found not only in granular media, but also in cosmology and material sciences.

6.2. Honoring hard connectivity metrics

Simulated annealing is not a suitable tool for performing simulations reproducing binary 0/1 hard connectivity information because no continuous objective function can be built from this information. Two techniques have been proposed to simulate binary fields honoring connectivity constraints such as those imposed by a tracer test. Allard [138] proposed a Markov chain Monte Carlo (MCMC) Gibbs sampler [139] approach for the truncated Gaussian model. The initial field is generated without accounting for spatial correlation, but ensures that the connectivity constraints are respected. The Gibbs sampler is then used to obtain, after a certain number of iterations, a field that respects both the covariance function and the connectivity (or disconnectivity) information. At each site, a new (truncated) Gaussian random variable is drawn, conditionally on the fact that no condition is violated. On the examples presented, convergence is reached in about 400 full passage of the image. This algorithm can handle connectivity constraints of the type $x \leftrightarrow y$ as well as non-connectivity constraints of the type $x \not\leftrightarrow y$. The respect of first and second moment of the Gaussian field depends on how likely the set of connectivity constraints is. Let us denote S this set, and suppose for the moment that S is the unique condition $x \leftrightarrow x + h$. The probability of S is simply the unconditional connectivity function $P(S) = \tau(h)K(h)$. This probability will be very low if $p \ll p_c$ and increases dramatically around p_c. If S is made of several connection and non-connection constraints, such as in the example illustrated in Fig. 17, $P(S)$ reaches a maximum for a certain proportion, which will be around the percolation threshold where the variability of the connectivity metrics is maximum.

Fig. 18 shows the solutions obtained for the conditions of Fig. 17 for a truncated Gaussian model with a factorized exponential covariance function and range parameter = 3 pixels. When the proportion is close to or larger than the percolation threshold (middle and right panel), one can observe that the parameters of the non-conditional model are well reproduced. On the contrary, when the proportion is well below the percolation threshold (left panel), the histogram of the Gaussian values and the sill of the experimental variogram are shifted to higher values.

A very similar approach is implemented in the industrial version of the Boolean model presented in [133]. An initial configurations of objects is first built according to some ad hoc procedure such that all connectivity constraints are verified. Then, objects are added or removed according to an MCMC birth-and-death process. Additions or deletions of objects are only allowed if no connectivity constraints are violated. In this implementation, agreement between the model parameters (proportion, size of objects, anisotropy, attraction or repulsion) and the connectivity constraints is of paramount importance for achieving convergence and respecting the input parameters.

These approaches, similar in spirit, present two difficulties. The first one is that one must devise ad hoc methods for building an initial configuration verifying the connectivity constraints. The second is that, as it is the case for all MCMC algorithms, finding a criterion for deciding if the algorithm has reached convergence is difficult. Note that for both implementations, lack of convergence should not be considered as a failure of the algorithm, but should rather be interpreted as the inadequacy between the model parameters and the connectivity constraints.

In a very recent paper, Renard et al. [140] proposed an algorithm related to direct sampling [49]. The general idea is to borrow connected paths from a training image instead of iteratively building a simulation that satisfies both the connectivity and the structural constraints. The consistency is imposed by using as training image (to search for connected paths) either an unconditional simulation constructed with the method that will subsequently be used to generate the realizations, or the one used as input in a multiple-point algorithm such as *impala* [50], *snesim* [40] or the direct sampling method [49]. The general algorithm is now described (see Fig. 19) for the single two-point connectivity constraint $x \leftrightarrow x + h$. More details and generalization to multiple-point constraints can be found in [140]. First, the cluster function $C(x)$ is computed on the training grid $\mathcal{G}_t$. Then it is scanned in order to find all grid cells x_i such that $C(x_i) = C(x_i + h) \neq 0$. Let n be the number of those points. If $n = 0$ the algorithm is stopped; obviously the training image $\mathcal{G}_t$ is not compatible with the constraint. A new image must be sought.

At this point, if $n > 0$ all the preprocessing is done and we enter a loop that is applied for each simulation. The simulation grid is denoted $\mathcal{G}_s$.

1. One value i is chosen at random between 1 and n. It corresponds to the random, uniform selection of one replicate of a connected pattern.
2. The whole cluster $C(x_i)$ is identified, but not copied as such because: (i) it is not necessary to copy the whole cluster from $\mathcal{G}_t$ to $\mathcal{G}_s$ to ensure the connection, and (ii) it is possible that the selected cluster is larger than $\mathcal{G}_s$.
3. A path is selected within $C(x_i)$; in order to minimize the number of pixels which will play the role of conditioning data in the next step, a path is selected by applying propagation algorithms within $C(x_i)$.
4. The path is copied between x and $x + h$, thus creating the connection between these two points.
5. All these cells are taken as conditioning data for the simulation algorithm which is applied as usual, whatever the technique and the model.

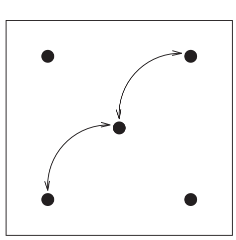

Fig. 17. 2D connectivity constraints: the three points on the first diagonal must be connected; the two other points must be disconnected.

Fig. 19 illustrates the algorithm. Adaptations of this base-case algorithm are able to account for hard conditioning data and for multiple-point connectivity.

This algorithm is very efficient, much more than MCMC ones. It is very general because it does not depend on the actual simulation technique. It can for example be applied without any modification to sequential indicator simulation or to truncated Gaussian or plurigaussian methods [100]. It is however not adapted to object based models, but extensions of the same idea, applied on the graph of intersecting objects instead of the graph imposed by the grid could probably be developed.

6.3. Honoring connectivity metrics and inverse problems

From a very broad perspective, the dynamic connectivity metrics are related to flow and transport state variables that can be either measured in the field or derived from numerical simulations. When a numerical model of an aquifer system is made, there is almost systematically a phase in which an inverse problem is posed so that the model reproduces the field measurements of state variables. Because the dynamic metrics are not the input parameters of the model, honoring them is nothing else than a special type of inverse problem.

In certain circumstances, when the connected features are well defined and when the data are sufficient, classical inverse techniques are able to identify correctly the connected channels [141] without any modification of the inverse method. There are however other situations in which the presence of the channels is not identified properly because the field observations are not sufficient or because the underlying geostatistical model does not assume the possibility of the existence of connected features [95].

Solving the inverse problem to reproduce dynamic connectivity metrics or simply to insure that a groundwater model reproduces the field observations usually involves a complex iterative procedure that aims at minimizing the discrepancy between the field observations and the model results [142,143]. This is similar to the optimization procedure that can be used to constrain the stochastic simulations of the media by static constraints as discussed in Section 6.2. The difficulty here is that the forward computation of the flow and transport responses allowing to infer the dynamic connectivity metrics and to define the functional that has to be minimized requires solving partial differential equations in transient state. This implies significant computational resources and computing times. The problem is therefore normally much harder than the one discussed for static connectivity.

Two alternative approaches have been followed over the last years. One approach is to use the information available on static connectivity to condition the stochastic model to this information to accelerate and facilitate the resolution of the inverse problem. The other idea is to develop methods able to directly generate models that are coherent with the dynamic connectivity information.

In the following, we describe these approaches in more detail.

6.3.1. Use of connectivity for model ranking and selection

As a preliminary step, to investigate the degree of variability of the possible responses of an aquifer or a reservoir, one of the most general approach is the use of Monte Carlo simulations. This is conceptually very simple. Multiple realizations of the geological heterogeneity and of the parameters are produced, and for each of them

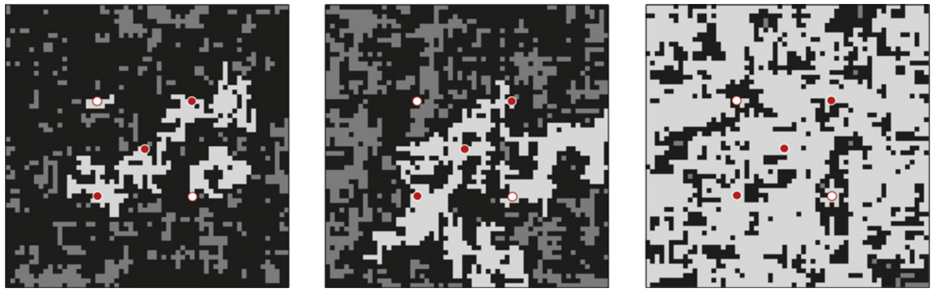

Fig. 18. Realizations obtained with the Gibbs sampler approach. The five conditional points are represented by the circles; the circles filled in red must be connected and the circles filled in white must be disconnected from the red circles; the cluster connecting the three points is in gray; the background is in black. From left to right the input proportions are $p = 0.2$, $p = 0.4$ and $p = 0.7$. Exponential covariance function with a practical range of 9 pixels [138]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

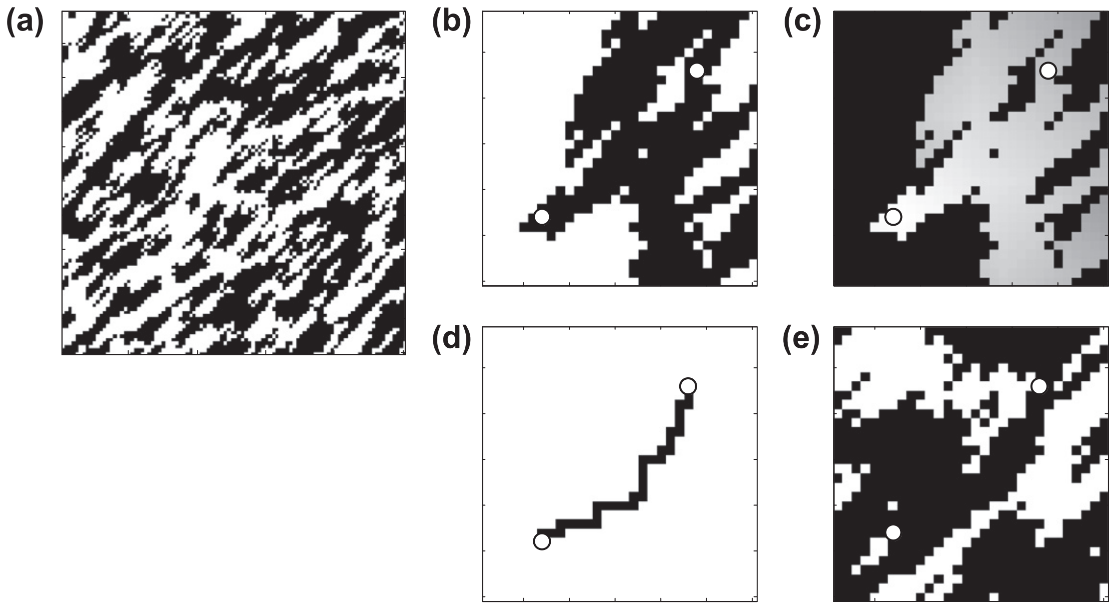

Fig. 19. Step by step description of the base algorithm. (a) is a training image (100×100 cells), it has been generated by sequential indicator simulation with a spherical variogram, an anisotropy oriented at 45 degrees, with variogram ranges equal to 10 and 3. (b) is an image of the simulation grid (30×30 cells) after one replicate of a connected body has been translated and pasted in the simulation. The two white disks represents the points that must be connected. (c) represents the distance function which is calculated inside the geobody to draw the random paths. (d) is one of these shortest paths. (e) is the final results of the simulation.

the physical response is computed. By analyzing the variability of the responses one gets a measure of the uncertainty. The method is general, no assumption is made on the type of heterogeneity nor on the type of physical models. But the method is slow. Many model runs are required.

To speed-up the procedure, one can rank the parameters or geological models and select a few representative models before running the costly simulations. By doing so, one can avoid running several models that are expected to have similar responses. In that procedure, static or dynamic connectivity metrics can be used if their calculation is fast [51,144–146].

de Jager et al. [147] propose a modeling flow chart that includes a step in which the connectivity of the geological models of a channelized reservoir is assessed through the use of experimental design and surface response mapping. They find a weak relation between the input parameters of the stochastic model of channels and the connectivity metrics, while they obtain a good correlation between the connectivity and the flow behavior. In a second step, they show how the knowledge of the strong correlation between flow and connectivity can be used to select models.

One recent progress in the development of model selection techniques is to use distances between the approximated model responses, e.g. a distance in terms of connectivity metrics between two models. Streamline computations are used for example by Park et al. [124] or Scheidt and Caers [148] to obtain rapidly an approximation of the recovery curve at the producer. Then, instead of using a single value to describe the connectivity of a given permeability field, they take the complete recovery curve $f_i(t)$ and they consider that two media i and j having a similar curve have a similar connectivity. In that manner, they do not consider a single value of connectivity but rather they compute the distance $d(i,j)$ between the two responses of the two media:

$$d(i,j) = \int_0^{t_{end}} \left(f_i(t) - f_j(t)\right)^2 dt. \quad (53)$$

Because the streamline computations are very fast, this technique allows to compare and group very rapidly a large number of different permeability fields. The distances between the models allow to map the models in an abstract space using the multidimensional scaling technique. One can then use the proximity or distance between the models in this space to accelerate model selection, uncertainty analysis or inverse problem solving [148]. Impressive results are obtained by this approach. Using only a very small number of models (less than 10) selected using distance mapping, Scheidt and Caers [149] are able to reproduce very accurately (and more accurately than traditional ranking approaches) the uncertainty ranges (quantified by the 10, 50 and 90 percentiles) of the cumulative oil production for a field in West Africa which would require otherwise a much higher number of model runs.

6.3.2. Accelerating inverse problem solving

There is a very wide range of techniques that have been developed to solve the inverse problem [142,143]. But from a conceptual point of view, all the methods aim at searching one or an ensemble of models or model parameters such that the physical response is close to the observations. We can then see these techniques as tools to a search in a very high dimensional space of model geometries and parameters. In this space, all kind of models with different degrees of complexities, physics and connectivity exist.

If some information is available about the connectivity then Alcolea and Renard [150] showed that it should be used to enhance convergence rates. The setup is a regional flow in a channelized aquifer. A well is pumped and the response is measured in several piezometers. A Monte Carlo Markov Chain sampling technique is used to modify iteratively a multiple-point simulation of channels with the constrain to reproduce the head response in the piezometers. Two configurations are compared. In the first, the algorithm is applied without considering the connectivity information between the wells. In the second configuration, the technique described in [140] is used to generate only permeability fields which are conditioned by the connectivity. The results of the two approaches are similar in terms of ensemble of simulations that have been retained to fit the data and represent the uncertainty. However, the procedure which included the connectivity information was faster because it was searching in the right subspace of possible models.

6.3.3. Direct method

Based on the work of Trinchero et al. [58] who defined a connectivity metric for radial flow and one for transport, Fernandez-Garcia et al. [151] extended those definitions to more general flow conditions. They express the point to point flow connectivity metric as a space integral of the transmissivity involving a weighting function proportional to the sensitivity of the heads with respect to the natural logarithm of the transmissivity. The tracer connectivity metric is expressed as line integral along the flow path between the two points. In that framework, Fernandez-Garcia et al. [151] express analytically the cross covariances between the local transmissivity values, the flow connectivity metrics and the transport metrics (travel times) as a function of the covariance of the logarithm of the transmissivities. This allows them to generate directly simulations of the three variables conditional to field observations either of transmissivity, travel times, or flow connectivity. This is extremely powerful since it does not require to solve explicitly any flow or transport equations with a numerical method. As shown by Fernandez-Garcia et al. [151] this method allows to delineate accurately and in a very straightforward manner the capture zone of a well in a multi-Gaussian framework. But one of the very interesting part of their results is that it also allows to analyze theoretically the impact of data conditioning. It is shown for example, that the impact of tracer data is maximum when they are obtained at locations different from the transmissivity measurements.

7. Discussion and conclusions

7.1. How to define connectivity?

In this paper, we focused on the connectivity of hydraulic conductivity or transmissivity fields. We provided a broad definition of connectivity as a concept in relation with the existence of a path for flow and transport from one location to another. To quantify the connectivity, we have reviewed several static and dynamic metrics. The static metrics are essentially derived from integral geometry and percolation theory. We have shown that using a single indicator of connectivity is often insufficient. Simultaneous consideration of several metrics is necessary. Furthermore, we have shown strong evidence against the use of the Euler number as a connectivity characteristic. As pointed out several times, the Euler number is essentially a local characteristic strongly related to the regularity of the random set, or to the regularity of the continuous field. But it is not a global characteristic and it is not related to connectivity at long distance. We recommend the use of the *global* probability of connection Γ between two cells, and we recommend to compute those metrics both for the permeable and the impermeable phases.

We showed on thresholded continuous fields that this metric is strongly related to the percolation transition. Roughly speaking, if Γ and Γ_c belong both to the interval [0.2, 0.8], the binary field is in the percolation transition domain.

The dynamic connectivity metrics are related to physical processes such as flow and transport. Some of them can be estimated from field experiments. They allow quantifying the departure of a phenomenon from what is expected to occur in a standard situation, most often a multi-Gaussian field. Most dynamic connectivity metrics are based on effective parameters such as the effective hydraulic conductivity for flow. For solute transport the connectivity does not only affect the value of the effective parameters, but it also implies a departure from the physical model used at the small scale. Mass transfer between the connected mobile phase and the less mobile water phase implies non-Fickian conditions which require the use of alternative transport models able to describe fast arrivals and long tailing. Overall, the dynamic connectivity metrics are often very specific and not necessarily easy to interpret since there is not always a clear standard to which it should be compared.

7.2. Is connectivity the key?

We have shown in this review that connectivity influences very strongly a wide range of groundwater flow and transport processes. With time, more and more configurations and types of processes are investigated, and naturally the effect of connectivity appears to be important for most of them. Why then, researchers did not consider it before? In fact, if we consider the binary fields which have been intensively studied, we see that connectivity does not play a major role for very low or very high proportions. In those situations the proportion is the main controlling factor. However in the range of the intermediate proportions, the exact arrangement of the connected bodies becomes very important. In summary, connectivity itself is not sufficient to forecast the behavior of a medium, but it should not be forgotten if one wants to make reasonable forecasts of the behavior of a groundwater system.

7.3. A new vision of field hydrogeology and modeling

If connectivity is an important concept, there is a need to include its characterization as a specific step which should be carried out in aquifer or reservoir studies.

More precisely, in addition to what is normally done in terms of aquifer characterization, it seems clear that one should try to characterize the dynamic connectivity metrics with field experiments. It implies that data should be gathered at different scales in a systematic manner. On the one hand, measurements of the physical properties of the aquifer at a local scale (laboratory test, slug tests, column experiment, etc.) must be used to investigate the variability of the local values. From these data, their univariate and bivariate distributions can be estimated. On the other hand, large scale parameters should be obtained from the interpretation of large scale experiments (long term pumping tests, tracer tests, etc.). The comparison between the large scale effective values and the forecasts of the effective parameters obtained from local measurements allows then to quantify the degree of connectivity.

Such an analysis will not allow to define precisely where the connected structures are and therefore it must be completed with additional localized information. One approach is to use the dynamic connectivity metrics which relates the response of the aquifer at different locations (travel times, local diffusivity), and use this specific information to condition the models.

Finally, this approach must integrate classical investigations such as geological analysis or geophysical mapping with the specific aim of identifying and quantifying the presence of connected or disconnected features. Quantifying the proportions of the different facies, the geometry of the bodies, and defining a clear conceptual geological model (channels, lenses, etc.) is extremely important because it controls indirectly the degree of connectivity as we have seen when analyzing the properties of various binary models.

In terms of modeling, constraining the connectivity of the models should also become an important step. Ideally, all this work should allow finally to exclude heterogeneity models whose connectivity features are not compatible with the observations.

7.4. Research needs

All the research conducted so far and partly reviewed here allows to better understand the role of connectivity and its quantification. From a very broad perspective, the major questions that need to be answered now are the following:

1. What is the impact of connectivity on 3D continuous fields for various standard models? Very few studies considered 3D configurations [152,117]. From partial studies and theoretical considerations, we expect the connectivity behavior to be quantitatively and perhaps qualitatively different in 3D than in 2D. For example, percolation of the permeable phase implies that the background is not connected in 2D while it can be connected in 3D.
2. How to quantify the relation between static and dynamic connectivity metrics? This quest must be pursued to better understand the fundamental relations between heterogeneity, anisotropy, connectivity and the effective behaviors of heterogeneous materials.
3. What are the different connectivity structures that one should expect underground? This is a very difficult question since aquifers are normally not fully accessible. Recent advances in image acquisition such as lidar technology can bring new insights into this question [153,154]. However, the true connectivity of 3D fields cannot be directly inferred from 2D images since there is no stereological relations between 2D and 3D connectivity. Therefore new approaches need to be developed.
4. How a general methodology including new experimental procedures and modeling approaches to quantify the connectivity in the field and to constrain the models will improve forecast? So far the research that is the closest to reach that aim is the work of Fernandez-Garcia et al. [151].

Acknowledgments

The authors thank A. Barry, editor of Advances in Water Resources, for inviting us to prepare this contribution; D. Ginsbourger, G. Mariethoz, P. Brunner and I. Neuweiler who provided an early feed-back on the content of the paper; as well as P. Trinchero, S. Schlütter and two anonymous reviewers for their very detailed and constructive reviews.

Appendix A

Most of the connectivity metrics presented in this review can be computed with freely available software or can be coded with minimum effort in MATLAB (The Mathworks, Inc.). In this Appendix A, we provide a brief list of links allowing the reader to find the relevant references.

The basic functions required to compute the connectivity indicators in MATLAB are the functions *bwlabel ()* to compute the cluster identification function $C(x)$ and *bweuler ()* to compute the Euler number. The computation of the connectivity function $\tau(h)$ is very simple to code once the cluster function is computed. Some pseudo-code is given in Western et al. [13] and other details in Ali and Roy [19]. Computing the scalar metrics Γ (Eq. (4)) is also very simple, one has simply to sum the number of pixels having the same value of $C(x)$ to get the size n_i of each connected component i. The sum of all the square values of these numbers divided by the square of the total number of permeable pixel n_p is equal to Γ.

For those who prefer to use open source codes, they can refer to Deutsch [51] for a FORTRAN code to compute the cluster identification function $C(x)$ on 3D grids and to rank simulations. This program computes also some basic statistics such as the number, dimension or tortuosity of the connected components. Pardo-Iguzquiza and Dowd [109] provide another FORTRAN code (CONNEC3D) to compute the connectivity function $\tau(h)$ of 3D grids as well as a number of statistics related to the dimension of the connected components, the cluster identification function and the percolating connected components. For the Euler number, one can use the code developed by Vogel et al. [112] to compute the Minkowski functions. Details about the algorithm are also provided in Legland et al. [155].

References

[1] Bakr AA, Gelhar LW, Gutjahr AL, MacMillan JR. Stochastic analysis of spatial variability in subsurface flows – 1. Comparisons of one and three-dimensional flows. Water Resour Res 1978;14(2):263–71.
[2] Delhomme JP. Spatial variability and uncertainty in groundwater flow parameters. Water Resour Res 1979;15(2):281–90.
[3] Dagan G. Flow and transport in porous formations. New York: Springer-Verlag; 1989.
[4] Gelhar LW. Stochastic subsurface hydrology. Engelwood Cliffs: Prentice-Hall; 1993.
[5] Zhang D. Stochastic methods for flow in porous media. San Diego, CA: Academic Press; 2002.
[6] Rubin Y. Applied stochastic hydrogeology. New York: Oxford University Press; 2003.
[7] Journel AG, Alabert FG. New method for reservoir mapping. J Petrol Technol 1990;42(2):212–8.
[8] Journel AG, Deutsch CV. Entropy and spatial disorder. Math Geol 1993;25(3):329–55.
[9] Sánchez-Vila X, Carrera J, Girardi JP. Scale effects in transmissivity. J Hydrol 1996;183(1–2):1–22.
[10] Gómez-Hernández JJ, Wen XH. To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology. Adv Water Resour 1998;21(1):47–61.
[11] Koltermann CE, Gorelick SM. Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches. Water Resour Res 1996;32(9):2617–58.
[12] de Marsily G, Delay F, Goncalves J, Renard P, Teles V, Violette S. Dealing with spatial heterogeneity. Hydrogeol J 2005;13(1):161–83.
[13] Western AW, Blöschl G, Grayson RB. Toward capturing hydrologically significant connectivity in spatial patterns. Water Resour Res 2001;37(1):83–97.
[14] Grayson RB, Blöschl G, Western AW, McMahon TA. Advances in the use of observed spatial patterns of catchment hydrological response. Adv Water Resour 2002;25(8–12):1313–34.
[15] Mueller EN, Wainwright J, Parsons AJ. Impact of connectivity on the modeling of overland flow within semiarid shrubland environments. Water Resour Res 2007:43.

[16] Meerkerk AL, van Wesemael B, Bellin N. Application of connectivity theory to model the impact of terrace failure on runoff in semi-arid catchments. Hydrol Process 2009;23(19):2792–803.
[17] Lexartza-Artza I, Wainwright J. Hydrological connectivity: linking concepts with practical implications. Catena 2009;79(2):146–52.
[18] Michaelides K, Chappell A. Connectivity as a concept for characterising hydrological behaviour. Hydrol Process 2009;23(3):517–22.
[19] Ali GA, Roy AG. Shopping for hydrologically representative connectivity metrics in a humid temperate forested catchment. Water Resour Res 2010;46.
[20] Appels WM, Bogaart PW, van der Zee S. Influence of spatial variations of microtopography and infiltration on surface runoff and field scale hydrological connectivity. Adv Water Resour 2011;34(2):303–13.
[21] Stanford JA, Ward JV. An ecosystem perspective of alluvial rivers – connectivity and the hyporheic corridor. J N Am Benthol Soc 1993;12(1):48–60.
[22] Poole GC, Stanford JA, Frissell CA, Running SW. Three-dimensional mapping of geomorphic controls on flood-plain hydrology and connectivity from aerial photos. Geomorphology 2002;48(4):329–47.
[23] Brierley G, Fryirs K, Jain V. Landscape connectivity: the geographic basis of geomorphic applications. Area 2006;38(2):165–74.
[24] Wainwright J, Turnbull L, Ibrahim TG, Lexartza-Artza I, Thornton SF, Brazier RE. Linking environmental regimes, space and time: interpretations of structural and functional connectivity. Geomorphology 2011;126(3–4):387–404.
[25] Schumaker NH. Using landscape indices to predict habitat connectivity. Ecology 1996;77(4):1210–25.
[26] Moilanen A, Nieminen M. Simple connectivity measures in spatial ecology. Ecology 2002;83(4):1131–45.
[27] Pringle C. What is hydrologic connectivity and why is it ecologically important? Hydrol Process 2003;17(13):2685–9.
[28] Freeman MC, Pringle CM, Jackson CR. Hydrologic connectivity and the contribution of stream headwaters to ecological integrity at regional scales. J Am Water Resour Assoc 2007;43(1):5–14.
[29] Vogel HJ, Roth K. Quantitative morphology and network representation of soil pore structure. Adv Water Resour 2001;24(3–4):233–42.
[30] Okabe H, Blunt MJ. Pore space reconstruction of vuggy carbonates using microtomography and multiple-point statistics. Water Resour Res 2007;43(12).
[31] Schaap JD, Lehmann P, Kaestner A, Vontobel P, Hassanein R, Frei G, et al. Measuring the effect of structural connectivity on the water dynamics in heterogeneous porous media using speedy neutron tomography. Adv Water Resour 2008;31(9):1233–41.
[32] Lehmann P, Berchtold M, Ahrenholz B, Tolke J, Kaestner A, Krafczyk M, et al. Impact of geometrical properties on permeability and fluid phase distribution in porous media. Adv Water Resour 2008;31(9):1188–204.
[33] Morin RH, LeBlanc DR, Troutman BM. The influence of topology on hydraulic conductivity in a sand-and-gravel aquifer. Ground Water 2010;48(2):181–90.
[34] Schlüter S, Vogel HJ. On the reconstruction of structural and functional properties in random heterogeneous media. Adv Water Resour 2011;34(2):314–25.
[35] Herron N, Wilson C. A water balance approach to assessing the hydrologic buffering potential of an alluvial fan. Water Resour Res 2001;37(2):341–51.
[36] Brunner P, Simmons CT, Cook PG. Spatial and temporal aspects of the transition from connection to disconnection between rivers, lakes and groundwater. J Hydrol 2009;376(1–2):159–69.
[37] Moilanen A, Hanski I. On the use of connectivity measures in spatial ecology. Oikos 2001;95(1):147–51.
[38] Mariethoz, G. Geological stochastic imaging for aquifer characterization. PhD thesis. University of Neuchatel; 2009.
[39] Western AW, Blöschl G, Grayson RB. How well do indicator variograms capture the spatial connectivity of soil moisture? Hydrol Process 1998;12(12):1851–68.
[40] Strebelle S. Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 2002;34(1):1–21.
[41] Krishnan S, Journel AG. Spatial connectivity: from variograms to multiple-point measures. Math Geol 2003;35(8):915–25.
[42] Knudby C, Carrera J. On the relationship between indicators of geostatistical, flow and transport connectivity. Adv Water Resour 2005;28(4):405–21.
[43] Stauffer D, Aharony A. Introduction to percolation theory. 2nd ed. London: Taylor and Francis; 1994.
[44] Grimmett G. Percolation. Berlin: Springer; 2000.
[45] Bour O, Davy P. Connectivity of random fault networks following a power law fault length distribution. Water Resour Res 1997;33(7):1567–83.
[46] de Dreuzy JR, Davy P, Bour O. Hydraulic properties of two-dimensional random fracture networks following a power law length distribution 2. Permeability of networks based on lognormal distribution of apertures. Water Resour Res 2001;37(8):2079–95.
[47] Manzocchi T. The connectivity of two-dimensional networks of spatially correlated fractures. Water Resour Res 2002;38(9).
[48] Xu C, Dowd PA, Mardia KV, Fowell RJ. A connectivity index for discrete fracture networks. Math Geol 2006;38(5):611–34.
[49] Mariethoz G, Renard P, Straubhaar J. The direct sampling method to perform multiple-point geostatistical simulations. Water Resour Res 2010;46.
[50] Straubhaar J, Renard P, Mariethoz G, Froidevaux R, Besson O. An improved parallel multiple-point algorithm using a list approach. Math Geosci 2011;43(3):305–28.
[51] Deutsch CV. Fortran programs for calculating connectivity of three-dimensional numerical models and for ranking multiple realizations. Comput Geosci 1998;24(1):69–76.
[52] Boisvert JB, Pyrcz MJ, Deutch C. Multiple point metrics to assess categorical variable models. Nat Resour Res 2010;19(3):165–75.
[53] De Iaco S, Maggio S. Validation techniques for geological patterns simulations based on variogram and multiple-point statistics. Math Geosci 2011;43(4):483–500.
[54] Comunian A, Renard P, Straubhaar J. 3D multiple-point statistics simulation using 2D training images. Comput Geosci, in press; doi:10.1016/j.cageo.2011.07.009.
[55] Neuweiler I, Vogel HJ. Upscaling for unsaturated flow for non-Gaussian heterogeneous porous media. Water Resour Res 2007;43(3).
[56] Knudby C, Carrera J. On the use of apparent hydraulic diffusivity as an indicator of connectivity. J Hydrol 2006;329(3–4):377–89.
[57] Hovadik JM, Larue DK. Static characterizations of reservoirs: refining the concepts of connectivity and continuity. Petrol Geosci 2007;13(3):195–211.
[58] Trinchero P, Sánchez-Vila X, Fernandez-Garcia D. Point-to-point connectivity, an abstract concept or a key issue for risk assessment studies? Adv Water Resour 2008;31(12):1742–53.
[59] Fiori A, Boso F, de Barros FPJ, De Bartolo S, Frampton A, Severino G, et al. An indirect assessment on the impact of connectivity of conductivity classes upon longitudinal asymptotic macrodispersivity. Water Resour Res 2010;46.
[60] Larue DK, Hovadik J. Connectivity of channelized reservoirs: a modelling approach. Petrol Geosci 2006;12(4):291–308.
[61] King PR. The connectivity and conductivity of overlapping sand bodies. In: Buller AT, editor. North Sea oil and gas reservoirs II. Graham and Trotman; 1990. p. 353–62.
[62] Allard D, Heresim Group. On the connectivity of two random set models: the truncated Gaussian and the Boolean. In: Soares A, editor. Geostatistics Tria'92, vol. 1. Norwell, MA: Kluwer; 1994. p. 467–78.
[63] Nurafza PR, King PR, Masihi M. Facies connectivity modelling: analysis and field study. In: SPE Europe/EAGE annual conference and exhibition, SPE 100333; 2006.
[64] Madden TR. Microcrack connectivity in rocks – a renormalization-group approach to the critical phenomena of conduction and failure in crystalline rocks. J Geophys Res 1983;88(NB1):585–92.
[65] de Marsily G. Flow and transport in fractured rocks: connectivity and scale effect. In: Hydrogeology of rocks of low permeability. Memoires – International Association of Hydrogeologists, Tucson, AZ, USA, vol. 17; 1985. p. 267–77.
[66] Odling NE. Scaling and connectivity of joint systems in sandstones from Western Norway. J Struct Geol 1997;19(10):1257–71.
[67] Guimera J, Carrera J. A comparison of hydraulic and transport parameters measured in low-permeability fractured media. J Contam Hydrol 2000;41(3–4):261–81.
[68] Berkowitz B, Bour O, Davy P, Odling N. Scaling of fracture connectivity in geological formations. Geophys Res Lett 2000;27(14):2061–4.
[69] de Dreuzy JR, Davy P, Bour O. Hydraulic properties of two-dimensional random fracture networks following a power law length distribution 1. Effective connectivity. Water Resour Res 2001;37(8):2065–78.
[70] Ji SH, Park YJ, Lee KK. Influence of fracture connectivity and characterization level on the uncertainty of the equivalent permeability in statistically conceptualized fracture networks. Transport Porous Med 2011;87(2):385–95.
[71] Berkowitz B. Analysis of fracture network connectivity using percolation theory. Math Geol 1995;27(4):467–83.
[72] Tsang CF, Neretnieks I. Flow channeling in heterogeneous fractured rocks. Rev Geophys 1998;36(2):275–98.
[73] Berkowitz B. Characterizing flow and transport in fractured geological media: a review. Adv Water Resour 2002;25(8–12):861–84.
[74] Chiles JP. Stochastic modeling of natural fractured media: a review. Geostatistics Banff 2004, vol. 14. Berlin: Springer; 2005. p. 285–94.
[75] Allen JRL. Studies in fluviatile sedimentation – an exploratory quantitative model for the architecture of avulsion-controlled alluvial suites. Sediment Geol 1978;21(2):129–47.
[76] Allen JRL. Studies in fluviatile sedimentation – an elementary geometrical model for the connectedness of avulsion-related channel sand bodies. Sediment Geol 1979;24(3–4):253–67.
[77] Fogg GE. Groundwater-flow and sand body interconnectedness in a thick, multiple-aquifer system. Water Resour Res 1986;22(5):679–94.
[78] Poeter E, Townsend P. Assessment of critical flow path for improved remediation management. Ground Water 1994;32(3):439–47.
[79] LaBolle EM, Fogg GE. Role of molecular diffusion in contaminant migration and recovery in an alluvial aquifer system. Transport Porous Med 2001;42(1–2):155–79.
[80] Ronayne M, Gorelick S, Caers J. Identifying discrete geologic structures that produce anomalous hydraulic response: an inverse modeling approach. Water Resour Res 2008;44(8).
[81] Caers J, Hoffman T. The probability perturbation method: a new look at bayesian inverse modeling. Math Geol 2006;38(1):81–100.
[82] Silliman SE, Wright AL. Stochastic-analysis of paths of high hydraulic conductivity in porous-media. Water Resour Res 1988;24(11):1901–10.

[83] Tidwell VC, Wilson JL. Upscaling experiments conducted on a block of volcanic tuff: results for a bimodal permeability distribution. Water Resour Res 1999;35(11):3375–87.
[84] Zheng CM, Gorelick SM. Analysis of solute transport in flow fields influenced by preferential flowpaths at the decimeter scale. Ground Water 2003;41(2):142–55.
[85] Kiraly L. Rapport sur l'tat actuel des connaissances dans le domaine des caractres physiques des roches karstiques [report on the present knowledge of the physical characters of karstic rocks]. In: Burger A, Dubertret L, editors. Hydrogeology of karstic terrains, vol. 8, no. 3. Paris, France: International Association of Hydrogeologists, International Union of Geological Sciences; 1975. p. 53–67.
[86] Lallemand-Barres P, Peaudecerf P. Recherche des relations entre la valeur de la dispersivité macroscopique d'un milieu aquifère, ses autres caractristiques et les conditions de mesures, étude bibliographique. Bull Bur Rech Geol Min 1978;3(4):277–87.
[87] Schulze-Makuch D, Cherkauer DS. Variation in hydraulic conductivity with scale of measurement during aquifer tests in heterogeneous, porous carbonate rocks. Hydrogeol J 1998;6:204–15.
[88] Clauser C. Permeability of crystalline rocks. EOS Trans 1992;73(21): 233–40.
[89] Gelhar LW, Welty C, Rehfeldt KR. A critical review of data on field-scale dispersion in aquifers. Water Resour Res 1992;28(7):1955–74.
[90] Neuman SP. Generalized scaling of permeabilities: validation and effect of support scale. Geophys Res Lett 1994;21(5):349–52.
[91] Wen XH, Gómez-Hernández JJ. Numerical modeling of macrodispersion in heterogeneous media: a comparison of multi-Gaussian and non-multi-Gaussian models. J Contam Hydrol 1998;30(1–2):129–56.
[92] Zinn B, Harvey CF. 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 Resour Res 2003;39(3):1051.
[93] Neuweiler I, Cirpka OA. Homogenization of richards equation in permeability fields with different connectivities. Water Resour Res 2005;41(W02009).
[94] Neuweiler I, Papafotiou A, Class H, Helmig R. Estimation of effective parameters for a two-phase flow problem in non-Gaussian heterogeneous porous media. J Contam Hydrol 2011:141–56.
[95] Kerrou J, Renard P, Hendricks Franssen H, Lunati I. Issues in characterizing heterogeneity and connectivity in non-multigaussian media. Adv Water Resour 2008;31(1):147–59.
[96] Zhou HY, Gomez-Hernandez JJ, Franssen HJH, Li LP. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Adv Water Resour 2011;34(7):844–64.
[97] Hoshen J, Kopelman R. Percolation and cluster distribution I. Cluster multiple labeling technique and critical concentration algorithm. Phys Rev B 1976;14(8):3438–45.
[98] Newman MEJ, Ziff RM. Fast monte carlo algorithm for site or bond percolation. Phys Rev E 2001;64(1):016706. doi:10.1103/PhysRevE.64.016706.
[99] Allard D. Connexité des ensembles aléatoires: application à la simulation de réservoirs pétroliers hétérogènes. PhD. Thesis. Ecole des Mines de Paris, Fontainebleau; 1993.
[100] Lantuéjoul C. Geostatistical simulations. Berlin: Springer; 2002.
[101] Matheron G. Random heterogeneous materials: microstructure and macroscopic properties. New-York: Springer; 2002.
[102] Vogel HJ. Topological characterization of porous media. In: Mecke K, Stoyan D, editors. Morphology of condensed matter: physics and geometry of spatially complex systems. Lecture Notes in Physics, vol. 600. New York: Springer; 2002. p. 7592.
[103] Matheron G. Random sets and integral geometry. New York: Wiley; 1975.
[104] Santalo L. Integral geometry and geometric probability. Reading, MA: Addison–Wesley; 1976.
[105] Serra J. Image analysis and mathematical morphology, vol. 1. Academic Press; 1984.
[106] Mantz H, Jacobs K, Mecke K. Utilizing Minkowski functionals for image analysis: a marching square algorithm. J Stat Mech Theor Exp 2008.
[107] Adler R, Taylor J. Random fields and geometry. Boston: Springer; 2007.
[108] Torquato S, Beasley JD, Chiew YC. 2-Point cluster function for continuum percolation. J Chem Phys 1988;88(10):6540–7.
[109] Pardo-Iguzquiza E, Dowd PA. CONNEC3D: a computer program for connectivity analysis of 3D random set models. Comput Geosci 2003;29(6): 775–85.
[110] Knudby C, Carrera J, Bumgardner JD, Fogg GE. Binary upscaling – the role of connectivity and a new formula. Adv Water Resour 2006;29(4):590–604.
[111] Soille P. Morphological image analysis. Principles and applications. Berlin: Springer; 2004.
[112] Vogel HJ, Weller U, Schlüter S. Quantification of soil structure based on Minkowski functions. Comput Geosci 2010;36(10):1236–45.
[113] Samouelian A, Vogel HJ, Ippisch O. Upscaling hydraulic conductivity based on the topology of the sub-scale structure. Adv Water Resour 2007;30(5):1179–89.
[114] Soille P. Constrained connectivity for hierarchical image partitioning and simplification. IEEE Trans Pattern Anal Mach Intell 2008;30(7):1132–45.
[115] Matheron G. Composition des perméabilités en milieu poreux hétérogène: méthode de Schwydler et règles de pondération. Rev Inst Fr Petrol 1967;23:443–66.
[116] Nœtinger B. The effective permeability of a heterogeneous porous medium. Transport Porous Med 1994;15:99–127.
[117] Bianchi M, Zheng CM, Wilson C, Tick GR, Liu GS, Gorelick SM. Spatial connectivity in a highly heterogeneous aquifer: from cores to preferential flow paths. Water Resour Res 2011:47.
[118] Wen XH, Gómez-Hernández JJ. Upscaling hydraulic conductivities in heterogeneous media: an overview. J Hydrol 1996;183(1–2):ix–xxxii.
[119] Renard P, Marsily Gd. Calculating equivalent permeability: a review. Adv Water Resour 1997;20(5–6):253–78.
[120] Sánchez-Vila X, Guadagnini A, Carrera J. Representative hydraulic conductivities in saturated groundwater flow. Rev Geophys 2006;44(3). Article No. RG3002.
[121] Ababou R. Random porous media flow on large 3-D grids: numerics, performance and application to homogenization. In: Wheeler MF, editor. IMA volumes in mathematics and its applications. Environmental studies: mathematical, computational and statistical analysis, vol. 79. New York: Springer-Verlag; 1996. p. 1–25.
[122] Bruderer-Weng C, Cowie P, Bernabe Y, Main I. Relating flow channelling to tracer dispersion in heterogeneous networks. Adv Water Resour 2004;27(8):843–55.
[123] Frippiat CC, Illangasekare TH, Zyvoloski GA. Anisotropic effective medium solutions of head and velocity variance to quantify flow connectivity. Adv Water Resour 2009;32(2):239–49.
[124] Park K, Scheidt C, Caers J. Simultaneous conditioning of multiple non-Gaussian geostatistical models to highly nonlinear data using distances in kernel space. In: Ortiz J, Emery X, editors. Proceedings of 8th international geostatistical congress, Gecamin, Santiago, Chile, SPE 100333; 2008. p. 247–56.
[125] Le Goc R, de Dreuzy JR, Davy P. Statistical characteristics of flow as indicators of channeling in heterogeneous porous and fractured media. Adv Water Resour 2010;33(3):257–69.
[126] Meier PM, Carrera J, Snchez-Vila X. An evaluation of Jacob's method for the interpretation of pumping tests in heterogeneous porous media. Water Resour Res 1998;34(5):1011–25.
[127] Sánchez-Vila X, Meier PM, Carrera J. Pumping tests in heterogeneous aquifers: an analytical study of what can be obtained from their interpretation using Jacob's method. Water Resour Res 1999;35(4):943–52.
[128] Bernabe Y, Mok U, Evans B, Herrmann FJ. Permeability and storativity of binary mixtures of high- and low-permeability materials. J Geophys Res Solid Earth 2004;109(B12).
[129] Hashin Z, Shtrikman S. A variational approach to the theory of the effective magnetic permeability of multiphase materials. J Appl Phys 1962;33(10):31253131.
[130] McKenna SM, Ray J, Marzouk Y, van Bloemen Waanders B. Truncated multigaussian fields and effective conductance of binary media. Adv Water Resour 2011;34:617–26.
[131] Mecke KR, Wagner H. Euler characteristic and related measures for random geometric sets. J Stat Phys 1991;64(3–4):843–50.
[132] Willmann M, Carrera J, Sánchez-Vila X. Transport upscaling in heterogeneous aquifers: what physical parameters control memory functions? Water Resour Res 2008;44(12).
[133] Allard D, Froidevaux R, Biver P. Conditional simulation of multi-type non stationary markov object models respecting specified proportions. Math Geol 2006;38(8):959–86.
[134] Mariethoz G, Kelly BFJ. Modeling complex geological structures with elementary training images and transform-invariant distances. Water Resour Res 2011:47.
[135] Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science 1983;220(4598):671–80.
[136] Robert C, Casella G. Monte Carlo statistical methods. Springer; 1999.
[137] Jiao Y, Stillinger FH, Torquato S. A superior descriptor of random textures and its predictive capability. Proc Natl Acad Sci 2009;106(42):17634–9.
[138] Allard D. Simulating a geological lithofacies with respect to connectivity information using the truncated gaussan model. In: Armstrong M, Dowd P, editors. Geostatistical simulations: proceedings of the geostatistical simulation workshop, Fontainebleau, France, 27–28 May, 1993. Norwell, MA: Kluwer; 1994. p. 197–211.
[139] Geman S, Geman D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans Pattern Anal Mach Intell 1984;6(6):721–41.
[140] Renard P, Straubhaar J, Caers J, Mariethoz G. Conditioning facies simulation with connectivity data. Math Geosci 2011;43(8):879–903.
[141] Alcolea A, Carrera J, Medina A. Pilot points method incorporating prior information for solving the groundwater flow inverse problem. Adv Water Resour 2006;29:16781689.
[142] Hendricks-Franssen HJ, Alcolea A, Riva M, Bakr M, van der Wiel N, Stauffer F, et al. A comparison of seven methods for the inverse modelling of groundwater flow. application to the characterisation of well catchments. Adv Water Resour 2009;32(6):851–72.
[143] Oliver D, Chen Y. Recent progress on reservoir history matching: a review. Comput Geosci 2011;15:185–221.
[144] Hird KB, Dubrule O. Quantification of reservoir connectivity for reservoir description applications. SPE Reserv Eval Eng 1998;1(1):12–7.
[145] Brandsaeter I, Wist HT, Naess A, Lia O, Arntzen OJ, Ringrose PS, et al. Ranking of stochastic realizations of complex tidal reservoirs using streamline criteria. Petrol Geosci 2001;7:S53–63.

[146] McLennan J, Deutch C. Ranking geostatistical realizations by measures of connectivity. In: SPE/PS-CIM/CHOA 98168. PS2005-437; 2005. p. 13.
[147] de Jager G, Van Doren JFM, Jansen JD, Luthi SM. An evaluation of relevant geological parameters for predicting the flow behaviour of channelized reservoirs. Petrol Geosci 2009;15(4):345–54.
[148] Scheidt C, Caers J. Representing spatial uncertainty using distances and kernels. Math Geosci 2009;41(4):397–419.
[149] Scheidt C, Caers J. Uncertainty quantification in reservoir performance using distances and kernel methods – application to a West Africa deepwater turbidite reservoir. SPE J 2009;14(4):680–92.
[150] Alcolea A, Renard P. Blocking moving window algorithm: conditioning multiple-point simulations to hydrogeological data. Water Resour Res 2010;46.
[151] Fernandez-Garcia D, Trinchero P, Sánchez-Vila X. Conditional stochastic mapping of transport connectivity. Water Resour Res 2010;46.
[152] Vassena C, Cattaneo L, Giudici M. Assessment of the role of facies heterogeneity at the fine scale by numerical transport experiments and connectivity indicators. Hydrogeol J 2010;18(3):651–68.
[153] Klise KA, Weissmann GS, McKenna SA, Nichols EM, Frechette JD, Wawrzyniec TF, et al. Exploring solute transport and streamline connectivity using lidar-based outcrop images and geostatistical representations of heterogeneity. Water Resour Res 2009:45.
[154] Nichols EM, Weissmann GS, Wawrzyniec TF, Frechette JD, Klise KA. Processing of outcrop-based lidar imagery to characterize heterogeneity for groundwater models. SEPM concepts in sedimentology and paleontology, vol. 10; 2011. p. 239–47.
[155] Legland D, Kie~u K, Devaux MF. Computation of Minkowski measures on 2D and 3D binary images. Image Anal Stereol 2007;26(2):83–92.