



INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus

Géoscience

Sciences de la Planète

Benoît Noetinger

Random fields and up scaling, towards a more predictive probabilistic quantitative hydrogeology


Volume 355, Special Issue S1 (2023), p. 559-572

Online since: 27 February 2023

Part of Special Issue: Geo-hydrological Data & Models

Guest editors: Vazken Andréassian (INRAE, France),
Valérie Plagnes (Sorbonne Université, France), Craig Simmons (Flinders University,
Australia) and Pierre Ribstein (Sorbonne Université, France)

<https://doi.org/10.5802/crgeos.188>

 This article is licensed under the
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



*The Comptes Rendus. Géoscience — Sciences de la Planète are a member of the
Mersenne Center for open scientific publishing*
www.centre-mersenne.org — e-ISSN : 1778-7025



Research article

Geo-hydrological Data & Models

Random fields and up scaling, towards a more predictive probabilistic quantitative hydrogeology

Benoît Noetinger^{©,a}

^a IFP Energies Nouvelles, France

E-mail: benoit.noetinger@ifpen.fr

Abstract. Random fields are becoming a mature tool sharing applications in many area of physics, mechanics and geosciences. In the latter, it is commonly used under the name of geostatistics. Continuous enrichment of geological/geostatistical models leads to manipulating hydrogeological models characterized by many parameters or hyperparameters corresponding to statistical aggregates that may be poorly estimated due to the scarcity of field data. Those parameters are generally support-scale-dependent and uncertain, so some inverse problem and uncertainty analysis must be carried out in practical applications that involve generally some forward calculation for example a fluid flow simulation if one is interested in transfers in the subsurface. Up scaling techniques are still required to find and to restrict in a controlled manner the more relevant parameters, allowing to lower the dimension of the parameter space. In the stochastic case, the interaction between the conductivity spatial distribution and the flow pattern can lead to non trivial behaviours that will be discussed. Fractured media will not be considered. That note does not present original results, but a selection of some potentially fruitful research avenues suggested by previous works.

Keywords. Applied geosciences, Porous media, Disorder, Upscaling, Geostatistics, Quenched disorder.

Manuscript received 11 April 2022, revised 4 November 2022, accepted 22 November 2022.

1. Introduction

The use of numerical models for studying subsurface flow has become common practice in hydrology and petroleum engineering over the last 50 years [Renard and De Marsily, 1997, De Marsily et al., 2005]. However, one of the major questions that still poses a problem is that the data needed to build an accurate model are incomplete in essence. Although contemporary computers are growing ever more powerful and capable of describing the relevant flows with increasing details, the required input data growth imposes a major increase in computing facilities if one wants to explore correctly the uncertain input parameter space in order to make the best decisions [Gorell et al., 2001]. In many applied geosciences issues, the basic concern is to recover, to store or to fol-

low some fluid, waste or even thermal energy in subsurface formations at good economical conditions, while minimizing overall environmental risks. Most of the associated transport processes occur in the porosity of natural rocks, perhaps fractured, so studying flow in porous media at several scales is a major issue. In order to proceed, a concise macroscopic description of flow in porous media encompassing several scales without knowing the exact structure of the subsurface at lower resolution scales must be obtained.

Basically one wants to be able to estimate with some accuracy the average value, as well as the associated variance of any simulated value of interest. Knowledge of the full probability distribution is in general not required, although knowing if the re-

sulting distribution is approximately Gaussian or not may be important. In the general case, when the driving equations relating input data to output observables are strongly non-linear, extensive Monte-Carlo sampling of the input parameter space is required to get estimations of these output statistical quantities. A typical subsurface model is generally described by some random fields of say porosity, permeability maps conditioned to several field observations, sedimentary layer shapes, aquifer positions, etc... These random maps depend in turn on some hyperparameters of statistical nature such as variance, drift, correlation scale: that is the aim of geostatistics that will be presented with more details in Section 2.1. In addition, some man-controlled parameters such as the position of wells and the associated flow rates enter as input data in the model. For applications, that influence of that set of parameters may be sampled by simulation to select the best position and management of the wells.

Using such models, we are led to consider high dimensional parameter spaces, both for the random set of natural data, and for the man-controlled parameters. Consider a parameter space of $N = 100$ dimensions, and by hypothesis, let us assume that each parameter can take only 2 values. An exhaustive exploration of that parameter space will imply being able to perform 2^{100} independent runs, that is clearly impossible even with hypothetical quantum computers. In addition, in current models, most parameters belong to some continuous interval, increasing the effective dimensions of the parameter space. So following any exhaustive approach is a dead end. Many alternative sampling techniques were developed [Zhang, 2001]. Experimental design techniques coupled with machine learning methods are being developed [Scheidt et al., 2007, Santos Oliveira et al., 2021, and references therein].

That implies that some systematic workflow is to be built in order to reduce the number of dimensions of the effective working space under consideration. As that situation presents some similarities with statistical physics, some bridges between both disciplines can be expected: a description of these connections may be found in Noetinger [2020]. To sum-up, similar issues are at the heart of one the most active area of complex systems understanding, the so-called spin glasses for which G Parisi was awarded the 2021 Nobel prize in physics [Mézard et al., 1987, Zde-

borová and Krzakala, 2007, Zdeborová and Krzakala, 2016]. In the geoscience side, some reviews illustrate similar issues [Gorell et al., 2001, Floris et al., 2001, De Marsily et al., 2005].

The goal of that short paper is to suggest some avenues of research in order to provide answers to the following questions:

- (i) What is the amount of information that is required to build a useful numerical model with a degree of sophistication consistent with the input knowledge.
- (ii) is it possible to build a model that may be continuously improved as our knowledge of the particular aquifer increases along with the volume of data acquisition.

The paper is organised as follows: in Section 2, we provide some general elements. Then a focus is given in Section 2.1 about geostatistics, before giving some elements about information quantification in Section 2.2. Single phase flow problem serving as a basis for most of our discussion is introduced in Section 2.3. Then we discuss up-scaling applications in Section 3.1 by discussing the scale and support effects. Non linear issues implying a strong feed-back between the medium structure and the flow are discussed in the case of two-phase flows in Section 3.2. In Section 4, we present some ideas allowing to merge the preceding considerations between geostatistics, up-scaling and information theory for practical applications. Finally some fruitful research avenues are suggested for future investigations.

2. Medium description, up-scaling issues

We consider a subsurface formation of typical kilometeric scales. In Figure 1, the overall scales of the considered aquifer are represented, as well as the corresponding discretization mesh of interest, chosen cartesian for sake of simplicity. Each grid block may be populated by parameters such as porosity $\Phi(\mathbf{r})$, permeability $k(\mathbf{r})$ and any other relevant parameter. That can be drawn using some deterministic or geostatistical software.

2.1. About geostatistics

Modelling subsurface formation using numerical models implies to account for the fact that any local

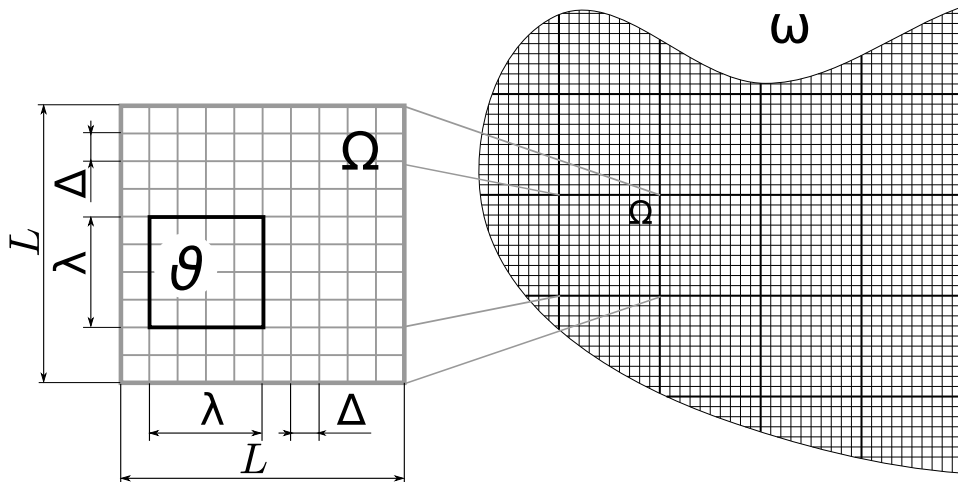


Figure 1. Geometry of the problem on a simplified 2D section of a subsurface model domain ω . A coarse grid made of blocks Ω of typical size L is super-imposed to a geological fine grid of typical size Δ in order to solve the conservation equations of interest. The intermediate scale λ serves in posterior treatments for checking the overall consistency of the model. Reprinted (adapted) from Colecchio et al. [2020] Copyright 2020 with permission of Elsevier.

quantity of interest may depend on position vector \mathbf{r} . In most cases, for example the quenched¹ positive permeability $k(\mathbf{r})$, is represented as being a random function of position. It is characterized by some

mean-value and fluctuations that are measured using field data (an issue in itself!). That is the aim of the so-called geostatistical approach, originally founded by Matheron [1963], Krige [1976], De Marsily et al. [2005], Chiles and Delfiner [2009] and developed by several others, that form the conceptual basis of several popular commercial softwares. The basic idea is to interpolate known measurements on a given support on several locations by inferring some underlying statistical structure. Notice that the primary random mathematical objects are functions or fields that are characterized by an infinite number of degrees of freedom. So the most natural embedding formalism that encompasses it is the so called field-theory employed in both statistical mechanics or particle physics [Noetinger, 2020, Hristopulos, 2020]. Another related issue that was pointed-out by Tarantola [2005] is that working with continuous variables implies that there is not a reference entropy measure allowing to quantify easily the information of a geostatistical distribution: the overall results are sensitive to the coordinate system used by the modeller. At this stage, the associated reference “non-informative” reference distribution [Tarantola, 2005, Abellan and Noetinger, 2010] is not determined. That explains some deep theoretical and practical difficulties for generalizing the approach to curved

¹The term quenched is employed in statistical physics of disordered systems, it corresponds to systems having a fixed but unknown disorder without thermal fluctuations that reorganise it, a glass or an amorphous solid are typical quenched material.

spaces or to non-uniform grid, Mourlanette et al. [2020]: such a situation may occur if tectonics imply strong deformation of a sedimentary basin. In most practical cases, some implicit cartesian assumption is made. Probably it could be justified using the sedimentary nature of most aquifers. In the similar statistical physics case, it may be shown that the continuous Boltzmann–Gibbs distribution that is the basis of most applications may be derived by a suitable limit of the underlying quantum mechanics processes that poses paradoxically less difficulty owing to their discrete nature (Setting-up a natural probability measure to a 6 face dice rather than to a continuous process). In geoscience, approximate Log-normal distributions (the logarithm of the conductivity is a Gaussian distributed variable) were observed at the core scale within several geological environments [Gelhar, 1993]. Note that in most cases, the amount of data is not sufficient to provide the form of the probability distribution function (pdf) of k , and higher order correlation functions are almost impossible to be determined from field data. In the current modelling practice, they are implicitly fixed by choosing some additional assumptions based on geological expertise on similar environments. As a consequence, even the input stochastic model is questionable.

Another class of models that intend to add information on high order correlation functions are multipoint or so-called process-based models, intending to reconstruct the apparent randomness of the subsurface by simulating the long term history of the formation using some dynamical physically based model: Diggle et al. [1998], Granjeon and Joseph [1999], Hu and Chugunova [2008], Michael et al. [2010]. These models introduce more physically based parameters and are more realistic. However, they may depend on assumptions regarding the time-dependant external forcing at their boundaries, introducing stochasticity in a somewhat another manner. An essential specificity of geosciences is the so-called measurement support effect: any data is obtained using some instrumentation that averages the details of the process at hand on a characteristic length that depends on the physical process. That scale may vary from some centimeters for laboratory/logging tools, to hectometers in geophysics corresponding to the typical wavelength of geophysical waves, or to the support of transient well tests in-

terpretation. Any measurement is characterized by a support scale characterizing the measurement itself, and by another spatial scale that is the typical resolution.

Finally, due to the data scarcity, it is necessary to up-date any prior model of the subsurface by accounting for the additional data that are continuously enriching the available information. That new data can be provided by pressure and rates measurements at some producing wells, time-dependant tracer concentration showing connections between several regions of the subsurface etc... This is the so called inverse problem issue that was addressed by many authors [Jacquard, 1965, LaVenue et al., 1995]. It is well-known that such problems are ill conditioned and that some regularization is required [Tarantola, 2005, Oliver and Chen, 2011, and references therein]. That can be achieved accounting for a prior geostatistical structure [LaVenue et al., 1995, Le Ravalec et al., 2000, Hu and Chugunova, 2008] that is easier to carry-out considering gaussian priors.

2.2. Quantification of information

As data may be scarce (but even redundant), it is essential to be able to quantify the amount of information provided by a measurement, as well as the quantity of information that is required to make faithful predictions. That idea can be formalized using information theory, as attempted in Abellan and Noetinger [2010, and references therein]. A quantitative estimate can be defined introducing the relative entropy (Kullback–Leibler divergence) between posterior and prior distribution that serves as a non-informative reference distribution. That quantity is defined by:

$$I(p) = I(p|\mu) = \int p(\mathbf{m}) \log \frac{p(\mathbf{m})}{\mu(\mathbf{m})} d\mathbf{m}. \quad (1)$$

Here, $p(\mathbf{m})$ is the posterior distribution of \mathbf{m} given that some measurement was carried out on the system. $\mu(\mathbf{m})$ denotes the prior distribution \mathbf{m} . For clarity, it is easier to introduce some discretization of \mathbf{m} on a regular grid. So \mathbf{m} is represented as a realization of the M dimensional vector, so $\mathbf{m} = (m(\mathbf{x}_1), \dots, m(\mathbf{x}_M))^t$. The corresponding $\mu(\mathbf{m})$ is assumed to be centered and multigaussian so the corresponding probability measure is given by $\mu(\mathbf{m}) \sim$

$\exp -1/2(\mathbf{m} \cdot \mathbf{C}_p \cdot \mathbf{m})$ with a gaussian covariance matrix with elements that are given by:

$$\mathbf{C}_p(\mathbf{x}_i, \mathbf{x}_j) = \sigma_0^2 \exp\left(\frac{-|\mathbf{x}_i - \mathbf{x}_j|^2}{\ell_c^2}\right) \quad (2)$$

with ℓ_c^2 is the correlation length, and σ_0^2 the associated variance. Then the relative entropy I_{MN} of a posterior distribution corresponding to N linear measurements with independant errors σ_d^2 gathered as a N dimensional vector $\mathbf{F} \cdot \mathbf{x}$ is given by I_{MN} :

$$I_{MN} = \frac{1}{2} \log |\mathbb{I}_N + \mathbf{C}_v^{-1} \cdot \mathbf{F}^t \mathbf{C}_p \mathbf{F}| \quad (3)$$

where $|\dots|$ is the determinant of the square matrix \dots . Considering the specific problem of local measurements of the random field \mathbf{m} at N different locations $(\mathbf{x}_1, \dots, \mathbf{x}_N)$, $N < M$ with measurement errors σ_d^2 , one is led to evaluate:

$$I_{MN}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \frac{1}{2} \log \left| \left(\delta_{ij} + \frac{\sigma_0^2}{\sigma_d^2} \exp \frac{-|\mathbf{x}_i - \mathbf{x}_j|^2}{\ell_c^2} \right)_{1 \leq i, j \leq N} \right|, \quad (4)$$

when the measurement locations are far apart (typically all distances greater than ℓ_c), one gets the estimation:

$$I_{MN}(x_1, \dots, x_N) = \frac{N}{2} \log(1 + \mu^2) \quad \text{with } \mu^2 = \frac{\sigma_0^2}{\sigma_d^2}. \quad (5)$$

Figure 2 illustrates present results that were also generalised to pumping test analysis and non-linear flow driven issues in Abellan and Noetinger [2010].

2.3. Fluid-flow modelling

In most cases, modelling fluid flow in a porous formation [De Marsily, 1986] leads to solve a diffusion-like equation that reads:

$$\phi c_t \frac{\partial p(\mathbf{r}, t)}{\partial t} = \nabla \cdot (k(\mathbf{r}) \nabla p(\mathbf{r}, t)) + f(\mathbf{r}). \quad (6)$$

Here, the parameters $p(\mathbf{r}, t)$, c_t and $f(\mathbf{r})$ are respectively the time-dependent potential of interest, the total compressibility, and a source term. Dirichlet or Neumann Boundary conditions are known at the boundary of the domain ω . Once that potential is known, associated transport problems can be solved, including multiphase flow issues that are not in the scope of present paper.

The basic issue is to be able to solve such equations, accounting for both the the effect of heterogeneities, and to be able to quantify the related uncertainties. In that context, analytical methods were

developed by hydrogeologists [Gelhar, 1993, Dagan, 1989, Indelman and Abramovich, 1994, Abramovich and Indelman, 1995, Renard and De Marsily, 1997], mathematicians [Jikov et al., 2012, Armstrong et al., 2019] and physicists, among other [King, 1989, Noetinger, 1994, Hristopulos and Christakos, 1999, Noetinger, 2000, Attinger, 2003, Eberhard et al., 2004, Teodorovich, 2002, Stepanyants and Teodorovich, 2003].

On the numerical side, a possible strategy is to transfer the small scale spatial fluctuations to a large scale support that encompasses the low spatial frequency components of the fields of interest. The calculations are carried out in practice using some numerical model in which the Laplace equation is solved using a grid of resolution L generally considerably much coarser than the input fine geological grid Δ (notations in Figure 1), because the available computing power leads also to continuous improvement of local geological 3D representations. This implies obtaining a coarse grained Laplace equation with a renormalized conductivity map accounting as best as possible to the local subgrid variations. In the case of multimodal distributions, in which connectivity aspects are dominant, specific methods can be developed within the framework of percolation theory, such as real-space renormalization techniques [Berkowitz and Balberg, 1993, Hunt et al., 2014, Hristopulos, 2020].

In the stochastic context, a strategy is to average the solution of Equation (6) to get the average head (or pressure) $\langle p(\mathbf{r}, t) \rangle$ in which the ensemble-average $\langle \dots \rangle$ is to be taken on the disorder of the conductivity field $k(\mathbf{r})$ corresponding to a Monte-Carlo average over the disorder of the medium. Then, one can look for the equation that may drive that average head² [Noetinger and Gautier, 1998].

3. About flow up-scaling

3.1. Single phase steady state up-scaling

It can be shown that under quite general hypothesis (statistical stationarity and convergence conditions)

²At first sight, averaging the driving equation looks to be simpler, and would evidently provide the arithmetic average for the conductivity, a clearly wrong result.

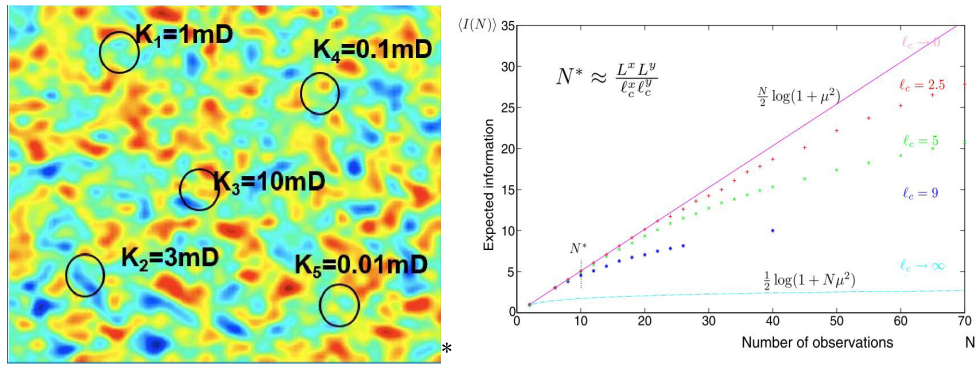


Figure 2. Left: A typical realization of a random map 2D section of a subsurface model on a domain Ω , representing $\text{Log}(k)$. The superimposed circles correspond to $N = 5$ physical measurements of permeability on a support scale with the corresponding radius, with several realistic values. The average distance between the circles corresponds to the coverage of the area. In a perfect situation, all the area should be covered, that is seldom the case. Right: the amount of information provided by N local measurements defined by the relative entropy of posterior distribution versus prior. Note the saturation of the information once the medium is well covered by the measurements, μ is defined in the main text. L^x and L^y are the overall size of the domain in the x and y directions, ℓ_0^x and ℓ_0^y denote the correlation length in corresponding directions. N^* is thus the number of correlated volumes inside the overall domain. It corresponds to the crossover the transition of information content between independent measurements and redundant ones. Copyright 2020 with permission of Springer-Nature.

that the low frequency components of the average potential $\langle p(\mathbf{r}, t) \rangle$ is driven by an effective equation that reads:

$$\langle \phi \rangle_{c_t} \frac{\partial \langle p(\mathbf{r}, t) \rangle}{\partial t} = \nabla \cdot (K_{\text{equ}} \nabla \langle p(\mathbf{r}, t) \rangle) + f(\mathbf{r}). \quad (7)$$

Here, the average $\langle \dots \rangle$ corresponds to an averaging over the randomness of the medium. The parameter K_{equ} is called the equivalent conductivity. It corresponds to the “natural” large scale relation between the mean flux and the large scale pressure gradient that can be provided by homogenization theories solving the so-called “auxiliary problem” to be solved numerically in x , y and z directions [Armstrong et al., 2019, and references therein], as it is illustrated Figure 3.³ Many expressions were proposed to relate K_{eff} to the underlying disorder [Renard and

De Marsily, 1997]. For one dimension, an elementary analytical calculation provides the harmonic average $K_{\text{eff}} = \langle k^{-1} \rangle^{-1}$. Generalizations of such a simple formula were proposed by many authors [Landau and Lifshitz, 1960, Matheron, 1967, King, 1987, 1989, Dagan, 1993, Neuman and Orr, 1993, Noetinger, 1994, Indelman and Abramovich, 1994, De Wit, 1995, Teodorovich, 2002, Stepanyants and Teodorovich, 2003, Colechio et al., 2020, Nøetinger, 2000]. Looking at the evolution of the permeability distribution, in 2 dimensions, numerically the log normal distribution appears to be stable under the up-scaling transformation, Figure 4 from Colechio et al. [2020], analogous to the central limit theorem. That may be related to the duality argument of Matheron [1967] that justifies the geometric mean in 2D. In the 3D case, studying the emergence of a “stable” conductivity distribution invariant on the up scaling transformation would also be useful for studying strongly correlated systems having conductivity correlations decaying as a power law with the lag distance. This may be illustrated on Figure 4 below: in the practical side, in most cases, at present times,

³Notice that in the current practice, the result of the evaluation of these formula, or of the numerical solution of the auxiliary problem is used to populate the coarse grid of Ω blocks, using the underlying local fine grid data. So the equivalent K_{equ} may still depend on position rather than the so-called K_{eff} that is intrinsic, it would correspond to working with an infinite volume Ω .

using a numerical technique is sufficient. In case of extremely heterogeneous media (that can correspond to bimodal media) at percolation threshold, a percolation transition may occur [Charlaix et al., 1987, Berkowitz and Balberg, 1993, Hunt et al., 2014, Stauffer and Aharony, 2014, Colecchio et al., 2020]. Its main consequence, a stabilization of the effective conductivity over a very large scale can be observed on Figure 5. In both situations, it appears that at some scale (becoming infinite just at the percolation threshold for an infinite small-scale conductivity contrast), the formation can be treated as an almost continuous medium: that is the so called self-averaging property that manifests itself once the characteristic size of the flow is larger than the typical correlation scale. Long ranged media, as well as media close to a percolation threshold remain a notable exception. The results presented in Figures 4 and 5 were obtained using a suitable post treatment of a single phase flow simulation on a large domain, without adding any external expansive Monte Carlo loop: that is sufficient to capture the dominant characteristic scales of the flow. In addition, in Figure 5, middle, it is illustrated the emergence of the so-called localization phenomena that is a consequence of high permeability contrasts [Arnold et al., 2019] that localizes high velocity regions in tiny regions, even in non-fractured media. Such flow-based maps are well suited for post-treatment: helping to build a coarse mesh, estimating critical paths that may control the overall uncertainties.

3.2. *About the interaction between the conductivity spatial distribution and the flow pattern*

In several situations, such as when considering multiphase flow flows in heterogeneous media, some strong coupling between the flow and the underlying medium may occur. This can be the case when considering water displacing oil, CO₂ injection etc... This may give rise to an instability driven by a mechanism is similar to the well-known Saffman–Taylor instability that leads to viscous fingering if the displacing fluid is more mobile than the fluid initially in place. This emergence problem was studied by many authors such as Saffman and Taylor [1958], Homsey [1987], Tabeling et al. [1988], Tang [1985], Shraiman [1986]. A generalization to immiscible two phase flows

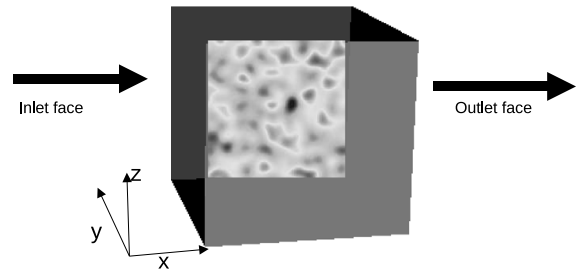


Figure 3. Up-scaling geometry. The coarse block of size L have a detailed conductivity map given by geology. It is up-scaled by solving a steady-state quasi Laplace equation to determine an effective conductivity in the mean flow direction. Boundary conditions are usually no-flux parallel to the imposed mean flow, or periodic.

was carried-out in the paper of King and Dunayevsky [1989, and references therein]. In that case, it appears a well defined sharp front separating both phases moves at the local velocity of the fluid, as shown in Figure 6, top. That front may be in unstable, by the same mechanism than the Saffman–Taylor instability, as shown in Figure 6, bottom. In the unstable case, an amplification of the heterogeneity effect is expected, while some smoothing arises in the stable case. The intuition suggests that a highly mobile

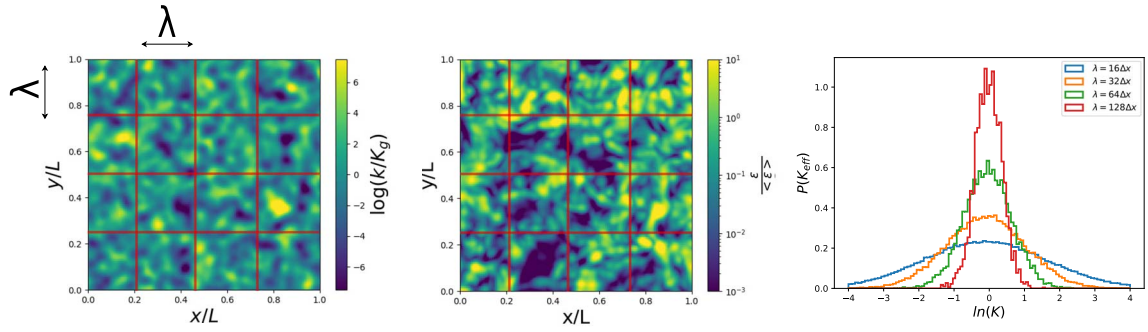


Figure 4. Monte Carlo study of the dependence of the effective conductivity pdf with coarsening scale λ . The overall flow is solved using several realizations of the input log conductivity map (left). The associated local dissipation map (center) allows to evaluate a distribution of coarsened effective conductivities averaged at intermediate scale λ [Colecchio et al., 2020]. The associated pdf's are plotted (right). The self averaging feature is highlighted by the sharply peaked distribution around the geometric average for $\lambda = 128$ units. Reprinted (adapted) from Colecchio et al. [2020] Copyright 2020 with permission of Elsevier.

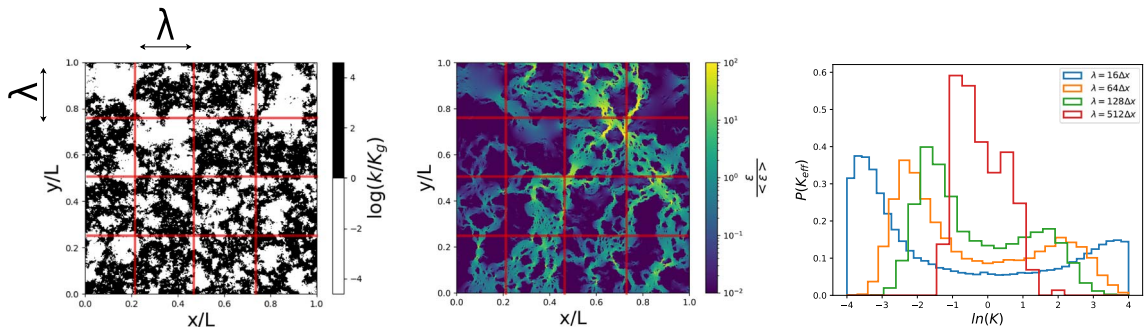


Figure 5. Monte Carlo study of the effective conductivity pdf with coarsening scale λ . The overall flow is solved using several realizations of the input bimodal map with conductivity contrast of 10^4 . The associated local dissipation map (center) highlights the localization phenomenon and allows to evaluate pdf's of coarsened effective conductivities averaged at scale λ [Colecchio et al., 2020]. The resulting scale-dependant pdf's are plotted (right). As the scale is increasing, the two peaks merge, the bimodal distribution disappears and becomes close to a log-normal distribution. The convergence to this asymptotic distribution shows critical slowing-down when the facies proportions are close to percolation threshold. In the infinite contrast case, scaling-laws are recovered [Stauffer and Aharony, 2014]. Reprinted (adapted) from Colecchio et al. [2020] Copyright 2020 with permission of Elsevier.

fluid will follow high velocity paths, its presence amplifying thus that advantage by a positive feedback loop. This is the so called channeling issue. This problem was addressed by De Wit and Homsy [1997a,b], and revisited in the stochastic context by Artus et al. [2004], Noetinger et al. [2004], Artus and Noetinger [2004], merging perturbation theory within the framework developed by King and

Dunayevsky [1989]. In order to be more specific, the underlying equations reads:

$$\nabla \cdot [\lambda(S(\mathbf{r}, t))k(\mathbf{r})\nabla p(\mathbf{r}, t)] = 0 \quad (8)$$

$$\phi \frac{\partial S(\mathbf{r}, t)}{\partial t} + \nabla \cdot (f(S(\mathbf{r}, t))\mathbf{U}(\mathbf{r}, t)) = 0 \quad (9)$$

$$\mathbf{U}(\mathbf{r}, t) = -\lambda(S(\mathbf{r}, t))k(\mathbf{r})\nabla p(\mathbf{r}, t). \quad (10)$$

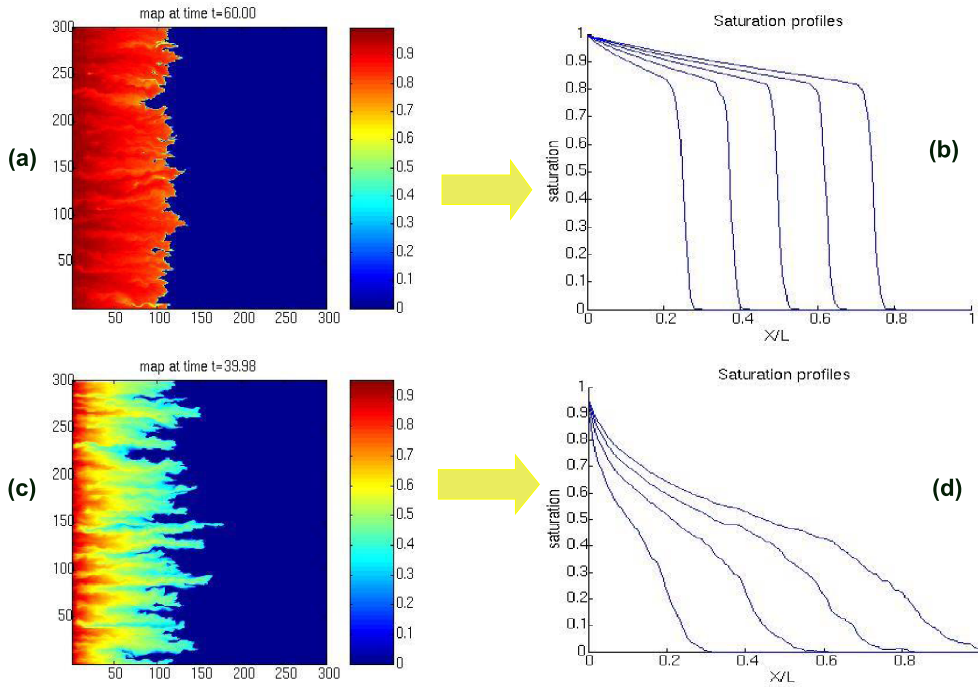


Figure 6. Dynamics of a two phase flow front moving in a heterogeneous rock, imposed mean flow from left to right, underlying random log normally distributed conductivity map top (a) stable case (b) Y averaged saturation at different times, bottom (c) fingering in the unstable case, (d) associated Y -averaged saturation at different times.

Here, $S(\mathbf{r}, t)$, $\lambda(S(\mathbf{r}, t))$ and $f(S(\mathbf{r}, t))$, S denote respectively the water saturation (local % of water, total mobility and the so-called fractional flow of water). This set of coupled equations may be solved numerically (it is at the heart of any multiphase flow in porous media simulator, to which additional complexities such as phase transitions and boundary condition management must be added). The saturation equation (9) is hyperbolic, leading to the formation of a shock front, whose stability is controlled by the jump of the total mobility $\lambda(S(\mathbf{r}, t))$ at the front.

The technical difficulty for setting-up a perturbation expansion comes from the presence of the front that implies a mobility jump that renders perturbation theory a bit tricky [King and Dunayevsky, 1989, Nøttinger et al., 2004]. The difficulty may be avoided by a suitable change of variable, using a working variable $x(S, y, t)$ rather than $S(x, y, t)$. It is thus possible to introduce the function $x(S_f, t)$, in which S_f is the saturation of water just behind the front

M_f is the corresponding total mobility jump $M_f = \lambda(S_f)/\lambda(S=0)$. The randomness of the underlying conductivity field propagates to the randomness of $x(S_f, t)$, of average value $\langle \mathbf{U} \rangle t$. At long times, the associated two point correlation function can be shown to converge to a well defined function in the stable case, while in unstable case it diverges, a manifestation of the spreading of the front [Tallakstad et al., 2009, Toussaint et al., 2012]. A possible approach of practical interest close to the single phase flow approach would be to look at an effective equation driving the ensemble-averaged water saturation $\langle S(\mathbf{r}, t) \rangle$, or the Y -averaged saturation $S(x, t)$. A diffusive-like regime arises in the case $M_f = 1$, that corresponds to a marginal stability criterion. In the general case, several proposals were reported long ago for characterizing the emerging large scale transport equation [Koval et al., 1963, Todd et al., 1972, Yortsos, 1995, Blunt and Christie, 1993, Sorbie et al., 1995]. In the unstable case, one can consider that long fingers parallel to the imposed flow may be treated as a stratified medium.

This leads to modify the fractional flow function with an ad-hoc change [Koval et al., 1963]. In the stable case, it can be shown that the competition between the disorder that distorts the front and the viscous forces that tends to sharpen the front [Noetinger et al., 2004] must lead to another form of the effective fractional flow, including some macrodispersion representing the net effect of the averaged disorder. In the infinite contrast case (e.g. immiscible gas injection) Diffusion Limited Aggregation (DLA) models were proposed, leading to a very rich literature involving percolation invasion models, fractals [Witten Jr and Sander, 1981, Wilkinson and Willemsen, 1983, Paterson, 1984, Masek and Turcotte, 1993] with many contributions of SP.

4. Geostatistics, up-scaling and inverse modelling

Our original issue was to know the relevance of a complex model in regard to the input knowledge. It appears that if one is interested by predicting the average and variance of the output of some transport process, a compromise must be found between the overall accuracy of the calculation involving a high resolution grid implying large computing costs, and the statistical accuracy that requires being able to perform a huge number of Monte-Carlo simulations⁴ required to get representative statistical averages of the process under consideration. The overall simulation budget $B(\text{Size}_{\text{model}}, N_{\text{MC}})$ in terms of hours of CPU depends on the model size $\text{Size}_{\text{model}}$ and the number of Monte Carlo simulations N_{MC} . The latter depends on the number of relevant input parameters that control the overall behaviour of the system. It is possible to determine the couple $(\text{Size}_{\text{model}}, N_{\text{MC}})$ such that the variance $\sigma^2(\text{Size}_{\text{model}}, N_{\text{MC}})$ of the prediction errors of the simulation output is minimized under the constraint of fixed $B(\text{Size}_{\text{model}}, N_{\text{MC}})$. As presented in the preceding section, up-scaling techniques may provide some tools allowing to get a case-dependent prior estimate of that variance without running the whole set of simulations.

Practitioners want to get estimations of the average value and of the variance of the output of interest

(water supply, heat recovery etc...). So the approach developed by Abellan and Noetinger [2010] could be carried-out in the prediction space. The advantage of such an approach will be to be closer to operational quantities of practical interest, and also to be more independent on the dimension of the input parameter space. Such an avenue needs to be developed.

5. Conclusions and some perspectives

In that short note, we discussed some ideas and previous findings illustrating mainly the interaction of the quenched disorder of the natural media present at all scales to the flow. It appears that hydrologists could be greatly helped by developing an approach combining pragmatic practice, advanced averaging techniques and optimization techniques merged within a framework of physically guided artificial intelligence techniques. An approach is to draw the “phase diagram” of the problem at hand, i.e. the set of dominant parameters controlling the overall behaviour of the system, and providing descriptions of the critical behaviour of the system between the different regions of the phase diagram. Information theory concepts can be helpful to quantify the disorder and the amount of information of the system, providing useful links to statistical physics concepts and methods.

Such an approach is evidently process-dependant and must be adapted for heat or tracer transport and mixing, as well as reactive flows, issues that were not discussed in that short paper.

A promising approach would be to map the linear systems to be solved that are essentially of Laplacian nature into random matrix theory close to random graphs theory [Mohar, 1997, Bordenave and Lelarge, 2010, Potters and Bouchaud, 2019]. That area of research is a fast growing field with applications for modelling networks such as the Internet, urban and road traffic, smart electrical networks, brain modelling, finance etc... Some applications may be found in the modelling of flow in fractured rocks [Hyman et al., 2017, 2019]. A possible mathematical approach is to study the spectrum of the large matrices obtained for solving the quasi Darcy flow equation that depends randomly on the geological input parameters. That may provide information about the effective number of relevant degrees of freedom that must be retained to describe the system [Biroli et al., 2007].

⁴We implicitly consider that fully analytically solvable problems are very scarce!

Depending on the degree of disorder, some eigenvalues of the associated Laplacian matrix can be expected to provide information while other eigenvalues may follow some universal distribution such as a Marchenko–Pastur distribution [Marchenko and Pastur, 1967]. Such topics are deeply connected to classification methods by neural nets [Louart et al., 2018, Dall’Amico et al., 2019].

Finally, as illustrated by De Marsily [2009], water supply issues will become more and more critical, implying great applied and academic research efforts. As it was explained in present paper, expecting tremendous computing power cannot be the single option. The community will have to rely on innovative approaches at the cutting edge of modern science, as our mentors did [Matheron and De Marsily, 1980, De Marsily et al., 2005].

Conflicts of interest

The author has no conflict of interest to declare.

Acknowledgements

The author is indebted to MM Alexandre Abellan, Alejandro Boschan, Ivan Colecchio Pauline Mourlanette and Alejandro Otero for their useful contributions.

References

- Abellan, A. and Noetinger, B. (2010). Optimizing subsurface field data acquisition using information theory. *Math. Geosci.*, 42(6), 603–630.
- Abramovich, B. and Indelman, P. (1995). Effective permittivity of log-normal isotropic random media. *J. Phys. A: Math. Gen.*, 28(3), 693–700.
- Armstrong, S., Kuusi, T., and Mourrat, J.-C. (2019). *Quantitative Stochastic Homogenization and Large-scale Regularity*, volume 352. Springer, Berlin.
- Arnold, D. N., David, G., Filoche, M., Jerison, D., and Mayboroda, S. (2019). Localization of eigenfunctions via an effective potential. *Commun. Partial Differ. Equ.*, 44(11), 1186–1216.
- Artus, V. and Noetinger, B. (2004). Up-scaling two-phase flow in heterogeneous reservoirs: current trends. *Oil Gas Sci. Technol.*, 59(2), 185–195.
- Artus, V., Noetinger, B., and Ricard, L. (2004). Dynamics of the water–oil front for two-phase, immiscible flow in heterogeneous porous media. 1–stratified media. *Transp. Porous Media*, 56(3), 283–303.
- Attinger, S. (2003). Generalized coarse graining procedures for flow in porous media. *Comput. Geosci.*, 7(4), 253–273.
- Berkowitz, B. and Balberg, I. (1993). Percolation theory and its application to groundwater hydrology. *Water Resour. Res.*, 29(4), 775–794.
- Biroli, G., Bouchaud, J.-P., and Potters, M. (2007). Extreme value problems in random matrix theory and other disordered systems. *J. Stat. Mech.*, 2007(07), article no. P07019.
- Blunt, M. and Christie, M. (1993). How to predict viscous fingering in three component flow. *Transp. Porous Media*, 12(3), 207–236.
- Bordenave, C. and Lelarge, M. (2010). Resolvent of large random graphs. *Random Struct. Algorithms*, 37(3), 332–352.
- Charlaix, E., Guyon, E., and Roux, S. (1987). Permeability of a random array of fractures of widely varying apertures. *Transp. Porous Media*, 2(1), 31–43.
- Chiles, J.-P. and Delfiner, P. (2009). *Geostatistics: Modeling Spatial Uncertainty*, volume 497. John Wiley & Sons, New-York.
- Colecchio, I., Boschan, A., Otero, A. D., and Noetinger, B. (2020). On the multiscale characterization of effective hydraulic conductivity in random heterogeneous media: a historical survey and some new perspectives. *Adv. Water Resour.*, 140, article no. 103594.
- Dagan, G. (1989). *Flow and Transport in Porous Formations*. Springer-Verlag GmbH & Co. KG, Berlin.
- Dagan, G. (1993). Higher-order correction of effective permeability of heterogeneous isotropic formations of lognormal conductivity distribution. *Transp. Porous Media*, 12(3), 279–290.
- Dall’Amico, L., Couillet, R., and Tremblay, N. (2019). Classification spectrale par la laplacienne déformée dans des graphes réalistes. In *GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images, August 2019, Lille, France*. fffhal-02153901.
- De Marsily, G. (1986). Quantitative hydrogeology. Technical report, Paris School of Mines, Fontainebleau.
- De Marsily, G. (2009). *L’eau, un trésor en partage*. Dunod, Paris.

- De Marsily, G., Delay, F., Gonçalves, J., Renard, P., Teles, V., and Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeol. J.*, 13(1), 161–183.
- De Wit, A. (1995). Correlation structure dependence of the effective permeability of heterogeneous porous media. *Phys. Fluids*, 7(11), 2553–2562.
- De Wit, A. and Homsy, G. (1997a). Viscous fingering in periodically heterogeneous porous media. I. Formulation and linear instability. *J. Chem. Phys.*, 107(22), 9609–9618.
- De Wit, A. and Homsy, G. (1997b). Viscous fingering in periodically heterogeneous porous media. II. Numerical simulations. *J. Chem. Phys.*, 107(22), 9619–9628.
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *J. R. Stat. Soc.: Ser. C Appl. Stat.*, 47(3), 299–350.
- Eberhard, J., Attinger, S., and Wittum, G. (2004). Coarse graining for upscaling of flow in heterogeneous porous media. *Multiscale Model. Simul.*, 2(2), 269–301.
- Floris, F. J., Bush, M., Cuypers, M., Roggero, F., and Syversveen, A. R. (2001). Methods for quantifying the uncertainty of production forecasts: a comparative study. *Pet. Geosci.*, 7(S), S87–S96.
- Gelhar, L. W. (1993). *Stochastic Subsurface Hydrology*. Prentice-Hall, Upper Saddle River, NJ.
- Gorell, S., Bassett, R., et al. (2001). Trends in reservoir simulation: Big models, scalable models? Will you please make up your mind? In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, Houston.
- Granjeon, D. and Joseph, P. (1999). Concepts and applications of a 3-D multiple lithology, diffusive model in stratigraphic modeling. In *Numerical Experiments in Stratigraphy: Recent Advances in Stratigraphic and Sedimentologic Computer Simulations*. SEPM Society for Sedimentary Geology, Oklahoma City.
- Homsy, G. M. (1987). Viscous fingering in porous media. *Annu. Rev. Fluid Mech.*, 19(1), 271–311.
- Hristopulos, D. and Christakos, G. (1999). Renormalization group analysis of permeability upscaling. *Stoch. Environ. Res. Risk Assess.*, 13(1–2), 131–161.
- Hristopulos, D. T. (2020). *Random Fields for Spatial Data Modeling A Primer for Scientists and Engineers*. Springer, Berlin.
- Hu, L. and Chugunova, T. (2008). Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. *Water Resour. Res.*, 44(11), article no. W11413.
- Hunt, A., Ewing, R., and Ghanbarian, B. (2014). *Percolation Theory For Flow in Porous Media*, volume 880. Springer, Berlin.
- Hyman, J. D., Dentz, M., Hagberg, A., and Kang, P. K. (2019). Emergence of stable laws for first passage times in three-dimensional random fracture networks. *Phys. Rev. Lett.*, 123(24), article no. 248501.
- Hyman, J. D., Hagberg, A., Srinivasan, G., Mohd-Yusof, J., and Viswanathan, H. (2017). Predictions of first passage times in sparse discrete fracture networks using graph-based reductions. *Phys. Rev. E*, 96(1), article no. 013304.
- Indelman, P. and Abramovich, B. (1994). A higher-order approximation to effective conductivity in media of anisotropic random structure. *Water Resour. Res.*, 30(6), 1857–1864.
- Jacquard, P. (1965). Permeability distribution from field pressure data. *Soc. Pet. Eng. J.*, 5(04), 281–294.
- Jikov, V. V., Kozlov, S. M., and Oleinik, O. A. (2012). *Homogenization of Differential Operators and Integral Functionals*. Springer Science & Business Media, Berlin.
- King, M. J. and Dunayevsky, V. A. (1989). Why waterflood works: a linearized stability analysis. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, Houston.
- King, P. (1987). The use of field theoretic methods for the study of flow in a heterogeneous porous medium. *J. Phys. A: Math. Gen.*, 20(12), 3935–3947.
- King, P. (1989). The use of renormalization for calculating effective permeability. *Transp. Porous Media*, 4(1), 37–58.
- Koval, E. et al. (1963). A method for predicting the performance of unstable miscible displacement in heterogeneous media. *Soc. Pet. Eng. J.*, 3(02), 145–154.
- Krige, D. (1976). A review of the development of geostatistics in South Africa. In *Advanced Geostatistics in the Mining Industry*, pages 279–293. Springer, Berlin.
- Landau, L. and Lifshitz, E. (1960). *Electrodynamics of Continuous Media*, volume 8. Mir, Moscow.
- LaVenue, A. M., RamaRao, B. S., De Marsily, G., and Marietta, M. G. (1995). Pilot point methodology for automated calibration of an ensemble of con-

- ditionally simulated transmissivity fields: 2. Application. *Water Resour. Res.*, 31(3), 495–516.
- Le Ravalec, M., Noetinger, B., and Hu, L. Y. (2000). The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning Gaussian simulations. *Math. Geol.*, 32(6), 701–723.
- Louart, C., Liao, Z., Couillet, R., et al. (2018). A random matrix approach to neural networks. *Ann. Appl. Probab.*, 28(2), 1190–1248.
- Marchenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mat. Sb.*, 114(4), 507–536.
- Masek, J. G. and Turcotte, D. L. (1993). A diffusion-limited aggregation model for the evolution of drainage networks. *Earth Planet. Sci. Lett.*, 119(3), 379–386.
- Matheron, G. (1963). Principles of geostatistics. *Econ. Geol.*, 58(8), 1246–1266.
- Matheron, G. (1967). *Éléments pour une théorie des milieux poreux*. Masson, Paris.
- Matheron, G. and De Marsily, G. (1980). Is transport in porous media always diffusive? a counterexample. *Water Resour. Res.*, 16(5), 901–917.
- Mézard, M., Parisi, G., and Virasoro, M. (1987). *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, Singapore.
- Michael, H., Li, H., Boucher, A., Sun, T., Caers, J., and Gorelick, S. (2010). Combining geologic-process models and geostatistics for conditional simulation of 3-d subsurface heterogeneity. *Water Resour. Res.*, 46(5), article no. W05527.
- Mohar, B. (1997). Some applications of laplace eigenvalues of graphs. In *Graph Symmetry*, pages 225–275. Springer, Berlin.
- Mourlanette, P., Biver, P., Renard, P., Noetinger, B., Caumon, G., and Perrier, Y. A. (2020). Direct simulation of non-additive properties on unstructured grids. *Adv. Water Resour.*, 143, article no. 103665.
- Neuman, S. P. and Orr, S. (1993). Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resour. Res.*, 29(2), 341–364.
- Noetinger, B. (1994). The effective permeability of a heterogeneous porous medium. *Transp. Porous Media*, 15, 99–127.
- Noetinger, B. (2000). Computing the effective permeability of log-normal permeability fields using renormalization methods. *C. R. Acad. Sci. - Ser. IIA - Earth Planet. Sci.*, 331(5), 353–357.
- Noetinger, B. (2020). Statistical physics and applied geosciences: some results and perspectives. *C. R. Phys.*, 21(6), 539–560.
- Noetinger, B., Artus, V., and Ricard, L. (2004). Dynamics of the water–oil front for two-phase, immiscible flow in heterogeneous porous media. 2–isotropic media. *Transp. Porous Media*, 56(3), 305–328.
- Noetinger, B. and Gautier, Y. (1998). Use of the Fourier-Laplace transform and of diagrammatical methods to interpret pumping tests in heterogeneous reservoirs. *Adv. Water Resour.*, 21(7), 581–590.
- Oliver, D. S. and Chen, Y. (2011). Recent progress on reservoir history matching: a review. *Comput. Geosci.*, 15(1), 185–221.
- Paterson, L. (1984). Diffusion-limited aggregation and two-fluid displacements in porous media. *Phys. Rev. Lett.*, 52(18), 1621–1625.
- Potters, M. and Bouchaud, J.-P. (2019). *A First Course in Random Matrix Theory*. Cambridge University Press, Cambridge.
- Renard, P. and De Marsily, G. (1997). Calculating equivalent permeability: a review. *Adv. Water Resour.*, 20(5), 253–278.
- Saffman, P. G. and Taylor, G. I. (1958). The penetration of a fluid into a porous medium or Hele-Shaw cell containing a more viscous liquid. *Proc. R. Soc. Lond. A. Math. Phys. Sci.*, 245(1242), 312–329.
- Santos Oliveira, D., Horowitz, B., and Rojas Tueros, J. A. (2021). Ensemble-based method with combined fractional flow model for waterflooding optimization. *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles*, 76, article no. 7.
- Scheidt, C., Zabalza-Mezghani, I., Feraille, M., and Collombier, D. (2007). Toward a reliable quantification of uncertainty on production forecasts: adaptive experimental designs. *Oil Gas Sci. Technol. - Revue de l'IFP*, 62(2), 207–224.
- Shraiman, B. I. (1986). Velocity selection and the Saffman–Taylor problem. *Phys. Rev. Lett.*, 56(19), 2028–2032.
- Sorbie, K., Zhang, H., and Tsibuklis, N. (1995). Linear viscous fingering: new experimental results, direct simulation and the evaluation of averaged models. *Chem. Eng. Sci.*, 50(4), 601–616.
- Stauffer, D. and Aharony, A. (2014). *Introduction to*

- Percolation Theory: Revised Second Edition*. CRC Press, London.
- Stepanyants, Y. A. and Teodorovich, E. (2003). Effective hydraulic conductivity of a randomly heterogeneous porous medium. *Water Resour. Res.*, 39(3), article no. 1065.
- Tabelling, P., Zocchi, G., and Libchaber, A. (1988). An experimental study of the Saffman–Taylor instability. In *Dynamics of Curved Fronts*, pages 219–234. Elsevier, Amsterdam.
- Tallakstad, K. T., Knudsen, H. A., Ramstad, T., Løvoll, G., Måløy, K. J., Toussaint, R., and Flekkøy, E. G. (2009). Steady-state two-phase flow in porous media: statistics and transport properties. *Phys. Rev. Lett.*, 102(7), article no. 074502.
- Tang, C. (1985). Diffusion-limited aggregation and the Saffman–Taylor problem. *Phys. Rev. A*, 31(3), 1977–1987.
- Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*, volume 89. SIAM, Philadelphia.
- Teodorovich, E. (2002). Renormalization group method in the problem of the effective conductivity of a randomly heterogeneous porous medium. *J. Exp. Theor. Phys.*, 95(1), 67–76.
- Todd, M., Longstaff, W., et al. (1972). The development, testing, and application of a numerical simulator for predicting miscible flood performance. *J. Pet. Technol.*, 24(07), 874–882.
- Toussaint, R., Måløy, K. J., Méheust, Y., Løvoll, G., Jankov, M., Schäfer, G., and Schmittbuhl, J. (2012). Two-phase flow: Structure, upscaling, and consequences for macroscopic transport properties. *Vadose Zone J.*, 11(3), article no. vzj2011.0123.
- Wilkinson, D. and Willemsen, J. F. (1983). Invasion percolation: a new form of percolation theory. *J. Phys. A: Math. Gen.*, 16(14), 3365–3376.
- Witten Jr, T. and Sander, L. M. (1981). Diffusion-limited aggregation, a kinetic critical phenomenon. *Phys. Rev. Lett.*, 47(19), 1400–1404.
- Yortsos, Y. C. (1995). A theoretical analysis of vertical flow equilibrium. *Transp. Porous Media*, 18(2), 107–129.
- Zdeborová, L. and Krzakala, F. (2016). Statistical physics of inference: Thresholds and algorithms. *Adv. Phys.*, 65(5), 453–552.
- Zdeborová, L. and Krzakala, F. (2007). Phase transitions in the coloring of random graphs. *Phys. Rev. E*, 76(3), article no. 031131.
- Zhang, D. (2001). *Stochastic Methods for Flow in Porous Media: Coping With Uncertainties*. Elsevier, Amsterdam.